Econ272 Foundations of Economic Analysis Text: An Economic Way of Thinking

4th Edition

Table of Contents

Chapter 1: An Economic Way of Thinking

Economists and Normal People

Economists are weird. They don't think like normal people. Because you are most likely a normal person, and you will encounter a great deal of abnormal thinking in this text, I'd like to start by confronting this issue directly, rather than letting it lie dormant for a few important chapters. The weirdness of economic thinking often makes economics hard to understand, and I want you to understand basic economics. You don't have to like it. You don't even have to think like an economist after the final exam, but I hope in this text to show you that it might be a good idea once in awhile. Let's now look at a few characteristics of their thinking that make economists a little different, a little strange, a little abnormal.

Some Keys to Understanding the Way Economists Think

Economists assume people try to do the best they can with the resources they have. We call this rational behavior. Our definition of "best they can" usually means they do the best they can for themselves. This doesn't mean that people aren't concerned about the welfare of others. Take poverty for example. In economics we would assume that an individual could want fewer poor people because it would make him or her feel better to have less poverty. It's the well being of the individual thinking about poverty that is important in a definition of rationality, not the level of poverty per se.

Economists focus on individuals rather than groups of individuals. Even though we study groups of individuals such as businesses, governments, and families, we always realize that individuals make up these groups, and that these individuals have a paramount interest in their own well being. The idea of a "society" or "community" wanting something is foreign to the economist. Societies and communities do not choose and act. People do.

Economists don't question the validity of individual preferences. Economists are not normative in an "in your face" sort of way. A normal person might question the ethics of driving a gas-guzzler in an energy crisis, or drinking alcohol, smoking tobacco, or viewing internet pornography. An economist would consider these actions simply the choices of a rational individual. “Different strokes for different folks” could be an economist’s anthem.

Economists emphasize making the economic pie bigger and worry less about the relative size of an individual’s piece of it. This is sometimes referred to as a concern for efficiency over equity. Or like the character Peekay in Bryce Courtney’s Power of One, economists think first with the head, then with the heart. Normal people think first with the heart, then with the head.

Economists love abstraction, theory and models. In order to understand how the economic world works, economists form theories and use models to represent these theories. Take crime for example. Why does someone rob a liquor store? A normal person might posit that a criminal is the product of a broken poverty-stricken home, a person raised without instruction in proper values and modes of social conduct. Not an economist. To an economist, one robs a liquor store if the expected benefits exceed the expected costs. Irrespective of one’s background, one has to choose to commit the robbery. Likewise, an economist would model marriage as an exchange between two rational individuals each trying to maximize the expected present value of a lifetime stream of benefits and costs.

Because they assume people act rationally in their own interest, economists often point out the self-interest in seemingly lofty human behavior. Need-based financial aid at colleges is price discrimination. The NCAA is a successful buyers cartel, exploiting college athletes for the good of the institution. Bureaucrats are not just passive conduits through which legislative decisions are channeled into our lives. Bureaucrats have their own agenda. Churches can best be understood as profit maximizing institutions, and so on. Normal people see this as cynicism and avoid associating with economists.

Economists have a strong commitment to the market economy as a way of addressing the basic economic problem. To the economist, trade is mutually beneficial. Normal people distrust markets, and often think that trade is a way for the powerful to take advantage of the weak.

Economists see the benefit to others from self-interested behavior. This is the famous invisible hand principle. In 1776 Adam Smith wrote a book inquiring into the causes of wealth and prosperity. One of his conclusions was that self-interested pursuit of profit by the farmer, the rancher, butcher and baker puts food on the table of others. They intend to maximize their own well being, but are guided as if by an invisible hand to provide for the well being of others. Normal people are more romantic than economists, believing that good in society comes from a noble perfection of the individual, not through the vulgarity of self-interest.

Economists emphasize that for every benefit there is a cost. Normal people don't like to think about the "down side" of actions, and we economists are quick to remind normal people about opportunity cost, the value of opportunities foregone when making choices. Normal people interpret this as negativity (perhaps it is) and don't invite economists to parties. Who wants to eat cheese, drink wine and chat with someone who thinks that rivers and lakes can be too clean, and that we can have too little crime?

I could give you other examples of economists' strange ideas, but if I did you might abandon the book in this early chapter. I don't want to overload you with economic weirdness too soon, but this little warning might make the rest of the book make a little more sense.

The Basic Economic Problem Scarcity, Choice and Cost

At its most fundamental level economics is about a conflict between ends and means. The ends involve human well being and the means are the resources humans can muster to provide it. In economics we define human well being in terms of the value of satisfied human wants. The higher the value of satisfied human wants, the higher the level of human well being.

The means to satisfy human wants are resources. These are labor, capital, human capital and natural resources. Labor is the productive power people have through the use of their time. In economics we focus a lot on the amount of this time people sell to others, but the scope of economics also recognizes time used to cook a meal, mow a lawn, play a piano, or take a hike.

The people in an economy must create capital resources. People create capital goods by saving. Capital includes any good that doesn't satisfy human wants directly now, but rather is used to produce goods and services that satisfy human wants over time. For example, factories, warehouses and dump trucks do not provide direct want satisfaction. Oh, perhaps an operations management major might see the beauty in an assembly line, but the main purpose of an assembly line is to give us beer, bread, and beauty aids, things we consume to provide our well being. Individuals have their own capital goods such as houses, refrigerators, stoves and lawnmowers, and can own a portion of other capital goods by purchasing shares of a company's stock through a stock market. We also own collectively capital goods through governments, such as roads, parks, fighter jets and school buildings.

Human capital is the resource people create by augmenting the productivity of their labor time by acquiring skill and technique. Economists are interested mostly in the enhanced productivity that people sell in the marketplace, what we call marketable human capital. Most college students understand that acquiring human capital requires that we forego want satisfaction today in exchange for more want satisfaction in the future. Learning about Asian history, French impressionist painting, and the 20th century American novel all create human capital as well, albeit usually a less marketable form.

Natural resources are gifts from nature, such as land to grow food and trees on, to dig minerals out of, to build on, or to dispose of waste on. Oceans, lakes and rivers provide sources of food, energy, transportation and recreation. While the physical resource base is limited, the value of that resource base in terms of want satisfaction is determined by the amount of labor, capital and human capital we have to use with natural resources. We also receive other life support services from natural resources, such as the protection from ultraviolet radiation by the ozone layer, a hospitable climate, and a rich and diverse biota.

The implications of the basic economic problem are illustrated in Figure 1-1. The fundamental conflict between wants and resources is shown buy the two solid arrows running into each other. This conflict results in scarcity, the most fundamental economic fact of life. Our wants exceed our ability to satisfy them with our resources. Sometimes economists assume unlimited wants, and this might be a correct empirical observation, but unlimited wants are not necessary for scarcity. As long as wants exceed what resources can provide we have scarcity.

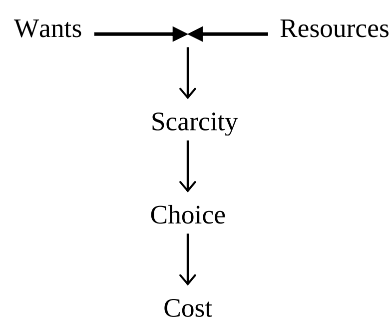


Figure 1-1
The Basic Economic Problem

The fact of scarcity leads to the next implication shown in Figure 1-1, choice. Because we cannot satisfy all wants with our resources we have to choose among wants, and choose how to use our resources in satisfying them. Choice is such a fundamental concept that economics is sometimes called the science of choice. Individuals certainly have to choose. Do I go to a movie or study economics? Do I buy a new computer or new ski equipment? Do I forego a better car today to invest in human capital by going to college?

Any organized group of people also faces scarcity and is forced to choose. A family chooses how many of its members work outside the home, and who cooks or does the dishes or mows the lawn. Campus organizations choose among alternative party and public service projects. Businesses choose which goods and services to produce and how much of each resource to use in the production process. Local, state and national governments face wants of constituents that exceed what limited tax revenues can provide. They, too, have to choose.

Finally, we see the most unfortunate implication of the basic economic problem at the bottom of Figure 1-1, cost. When we choose one use of resources, we give up another. The cost of that choice is the value of the best foregone alternative. If you can buy only a computer or new ski equipment, the cost of choosing the computer is the ski equipment you didn't buy. Because choice satisfies a particular want and foregoes the opportunity to satisfy others, we often refer to cost as opportunity cost in economics.

Please note that we have found the source of cost in economics without talking about money or prices at all. Cost can result from spending choices in a marketplace, but it doesn't have to. Every spring students understand opportunity cost when they are confronted with the choice of going to class or playing in the warm sunshine. The foregone value of tossing a Frisbee is the cost of attending a history class. No market transaction arises, but cost occurs just the same.

Alternative Approaches to the Basic Economic Problem

We will be studying the standard economic approach to the basic economic problem in this text. In standard economics we examine ways to maximize the value of human wants satisfied. Using more and more resources and using them in the best way are ways to do this. Generally in standard economics the more human wants satisfied the better. More is preferred to less.

The basic economic problem does not require that we approach it this way. We could scale back our wants to live within the means provided by our resources. Perhaps we humans could adopt the teachings of many of the world's religions. It doesn't say in the Bible or the Koran "He who ends with the most wants satisfied wins." E.F. Schumacher in his 1973 book *Small is Beautiful* even laid out the principles of a "Buddhist Economics," where the primary want is being free of unlimited wants. Modern day members of Green political parties around the world advocate scaling back certain human wants so that nonhuman members of the planet and their ecosystems could do a little better. And there is nothing wrong with this approach. It's just not standard economics.

Alternative approaches to the basic economic problem are usually more personally normative than the standard economic approach. By normative we mean based on value judgments. Usually normative statements appear with words like should and ought. For example, someone with Green values would suggest that we should address the basic economic problem by using less fossil fuels and more solar energy, by recycling and reusing goods and packaging, and by generally reducing the human-centered approach to natural resource use. In standard economics any want an individual has is O.K., and no particular resource has any special priority in terms of use.

The standard economic approach does not, however, guarantee that we will trash our natural environment to satisfy material wants. People have preferences for environmental quality, and reducing air and water pollution and preserving endangered species are actions we might choose to do, if these choices improve our well being. But make no mistake about it, in standard economics wants of people drive the system, not those of the blue whale or the bald eagle. Standard economics is strictly human-centered.

Economic Systems and the Basic Economic Problem

An economic system is a set of social rules and institutions that guide the way a group of people address the basic economic problem. One such group we will call the economy of a country. In the face of scarcity, people in an economy have to answer three basic economic questions. What will we produce? How will we produce it? And for whom will the goods and services be produced, i.e., who gets to consume what portion of the national economic output? When these decisions are answered economists say the economy has allocated its resources.

One way to classify economic systems is illustrated in Figure 1-2. We use only two characteristics of an economic system here, and both have to do with resources. Across the top of the figure we find resource allocation, with two possible subcategories, market and command. Down the side of the figure we have two categories of resource ownership, private and government. The words collective, public, and state often are substituted for the word government. The two possible forms of resource allocation and two possible forms of resource ownership give four combinations that describe four types of economic systems. But before we address these economic systems, let's discuss the two characteristics of economic systems in more detail.

		Resource Allocation	
		Market	Command
Resource Ownership	Private	Market Capitalism	Command Capitalism
	Government	Market Socialism	Command Socialism

Figure 1-2
A Classification of Economic Systems

Market allocation is something we will spend a lot of time on in this book. With market allocation an economy uses the decentralized process of people interacting in markets to answer the three fundamental economic questions. People answer the WHAT question by deciding what they want to consume and by their willingness to buy it in a market. If sellers can produce and sell something at a price that people are willing to pay, it gets produced. If not, the good is not produced. For example, take platinum lawn sprinklers. At the price a producer would have to receive to make platinum lawn sprinklers profitable, no one would buy them. Platinum lawn sprinklers don't get produced. By contrast, people are willing to pay $20,000 – $40,000 for sport utility vehicles in the U.S. We can find them for sale almost everywhere.

Market resource allocation addresses the HOW question in a similar decentralized manner. Producers want to produce goods at the lowest possible cost, because if they keep costs down there will be more profit for the owners of the company. Producers have an incentive to use just the right amount of labor, human capital,

capital and natural resources. Owners of resources wish to sell them at a price that yields the highest amount of income. With market allocation buyers of resources and sellers of resources come together in the market, and in the process of buying and selling prices are established that guide both buying and selling behavior.

In market allocation the FOR WHOM question is answered by market processes as well. People receive income from selling resources in the market. If you have a lot of resources to sell, or if you just get a high price for what you do sell, you have a high income and can buy lots of goods and services in the marketplace. If you don't own many resources, or if you receive a low price for what you do sell, you can't buy as much in the marketplace. Your portion of national economic output is much smaller. In market allocation your portion of national output depends on the value of the resources you sell in the marketplace.

Command resource allocation relies on centralized political decisions about resource use rather than individual decisions coordinated by markets. The government decides what will be produced, how it will be produced, and the incomes of people in the economy. Political incentives rather than price incentives guide the allocation of resources.

No economy relies totally on either market or command allocation. Some form of markets exists even in the most extreme command economies. Likewise, some government involvement in resource allocation exists in all market-oriented economies, usually in the form of regulation of product and resource markets or in the form of income redistribution programs, such as food and housing subsidies for low-income people. Governments might restrict certain market transactions in prostitution, pornography, drugs, and gambling. Liquor stores are closed on Election Day many places in the U.S. Businesses often need to obtain licenses even in market economies. Zoning laws restrict the type of business you can have in a certain area. Child labor, hazardous working conditions, uncontrolled air and water pollution and misleading financial statements might be illegal. And, of course, we are all familiar with taxes, part of our income that we are not allowed to buy goods and services with.

The meaning of private resource ownership is pretty straightforward. Individuals or voluntary associations of them in businesses own resources. And if you own a resource you have a right not to use that resource in the economy. If you don't want to work, you don't have to. Private ownership of capital can be seen in the different forms of businesses and in housing. In a sole proprietorship, the business owner owns capital. In a partnership a group of individuals owns the capital. The corporation is a form of business where capital is owned by whomever purchases shares of stock in the company. The existence of a stock market in a country indicates private ownership of capital. Owner-occupied housing and private owners of apartment buildings are also examples of private ownership in an economy.

Private individuals or voluntary collections of them also own natural resources. Oyster beds, hydropower sites, and frequencies of the radio spectrum can be privately owned. Those who have a certain kind of water right can own ground water and surface water, as well as timber lands, farms, mines and oil wells.

Government ownership means that we own a resource in common through the government. State universities, local public schools, dams in public utility districts, national and state forests, federal, state, county and city parks, highways, and roads and streets are all examples of resources owned collectively through government.

The four quadrants of Figure 1-2 comprise our classification of economic systems. Most economic systems lie on a continuum from the upper left to the lower right in the figure, along the main diagonal of this 2 by 2 matrix. Economic systems with a lot of private ownership and market allocation of resources we call market capitalism. Countries like the United States, Ireland, and Estonia would solidly fit in this box. In the lower right of the figure, with high amounts of government ownership and command allocation of resources we would find command socialist economic systems. Countries such as Cuba, North Korea and Vietnam would occupy this position.

The "off-diagonal" economic systems are a little less common. Command capitalism is best illustrated by economies that are no longer with us, such as Nazi Germany and Italy under Mussolini. Japan, with much private ownership, but a great deal of government direction of the economy might be modestly command capitalist. Market socialism has government ownership and market allocation. The former Yugoslavia under Tito was an example of this kind of economic system. As China continues its transition away from command socialism under Mao, it can be considered as moving through market socialism toward market capitalism.

Later in this book we will be discussing the political economy of economic policy. Political economy is the interaction of politics and economics. When we make a meal from these two ingredients, politics and economics, not all will find the dishes to their liking. In fact, a great many political issues around the world today have much to do with our two ways of classifying economic systems, the methods of resource allocation, market or command, and the modes of ownership, private or government.

The Science of Economics: Observation, Theory and Models

Just as chemists, physicists and geologists wish to explain the natural world, economists wish to explain individual behavior, social relationships and the causes of economic well being. In observing the economic world, we see all kinds of economic phenomena. Prices of goods and services rise and fall, and when they do people change how much of them they buy and sell. For some goods lines of buyers form where the good is sold. For other goods, chronic surpluses develop. New products come on the market and others vanish. Some people are rich, others are poor. Economies grow, but they also decline, sometimes by a large amount like in the Great Depression from 1929-1933. In some economies, people become more and more well off over time, while in others they remain locked in poverty. Some businesses have the power to set prices at whatever level they want to, while others appear to have little pricing power. Some businesses succeed and some businesses fail. Democratic governments seem always to spend more than they collect in taxes, resulting in chronic budget deficits. Goods like national defense, wildlife species, and clean air and water are under produced in an economy unless the government is involved. This is only a sampling from the list of economic phenomena. Why do these events happen? How does the economic world work?

One common method of explaining real world phenomena is the use of theory and observation. Economics is not unique here. The theory of relativity, the gas laws, and the laws of thermodynamics are examples of theory from the natural sciences. After observing phenomena in the real world, we develop possible reasons for their occurrence. These speculations take the form of cause and effect relationships called theories. If A happens, then B results. For example, we will study in Chapter 3 a very important economic theory, the theory of demand, which posits an inverse relationship between the price of a good and the amount buyers will buy, other things held constant. But is this a good theory?

Sometimes we evaluate a theory according to its logic and internal consistency, but more often in modern science we have to test theories to see if they work. Is the pudding tasty? Try it and see if you like it. “The proof of the pudding is in the eating.”

Theories are harder to test in economics than in the natural sciences, because laboratory experiments are difficult to arrange. If a chemist wants to test the theory about the pressure in a closed container with a given volume of a gas when she turns up the heat, she can go to the laboratory and do a controlled experiment. She holds the volume of gas constant in a closed container, turns up the heat and measures the pressure change. In economics it’s much more difficult to do controlled experiments. Can you imagine the national outcry if an economist received a grant to test what would happen if a government rapidly and dramatically increased the amount of money in the economy. The rapid increase in the average level of prices would explain an interesting cause and effect relationship, something we’ll call inflation in Chapter 11, but that would be the last economics experiment performed.

One way of exploring the implications of theories about reality is the use of models. The U.S. Army Corps of Engineers, a government agency actively involved in flood control and river transportation, has a scale model of the Mississippi River system. By running a large flow of water through the scale model, they can get an idea of how floods will affect natural and human activities along the banks of the Mississippi without actually flooding out the folks who live there. Probably better than testing the theory in the real world, don’t you think? The important thing about this hypothetical flood in the model of the Mississippi is that Corps of Engineers can control what they want to do with the model. They have the power to control the uncontrollable and see implications they would have missed by just staring out the window at reality. Models are very useful.

Scale models are rare in economics. We mainly use mathematical models in economics, and representations of them with graphs. For some beginning students of economics abstraction is difficult and the models we use to explore theories are a barrier to learning. But the simplification of reality is absolutely necessary if we are going to understand it.

Because the real economic world is so complex we can never capture economic reality in its entirety, nor would we want to. Economists abstract from reality in order to understand it better. When we create an abstraction, we assume part of the world away. It is possible that the grape harvest in the Rhone Valley of France could affect the amount of bread people buy in Dayton, Ohio, but for the purpose of understanding the demand for bread in Dayton, it's best to assume away the Rhone grape harvest.

A Model of an Economy: The Production Possibility Frontier I

When we discussed the basic economic problem in the last section we did it in a very general manner. By using a very simple model of an economy we can be a little more specific about particular aspects of the basic economic problem and can develop some economic theories related to it. A classic production possibility frontier (ppf) model is shown in Figure 1-3. A large amount of abstraction from reality is apparent at the outset. We assume the economy produces only two goods, guns and butter, and represent each with an axis on the graph. A larger amount of butter is represented as a movement rightward along the horizontal axis of the graph, e.g., the amounts B_1, B_2, and B_{max}. Likewise, a movement up the vertical axis represents a larger amount of guns, e.g., the amounts G_2, G_1, and G_{max}. The relevance of the notation of the model will become apparent in a moment.

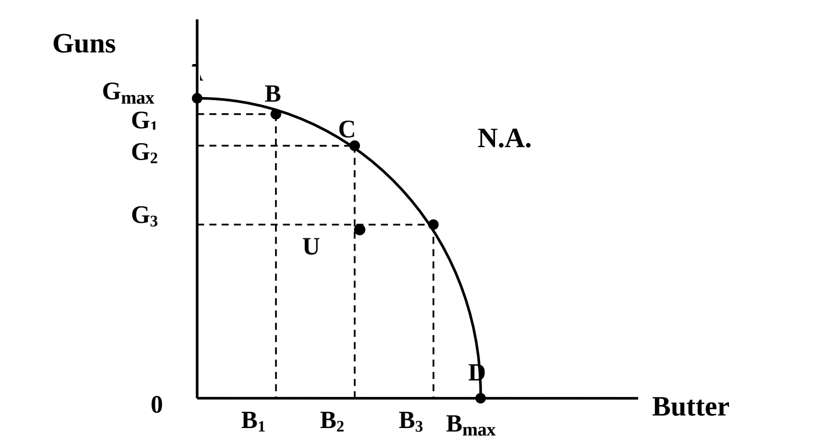


Figure 1-3
The Production Possibility Frontier: Guns and Butter

Because resources in the economy are limited, we can't have all the butter we want. If we put all of the economy's resources into producing butter we would be limited to some amount, say B_{max}. If we put all our resources into production of guns we would also be limited to an amount, say G_{max}. Note that at G_{max} we are producing only guns and no butter (point labeled A), and at B_{max} we are producing only butter and no guns (point labeled D). These two combinations of goods are two possible points on the economy's ppf, but only two of the possible points.

In an extremist society of either butter crazies or gun nuts we would be finished with the model at this point. But in most economies people would want to produce some of both goods, because they would want to consume some of each. Points B and C on the graph represent two additional combinations of guns and butter produced by the economy. At point B we produce G_1 guns and B_1 butter. Note that we have now represented a theory about the implications of scarcity in the model. If we are using all of our limited resources at point A, producing only guns, and we want some positive amount of butter, we must transfer some of the resources from gun production to butter production. At point B we have $B_1 – 0$ more butter production than we had at point A, but $G_{max} – G_1$ less gun production. The only way we can produce (and consume) more butter is to choose to produce fewer guns. We also can say that the cost of choosing B_1 butter is the $G_{max} – G_1$ guns given up. Fewer guns is the opportunity cost of more butter. If scarcity exists, then people choose and incur cost.

Point C is another possible combination of guns and butter production for the economy, G_2 guns and B_2 butter. Again, to produce more butter we have to give up more guns. Now the cost of $B_2 – B_1$ butter is $G_2 – G_1$ guns. Note that the model has been constructed such that butter has become more costly as we have more and more of it. We can see this in the model because the two increments of butter, $B_1 – 0$ and $B_2 – B_1$ are equal. The costs of the two equal amounts of butter, however, are not equal, as $G_2 – G_1 > G_1 – G_{max}$. We call this the law of increasing cost, our first encounter with what economists call diminishing returns.

One way to understand the law of increasing cost is to realize that all resources are not equally suited for the production of both goods. In the U.S., for example, Wisconsin would be a much better place to produce butter than guns. In Wisconsin, we have abundant good pastureland and many pacifists who study at the University of Wisconsin. Utah, however, would be a much better place for military operations, as the mountain and desert terrain are much better for bombing ranges than for cows, and students at Brigham Young University are more likely to support a strong national defense. At point A, with only gun production, we're producing guns in both Utah and Wisconsin. As we transfer land and labor from gun production to butter production, we don't do it in Utah, but rather we expand butter production in Wisconsin. We leave Utah land (mountains and desert) and labor (BYU students) in gun production and convert Wisconsin bombing ranges to pasture, and make dairy farmers out of reluctant soldiers. We get a big increase in butter, $B_1 - 0$, for a small cost in guns, $G_1 - G_{max}$. But as we convert more and more land and labor from guns to butter, we eventually have to start making dairy farmers out of BYU graduates and start milking cows in the desert of the Great Basin. Note that the final increment of butter, $B_{max} - B_2$, requires a much greater opportunity cost in guns. $G_2 - 0$.

Points A, B, C, and D are just four possible combinations of guns and butter available in the economy. We can imagine an almost infinite number of points between A and B and between B and C and C and D. If we connected all these points we would have the production possibility frontier ABCD of Figure 1-3. The concavity to the origin, (it's bowed outward) represents the law of increasing cost Scarcity requires that people in the economy choose points on or inside the production possibility frontier. We would like to be farther to the upper right in Figure 1-3, because we assume more is preferred to less, but we have to label points in this area not attainable (N.A.), because of limited resources and the ensuing scarcity they generate.

One point inside the ppf is point U. Here the economy is not using all its resources, or it is using them inefficiently. Because we assume people in the economy prefer more to less of each of the goods in our model, we can safely say that the combination of goods represented by point U is a bad one. Any point within the approximate triangular area UBC is preferable to point U, because we can have more of one good without sacrificing any of the other. At many points we could have more of both goods. In Chapter 9 we will have more to say about moves like the one from point U to a point on the ppf. We can see here what economists find objectionable about an economy with unemployed resources or inefficient uses of them. When resources are unemployed people in the economy forego the satisfaction of valuable wants. That can't be good when more is preferred to less.

A Model of a College Student's Basic Economic Problem: The Production Possibility Frontier II

We noted earlier in this chapter that the basic economic problem applied at all levels of economic existence. The guns and butter ppf model applied to the total macroeconomic economy. In Figure 1-4 we see that the same type of model can be applied to an individual college student's situation. Instead of guns and butter on the axes of the ppf, we now have two ends of student resource allocation, grades and goodies, representing academic and social achievements, respectively. We measure grades on the vertical axis of the graph with the familiar 0 - 4.0 grade point average (g.p.a.) scale. On the horizontal axis we have a 10-point scale running from 0, indicating abject misery, to 10, a blissful Nirvana.

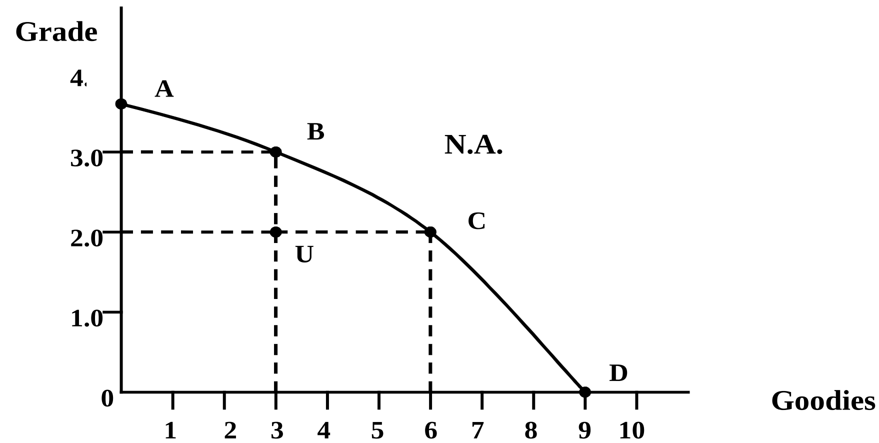


Figure 1-4
The Production Possibility Frontier: Grades and Goodies

The college student has labor, capital and human capital to allocate between these two pursuits. The ppf shown in Figure 1-4 represents the possibilities for a rather average student, both academically and socially. While the student would like to be way up in the upper right of the ppf with a 4.0 g.p.a. AND a 10 on the Nirvana index, limited resources make this combination of grades and goodies nonattainable. If this student puts all her resources into studying and producing grades, she can get a 3.5 g.p.a. If she is a total party animal devoting all her resources to reaping goodies, she can get to a 9 on the bliss scale, but scarcity dictates that she cannot have both.

If our student decides not to pursue grades only or goodies only, she can be at a point like B with a 3.0 g.p.a. and 3 points on the Nirvana index. The first 3-point increase in goodies, moving from 0 to 3, involves a reallocation of resources away from grades toward goodies, and results in a cost of .5 g.p.a. points. Note that we assume the law of increasing cost also applies in this case, so the next three point gain in the goodies index

(6 – 3) requires a full 1 point reduction in g.p.a. This is represented by a movement from point B to point C on the ppf. And the final net increase of three points on the goodies index requires a large 2 point cost in g.p.a. This is the movement from point C to point D.

It's not difficult to imagine that the law of increasing cost applies to the college student's basic economic problem. All resources are not equally productive in producing both ends. Take time on Saturday night for example. These hours would probably be the first hours transferred from studying to the pursuit of goodies. This is not the best time to study, but a good time to party, so the increase in goodies from point A to point B requires a minor sacrifice in grades. But in moving from C to D the student is out in the sunshine working on her tan during finals week, a very costly pursuit of goodies indeed.

Note that an individual student can have unemployed resources, or use her resources inefficiently. Point U represents this situation, a point inside the ppf. She is wasting her time at point U. By not wasting time, or managing it more effectively, she could have an increase in grades without any sacrifice in goodies, and vice versa. University guidance councilors attempt to provide this type of advice on resource allocation.

Saving, Capital and Growth: The Production Possibility Frontier III

An additional application of the ppf model appears in Figure 1-5. Now, instead of two consumption goods on the two axes of the graph, we have one capital good and one consumption good. In economic models, we denote capital goods with a K, most likely in honor of Karl Marx, whose 19th century book *Das Kapital* has been a bible for heterodox radical economists for over 150 years. On the horizontal, axis we denote consumption goods with a C.

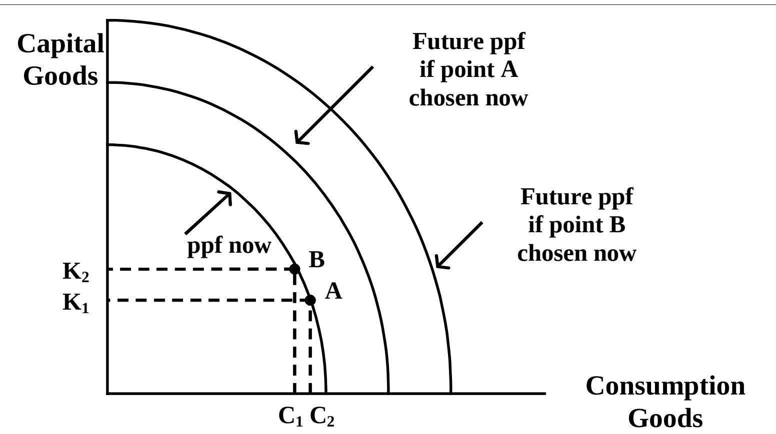


Figure 1-5
The Production Possibility Frontier: Capital Goods, Consumption Goods, and Economic Growth

Unlike our earlier models we now have three ppf lines in Figure 1-5. We have added a dynamic element to the model here by having two different time periods, now and the future. The most inward ppf, labeled ppf now, is the one representing present choices for the economy. The other two lines represent alternative future ppfs depending on the mix of capital and consumption chosen in the current period. Point B represents a choice in the current period where people consume less and save more than they do at point A. Saving is the source of capital. Because capital is a resource, and the amount of resources determines the position of the ppf, the future ppf corresponding to choosing point B now is farther out than the one corresponding to the choice of point A now. More saving produces more capital, and more capital creates economic growth.

The Circular Flow Economic Model

The production possibility frontier model represents the basic economic problem of scarcity, choice and cost. There is little in the ppf model, however, about economic systems used to address it, and especially about exchange (trade) relationships in market capitalism. The circular flow economic model of Figure 1-6 shows the interrelationship and interdependence of two types of economic entities in a market economy, households and businesses. We will spend a great deal of time examining models of both households and businesses in Chapters 4 and 6 of this book.

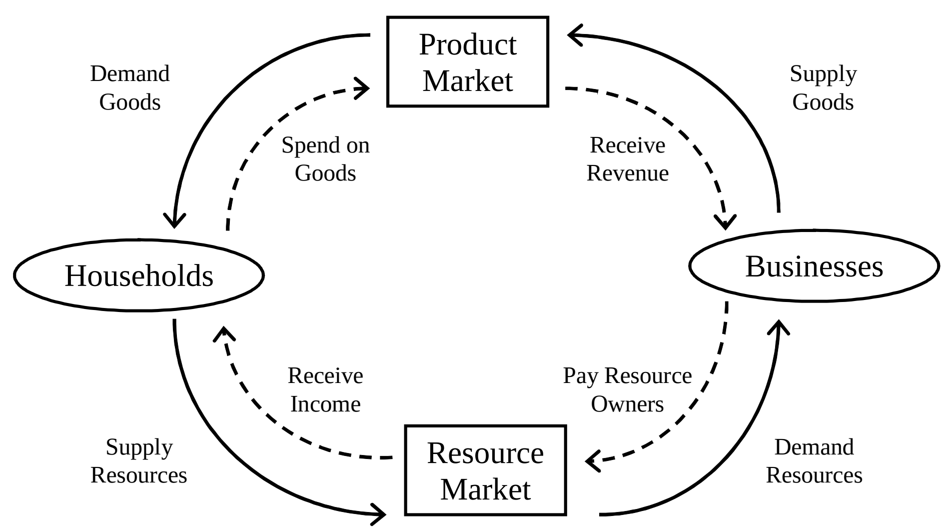


Figure 1-6
A Simple Circular Flow Economic Model

Two markets are also shown in the model, the product market where households and businesses exchange goods and services, and the resource market, where they come to buy and sell labor, human capital, capital and natural resources. In Chapter 3 we will develop a formal model of a market when we study supply and demand analysis.

Two circles of arrows show the interrelated flows of physical goods and resources and the income received from selling one, which is used to buy the other. For example, the solid outside arrow shows that households supply resources to the resource market where businesses demand them. Businesses then use those resources to produce goods that they supply to the product market. Households demand goods from the product market.

The dashed arrows in the model represent income flows. By supplying resources to the resource, market households receive income which they spend on goods in the product market. By selling goods in the product market, businesses receive revenue that they use to buy resources to produce those goods.

In addition to the summary of market relationships in an economy, the circular flow model also gives an initial insight into something that we will study in Chapters 9-13, macroeconomic instability. Suppose for example that businesses are fearful of the future and cut back on production to meet their forecasted reduced demand. If they produce less, they need fewer resources, so they hire less labor, invest less in capital equipment and buy fewer natural resources. But this is not the end of the circular flow story. If fewer resources are purchased by businesses, it means that households have lower incomes to spend. This creates a self-fulfilling prophecy on the part of businesses. Because of the actions they took in the face of lower demand they expected, it happens.

Likewise, if households see an economic rainy day coming, and cut back their spending in the product market to save for it, businesses will see the reduced spending and create the reduction in resource income that households initially forecast. This is why economic forecasters are so interested in something called consumer confidence, and why every president's state of the union message is optimistic.

The circular flow model is a very simple model, one that assumes away many real world economic relationships. But focusing on the important relationship between households and businesses gives us insights into both microeconomic and macroeconomic relationships. Adding more complexity to the model would make it more realistic, but the added reality might add more clutter than clarity. Such a model is shown in Figure 1-7. In this model we add only one more economic actor in the system and one more market, the financial market. We will deal with government in a market economy in Chapter 9 and financial markets in Chapter 11.

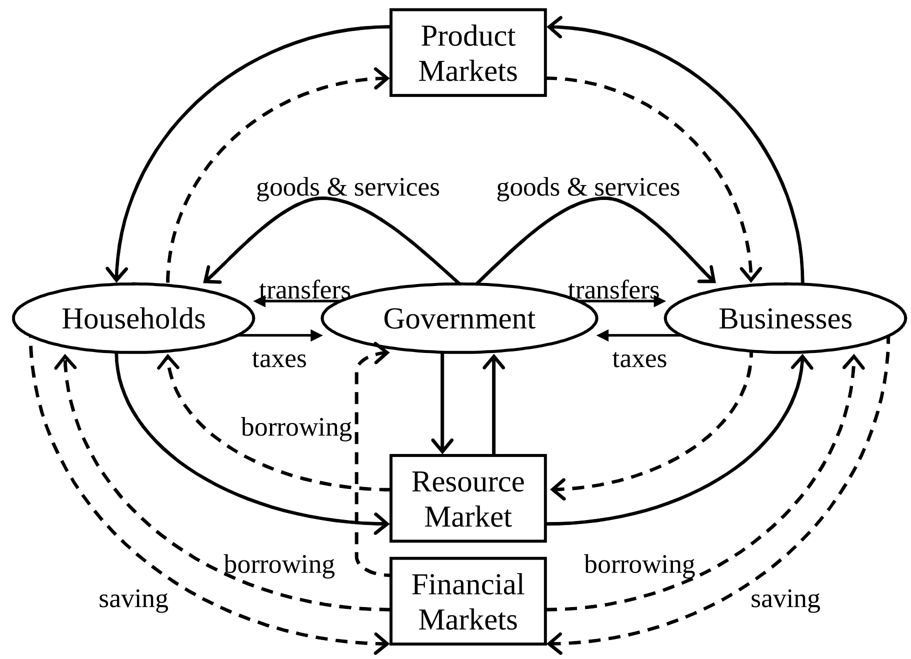


Figure 1-7
A More Complicated Circular Flow Economic Model

Note the complexity added by this move toward reality in the model. We see that government receives taxes from both businesses and households and transfers income to both. Examples of transfers to households include social security payments, unemployment compensation, food stamps and subsidized student loans. Transfer payments to businesses would include agricultural subsidies and various forms of corporate "bailouts." Government also produces goods and services as well, requiring additional arrows on the graph. Solid lines to and from the resource market box indicate that government demands resources and supplies them as well.

With the addition of a financial market, we allow for the consideration of borrowing and saving activities among households, businesses and government, with each relationship requiring a dashed line. If we had a mathematical model, each of these lines would be represented by one or more equations. Is the added benefit of reality worth the extra cost of complexity? Modeling is like love. Beauty is in the eyes of the beholder. I think you can see here, however, that by inserting more reality we gain little in the explanation of the basics of the circular flow model, coordination of economic activities through markets and the potential for macroeconomic instability because of it.

The appropriate level of simplicity and complexity in a model depends of the purposes to which you want to direct your model. In the next section we will see that the entire idea of circularity in the economic process itself can be called into question in some circumstances.

The Linear Throughput Economic Model

Someone with a more Green perspective on the basic economic problem would think that the circular flow model looks like an economic attempt to create the physical fallacy of perpetual motion. Where does the environment fit into the model? Where is the unrelenting entropic flow from valuable resources to useless waste, i.e., where is the garbage and pollution in the model?

Figure 1-8 is a Green model of the economic system. Rather that a circular flow of goods, resources, and income, we represent the economy as a linear throughput of resources to waste. And probably the most important aspect of the model is the fact that the environment is the source of all resources and the repository of all waste. A special role for and appreciation of solar energy also appears in the model.

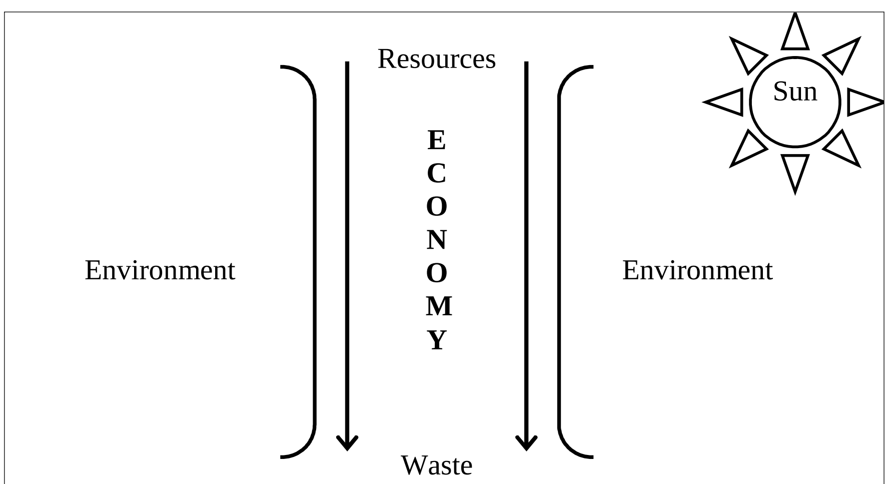


Figure 1-8
A Linear Throughput Economic Model

Is this a better model of an economy than is the circular flow model? Maybe, maybe not. The answer depends on the purposes you want your model to address. Each is appropriate in its own way, and each has its own unique abstraction from reality in order to understand reality better.

Chapter 2: Specialization, Exchange and Economic Well Being

Specialization, Exchange and Economic Well Being: A Nontechnical Introduction

The circular flow model we studied in the last chapter captured two of the main characteristics of market economies, interdependence and coordination through markets. But economies need not have interdependence and market coordination. Markets are human inventions and we could choose not to have them. We all could be self-sufficient. We could build our own houses, make our own clothes, grow our own food and provide our own transportation, health care and entertainment.

Self-sufficiency has a certain romantic ring to many normal people. We admire master gardeners, people who build their own houses and repair their own cars. But imagine a life of total self-sufficiency. I don't know about you, but my life of self-sufficiency would be in the terms of Thomas Hobbes, "nasty, brutish and short." Economists abhor self-sufficiency, because people have different abilities, and if they specialize in what they do best and trade the fruits of that specialization with others, we have a much better chance of increasing the value of satisfied human wants. In this chapter we will carefully and formally define "best" in terms of comparative advantage.

Specialization brings out fear in some normal people, because when we specialize we become dependent on others. That can be scary. Can we rely on others to provide the things we don't specialize in? Won't they most likely look after their own interests, rather than ours? But to an economist specialization is not scary at all, because self-interest coordinated by markets will guarantee availability of what we want. Those other people specializing in what they do best will be looking to us to provide what they don't produce. Markets will also provide information to help each of us decide what we should specialize in, and if this information is correct, and if self-interest guides us to specialize in activities that have the highest return or value to us, we will be making the most of the resources we own. When we maximize the value of the output from our resources, we create what we call efficiency in economics.

Absolute and Comparative Advantage

Some people are simply more productive than others are. If someone can produce more with a given amount of resources, or produce the same amount of something with fewer resources we say that person has an absolute advantage. For example, if you can pick more huckleberries than I can in a day (produce more with the same resource, one day of labor) you have an absolute advantage in huckleberry picking compared to me. The definition of absolute advantage is pretty straightforward.

To an economist, however, it's something called comparative advantage, not absolute advantage that is the more important concept. And the definition of comparative advantage is a little trickier. In fact, a wiseguy once asked Nobel Laureate economist Robert Solow to name one economic concept that was neither trivial nor immediately transparent. Professor Solow said, "the law of comparative advantage."

We define comparative advantage in terms of opportunity cost, not resource cost. Complete understanding of this difference must await the detailed numerical examples of the next three sections, but we can make a first attempt here. Let's look again at the huckleberry picking example. You may be able to pick more huckleberries in a day, but you also might be able to do more car repairs in a day as well. With the right value of car repairs to huckleberries the opportunity cost of you picking berries is much higher than the opportunity cost of me picking berries. Even though you have an absolute advantage in both huckleberry picking and car repairs, I likely have a comparative advantage in huckleberry picking. My opportunity cost of huckleberry picking is lower than yours, because I sacrifice fewer car repairs per gallon of berries. I am a very poor mechanic. I'm least-worst in huckleberry picking.

The notion of comparative advantage is related to the idea of specialization, exchange, and the maximization of value we discussed above. For if we all specialize in the activity in which we have a comparative advantage, and then trade with each other, we will be better off than if we try to be self-sufficient in both activities. The examples in the next sections will help clarify these important relationships.

Simple comparative advantage examples in economics often involve two individuals stranded somewhere with two goods needed for survival. Robinson Crusoe and Friday on a tropical island, or Gilligan and Marianne on Gilligan's Island are representative examples. We begin a series of comparative advantage stories in the next section by honoring this tradition of survival.

Ole and Lars at Frozen Fjord

Ole and Lars both have huts on Frozen Fjord in northern Norway where they go to fish and get away from the congestion and noise of the city. One late fall day a terrible snowstorm and cold spell hits Frozen Fjord unexpectedly. The snow gets so deep they cannot get out. They're stranded. It's cold and getting colder. Their thoughts turn to survival, as the relentless storm shows no signs of abating.

Ole and Lars need two goods to keep them alive until they are rescued, fish to eat and firewood to keep from freezing. Ole and Lars are not equally productive in catching fish and gathering firewood. The production possibilities for Ole and Lars are shown in Figure 2-1. In the rapidly shortening daylight hours at Frozen Fjord Ole can catch 10 pounds of fish OR gather 5 bundles of firewood. In the same amount of time Lars can catch 4 pounds of fish OR gather 4 bundles of firewood.

	Ole	Lars
Pounds of Fish	10	4
Bundles of Firewood	5	4

Figure 2-1
Output from One Day of Work by Ole and Lars at Frozen Fjord

The word OR is very important here. It means that if Ole devotes all his daylight hours to fishing he can catch 10 pounds of fish. If he does no fishing, and devotes all his time to gathering firewood, he can gather 5 bundles. The production possibilities are not such that he can catch 10 pounds of fish AND gather 5 bundles of firewood. This proviso applies to Lars, as well. If Lars devotes all his daylight hours to fishing he can catch 4 pounds of fish. If Lars does no fishing, and devotes all his time to gathering firewood, he can gather 4 bundles. Each of our Norwegians can also be less extreme and can allocate some time to each activity, but the more firewood they gather the less fish they catch, and vice versa. They face the basic economic problem of scarcity.

We can illustrate Ole and Lars'economic problem with production possibility frontiers such as the ones shown in Figure 2-2. For simplicity we assume that the law of increasing cost does not apply in these models. It's not an important issue for the task at hand, and by assuming constant cost we can make things easier by constructing straight line ppfs.

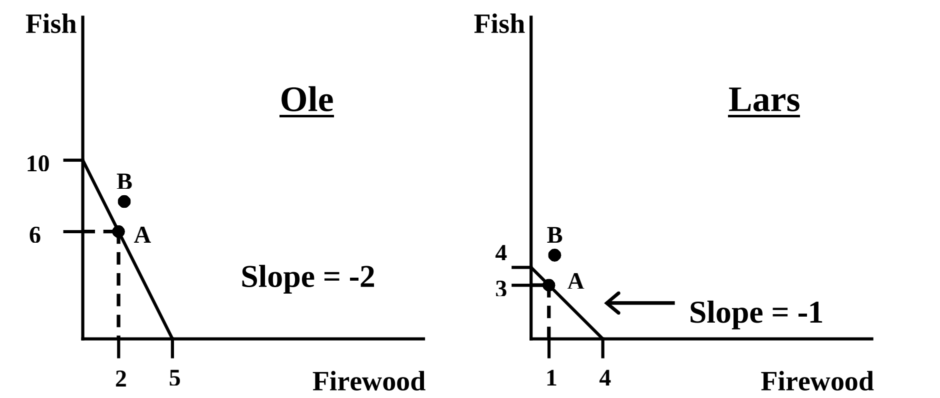


Figure 2-2
Production Possibility Frontiers for Ole and Lars at Frozen Fjord

Note that the intersection of the ppf with the fish and firewood axes comes at the maximum output levels given in Figure 2-2, 10 and 4 pounds of fish, and 5 and 4 bundles of wood for Ole and Lars, respectively. With the resource at their command, the number of daylight hours, Ole can catch more fish and gather more firewood than Lars. Ole has an absolute advantage in both fish and firewood. Let's assume that left to their own independent production under conditions of self-sufficiency, each would be at their own point A on their ppfs. Ole catches 6 pounds of fish and gathers 2 bundles of firewood. Lars is considerably hungrier, with 3 pounds of fish, and colder, at 1 bundle of firewood.

Poor Lars! It looks like he has nothing to offer in the Frozen Fjord survival economy, because Ole is so much more productive. Both might be tempted to go their separate ways, each fishing and gathering firewood and returning to their huts to await a rescue. But this independence and self-sufficiency would be economically foolish for both of them. Let's see why.

Self-sufficiency is foolish because it ignores the principle of comparative advantage and the gains from specialization and exchange. Remember that comparative advantage is defined in terms of the opportunity cost of the two goods. And here we see something different than we saw for absolute advantage. While Ole has an absolute advantage in both goods, he has a comparative advantage in only one, fish. The opportunity cost of each good for both Ole and Lars appear in Figure 2-3.

	Ole	Lars
Cost of Pound of Fish	½ Bundle of Wood	1 Bundle of Wood
Cost of Bundle of Firewood	2 Pounds of Fish	1 Pound of Fish

Figure 2-3
Opportunity Cost of Fish and Firewood at Frozen Fjord

Ole can substitute firewood for fish at a ratio of 5 bundles of firewood for every 10 pounds of fish. On a per unit basis, the cost of a pound of fish is 1/2 bundle of firewood. Once we calculate the opportunity cost of one good in these simple models, the other opportunity cost is always the reciprocal. Ole has to sacrifice 10 pounds of fish for 5 bundles of firewood. On a per unit basis the opportunity cost of one bundle of firewood is 2 pounds of fish. Likewise, Lars' substitution possibilities are 4 pounds of fish for 4 bundles of firewood and vice versa. The cost of a pound of fish to Lars is one bundle of firewood. The cost of a bundle of firewood is 1 pound of fish.

Note that the opportunity cost of fish is lower for Ole (1/2 < 1), but the opportunity cost of firewood is lower for Lars (1 < 2). Lars has an absolute advantage in neither good, but he has a comparative advantage in firewood. Ole has a comparative advantage in fish. Recognizing the differences in comparative advantage and the implications of these differences can make for a better sojourn for both Ole and Lars at Frozen Fjord.

Self-Sufficiency	Ole	Lars	Total
Pounds of Fish	**6**	**3**	**9**
Bundles of Firewood	**2**	**1**	**3**
Specialization			
Pounds of Fish	**10**	**0**	**10**
Bundles of Firewood	**0**	**4**	**4**

Figure 2-4
Total Production at Frozen Fjord with Self-Sufficiency or Specialization

Figure 2-4 shows the total output of the Frozen Fjord economy under two different economic organizations, self sufficiency and specialization. Under self-sufficiency Ole and Lars each produce both goods. Under specialization they each produce just one. Under self-sufficiency the total output of fish at Frozen Fjord is 9 pounds, 6 by Ole and 3 by Lars. Likewise, the total output of firewood is 3 bundles, 2 by Ole and 1 by Lars. But if they specialize in the good they are the best at producing, with best defined as comparative advantage, the total output is greater. If Ole devotes all his time to fishing he can catch 10 pounds of fish, more that the combined total of 9 under self-sufficiency. If Lars devotes all his time to gathering firewood, he can gather 4 bundles, more than the combined total of 3 under self-sufficiency.

Specialization by itself, however, is not sufficient for improvement in economic well being at Frozen Fjord, or anyplace else for that matter. For without trade with each other, Ole would freeze to death with a full stomach and Lars would starve. Both need to overcome their strong Norwegian independence and exchange goods with each other. One possible result is that they share the increase in the total output equally. Each gets ½ pound more fish and ½ bundle more firewood than they had under self-sufficiency. Instead of 6 pounds of fish and 2 bundles of firewood Ole could consume 6.5 pounds of fish and 2.5 bundles of firewood. Lars could consume the remainder 3.5 pounds of fish and 1.5 bundles of firewood. To get to this result, Ole produces only fish and trades it to Lars for firewood. Lars receives fish for firewood. Both are made better off from specialization and exchange.

We can also see the effects of this specialization and exchange in the ppf models of Figure 2-2. Ole and Lars can reach the point B in their ppf model. After specialization and exchange, they can have more of both goods and get into the previously nonattainable region of their ppfs. Recognizing their respective comparative advantage, and specializing in it and trading with the other has beaten back the problem of scarcity a little. As they have more of both goods, and more is preferred to less, both are better off through specialization and exchange. Now if the rescuers would just find them, life would be even better.

Harry and Joe at Priest Lake

In the example of Ole and Lars at Frozen Fjord, we assumed each Norwegian had a given amount of resources, daylight hours in this case. We then saw the amount of each good each of them could produce with this given amount of resources. Another way to set up a simple comparative advantage example is to fix the amount of the good to be produced, and specify the productivity of each person in terms of the amount of resources it takes to produce the good. This is the approach used here for the story of Harry and Joe at Priest Lake. Be sure to compare the initial information in both cases so you are sure you understand these different ways of expressing production information, output per amount of input, and input for a given output.

Harry and Joe like to go to Priest Lake in North Idaho. Their favorite activity is sitting on the beach in the sun drinking beer and reading trashy detective novels. But they can't spend all their time at the lake engaging in their favorite activity, because their wives expect them to bring home some Mackinaw trout and some morel mushrooms. Once they each acquire the required amount they can hit the beach.

Production information for Mackinaw fishing and morel gathering appears in Figure 2-5. As you can see in Figure 2-5, Joe is much more productive than Harry is in both activities. It takes Joe one hour to gather a wife's required morels, while it takes Harry a full 8 hours. Harry has less of an absolute disadvantage in fishing, but it still takes him 4 hours to catch a wife's requirement, while it takes Joe only 3. Joe has an absolute advantage in both activities. If Joe and Harry are normal people, they might just split up in the parking lot and agree to meet later on the beach. Certainly Joe wouldn't want anything to do with the hapless Harry. He would just put in his 4 hours and hit the beach to read and drink beer.

	Harry	Joe
Hours for Morels	**8**	**1**
Hours for MacKinaw	**4**	**3**

Figure 2-5
Hours Required for One Person's Daily Requirement of Morels and MacKinaw at Priest Lake

But just as the case with Ole and Lars at Frozen Fjord, this independent behavior would be a mistake. Even though Joe has an absolute advantage in both gathering mushrooms and fishing, Harry has a comparative advantage in fishing. Again, to figure out the comparative advantage of each lake visitor, we need to look at the opportunity cost of the two activities to Harry and Joe. We see in Figure 2-6 that Joe has a comparative advantage is gathering morels, because he has to sacrifice only 1/3 of a wife's requirement of Mackinaw trout to get them, where Harry has to sacrifice 2.

	Harry	Joe
Cost of One Wife's Requirement of Morels	2 Requirements Of MacKinaw	1/3 Requirement of MacKinaw
Cost of One Wife's Requirement of MacKinaw	1/2 Requirement of Morels	3 Requirements of Morels

Figure 2-6
Opportunity Cost of Morels and MacKinaw at Priest Lake

But in fishing, it's another story. Harry must sacrifice only ½ wife's requirement of morels to obtain one wife's requirement of Mackinaw, much less than Joe's opportunity cost of 3.

Again, we show the gains from specialization and exchange as we did in the case of Ole and Lars. This information appears in Figure 2-7. Under conditions of independent self-sufficiency, Harry spends 12 hours getting a wife's requirement of each good, 8 gathering morels and 4 hours fishing. Joe gathers morels for 1 hour and fishes for 3 to obtain his wife's requirement. Harry and Joe sacrifice 16 hours of total beach time to produce their goods.

But look what happens if Harry and Joe each specialize in the activity in which they have a comparative advantage. Each will produce two wife's requirements of the good in which they have a comparative advantage, and trade one to the other for the other good. Because Harry has a comparative advantage in fishing for Mackinaw, he devotes 8 hours to the task, 4 hours for each of the wife's requirements of fish. Joe specializes in gathering morels, using one hour for one wife's requirement and one more for an additional wife's requirement, a total of two hours gathering

Self-Sufficiency	Harry	Joe	Total Hours
Hours Gathering Morels	**8**	**1**	**9**
Hours Catching MacKinaw	**4**	**3**	**7**
Total Hours	**12**	**4**	**16**
Specialization			
Hours Gathering Morels	**0**	**2**	**2**
Hours Catching MacKinaw	**8**	**0**	**8**
Total Hours	**8**	**2**	**10**

Figure 2-7
Hours Worked at Priest Lake Under Self-Sufficiency or Specialization

morels. Look what happens to the total time required to make their wives happy. The total time away from the beach is now only 10 hours instead of 16, 8 for Harry and 2 for Joe. Note that each party shares in the gains from specialization and trade, this time in the form of fewer hours to produce a required amount of the two goods. Harry saves 4 hours and Joe saves 1. Specialization and exchange works again!

An International Example of Comparative Advantage, Specialization and Gains from Trade

Economists are not just "free-traders" within an economy, but across the borders of economies as well. They see the benefits from specialization and trade in an international context, and the reason for this advocacy of free trade is again comparative advantage. Let's look at an example of a two country model. Assume the world economy consists of only two countries and two goods (Ain't modeling great!) The two goods are clothing and cars, and the two countries are Country A and Country B. The production information for the two countries appears in Figure 2-8. With a unit of resources, Country a can produce 20 cars OR 2,000 units of clothing. With the same unit of resources, Country B can produce only 4 cars OR 1,200 units of clothing. By now you should begin to be able to see that again one entity has an absolute advantage in producing both goods.

	Country A	Country B
Number of Cars	20	4
Units of Clothing	2,000	1,200

Figure 2-8
Production of Cars and Clothing with a Unit of Resources in Countries A and B

That country is Country A, who can produce more cars or more clothing with a unit of resources than Country B can.

But as in our previous examples, Country A does not have a comparative advantage in both goods, and because of this gains from specialization and exchange exist. As seen in Figure 2-9, Country A has a lower opportunity cost of producing cars

	Country A	Country B
Cost of a Car	100 Units of Clothing	300 Units of Clothing
Cost of a Unit of Clothing	1/100 Car	1/300 Car

Figure 2-9
Opportunity Cost of Clothing and Cars in Countries A and B

than does Country B. Country A has to sacrifice only 100 units of clothing per car, where country B has to sacrifice 300 units of clothing. Country A has a comparative advantage in producing cars. The relative magnitudes of the opportunity costs of clothing, however, are reversed, 1/300 of a car for Country B and 1/100 of a car for Country A. Country B has a comparative advantage in producing clothing.

The production information in Figure 2-8 is given in terms of a unit of resources. Suppose we endow each country with 1,000 units of resources. We can now conveniently represent their production possibilities with production possibility frontiers. These appear in Figure 2-10. For simplicity, we again assume constant opportunity cost, which allows us to draw straight line ppfs. The black lines represent the ppfs in Figure 2-10. Ignore the blue lines for a moment.

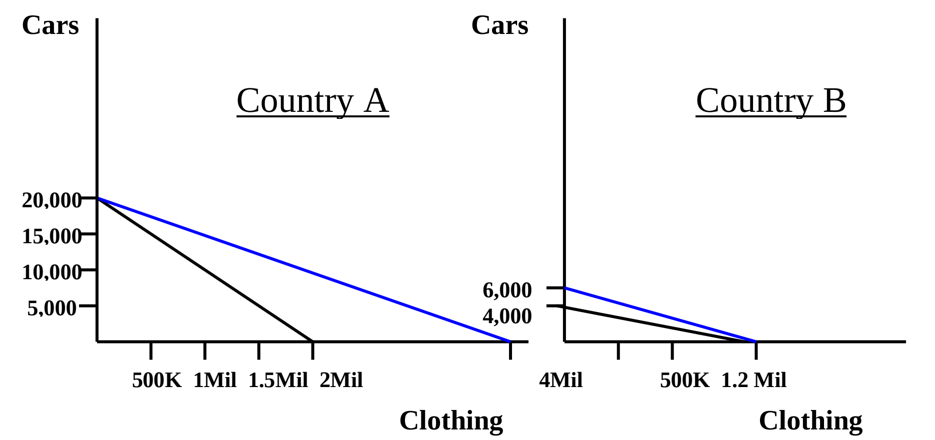


Figure 2-10
Production Possibility Frontiers and Trading Possibility Curves for Countries A and B

The intercepts on the axes of the ppfs are pretty straightforward. With one unit of resources Country A can produce 20 cars, so with 1,000 units of resources it can produce a maximum of 20,000 cars. With one unit of resources Country A can produce 2,000 units of clothing, so with 1,000 units of resources it can produce a maximum of 2 million units (2,000 X 1,000). By transferring resources from the production of one good to the production of the other, it can produce any combination of cars and clothing along its ppf. The ppf for Country B is constructed in a similar fashion. Under conditions of self-sufficiency, each country is limited by the constraint of its limited resources and must be on or interior to its ppf. It would like to be in the nonattainable region, because more is preferred to less. But it can't be there.

As is always the case, the slopes of the ppfs indicate the opportunity costs of one good in terms of the other. The slope of Country A's ppf is – 1/100, 1 car for every 100 units of clothing. Country B has to sacrifice

1/300 of a car for every unit of clothing, so the slope of its ppf is – 1/300. Suppose there is a market for cars and clothing in our world economy, and suppose the "terms of trade" in this market are 200 units of clothing per car, i.e., by trading 200 units of clothing on the market you could receive one car. We represent these terms of trade in Figure 2-10 by the blue trading possibility curves (tpc). Note that the slope of the blue lines in Figure 2-10 is the same for each country, -1/200, which reflects the terms of trade or trading ratio.

We could draw the tpc lines intersecting either of the intercepts on the axes of the ppfs, but we choose to connect them with the ppf on the axis of the good in which the country has a comparative advantage. Because Country A has a comparative advantage in producing cars, we start the tpc line for Country A at 20,000 cars and run it downward at a slope of 1 car per 200 units of clothing. This line intersects the clothing axis for Country A at 4 million units of clothing (20,000/4 million is the same ratio as 1/200.). Likewise we start the tpc line for Country B on the clothing axis of its model at 1.2 million units of clothing, and run it upward again at a ratio of 1 car for every 200 units of clothing. The tpc for Country B thus intersects the cars axis at 6,000 cars (6,000/1.2 million is the same ratio as 1/200.).

Note the implications of the ability to trade on the open market. Country A can specialize in cars, i.e., produce only cars, and trade cars on the world market along the blue tpc in Figure 2-10. Under self-sufficiency Country A could get only 100 units of clothing by transferring resources to clothing production from each car it stopped producing. Now, with the ability to trade at a ratio of 1 car for 200 units of clothing, it makes much more sense to trade for cars rather than produce them. Likewise, under self-sufficiency, along its ppf, Country B would have to free up resources by producing 300 fewer units of clothing for each car its people wanted to consume. Now, with the possibility of trade, Country B can specialize in clothing and buy cars on the open market at a much more favorable cost of 200 units of clothing per car.

The models of Figure 2-10 illustrate that specialization and trade allows a country to attain a point in the nonattainable region of its ppf. In other words, by specialization and exchange it is possible to move to points farther to the NE in the model and have more of both goods. Specialization and trade is surely a wonderful thing.

The 1/200 terms of trade ratio is just assumed here. Any ratio between the slopes of the ppfs (between –1/100 and –1/300) will work just as well. If the terms of trade lie outside these bounds it would not pay for the countries to specialize. It is also important to note that it is not specialization per se that creates the gains in economic well being. Rather each country has to specialize in the good in which it has a comparative advantage. To see this in the model, run the tpc line through the ppf intercept on the axis of the good in which the country does NOT have a comparative advantage. Note that this would run the tpc interior to the ppf. This would not be a very smart thing for a country to do!

Chapter 3: Demand, Supply and Elasticity

Introduction

In order to plow the traditional fields of economics you need a workhorse. The supply and demand model is just such an animal. No other model is more important to understanding basic economics. In the late 19^{th} century, Oscar Wilde said that you could teach a parrot economics. All you had to do was teach it to say "supply and demand."

The supply and demand model has special importance in economics, because it's a model of a market. We noted in Chapter 1 that economists differ from normal people in their fascination with markets. In the last Chapter, we began to see why. Markets allow trade, and trade allows specialization, and specialization means higher productivity and higher productivity means higher economic well being.

In this chapter, we will plow a little wider and deeper into the functioning of a market economy. We will see that markets create prices and that prices are very important pieces of information in an economy. Prices act as signals that guide resource sellers and buyers and product sellers and buyers. When something happens in the world, buyers and sellers react, and prices change, and buyers and sellers react again. For example, when tensions rise in the Mideast, the price of oil goes up. As we will see in this chapter, this is no accident. Buyers and sellers see a higher probability of an interruption in Mideast oil supply, so they sell less and buy more today in anticipation of higher prices in the future. These actions cause the price of oil to rise, and the world knows that oil is more scarce, even though not a drop has disappeared.

The Determinants of Demand and the Determinant of Quantity Demanded

We begin our discussion of the supply and demand model on the buying side. We focus initially on buyers because all of us buy many goods and services. Often we sell only one, our labor. The demand side of the supply and demand model is about buying behavior. But before we examine the mechanics of the buying side of our model of a market, we need to learn some confusing, yet important, terminology.

Few pieces of economic jargon confuse the beginning student of economics more than the terms demand and quantity demanded. Any normal person would conclude that these terms must mean pretty much the same thing. Unfortunately this is not the case, and we need to learn the real distinctions embodied in this terminology if we are going to be good supply and demand practitioners, if we are going to be good at using the supply and demand model.

These Affect What and How Much Buyers Want to Buy

- Price — Determinant of Quantity Demanded
- Population
- Tastes and Preferences
- Income and Wealth
- Prices of Related Goods:
  - Substitutes
  - Complements
- Expectation of Price in the Future

(Population through Expectation of Price in the Future: Determinants of Demand)

Figure 3-1
Determinants of Demand and Quantity Demanded

Look at the list of terms inside the brackets in Figure 3-1. These are all things that affect what and how much of a good people want to buy. Price is the first term on the list, and we will single it out later with a special honor, and give it its own axis on a graph. But for now just think about how price affects your buying behavior. Your behavior is probably like most other rational buyers. If the price of a good is lower, you are more likely to buy the good, and as the price falls you will buy more. If the price rises, you will want to buy less.

Marketing professionals point out that some goods do not follow this principle. With “prestige pricing,” you might get someone to buy a more of a good because its price is high. Hall of Fame heterodox economist Thorsten Veblen called these kinds of goods “snob goods.” We’ll ignore these types of goods in this chapter.

Note that just to the right of the term “Price” in the list, we see the words “Determinant of Quantity Demanded.” Price is the only term on the list that receives this distinction. We’ll soon see that this is related to receiving its own axis on a demand graph. Price is not like the other terms in the list. It is a determinant of quantity demanded not a determinant of demand. I warned you that this has the potential to be confusing, but don’t blame me, economics has been this way for about a century.

The remaining terms on the list are defined to be “Determinants of Demand.” The second term, population, has a straightforward relationship to how much buyers want to buy. Buyers are people, and if we have more people we have more buyers buying. I always like to think about the local Chamber of Commerce when I think of population and buying. Members of the Chamber of Commerce are never against growth in a community, because they sell things, and a higher population means more buyers to buy what they are selling. Population is a determinant of demand (but not of quantity demanded).

Tastes and preferences is a determinant of demand. If you like Brussels sprouts, you might buy some. If you don't like them, you likely won't buy any. Some people like Fords, some like Chevrolets, some really don't like American cars. This determinant of demand says that how much you like or dislike a good will have an effect on how much of it you buy. Tastes and preferences are determinants of demand (but not of quantity demanded).

If you don't have any income or wealth, you are probably not going to buy much of anything. Income and wealth are not the same. Income is what we economists call a flow. It has units of time associated with it. For example, if someone asks about your income, you will tell them that you earn, say, about $15,000 a year, or $300 a week, or some amount per hour. Note the time dimension in each of these cases. Flows in economics are just like flows in hydrology. If you want to know how much water is flowing by you in a stream, you want to know how many cubic feet per second are flowing by. Wealth doesn't have a time dimension. It's just a hunk of purchasing power. You can draw down or cash out some of your wealth and buy something with it.

For most of us, the more income and wealth we have, the more we will buy. Economists call goods that we buy more of when our income rises normal goods. Not all goods are normal goods, however. You might buy less Ramen noodles and K-mart basketball shoes when your income rises. We call these inferior goods. In any case, changes in income and wealth affect what and how much of something buyers want to buy. They are determinants of demand (but not of quantity demanded).

The prices of two kinds of related goods affect how much buyers want to buy. By related goods we mean goods that are related to the good in question. For example, suppose we are considering the cola drink, Coke. One type of related good is a substitute good. Pepsi is a substitute for Coke. If the price of Pepsi changes, it will affect how much Coke buyers will want to buy. If the price of a substitute good rises, people want to buy more of the good in question. If the price of a substitute falls, people will want to buy more of the substitute and less of the good in question, e.g., if the price of Pepsi falls, cola buyers will want to buy less Coke.

The other kind of related good is a complement. This is not a compliment with an "i" but complement with an "e." Complements go together when you consume them. Some think Beer and Pizza are complements. Hot dogs and hot dog buns, shoes and socks, canoes and canoe paddles, skis and ski poles, backpacks and sleeping bags, are all pairs of complementary goods. If the price of a complement falls, people want to buy more of the good in question. If the good in question is hot dog buns, a fall in the price of hot dogs will cause people to want to buy more buns. A rise in the price of hot dogs has just the opposite effect. Because a rise in the price of hot dogs leads people to buy fewer hot dogs, they will also buy fewer hot dog buns to go with the hot dogs they buy. Prices of related goods are determinants of demand (but not of quantity demanded).

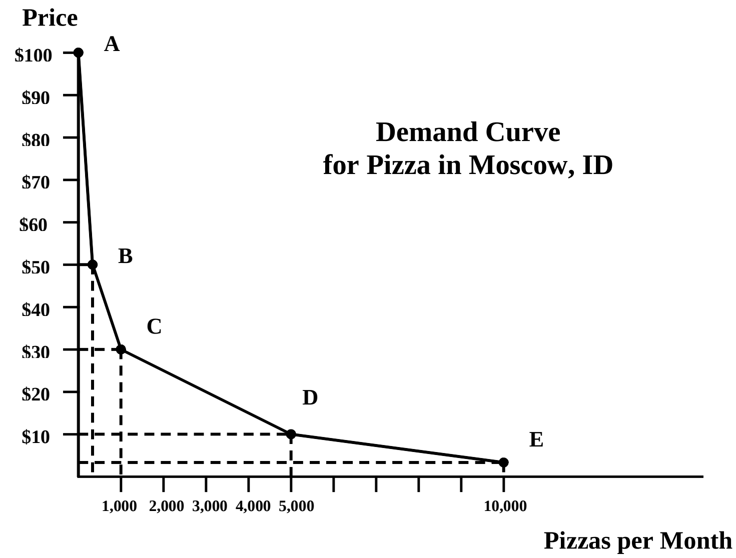
Finally, expectations of prices in the future affect how much of something people want to buy today. If people expect the price to rise, they will want to buy before the price rises. They will want to buy more today. If people expect the price of a good will fall in the future, some will postpone buying today, preferring to wait until the price falls in the future. Expectations of prices in the future is a determinant of demand (but not of quantity demanded).

Demand and Quantity Demanded: A More Technical Discussion

While the distinction between demand and quantity demanded is confusing at the outset, the problem is not insurmountable. The terminology merely refers to definitions. It would have been nice if the founders of modern economics had used terms that were a little more different, but alas, that is not the case. Just remember the definition. When we consider the effect of price on how much people want to buy, we call that amount quantity demanded. When we consider the effect of all the other items in the list in Figure 3-1, we say these items affect demand.

Let's move toward a model of the buyer side of a market and we will see why this distinction between demand and quantity demanded is so important. A demand schedule and the numerically illustrated demand curve corresponding to it appear in Figure 3-2. First look at the hypothetical demand schedule for pizza for

Moscow in the lower part of the figure. A demand schedule is simply a two-column table showing alternative prices for a good, in this case pizza, and the amount buyers in the market wish to buy at each particular price. I just made up these numbers, because many students find numerical illustrations helpful when first confronted with a demand model. We will dispense with the numerical illustrations and become more abstract shortly.



Demand Schedule for Pizza in Moscow, ID

Point on Demand Curve	Price	Quantity Demanded per Month
A	$100	0
B	50	500
C	30	1,000
D	10	5,000
E	5	10,000

Figure 3-2
A Demand Schedule and a Numerically Illustrated Demand Curve for Pizza

Note that in this demand schedule we have information only on price and quantity demanded. We have no information on tastes and preferences, population, income, prices of related goods and expectations. In a demand schedule we wish to isolate the effect of price on quantity demanded. For the demand schedule to have meaning, we can't allow these determinants of demand to change at the same time price is changing. If, say, income was changing at the same time price changed, we would not know if the change in quantity demanded was from price or income. We could make the same case for changes in the other determinants of demand. So we exercise a little modeling power to make our task easier. When we construct this numerical demand schedule we assume that the determinants of demand are held constant. Only price is allowed to change. Remember, we can hold ALL the determinants of demand constant and still consider a change in how much people want to buy, because price is not a determinant of demand, but rather a determinant of quantity demanded.

Note that we have assumed an inverse relationship between price and quantity demanded in our numerical demand schedule. In the Moscow pizza market, at a price of $100 per pizza, no one wants to buy a single pizza. At a price of $50, the quantity demanded is 500 pizzas a month. At $30 it's 1,000 per month, and so on. The lower the price, the higher the quantity of pizza demanded, assuming the determinants of demand are held constant.

If all the determinants of demand are held constant to construct our demand schedule, demand must not be changing. How can something change if all its determinants are held constant? It can't. But as price falls in our demand schedule, people want to buy more pizza. Isn't this an increase in demand? NO! This is a change in quantity demanded.

In the upper part of Figure 3-2, we simply plot the price and quantity combinations that appear in the demand schedule. At point A in the graph, the price is $100 and the quantity demanded is 0. At point B in the graph the price is assumed to be $50 and the corresponding quantity demanded is 500 pizzas per month. Each price and quantity combination in the demand schedule has a corresponding point on the price-quantity graph. Once we have the 5 points plotted in the graph, we can connect the dots. When we connect the dots representing a demand schedule we have what we call a demand curve. A demand curve is a graphical relationship on a price-quantity graph representing the relationship between price and quantity demanded, holding all the determinants of demand constant.

The information in our demand schedule is a little lumpy, e.g., the price falls from $100 to $50 to $30 to $10 to $5. A lumpy demand schedule leads to a kinky demand curve. If we had more price and quantity combinations in our demand schedule, we would have a smoother looking demand curve.

A demand curve represents one and only one level of demand. It has to, because along a demand curve all the determinants of demand are held constant. But along a demand curve the quantity demanded can change as price changes. In this sense, constant demand can represent many different amounts of quantity demanded, depending on what the price is. But what is the price in our model of this market? We don't know yet. We must introduce supply to the model first.

But before we introduce supply to our model of a market, we need to make our demand model a little more abstract, and we need to address the issue of what happens in the model when the constant determinants of demand are no longer constant, i.e., we need to develop a way to incorporate changes in demand into our model.

A More Abstract Demand Model

While economics students find numbers helpful when first confronted with a demand curve, numerical illustration becomes more and more cumbersome as their modeling ability increases. A little more abstraction makes the use of the demand model much easier. A more abstract demand model appears in Figure 3-3. Note that the price and quantity numbers have been replaced with the symbols P_1, P_2, and P_3, and Q_1, Q_2 and Q_3. We have preserved the most important characteristic of the demand curve. It slopes downward, indicating the inverse relationship between price and quantity demanded. P rises, Q falls. P falls, Q rises.

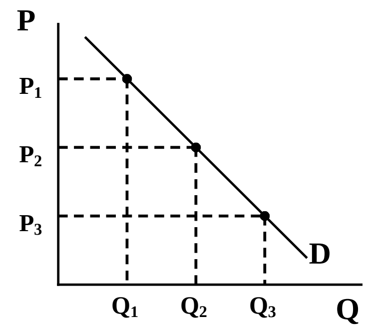


Figure 3-3
An Abstract Model of a Demand Curve

With a more abstract demand curve, we are not very interested in the actual shape of it. Straight-line demand curves like the one in Figure 3-3 are easy to draw, but curved ones will work just as well, AS LONG AS THEY SLOPE DOWNWARD!

Modeling Changes in Demand

The inverse relationship between price and quantity demanded is a very important concept, but we need additional modeling abilities to make our demand model useful. We will often use our demand model to examine more than the effect of a price change on quantity demanded, with the determinants of demand held constant. We need to know how to handle in our model the effects of changes in a determinant of demand. What happens when a determinant of demand is not held constant?

We see the demand model illustration of an increase in demand in Figure 3-4. The

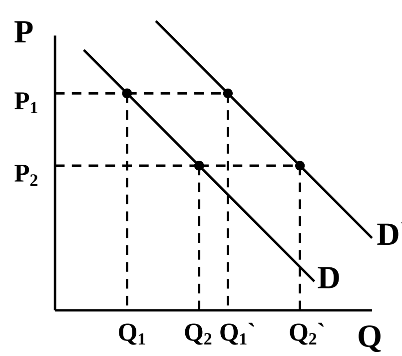


Figure 3-4
Modeling an Increase in Demand

demand curve labeled D is the original demand curve. At a price of P_1 buyers want to buy Q_1 units of the good. At the lower price P_2, they want to buy more, Q_2. Now suppose population grows in the market area. In our pizza example, suppose the population of Moscow increases. Now, because there are more buyers in the market, demand increases. At a price of P_1, buyers now want to buy Q_1' not Q_1. At the price P_2, before the population increase buyers wanted to buy Q_2, but with a higher population the quantity demanded at this price is Q_2'.

When learning to understand how to model a change in demand, start first with a particular price and ask the question, after a change in the determinant of demand, will buyers want to buy more or less of the good? Then go to another price and ask the same question, then to another price, then to another, etc. If you ask this question of enough prices, you can connect the dots of the new price-quantity combinations and you will have a new demand curve, and you will find that an increase in demand is represented by a shift to the right of the demand curve.

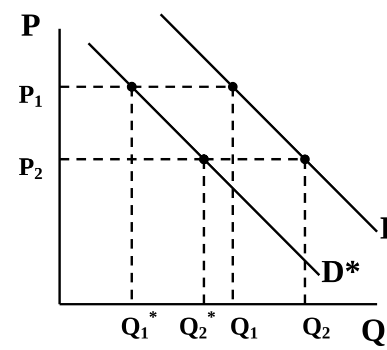


Figure 3-5
Modeling a Decrease in Demand

The demand model illustration of a decrease in demand appears in Figure 3-5. Here the new demand curve lies to the left of the original demand curve. Suppose population decreases. At a price of P_1, before the population decrease quantity demanded was Q_1. But now with fewer people in the area the quantity demand at a price of P_1 falls to Q_1*. Likewise, if the price were P_2 before the population decrease, quantity demanded would have been Q_2. But with a lower population the quantity demanded at the price P_2 is Q_2*. It doesn't really matter how much you shift the demand curve to illustrate a decrease in demand. It is important, however, that you shift the demand curve to the left.

Modeling Changes in Each Determinant of Demand

In the last section we saw how to show an increase and decrease in demand with the demand model. We shift the demand curve to the right for an increase in demand and to the left for a decrease in demand. But how do changes in each specific determinant of demand affect the demand curve? We address this in Figure 3-6.

We have six separate graphs in Figure 3-6, each one corresponding to one of the six determinants of demand. A + sign indicates that the determinant has increased, and a – sign indicates the determinant has decreased. Arrows associated with the + and – signs indicate the corresponding shift in the demand curve. For example, panel (a) in Figure 3-6 repeats the example of the last section. An increase in population (+) causes the demand curve to shift to the right. A decrease in population (-) shifts it to the left.

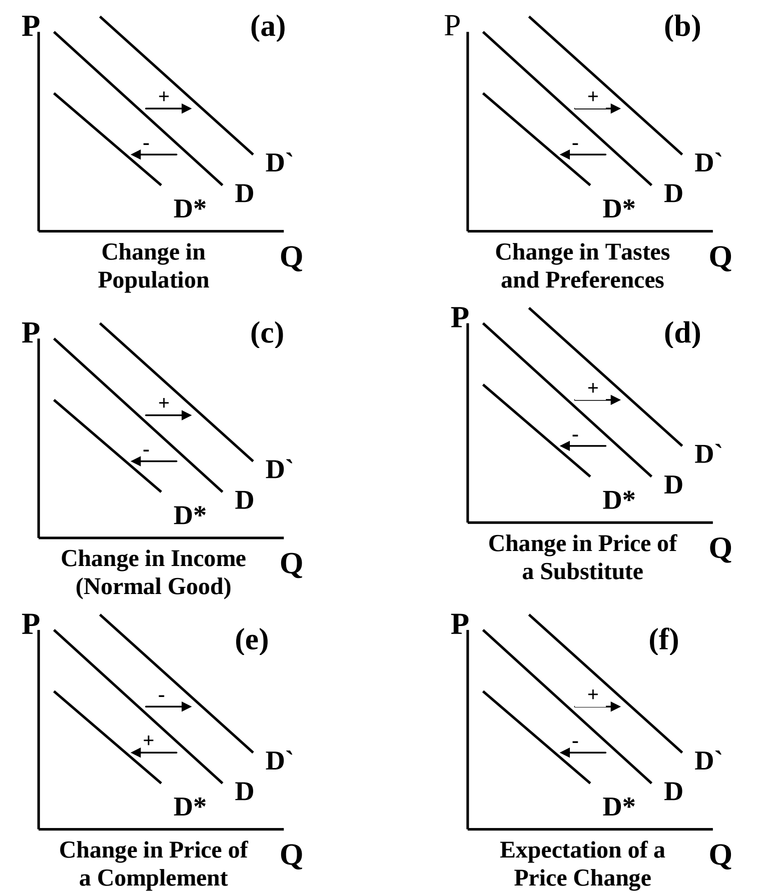


Figure 3-6
Changes in the Determinants of Demand and Shifts in the Demand Curve

In panel (b), we see that increasing preference for a good (+), i.e., we like the good more than we used to like it, increases demand and is represented by shifting the demand curve to the right. For example, suppose many people become believers in the Adkins low carbohydrate diet. In this diet one eats a lot of protein. People have a stronger preference for beef than they used to. The demand curve for beef shifts to the right. Staying with the beef example, suppose mad cow disease breaks out in the U.S. Tastes and preferences shift away from eating beef. Demand falls, and we represent this in the model by shifting the demand curve to the left.

Panel (c) is constructed for a normal good. Remember that a normal good is one you want to buy more of if your income goes up, and less of if your income goes down. An increase in income (+) increases demand for the good. We represent this by a shift to the right of the demand curve. Conversely, a decrease in income (-) causes a decrease in the demand for a normal good. We shift the demand curve to the left in this case. Cars are normal goods. If the incomes of households are increasing in an expanding economy, the demand for cars will increase. If the economy is contracting, a condition we will call a recession in Chapter 9, the demand for cars will fall. In this example, the shifts to the right and left in the demand curve for cars is a way to examine the effect of the ups and downs of the overall economy in the car market.

In panel (d), we show the effect of a rise (+) and a fall (-) in the price of a substitute good. If this were a demand model for Coke, a rise in the price of Pepsi would shift the demand curve for Coke to the right. A fall in the price of Pepsi would reduce the demand for Coke. We represent this as a shift of the demand curve for Coke to the left.

Panel (e) is the only case where an increase in something leads to a decrease in demand. If the price of a complement increases (+), the demand for the good in question falls. This is why the arrow in panel (e) with the (+) sign is associated with a reduction in demand and a shift of the demand curve to the left. The opposite occurs in the price of a complement falls. Think again of our hot dog and hot dog buns example. If we have a model of the hot dog bun market, a fall in the price of hot dogs (-) shifts the demand curve for hot dog buns to the right.

Finally, in panel (f) we model the effect of a change in expectations on demand. If people think the price of a good is going to rise in the future (+), they will want to buy more now, increasing demand. We shift the demand curve to the right in this case. If people think the price of a good will fall in the future, they will postpone purchases now, and reduce demand. We represent this in panel (f) by shifting the demand curve to the left.

I can't overemphasize the importance of understanding the effect on demand of changes in each of the determinants of demand, and knowing how to represent these changes in the demand model with the appropriate shift of the demand curve. Once you can do this easily, you know that you understand the demand model and how to use it. Practice here is a good investment, yielding a high rate of return.

The Determinants of Supply and the Determinant of Quantity Supplied

Our treatment of the selling side of the supply and demand model follows a pattern very similar to our treatment of the demand side. We begin with the distinction between the determinants of supply and the determinant of quantity supplied. The items inside the brackets of Figure 3-7 affect which and how much of a good sellers want to sell. As in our discussion of demand, we break the list into two parts. One item, price, is the only determinant of quantity supplied. The other six items inside the brackets are determinants of supply.

These Affect What and How Much Sellers Want to Sell

Price — **Determinant of Quantity Supplied**

Production Capacity
Price of Resources
Technology
Prices of Related Goods:
Substitutes
Complements
Expectations of Prices in the Future — **Determinants of Supply**

Figure 3-7
Determinants of Supply and Quantity Supplied

Imagine yourself a seller, or potential seller of some good. Would it be the case that you would want to sell more of a good the higher its price? The answer is yes. If you produce and sell pizzas for a living, you'd want to sell more if the price were \$20 than if the price were \$1. In fact, at a price of \$1, the quantity supplied of Pizza in the marketplace would probably be 0.

Production capacity affects how much sellers want to sell, but it's a determinant of supply, not a determinant of quantity supplied (This is picky, but important.). Production capacity takes two forms. One is the number of sellers, and the other is the size of the seller's operations. Both affect supply in the market. The more production capacity, the higher the supply. The lower the capacity, the lower the supply.

The next two determinants of supply combine to influence production cost. Sellers produce goods using resources and a particular technology or production process. If resource prices rise, technology held constant, production costs increase. Sellers will want to sell less. If resource prices fall, technology held constant, production costs will fall, and sellers will want to sell more. Likewise, if sellers experience improvements in technology, even with constant resource prices, costs will fall, and sellers will want to sell more.

As they did with demand, prices of related goods affect supply. But now we have to think about related production activities, and selling the good, rather than related buying opportunities. Suppose you are a farmer. You can plant either dry peas or lentils on your land. Let's consider the supply in the market for lentils. Holding all other determinants of supply constant, an increase in the price of peas will cause a decrease in the supply of lentils. The rise in the price of a substitute (peas) in production leads producers to plant more peas. If they plant more acres of peas, they can't plant as many acres of lentils, the good under consideration.

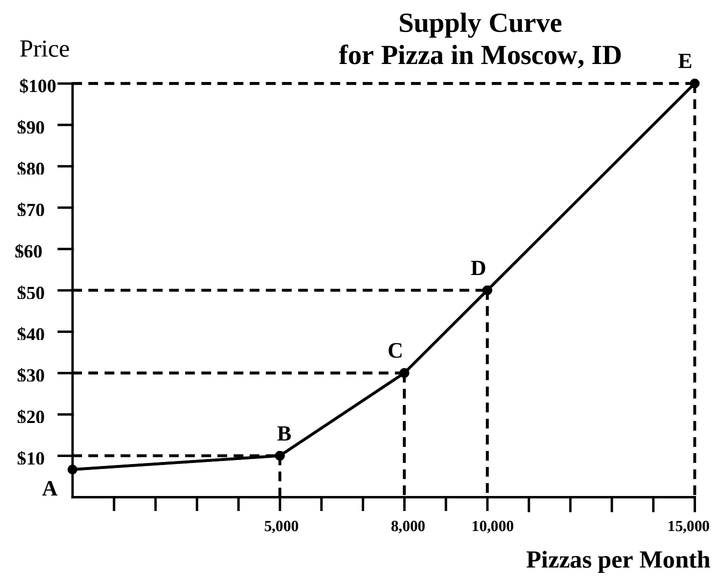
Examples of complements in production are trickier to understand. Honey and apples can be production complements. Honey requires bees and bees require blossoms and blossoms need bees for pollination. Consider the honey market. A rise in the price of apples leads orchardists to grow more apples. More apple trees mean more blossoms in need of pollination, more bees and more honey. A rise in the price of a complement in production (apples) increases the supply of the good in question, honey in this case. A fall in the price of a complement in production leads to the opposite effect.

Finally, expectations of future price is a determinant of supply. If sellers expect price to rise in the future, they will want to sell less today at the lower prices and sell more in the future. Expectation that the price will fall leads to an increase in the supply today, a get the sales while you can approach.

Supply and Quantity Supplied: A More Technical Discussion

As in our discussion of demand and quantity demanded, a move toward abstraction can draw out the differences between supply and quantity supplied. Figure 3-8 contains both a supply schedule and a supply curve for pizza in Moscow, Idaho. In order for our model to be applicable to the pizza market we must assume that pizza is a homogeneous product, i.e., all sellers sell the same type and quality of pizza. We will relax this requirement in Chapter 7 when we consider the topic market structure and implications of product differentiation.

The process of constructing a hypothetical supply schedule and related supply curve is very similar to what we did with the demand schedule and the demand curve. As we want to focus on the relationship between price and quantity supplied, we can't have the determinants of supply changing all over the place. So we hold them constant, give price one of the two axes on the graph and generate a numerically illustrated supply curve corresponding to the supply schedule.



Supply Schedule for Pizza in Moscow, ID

Point on Supply Curve	Price	Quantity Supplied per Month
A	$100	15,000
B	50	10,000
C	30	8,000
D	10	5,000
E	5	0

Figure 3-8
A Supply Schedule and a Numerically Illustrated Supply Curve for Pizza

Note a major difference exists between the supply curve and the demand curve. As the price of pizza rises the quantity supplied rises. Where price and quantity demanded are inversely related, price and quantity supplied are directly related. The higher the price, the higher the quantity supplied. The lower the price the lower the quantity supplied. This direct relationship between price and quantity supplied is reflected in an upward-sloping supply curve in Figure 3-8. Again, analogous to the relationship between the demand schedule and the demand curve, for every price-quantity combination in the demand schedule there is a corresponding price-quantity point on the supply curve. We have only five prices in this example. By adding additional prices and their corresponding quantities supplied, we could "smooth out" the kinks in our supply curve.

A More Abstract Supply Curve

Like learning about demand, you might find numbers helpful when first confronted with a supply curve. Numerical illustration, however, becomes more and more cumbersome as your modeling ability increases. A little more abstraction makes the use of the supply model much easier. A more abstract supply model appears in Figure 3-9. Note that the price and quantity numbers have again been replaced with the symbols P_1, P_2, and P_3, and Q_1, Q_2 and Q_3. We have preserved the most important characteristic of the supply curve. It slopes upward, indicating the direct relationship between price and quantity supplied. P rises, Q rises. P falls, Q falls.

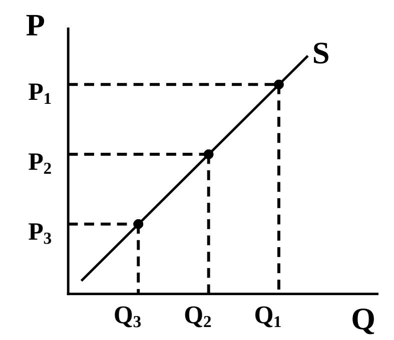


Figure 3-9
An Abstract Model of a Supply Curve

With a more abstract supply curve we are not very interested in the actual shape of it. Straight-line supply curves like the one in Figure 3-9 are easy to draw, but curved ones will work just as well, AS LONG AS THEY SLOPE UPWARD!

Modeling Changes in Supply

The direct relationship between price and quantity supplied is a very important concept, but we need additional modeling abilities to make our supply model useful. We will often use our supply model to examine more than the effect of a price change on quantity supplied, with the determinants of supply held constant. Just as on the demand side of the model, we need to know how to handle in our model the effects of changes in a determinant of supply. What happens when a determinant of supply is allowed to change?

We see the supply model illustration of an increase in supply in Figure 3-10. The supply curve labeled S is the original supply curve. At a price of P_1 sellers want to sell Q_1 units of the good. At the lower price P_2, they want to sell less, Q_2. Now suppose production capacity grows in the market area. In our pizza example, suppose the number and size of pizza business in of Moscow increases. Now, because there

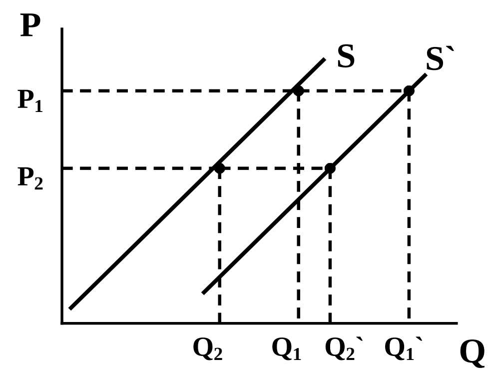


Figure 3-10
Modeling and Increase in Supply

are more and bigger sellers in the market, supply increases. At a price of P_1 sellers now want to sell Q_1' not Q_1. At the price P_2, before the production capacity increase sellers wanted to sell Q_2, but with a greater production capacity the quantity supplied at this price is Q_2'.

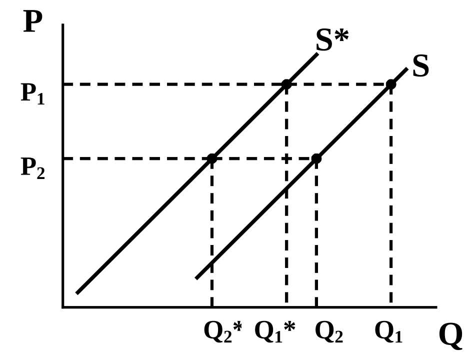


Figure 3-11
Modeling and Decrease in Supply

When learning to understand how to model a change in supply, start first with a particular price and ask the question, after a change in the determinant of supply, will sellers want to sell more or less of the good? Then go to another price and ask the same question, then to another price, then to another, etc. If you ask this question at enough prices, you can connect the dots of the new price-quantity combinations and you will have a new supply curve, and you will find that an increase in supply is represented by a shift to the right of the supply curve.

The supply model illustration of a decrease in supply appears in Figure 3-11. Here the new supply curve, S*, lies to the left of the original supply curve. Suppose production capacity decreases. At a price of P_1, before the decrease in production capacity, quantity supplied was Q_1. But now, with less production capacity in the area, the quantity supplied at a price of P_1 falls to Q_1*. Likewise, if the price were P_2 before the decrease in production capacity, quantity supplied would have been Q_2. But with a lower production capacity the quantity supplied at the price P_2 is Q_2*. It doesn't really matter how much you shift the supply curve to illustrate a decrease in supply. It is important, however, that you shift the demand curve to the left.

Modeling Changes in Each Determinant of Supply

In the last section, we saw how to show an increase and decrease in supply with our supply model. We shift the supply curve to the right for an increase in supply and to the left for a decrease in supply. But how do changes in each specific determinant of supply affect the supply curve? We address this in Figure 3-12. As I our discussion of determinants of demand, we have six separate graphs in Figure 3-12, each one corresponding to one of the six determinants of supply. A + sign indicates that the determinant has increased, and a – sign indicates the determinant has decreased. Arrows associated with the + and – signs again indicate the corresponding shift in the supply curve. For example, panel (a) in Figure 3-12 repeats the example of the last section. An increase in production capacity (+) causes the supply curve to shift to the right. A decrease in production capacity (-) shifts it to the left.

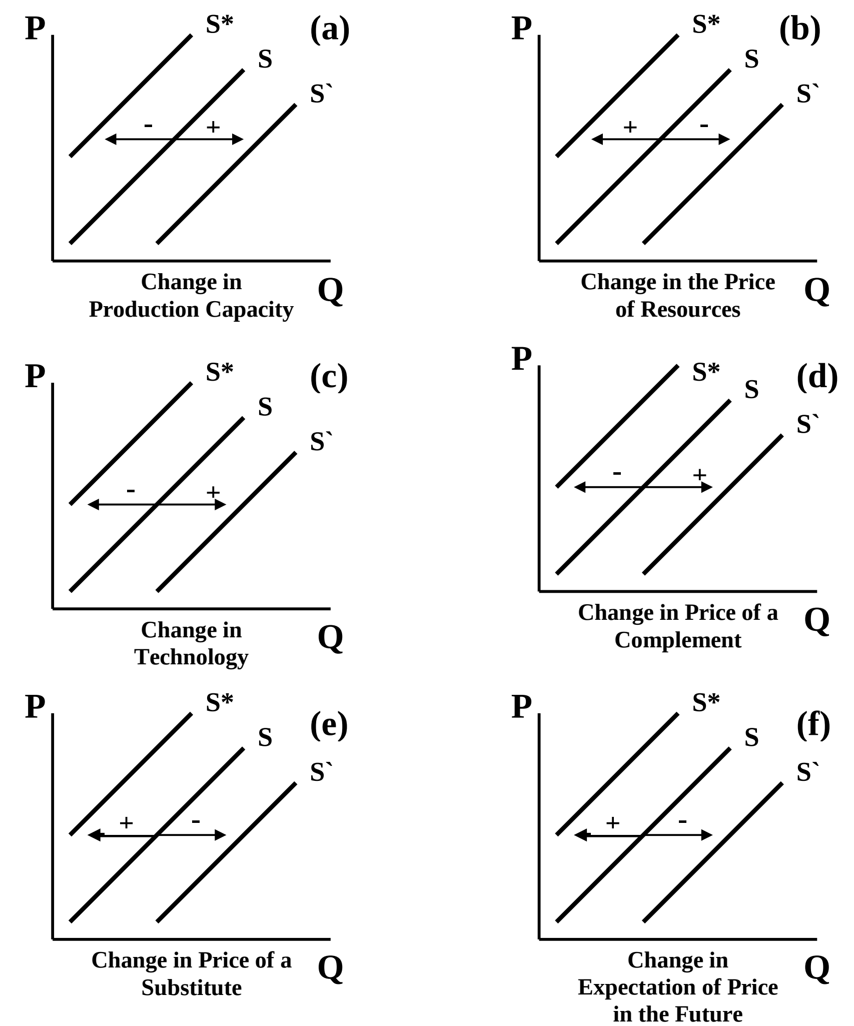


Figure 3-12
Changes in the Determinants of Supply and Shifts in the Supply Curve

In panel (b), we see that increasing the price of resources (+), decreases supply and is represented by shifting the supply curve to the left. For example suppose the wage (price) of labor increases. Without changes in technology, this will increase cost and supply falls, and we represent this in the model by shifting the supply curve to the left. An alternative way of thinking about the change in the price of resources is to think about the supply curve shifting up and down, rather than left and right. If costs rise, sellers need a higher price for any given quantity they sell, thus shifting the supply curve up.

Panel (c) shows the effect on supply of a change in technology. Favorable technological change will lower costs of producing any quantity. This will increase supply and is represented by a shift to the right of the supply curve. Alternatively, you could think of a lowered price required for sellers to sell a particular quantity when costs fall after favorable technological change. You could then think of the supply curve shifting down.

Panels (d) and (e) show the effects of changes in the prices of related production goods, substitutes and compliments. The fall in the price of a substitute increases the supply of the good in question, and a rise in the price of a substitute reduces the supply of the good in question. These effects are shown as a shift to the right and left of the supply curve, respectively. For complements in production, the results are just the opposite than they are for substitutes. This is reflected in panel (e) by the + and – signs being associated with arrows going in the opposite direction than those in panel (d).

Finally, expectations of a good's price in the future can affect supply today. Expecting price to rise will reduce supply today as sellers wait to sell at higher prices in the future. The reverse is true for expectations of a falling price.

Just as in the case of the determinants of demand, it is very important to know each determinant of supply, how changes in them affect supply, and how these changes in supply are represented abstractly in a supply model. Again, practice is a good idea here.

The Supply and Demand Model: Equilibrium and Market Clearing

We now need to merge the demand model with the supply model to form the supply and demand model. We are in luck because rules of algebra and geometry allow putting two lines on the same graph if the lines relate to the same items on the axes of the graph. In the demand model, the one with the demand curve, price of the good is on the vertical axis and quantity of the good is on the horizontal axis. Likewise, in the supply model, the one with the supply curve, price is also on the vertical axis and quantity is on the horizontal axis. If we merge the two models into one graph, we have the model pictured in Figure 3-13.

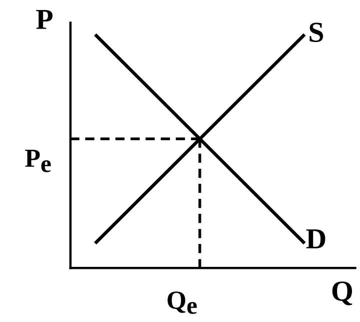


Figure 3-13
The Supply and Demand Model and the Equilibrium Price and Quantity

Usually two characteristics of an economics graph have the most importance, where two lines intersect, or where two lines have the same slope. In this case it's the former characteristic. The point where the demand curve intersects the supply curve in Figure 3-13 represents a unique and special combination of price and quantity. It is the only point where the quantity demanded equals the quantity supplied. We call this price, P_e, the equilibrium price and this quantity, Q_e, the equilibrium quantity. The term equilibrium is borrowed from the natural sciences, but it is an appropriate term in a model of a market. If the price is not at the equilibrium price, and because of this the quantity demanded in the market is not equal to the quantity supplied, something will happen in the market to return the market to an equilibrium, the condition where quantity demanded equals the quantity supplied.

We often use the term market clearing price instead of equilibrium price, because the terminology conveys the useful information that no excess amount of the good exists in the market, i.e., all that is offered for sale is purchased. Likewise, at the market-clearing price we see no buyers who want to buy at that price and cannot buy the good. At the market-clearing price, all who want to buy can buy just the amount they want and all who want to sell can sell just the amount they want to. In an equilibrium, when the market clears, we say there is no excess demand or excess supply. Such is not the case when the price is below or above the equilibrium price.

Excess Demand, Excess Supply, and Price Change to Clear a Market

Figure 3-14 shows what happens if the price is below the equilibrium price. At the price P_1, the quantity demanded, Q_d, is greater than the quantity supplied, Q_s, at that price. This is shown in the model as the bracketed distance $Q_d - Q_s$ labeled "excess demand." A price of P_1, and the excess demand as a result of it, will not survive in a market where price is allowed to change. As long as quantity demanded is greater than quantity supplied, price will rise. Buyers will be willing to bid up the price and sellers will be more willing to sell more at these higher prices.

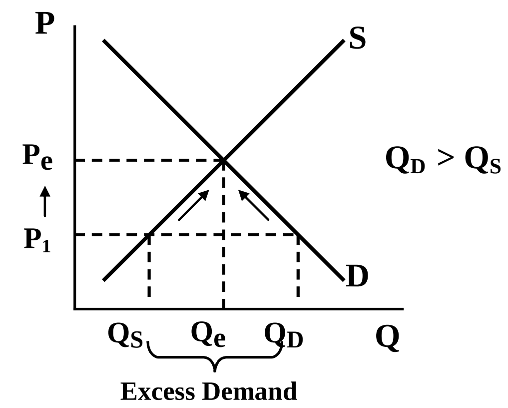


Figure 3-14
The Supply and Demand Model with Excess Demand and a Price Increase to Restore Equilibrium

The process of a common English auction, while a special case of fixed supply, shows the effect of excess demand on market price. At a low price for an item up for auction, many hands fly into the air to bid. As the price rises, more and more hands remain down until just the winning bidder herself remains. The excess demand has vanished in the face of a rising price.

The contrasting case of excess supply is shown in Figure 3-15. Here the price P_1 is above the market-clearing or equilibrium price P_e. Now a condition opposite than that in the excess demand case occurs. Now quantity supplied is greater than the quantity demanded. Sellers cannot sell all that they want to at that price. Excess supply is not sustainable either if price is allowed to change. Here the price will fall as sellers attempt to unload unsold goods to willing buyers at lower and lower prices. Again, the market clears where the quantity demanded equals the quantity supplied.

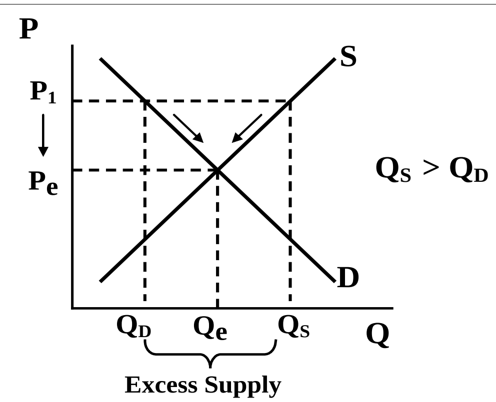


Figure 3-15
The Supply and Demand Model with Excess Supply and a Price Decrease to Restore Equilibrium

Changes in Demand and the Effects on Price an Quantity

In Figure 3-16 we see the effect on market price and quantity from an increase in demand. We start from a position of equilibrium in the market, the intersection of the supply curve and the demand curve D_1 at the point labeled 1 in the graph. The price in the market is P_1, and the quantity exchanged, the equilibrium quantity in this case, is Q_1. Quantity demanded equals quantity supplied. The market clears. Now suppose demand increases because one of the determinants of demand changes. Population could increase. If the good is a normal good, an increase in income and wealth cause such an increase in demand. A change in tastes and preferences, or the price of a related good, or of expectations of price in the future could all cause an increase in demand.

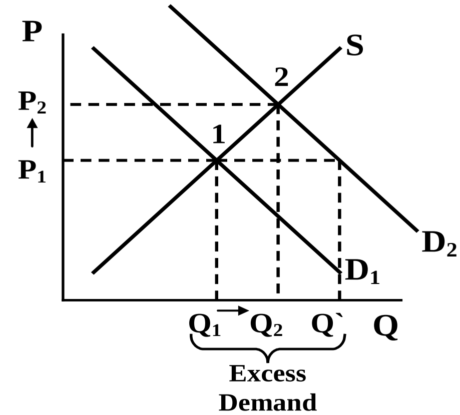


Figure 3-16
The Supply and Demand Model with an Increase in Demand

With an increase in demand, and with supply unchanged, the original market price P_1 is no longer the equilibrium price. The market will not clear at this price. After the increase in demand, represented by a rightward shift of the demand curve from D_1 to D_2, at a price of P_1 we have a temporary excess demand. Buyers want now to buy Q_1', but sellers only wish to sell Q_1. As we discussed in the last section, excess demand leads to an increase in the price of the good, and at the higher and higher prices the quantity supplied increases and the quantity demanded falls. The market will clear again when the price rises to P_2, found from the intersection of the supply curve and the new demand curve D_2. At the new equilibrium, the price has risen enough to induce sellers to supply Q_2 of the good, and has risen enough to dissuade just enough buying to get the quantity demanded to Q_2 as well.

The market has responded to an increase in demand with a price increase, from P_1 to P_2, and an increase in the amount bought and sold in the marketplace. If we think about this for a moment, isn't this what we would expect to happen in a market economy. Let's take a moment to reconsider resource allocation in a market economy. The increase in demand indicates that people want to buy more of the good at the current price P_1, so we would hope that a market economy could respond in a way to meet this new information. But at a price of P_1, sellers want to sell only Q_1. But sellers are willing to sell more if the price is higher, and at least some buyers are willing to pay a higher price to get more of the good, so the potential exists for some mutually beneficial new exchanges to occur. The information conveyed by the ensuing price change increases the amount sold and chokes off just enough of the excess demand so that the market clears again, but with more of the good exchanged, and a higher price. Suppliers have responded to the increased demand by selling more.

At this time we need to revisit that nasty terminological distinction between supply and quantity supplied. We have a situation in Figure 3-16, where demand has increased, price has increased, the amount bought and sold has increased, and supply has remained the same. How can more be bought and sold in the market and supply remain the same. The answer, of course, is that the quantity supplied has increased in response to a higher price. Remember only price is a determinant of quantity supplied. Because we did not change any of the determinants of supply, we have no shift in the supply curve in Figure 3-16, only a movement along it, a change in quantity supplied.

Perhaps a real world example is in order here. Suppose this is the market for apartment rentals in a college town. Suppose enrollment at the local college increases. Translated into the supply and demand model this is an increase in the determinant of demand population. At current rental rates, P_1, excess demand exists. Apartment owners see lines outside their offices, applications for apartments are way up, and their phones are ringing off the hook. Vacancy rates are very low. You get the picture. What are self-interested apartment owners going to do in this situation? Raise the rent.

As rental rates in the local housing market rise some students decide that apartment living is not for them, or they acquire more roommates, or they live at home, or in the dorm. You get the idea. Higher rental rates choke off some of the excess demand. And as rental rates rise, apartment owners start building more apartment buildings, and regular folks turn basement storage into student apartments. As apartment rental rates rise, the quantity of apartments supplied rises as well. When the market clears at the new price P_2, apartments rented in the area have increased to Q_2, just what we would expect to happen in a free rental market in the face of increased demand.

Figure 3-17 shows the effects of a decrease in demand on the equilibrium price and quantity. Now, instead of excess demand at the original price we have excess supply. The price falls from this excess supply. Sellers see that they cannot sell all the wanted to at the price P_1, but they can sell some of there excess at a lower price, a price at which buyers could be induced to pay. As the price falls, sellers adjust downward the amount they want to sell, and quantity supplied equals quantity demanded again at the Price P_2.

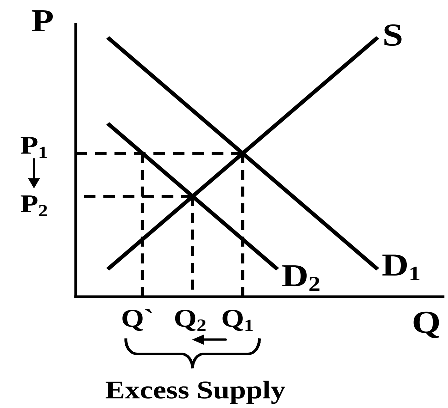


Figure 3-17
The Supply and Demand Model
with a Decrease in Demand

Again, think of how resources have been reallocated in response to a decrease in demand. Selling off an excess supply drives down the price and at the lower price sellers want to sell less, but in the new equilibrium just the amount buyers want to buy. We get less of the good exchanged in the market, just what we would expect to happen if buyers wanted less of it. Again note that supply has not changed. The supply curve has not shifted. We have, instead, a reduction in the quantity supplied in response to the decrease in price cause by the lower demand.

Now we have to run our apartment example in reverse. Suppose tuition and fees have gone up greatly at the local college, reducing enrollment and demand for apartments. Can you see that the price of a good related to apartments has risen? Apartments and total credit hours are complements in a college town. Now apartment owners see vacancy rates rise. Apartments sit empty for months at a time. What are apartment owners going to do? Lower the rent. Falling rents dampen the decline in apartment rentals but do not eliminate the decline. As rents fall apartment owners convert rental units to office space, and local homeowners turn former apartments into wine sellers and home theaters. The quantity of apartments supplied falls in response to the decline in demand and lower price.

Changes in Supply and the Effects on Price and Quantity

The case of an increase in supply is modeled in Figure 3-18. One of the determinants of supply has changed. To make it easy, assume entry to the market has occurred. More suppliers exist. There is more "production capacity" for apartments. We represent this in the model by shifting the supply curve to the right. At the original price P_1, we now have an excess supply. Unable to sell all they want to at the price P_1, sellers lower prices. As price falls in the market buyers are willing to buy more of the good. Also, as the price falls suppliers help reduce the excess supply by lowering the amount they want to sell. The market clears at the new lower price P_2, and a higher quantity Q_2. This is what we would expect in a market economy. If supply increases, it has to go somewhere. Where does it go? It goes to buyers who are now willing to buy it at lower prices.

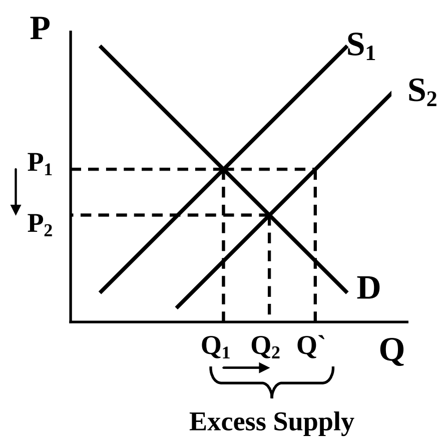


Figure 3-18
The Supply and Demand Model
with an Increase in Supply

In the apartment market example, we can think of this as a wave of new apartment building. Suppose new construction technologies, or less local regulation, or wage concessions by construction unions have lowered production costs. Apartments look like a good place to put new capital. But the increase in apartment building causes a glut of apartments. Vacancies rise. It's hard to find renters at the current rental rates. Rental rates fall in response to the excess supply, and this in turn brings students out of the dorm, out of Mom and Dad's house, and away from sharing rooms with roommates. These words describe the information from the model in Figure 3-18.

We model a decrease in supply in Figure 3-19. Again a determinant of supply has changed, but this time in such a manner that we shift the supply curve to the left. Imagine that the price of a substitute good in production rises, reducing the supply of the good in question. At the original price, P_1, we now have a temporary case of excess demand. With the reduced supply, quantity demand is greater than the quantity supplied. The excess demand causes a price increase, lowering the quantity demanded to the new equilibrium quantity Q_2.

Again the market accomplishes what we might expect it to do. We have a reduced supply of the good. How are we in the economy going to decide who gets to consume the good? By price of course. As price rises, buyers fall out of the market and remaining buyers buy less of the good. Seeing the higher prices, sellers help reduce the excess demand by willingly selling more as prices rise. The excess demand is eliminated from both sides of the market and the market clears again.

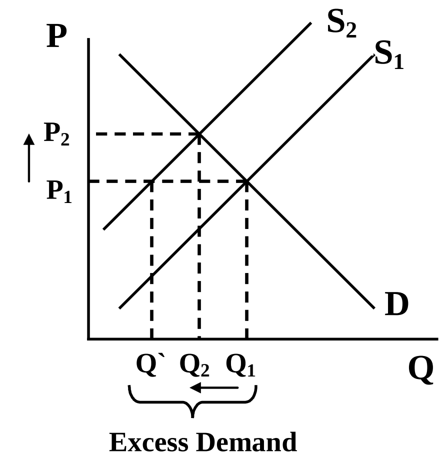


Figure 3-19
The Supply and Demand Model with a Decrease in Supply

Continuing the apartment story, assume rents for office space rise dramatically. This is an increase in the price of a substitute good in production. Apartments are converted to offices and the supply of apartments falls. As an economy we need to decide who gets apartments under the new conditions. In a market economy we decide to rent apartments to those who are willing to pay the new market-clearing price P_2. Again, some of the initial excess demand for apartments is also reduced from the supply side as rising prices dampen apartment conversion to offices and mother in laws are thrown out of the basement apartment into assisted living.

Changes in Both Demand and Supply

To this point in our supply and demand modeling, we have allowed only a determinant of demand or a determinant of supply to change. Unfortunately, this is sometimes too simplistic a case, as determinants of both demand and supply can change at the same time. Figure 3-20 shows the result of an increase in demand and an increase in supply. Both the demand curve and the supply curve shift to the right. In this case we can be sure that the equilibrium quantity will increase, as it does in Figure 3-20, from Q_1 to Q_2. This should come as no surprise, because increases in supply and demand are both signaling MORE in the marketplace. We cannot, however, be certain whether price will rise or fall in this case. An increase in supply puts downward pressure on price, but an increase in demand applies upward pressure. The direction of price change depends on the relative magnitude of the changes in demand and supply. As pictured in Figure 3-20, the price rises from P_1 to P_2, but just imagine the supply curve shifting out more, or the demand curve shifting out less, and price could fall.

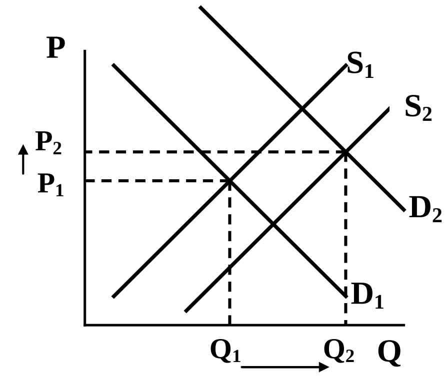


Figure 3-20
The Supply and Demand Model with an Increase in Demand and Supply

Figure 3-21 shows the supply and demand model with a decrease in both demand and supply. Here both actions signal LESS to the marketplace. We see that the equilibrium quantity unambiguously falls in this case, as supply and demand changes reinforce each other. Again, with supply and demand moving in the same direction, albeit downward this time, the effect on price can be either positive or negative. If the price-increasing effect of a reduction in supply outweigh the price-lowering effect of a decrease in demand, price will rise, as shown in Figure 3-21. But if demand decreases a large amount and supply decreases just a little, price could fall. You should be sure that you can demonstrate each case with your supply and demand modeling.

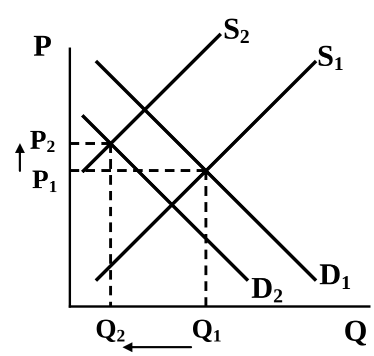


Figure 3-21
The Supply and Demand Model with a Decrease in Demand and Supply

Demand and supply need not move in the same direction, however. In Figure 3-22 we show the effects of a demand increase and a supply decrease. Now it is price for which we see an unambiguous effect, for both demand increases and supply decreases have a positive influence on price. The price rises in the model from P_1 to P_2. But now the effect on equilibrium quantity is ambiguous. We see quantity falling in the case shown in Figure 3-22, but we could construct a case where the supply shifts less to the left and the demand curve shifts more to the right, with the result an increase in equilibrium quantity. This ambiguous effect on quantity is due to the opposing forces demand and supply have on quantity in this case. The demand increase is signaling MORE to the marketplace, while the supply decrease is signaling less. It's a tug-o-war and the stronger force wins.

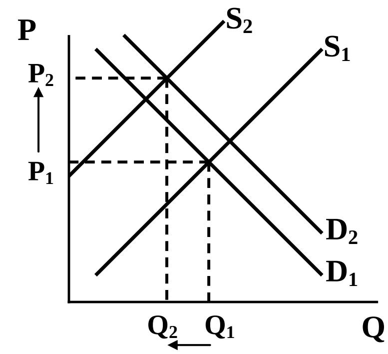


Figure 3-22
The Supply and Demand Model with an Increase in Demand and a Decrease in Supply

In Figure 3-23, we reverse the directions of Figure 3-22. Now we have a decrease in demand from D_1 to D_2, and an increase in supply from S_1 to S_2. Price falls unambiguously to P_2. As shown, quantity falls from Q_1 to Q_2, but this need not be the case. Again, when supply and demand move in opposite directions the direction of quantity change depends on the relative magnitude of the changes in demand and supply. A small shift in the demand curve and a bigger shift in the supply curve could result in an increase in the equilibrium quantity.

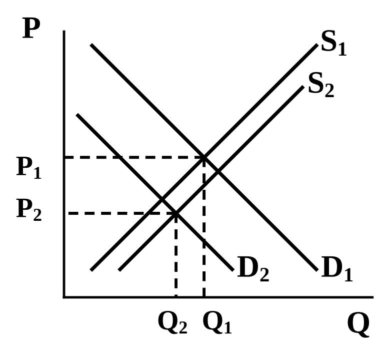


Figure 3-23
The Supply and Demand Model with a Decrease in Demand and an Increase in Supply

A Theory of Price Ceilings and Shortages

All of the cases of the supply and demand model examined so far have involved prices which move up or down according to changes in the determinants of demand or supply. The theory of markets, represented by the supply and demand model, predicts that markets will clear after changes in the determinants of demand or supply, if prices are allowed to adjust. Now we examine the case where the price is held artificially below the market-clearing price. This is called a price ceiling.

Because ceilings are usually above us, some economics students are confused initially when confronted with this terminology. How can something BELOW the market-clearing price be a ceiling? This confusion is easily cleared up with a better metaphor. Think of a price under conditions of excess demand as a helium-filled balloon. The balloon (price) wants to rise. Suppose we let go of the balloon inside some building. Where does it end up? Yes, at the ceiling. It can't go any higher.

Usually government is the institution that imposes this restriction on the market, but not always. Apartments, river float trip permits and household water are just three examples where price is often not allowed to rise to meet its equilibrium level.

City council members, in an attempt to protect renters (and get votes from them), might pass a law that controls the level of apartment rents below the market–clearing level. This is called rent control. The U.S. Forest Service charges a price for permits to float the Salmon River that is far below a market-clearing price. When drought hits cities in the West, government water agencies are reluctant to raise price to deal with the excess demand caused by a decrease in the supply of water and an increase in the demand for it. We'll return to these examples below, but first we demonstrate what happens in general when a price ceiling is applied to a market.

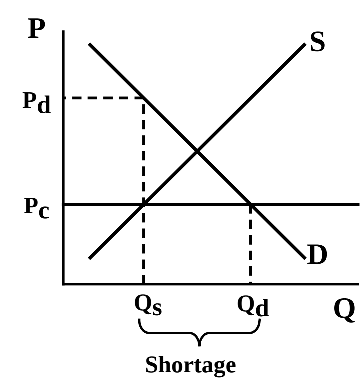


Figure 3-24
The Supply and Demand Model with a Price Ceiling

Figure 3-24 shows the effect of a price ceiling. The price for the price ceiling is labeled P_c in Figure 3-24. A price below the market-clearing price is nothing new. We've seen it in previous sections of this chapter. The quantity demanded is greater than the quantity supplied. In a market without a price ceiling this excess demand, a temporary phenomenon, would lead to a increase in price and the elimination of the excess demand as the price moved up to its new market-clearing (equilibrium) level. But now the excess demand is not temporary, but permanent. The permanent excess demand, the one that does not vanish, we call a shortage. The theory of price ceilings and shortages is simply that the former causes the latter.

Shortage is a technical term in economics. It does not mean reduction in supply, although a reduction in supply could generate an upward price movement that might be attacked with a price ceiling. A shortage arises only because of a price ceiling. Without the price ceiling, price will rise to a new market-clearing level where the quantity demanded equals the quantity supplied.

Another implication of a price ceiling is that a lower amount of the good is exchanged in the marketplace. While the quantity demanded is greater than the equilibrium quantity, some of this quantity demanded is unsatisfied demand. In Figure 3-24 the amount actually exchanged is only Q_s. Economists often describe this case as "the short side determines the market quantity." In the case of a price ceiling, the quantity supplied is the "short side" of the market. At the ceiling price, P_c, sellers are willing to sell only Q_s.

Another implication of a price ceiling and the shortage it creates is that we have to find some way to decide who gets the good other than willingness to pay the price. At the price P_c, more people are willing to pay the price for the good than can be supplied at that price. Economists often call this "non-price rationing." First come – first served and various forms of discrimination are just two examples of non-price rationing. Another method of distributing the good is use of a black market.

A black market is an unofficial market, often an illegal market. Price ceilings and their related shortages create incentives for black markets to form. Return to Figure 3-24. Note the price P_d in the model. If the price were P_d, buyers would want to buy the amount labeled Q_s. Another way to look at this is to realize that buyers are willing to pay P_d for the Q_s of the good. Because the price that buyers are willing to pay is greater than the price at which sellers are willing to sell, the potential for mutually beneficial exchanges exists. Unfortunately, because of the price ceiling, the transactions must occur in the unofficial, under-the-table or illegal black market. So the theory of price ceilings is also a theory of black markets.

Some Applications of the Supply and Demand Model with a Price Ceiling

Rather than create an additional model for each example, we can use the same general model, Figure 3-24, to apply to all cases. This use of a general model to apply to specific cases also illustrates an important benefit of theory and modeling. If we understand a more general theory, we can economize on the specific information we need to carry around in our brains. And helping to keep a brain uncluttered is no small accomplishment. Lets see if we can apply the theory of price ceilings and shortages represented by Figure 3-24 to the examples we mentioned above, rent control, float trip permits, and water.

For the apartment example consider the price to be the monthly apartment rental rate. With a price ceiling the quantity of apartments demanded is greater than the quantity supplied at the price ceiling. More people want apartments at that price than can be supplied at that price. Apartment owners have a choice among many potential apartment renters. When the market clears at the market-clearing price this list of potential renters does not exist. The apartment owner could decide not to rent to students, or professors, or women, or Indians, or black people, or people with long hair, of families with children, or old people, or old, black, female students with children. You get the picture. A shortage makes discrimination possible. When price is allowed to adjust, and the amount demanded is just equal to the amount supplied, we are more likely to ration by renting to people writing checks or waving $100 bills around.

Of course, black markets could emerge in rent controlled apartment markets. In Figure 3-24 some renters are willing to pay P_d for an apartment that apartment owners are willing to rent for P_c. Nonrefundable key deposits or outright bribes would help decide who gets the apartment.

River permits are subject to a price ceiling for the purposes of perceived fairness. The government decides that river permits are too important to be left up to the market for allocation and distribution. Those with higher incomes would be more able to buy river permits in a market setting, so the price is kept low to make them affordable to lower income households. Again, good intentions might drive the imposition of the price ceiling. But, of course, a shortage of river permits exists at the price ceiling. Nonprice rationing must be employed.

The forest service, perhaps in order to avoid a parking lot turned into a campground under a first-come first-served system, usually uses a random draw to allocate permits. They "put all the names in a hat" and draw out Q_s names. These people pay the ceiling price and float the river.

Economists often point out that it is not necessary to avoid a market process entirely when we are concerned about the level of income of river permit demanders. The government could create an initial distribution of permits by random draw, and then allow a period where permit holders could sell their permits to anyone who wanted to buy them. This is something called an “aftermarket.” It’s sort of like legalized “scalping” of river permits.

Finally, the model applied to drought in the West. We can imagine that a water shortage might occur when the water market, initially in equilibrium experiences a reduction in supply due to a low winter snowpack. The initial equilibrium price, which cleared the market in a normal water year, no longer clears the market. If the price is not raised, a de facto ceiling emerges, and the model again looks like Figure 3-24. Make no mistake about it, a reduction in supply need not cause a shortage if price is allowed to rise to a new market-clearing price.

Government water agencies are reluctant to use price as a way to deal with drought. The more common method of dealing with a water shortage is with nonprice rationing methods such as water ordinances that prohibit outside watering during the hot part of the day, or allowing watering on odd or even numbered days, depending on whether your street address is an odd or even number. Reduced water pressure is another nonprice rationing method. Water authorities could also have different parts of the city go without water every now and then. This would be equivalent to the “rolling blackouts” that occurred in California in response to reduced electricity supply and ceilings on electricity prices.

The above examples are all examples of government-imposed price ceilings. Other nongovernment examples exist as well. Tickets to the NCAA basketball Final Four tournament are priced well below the market-clearing price. We know this because the NCAA distributes tickets with a random draw method. Go to the site of any Final Four and you will probably see active black markets in action, where the prices at which buyers are willing to buy are higher than the minimum prices at which some ticket holders are willing to sell.

When country singer Garth Brooks was touring, he often reserved a certain number of low-priced tickets for regular folks who were willing to get them on a first-come first-served basis. Sidewalks in front to ticket windows turned into campgrounds. Any time you see a long line of people waiting to buy a good, it’s likely that the good is subject to an informal price ceiling. Free food at college picnics is a good example here. At a price ceiling of $0, you may have to experience the nonprice rationing of waiting in a long food line.

Parking on university campuses is usually a situation of a shortage caused by parking permits priced below the market-clearing price. Have you ever heard someone refer to a parking permit as a hunting license, not a parking permit? At 9:00 in popular morning hours on college campuses we give parking places to those who arrive early, or win a game of "chicken" in competition for a parking place. Both are forms of nonprice rationing.

A Theory of Price Floors and Surpluses

Like a price ceiling, a price floor comes from outside the market, often from government, and works to keep the price and quantity from reaching their market-clearing levels. But in the case of a price floor the price is set above the market-clearing price rather than below. Again this might cause some initial confusion. Aren't floors below us? How can a price floor be above the market-clearing price? To remove the confusion in the case of the price ceiling we thought of the price as a helium balloon trying to rise. In the case of a price floor we should think of the price as a rock (or any heavy object) that we drop from our hand. Where will the rock end up? On the floor, of course.

A price floor in the supply and demand model is shown in Figure 3-25. Because it is above the market-clearing price, the price floor (P_f) creates excess supply. We know from our theory of excess supply that price will fall under normal market conditions in this case, and the temporary excess supply will vanish. But with a price floor, the price cannot fall to its equilibrium level, and the temporary excess supply becomes permanent. When this happens we have the opposite of a shortage, a surplus. The quantity supplied at the price floor is greater than the quantity demanded.

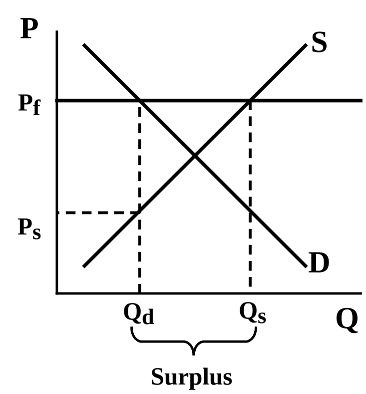


Figure 3-25
The Supply and Demand Model with a Price Floor

Note again the other result of our theory, similar to that in the case of a price ceiling. The amount exchanged with a price floor will be lower than would be under conditions where price is allowed to reach its equilibrium level. At the price P_f suppliers wish to sell Q_s, but the quantity demanded is not sufficient to buy this amount of the good. Buyers in the market only buy the amount Q_d. What happens to rest, the surplus? Often it just piles up.

The best examples of price floors come from government policy toward agriculture in the U.S. and almost in any country producing agricultural products around the world. In an attempt to bolster farmers' incomes, governments establish price floors to increase the price of agricultural commodities. Wheat, corn, dairy products, cotton, and peanuts are just a few crops that have experienced the surplus-generating affects of price floors.

Theory would suggest that a black market might develop in response to a price floor like it does in response to a price ceiling. At the quantity Q_d, sellers only require a price of P_s to offer that amount for sale. This price is below the price P_f that buyers are willing to pay, so the possibility for mutually beneficial black market trades exists. This is what happens when farmers try themselves to set minimum prices for their products. Such actions sow the seeds of their own destruction, however, as individual farmers cheat and engage in black market sales below the floor price. Something similar often occurs when oil producing nations who are members of the Organization of Petroleum Exporting Countries (OPEC) violate production quotas and sell below the official OPEC price.

Black markets are much less likely to develop in the case of price floors established by the government, because of the way the price floor is applied. Rather than passing a law saying the price can be no lower than the price floor, the government often enters the market as a buyer willing to pay the floor price. Now the surplus does not build up on the farm, but rather gets transferred to the government. The government then disposes of the surplus in some manner. In the U.S., we have dairy distribution programs to give or sell cheese at low prices to poor households. We also give food aid abroad using the surplus purchased in the price support program. This practice is becoming increasingly controversial, as developing countries receive lower prices for their agricultural outputs when the U.S. and other industrialized countries "dump" surplus commodities on the world market.

A Labor Market Application of a Price Floor

We will examine resource markets in more detail in Chapter 7, but we can't pass up the opportunity to examine one of the most well known uses of price floors, the minimum wage. The wage is the price of labor. We see the price floor model applied to a labor market in Figure 3-26. Now we have W, the wage, on the vertical axis, and L the amount of labor on the horizontal axis. The supply curve of labor is upward sloping, as people are willing to work more when wages are higher. The demand curve is downward sloping, as employers are willing to hire more workers if wages are lower. The equilibrium wage occurs at W_e where the quantity of labor demanded just equals the quantity supplied.

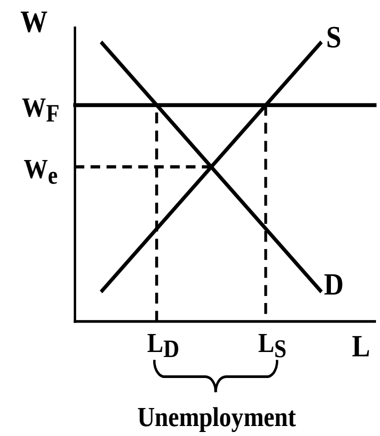


Figure 3-26
The Supply and Demand Model with a Price Floor

Now suppose the government, in an attempt to raise the incomes of lower income workers sets a wage floor, a minimum wage, at the level W_f. Now the quantity of labor supplied L_s is greater than the quantity of labor demanded L_d. A surplus of labor exists, which we call unemployment. Our theory of the minimum wage suggests that a minimum wage will increase the level of unemployment in the market in which it is applied.

The Price Elasticity of Demand

To this point our theory of demand has covered the distinction between demand and quantity demanded and the effect of a change in a determinant of demand or price. We now go one step further and consider the relative magnitudes of these changes. In economics we use the word elasticity to describe this kind of relationship.

A very important elasticity is the price elasticity of demand. Figure 3-27 shows two alternative definitions of the price elasticity of demand. The most common definition is the absolute value of the percent change in quantity demanded divided by the percent change in price. We compute price elasticity as an absolute value because price and quantity demanded always move in opposite directions. This makes the fraction in the elasticity formula negative, and comparing the magnitudes of negative numbers adds needless difficulty that can lead to confusion.

$$E_d \equiv \left| \frac{\%\Delta Q_d}{\%\Delta P} \right|$$

$$E_d \equiv \left| \frac{dQ_d}{Q_d} \Big/ \frac{dP}{P} \right| \equiv \left| \frac{dQ_d}{dP} \frac{P}{Q} \right|$$

Figure 3-27
Definitions of the Price Elasticity of Demand

Please note that the numerator and denominator in the elasticity formula are in percent or proportional terms. The price elasticity of demand is NOT the magnitude of the change in quantity demanded divided by the magnitude of the change in price. This is not elasticity, but rather just the slope of the demand curve. We'll have more to say about the relationship of the slope of a demand curve and the price elasticity of demand below.

The second form of the definition in Figure 3-27 is one using the notation of differential calculus. Two aspects of this definition are instructive. First, the differential calculus implies small changes in price and quantity demanded at a particular point on the demand curve. Sometimes we even call this formulation point elasticity, and distinguish it from the case where we have larger changes in P and Q. When we have larger changes in P and Q we use the term arc elasticity. This distinction is only important when we want to calculate a precise number for the price elasticity of demand. In the case of larger changes in P and Q we have the problem of which P and Q to use in the calculations, the initial values or the ending values. In arc elasticity, an average of the initial and ending value is used in the calculation. The second instructive aspect of the calculus formulation comes from the explicit appearance of the price and quantity magnitude, which indicates that even if the slope of the demand curve (dQ/dP) remains the same, the price elasticity of demand changes with different price and quantity combinations.

In Figure 3-28 we find the general categories of the price elasticity of demand. If the price elasticity of demand is greater than 1, demand is price elastic. If less than one, demand is price inelastic. If the price elasticity of demand is equal to 1, we say the elasticity is unitary. Common understanding of the word elasticity helps in understanding the price elasticity of demand. If something is elastic, it stretches when you pull on it. If it's inelastic, it stretches much less when you pull on it. Just think of a certain percent price change as the pull, or the tug. If the stretch (percent change in quantity demanded) is greater than the pull or tug, demand is price elastic. If the stretch is less than the pull or tug, again measured in percent terms, demand is price inelastic. Go back to Figure 3-27, look at the elasticity formula, and confirm for yourself that if the percent change in the quantity demanded is greater than the percent change in price, the fraction in the formula will be greater than 1, and when the percent change in quantity demanded is less that the percent change in price the fraction will be less than 1.

If $E_d > 1$, then price elasticity of demand is elastic

If $E_d < 1$, then price elasticity of demand is inelastic

If $E_d = 1$, then price elasticity of demand is unitary elastic

Figure 3-28
Elastic, Inelastic, and Unitary Elastic Demand

A very important application of the price elasticity of demand involves the relationship between the elasticity, price change, and the change in total revenue received from selling a good. Suppose you are a seller of a good and you have some pricing power, i.e., you can raise your price and not have the amount you sell go to zero. We will discuss this kind of power in more detail in Chapter 6 when we use pricing power to help in the definition of something called market structure. Normal people often mistakenly assume that raising a price always leads to an increase in revenue. This is only true if demand is price inelastic.

Figure 3-29 summarizes the relationship between price elasticity, price change and the change in revenue. Rather than memorize these relationships you should try to understand why they occur. Because total revenue from selling a good is the product of price and quantity (TR = P x Q), and price and quantity demanded

If	Then
$E_d > 1$	$P\uparrow \rightarrow Q\downarrow$ and $TR\downarrow$ $P\downarrow \rightarrow Q\uparrow$ and $TR\uparrow$
$E_d < 1$	$P\uparrow \rightarrow Q\downarrow$ and $TR\uparrow$ $P\downarrow \rightarrow Q\uparrow$ and $TR\downarrow$
$E_d = 1$	$P\uparrow \rightarrow Q\downarrow$ and TR same $P\downarrow \rightarrow Q\uparrow$ and TR same

Figure 3-29
Price Elasticity of Demand, Price Change, and Change in Total Revenue of Expenditure

move in opposite directions, the effect of a price change on total revenue is the result of a tug-o-war between opposing forces. Think first of a price increase. If price goes up relatively more than quantity falls, then revenue will rise. Suppose price is $10 and initial quantity is ten units. Initial revenue is $100. Now suppose we raise price to $12 and quantity falls to 9 units. Revenue is now higher, $108. Note that the price change won the tug-o-war. It rose 20%. Quantity demanded fell only 10%. We know that demand was inelastic. Just plug 10% into the numerator of the price elasticity formula and 20% into the denominator. The value is less than 1, inelastic.

If we reverse the percent changes and price rises 10% and quantity demanded falls 20%, then demand is elastic. With the same initial starting point ($10 and 10 units), price is now $11 and quantity demanded is 8 units. Quantity now wins the tug-o-war and revenue falls. In this case the bad effect, quantity falling, outweighed the good effect, price rising.

When we lower price to raise revenue, we are hoping that demand is price elastic. Price falling is the bad thing in this case. Quantity rising is the good thing. If the good thing is relatively larger in percent terms than the bad thing, the net effect is an increase in revenue. And if the percent change in quantity demanded is greater than the percent change in price, demand is price elastic, by definition.

We also see in Figure 3-29 that a price elasticity of 1 results in no change in revenue when we change price. By definition, a price elasticity of 1 means that the percent change in quantity demanded is exactly equal to the percent change in price. The percent change in the good thing is exactly balanced by the percent change in the bad thing in the opposite direction. The tug-o-war between price and quantity demanded is a standoff. Revenue remains the same.

Price Elasticity of Demand, the Slope of a Demand Curve, and Revenue Rectangles

When first confronted with the theory of the price elasticity of demand, students often want to model price elasticity as the slope of a demand curve. This is not a good idea, as shown in Figure 3-30. The demand curve in Figure 3-30 has the same slope throughout its entire length, but the elasticity of demand is different at each point on it. To see this, we need to introduce the revenue rectangle, the graphical representation of total revenue.

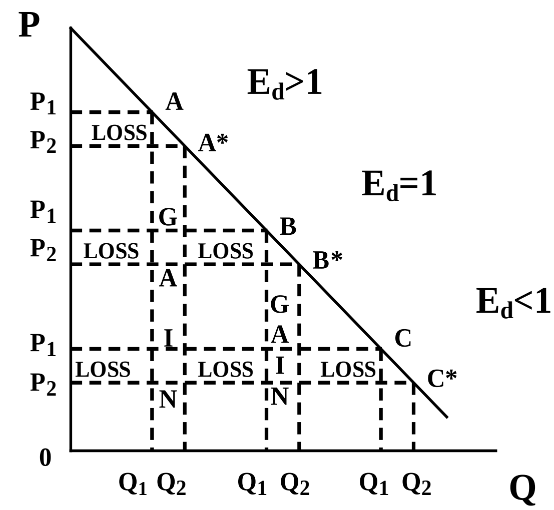


Figure 3-30
Price Elasticity of Demand at Different Points on a Straight Line Demand Curve

On a graph, revenue is represented as an area of a rectangle. The length of the vertical side of the rectangle is price, and the length of the horizontal side is quantity. For example, look at point A on the demand curve in Figure 3-30. The price is P_1 and the quantity demanded is Q_1. The initial revenue (P_1 x Q_1) is the area of the rectangle OP_1AQ_1. Now suppose the price falls to P_2, and quantity demanded increases to Q_2. The new revenue (P_2 x Q_2) is the area of the rectangle $OP_2A^*Q_2$. The new revenue rectangle is larger than the initial one.

It is often easier to see the net effect of a price change in the model by looking at rectangles representing gains and losses in revenue, rather than the relative size of the revenue rectangles themselves. Look at the initial point A again. At the price P_1, buyers are buying Q_1 of the good. If we lower the price to P_2, and we assume everyone pays the same price, we now sell the initial Q_1 quantity at a lower price, P_2. This is a loss in revenue indicated by the rectangle labeled LOSS. But this is only half the story. At a lower price, buyers by more of the good, $Q_2 – Q_1$ more to be precise. This is a gain in revenue indicated by the GAIN label in the vertical rectangle. At point A, it is easy to see that the gain in revenue from a price reduction is greater than the loss of revenue. The higher quantity at the lower price more than compensates for the reduction in price to those buying the initial quantity.

We now can infer the price elasticity of demand at point A. We lowered the price and revenue rose. The percent change in the good thing, quantity increasing, outweighs the percent change in the bad thing, price decreasing. Demand is price elastic. The percent changes in price and quantity can also be inferred from the magnitude of the change relative to the initial level. At point A, $P_1 - P_2$, the change in price is a small proportion (percent) of the initial price P_1, while the change in quantity demanded, $Q_1 - Q_2$, is a much larger proportion (percent) of the initial quantity Q_1.

Now look at point C in Figure 3-30. The result is strikingly different than that at point A. When the price falls from P_1 to P_2, it falls for a large initial quantity. The LOSS rectangle from the price reduction is much larger than the gain rectangle coming from more quantity demanded the lower price. At point C, demand is price inelastic. A reduction in price leads to a decrease in revenue. Note again that the slope of the demand curve is the same at point A and point C, but the price elasticity of demand differs greatly.

At point B in the middle of the demand curve the LOSS rectangle from a price reduction is just matched by the GAIN rectangle from the increase in quantity. Note that the proportional (percent) change in price and quantity are roughly the same. The fraction in the price elasticity formula is now equal to 1. Price elasticity of demand at point B is unitary.

While we looked only at price reductions in Figure 3-30, we could have posited price increases. Check for yourself that when we reverse the P_1 and P_2 the GAIN rectangles turn into LOSS rectangles and the effect on revenue changes. It is also instructive to compare the visual results of Figure 3-30 with the relationships summarized in Figure 2-29.

Other Demand Elasticities

Figure 3-31 contains the definition of two additional demand elasticities. The first, income elasticity of demand, measures the percent change in quantity demanded from a

Income Elasticity of Demand $$E_I \equiv \frac{\%\Delta Q_d}{\%\Delta I}$$

Cross Price Elasticity of Demand $$E_{1,2} \equiv \frac{\%\Delta Q^1_d}{\%\Delta P^2_d}$$

Figure 3-31
Other Elasticities of Demand

percent change in income, with price and the other determinants of demand held constant. Note that this elasticity, while a different concept than price elasticity, has the same general form, the ratio of two percent changes. With income elasticity, the key reference point is now 0 and not 1. Demand is income elastic if the elasticity is greater than zero. In this case, a positive increase in income leads to a positive increase in demand. A normal good has a positive income elasticity of demand. An inferior good has a negative income elasticity.

The cross price elasticity of demand is about the effect of changes in the price of related goods, ones we define as substitutes and complements earlier in the chapter. Here we have the notation for good 1 and good 2. $E_{1,2}$ is the percent change in the quantity demanded of good 1 divided by the percent change in the price of good 2. Remember that the demand for a good rises when the price of a substitute good rises. Think of the Coke and Pepsi example again. If the price of Pepsi rises, this causes an increase in the demand for Coke. These substitute goods have a positive cross price elasticity of demand. If the price of a substitute rises (magnitude in the denominator), the demand for the good in question (magnitude in numerator) moves in the same direction. In contrast, complementary goods have a negative cross price elasticity of demand.

The Price Elasticity of Supply

If we have a price elasticity of demand, we must also have a price elasticity of supply. This definition appears in Figure 3-32. Note that the definition of the elasticity of supply has the same form as the other elasticities we have studied, the ratio of two percent changes. Because price and quantity supplied are positively related, we don't need the absolute value notation in the definition as we did with the price elasticity of demand. However, the same conventions regarding elastic, inelastic, and unitary elasticity apply here. These appear in Figure 3-33.

$$E_s \equiv \left| \frac{\%\,\Delta Q_s}{\%\,\Delta\, P} \right|$$

$$E_s \equiv \left| \frac{dQ_s}{dP} \frac{P}{Q} \right|$$

Figure 3-32
Definitions of the Price Elasticity of Supply

We model two special extreme cases of the price elasticity of supply in figure 3-34. A vertical supply curve has a price elasticity of supply equal to zero. While a vertical demand curve is so rare that it is almost nonexistent, a vertical supply curve is relatively common, especially if the time period under consideration is short enough. Suppose you sell fresh raspberries in a farmers market in Phoenix. You bring a certain amount of raspberries to the market and are willing to sell them at whatever price they will bring. This is because if you don't sell them the raspberries will rot in the heat on the way home. You are going to sell those raspberries at any positive price. For this reason, this example is often referred to in economics as the market period case.

If $E_s > 1$, then price elasticity of supply is elastic

If $E_s < 1$, then price elasticity of supply is inelastic

If $E_s = 1$, then price elasticity of supply is unitary elastic

Figure 3-33
Elastic, Inelastic, and Unitary Elastic Supply

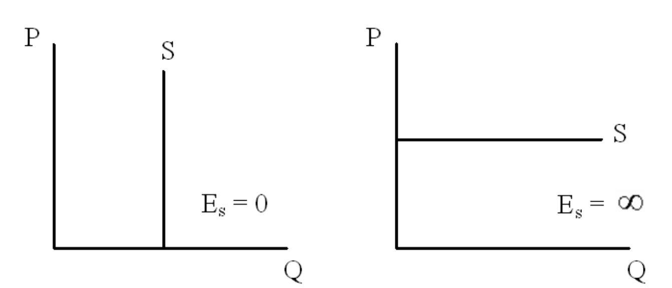


Figure 3-34
Two Special Cases of the Price Elasticity of Supply

The other supply curve in Figure 3-34 is totally flat, which means an infinite elasticity of supply. This case if often used to illustrate the long-run response to an increase in price. In the market period, quantity supplied is fixed, but in the long run, given enough time for supplies to respond by adding more production capacity and for new suppliers to enter a market, a horizontal supply is feasible.

The relationship of the time period and supply elasticity is shown in the context of a supply and demand model in Figure 3-35. The initial equilibrium is at price P_1 and quantity Q_1, at the intersection of the demand curve D_1

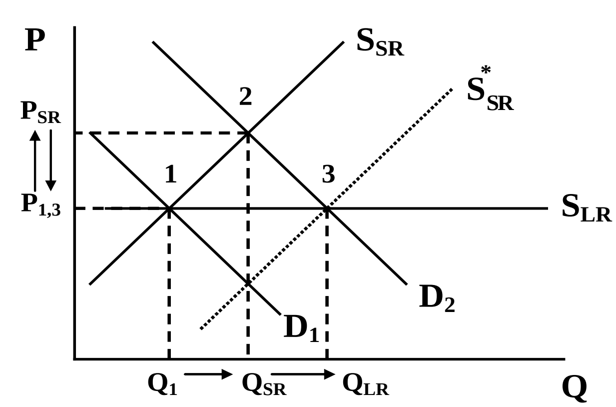


Figure 3-35
Long-run and Short-run Elasticity of Supply

and a short-run supply curve labeled S_{SR}. Now suppose demand increases, represented in the model by a shift of the demand curve to the right to D_2. The short-run response to the price increase is an increase in the quantity supplied to Q_{SR} and an increase in the price to P_{SR} and a new intersection of the supply and demand curves at point 2. But given enough time suppliers can add new production capacity and new suppliers can enter the market, shifting the supply curve to the right to the new curve $S_{SR}*$, with a new intersection at point 3. The price now has fallen back to $P_3 = P_1$. The horizontal line Labeled S_{LR} represents a long run supply curve with infinite price elasticity.

Differences between short-run and long-run elasticities are very important in the real world. Generally, time is on the side of price elasticity of supply, and the price elasticity of demand as well. A local apartment market again provides a good example. With higher enrollment in a college, demand for apartments increases in a college town. In the short-run, rents increase like the price in Figure 3-35, along with the grumbling of normal college students. Economics students, on the other hand, temper their grumbling with the realization that the increase in rents is necessary in the market to signal the increase in demand to suppliers and potential suppliers. After a wave of apartment construction, rents return to their original level.

Chapter 4: Marginal Analysis and a Theory of the Consumer

Introduction

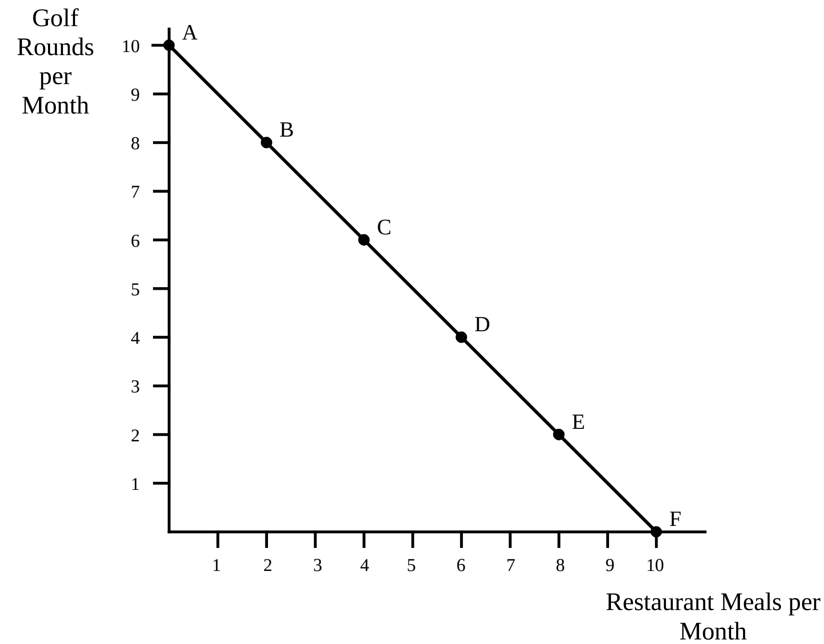
In this chapter, we move behind the demand side of the market and examine one important component of individual choice, consumer behavior. This is the most microeconomic topic in economics. No aggregation exists. As we saw in Chapter 1, economists assume that an individual practices rational behavior by acting to maximize his or her well being. While nonmarket actions such as dishwashing, lawn mowing, hiking and cooking occupy a large amount of an individual's time, we ignore these actions here and focus on choices about buying goods and services in the marketplace.

Please do not interpret this focus on buying behavior as a judgement that nonmarket activity is unimportant, either to the individual or to the economist. Indeed, Gary S. Becker won the Nobel Prize in Economics in 1992 for his pioneering work studying the allocation of time to all kinds of activities, both market and nonmarket.

Consistent with our emphasis on scarcity as the foundation of economics, we begin in the next section by looking at a model highlighting the constraint on consumer choice, the consumer's budget. We postpone until Chapter 6 a discussion of how the consumer obtains the income upon which this constraint is based.

The Budget Line Model

Figure 4-1 represents the consumption possibilities for a hypothetical individual, Professor Pete. As in all good models, we abstract from reality and assume that Professor Pete has only two goods in his world, rounds of golf and restaurant meals. We also look at his consumption choices for an entire month. In the bottom part of Figure 4-1 we see some numbers representing the amount of each good Professor Pete can purchase in a month, and the expenditure on each good. Of course, the amount of each good and the amount spent on each depends on three important pieces of information, the price of golf rounds, the price of restaurant meals, and Professor Pete's overall budget. We assume the price of each good is \$20, and that Professor Pete's monthly recreation budget is \$200.



Professor Pete's Consumption Possibilities with Recreation Budget of $200 per Month

	Golf Rounds at $20 Each		Restaurant Meals at $20 Each	
	Quantity	Expenditure	Quantity	Expenditure
A	10	$200	0	$ 0
B	8	$160	2	$ 40
C	6	$120	4	$ 80
D	4	$ 80	6	$120
E	2	$ 40	8	$160
F	0	$ 0	10	$200

Figure 4-1
The Budget Line Model

Down the leftmost column in the table at the bottom of Figure 4-1 we see the letters A, B, C, D, E, and F. These letters represent choice combinations of both goods that use up Professor Pete's entire monthly budget. We assume Professor Pete doesn't save any of his budget. For example, with choice A, Professor Pete buys 10 rounds of golf and no restaurant meals.

Note that this choice exhausts his budget ($20 X 10 = $200). To simplify the analysis we don't show all possible combinations, such as 9 rounds of golf and 1 restaurant meals, but rather show consumption possibilities in increments of two. With choice B, Professor Pete would buy 8 rounds of golf and two restaurant meals. At choice C, the combination is 6 rounds of golf and 4 restaurant meals, and so on, until we see that with choice F Professor Pete buys only restaurant meals and no rounds of golf. Note that at all choice combinations A through F Professor Pete spends his entire budget.

The graph at the top of Figure 4-1 simply represents in a graphical model the information in the table. We have assigned the vertical axis to golf rounds per month, and the horizontal axis to restaurant meals per month, but we could have switched the axes if we had wanted to. Note the nature of this budget line. It looks like a production possibility frontier without the law of increasing cost The similarity is no accident. The production possibility frontier is a constraint in production, either for an economy producing guns and butter or a student producing grades and goodies. The budget line is a constraint for the consumer, but this time not a constraint on production, but rather a constraint on consumption.

Professor Pete would like to be way up in the northeast corner of the model, but these choices are not attainable. He has to be on his budget line. He could spend his entire $200 budget on rounds of golf and consume 10 of them (point A), or he could spend all his budget on restaurant meals and consume 10 of them (point B). Or by trading off golf for meals, or meals for golf, he can be at other points on his budget line. He can't be inside his budget line because we have assumed that he spends his entire budget.

Note that the budget line is downward sloping reflecting the explicit opportunity cost to Professor Pete of one good in terms of the other. The slope of this budget line is equal to –1, because the prices of the two goods are the same. In order to buy another restaurant meal Professor Pete has to have $20. He can get this $20 by giving up one round of golf and freeing up $20 possible spending.

Representing Price and Budget Changes in the Budget Line Model

In the initial budget line model, we assumed a given budget of $200 per month and a price of $20 for each good. Later in this chapter, however, we will see that one of the important applications of the model involves changing the consumer's budget or the prices of the goods. The various panels of Figure 4-2 show how we incorporate such changes in the model.

In panel (a) we increase Professor Pete's monthly budget to $300 leaving both prices unchanged. Instead of a maximum of 10 golf rounds OR 10 restaurant meals, the intercepts on the axes of the original model, after the budget increase to $300 Professor Pete can buy 15 rounds of golf OR 15 restaurant meals. Note that the slope of the budget line remains the same, because neither price has changed. A decrease in Professor Pete's budget would be represented by a parallel shift inward of his budget line.

Panels (b) and (c) show the effect on the model of a change in one of the prices holding the total budget constant. In panel (b) the price of restaurant meals falls to $10 while the price of golf rounds remains the same. The budget line still intersects the golf axis of the model at 10 rounds, because we assume that price has not changed, i.e., if Professor Pete spends his entire $200 budget on rounds of golf, he would still be able to buy only 10. But now restaurant meals are cheaper. If Professor Pete chooses to spend his entire budget on restaurant meals, he can buy 20 ($200 budget/$10 per meal). Note that the change in the price of restaurant meals has altered the slope of the budget line. It is now flatter.

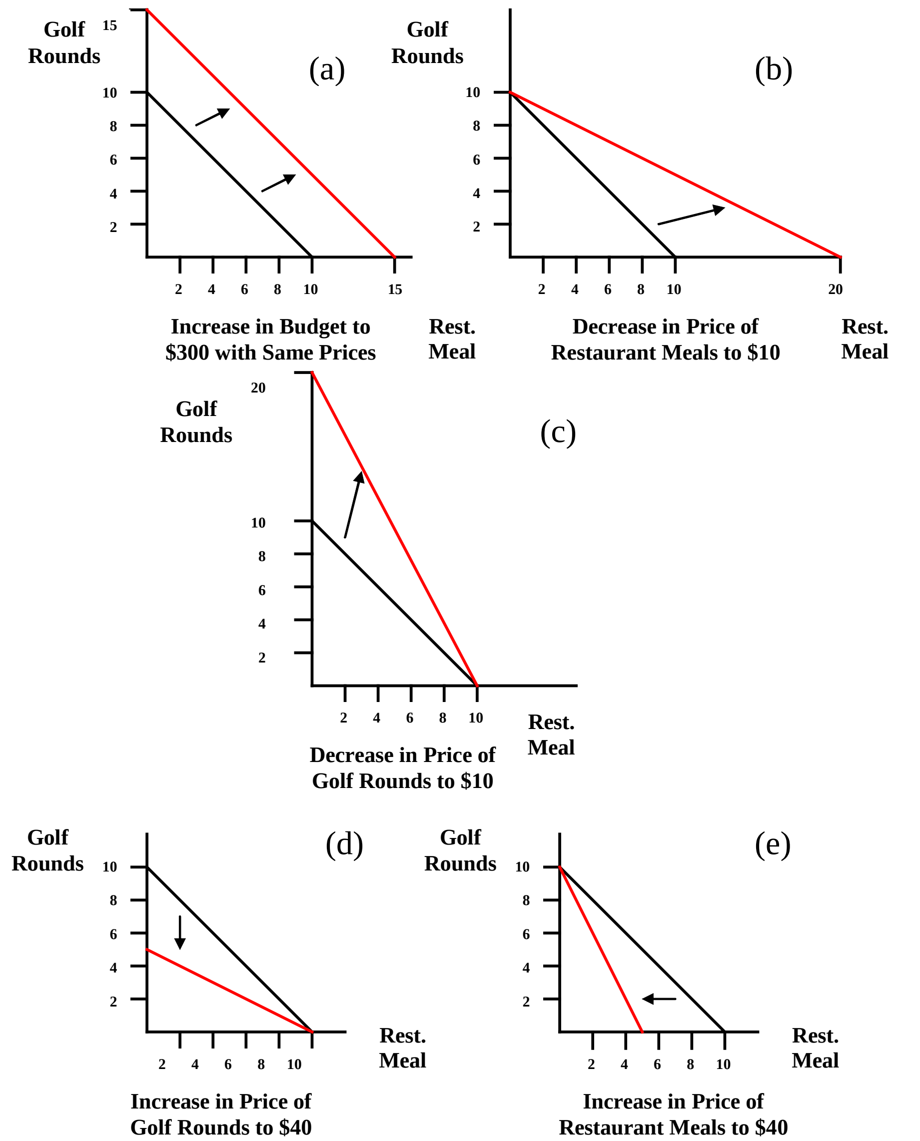


Figure 4-2
The Budget Line Model with Changes in Prices or Budget

Panel (c) shows the opposite case, where the price of restaurant meals remains the same and the price of golf rounds decreases to $10. Note that the budget line still intersects the restaurant meal axis at 10 meals. If he spent all of his $200 budget on restaurant meals, the maximum he could buy would still be 10. The intercept of the budget line on the golf rounds axis, however, now changes. At a price of $10 per round of golf, he can now buy 20 of them if he spends his entire budget on golf. Again the slope of the budget line changes because of the price change, but this time it becomes steeper, not flatter.

Panel (d) shows the effect of an increase in the price of golf to $40 per round, with the budget and the price of restaurant meals remaining unchanged. Panel (e) shows the effect of an increase in the price of restaurant meals to $40, again with the other price and the budget remaining unchanged.

Point on Budget Line	Q	Total Utility (TU)	Marginal Utility (MU)	MU/$
A	0	0	400	20
	1	400	300	15
B	2	700	200	10
	3	900	100	5
C	4	1000	80	4
	5	1080	60	3
D	6	1140	40	2
	7	1180	30	1.5
E	8	1210	20	1
	9	1230	10	.5
F	10	1240		

Figure 4-3
Total Utility, Marginal Utility, and Marginal Utility per Additional Dollars Spent on Restaurant Meals

Representing Consumer Preferences with a Utility Function

Figures 4-1 and 4-2 represented only the constraint on Professor Pete's choices. All of the points on his budget line are possible choices. But how can we model Professor Pete's particular choice that maximizes his well being? We need to know something about Professor Pete's preferences for golf and restaurant meals. We do this by introducing the abstract notion of utility.

Figures 4-3 and 4-4 give information about Professor Pete's preferences for restaurant meals. The first column in Figure 4-3 shows the points on the budget line in Figure 4-1. At point A, Professor Pete buys no restaurant meals. At point B, he buys 2, and so on, until at point F he is spending his entire $200 budget on restaurant meals. The

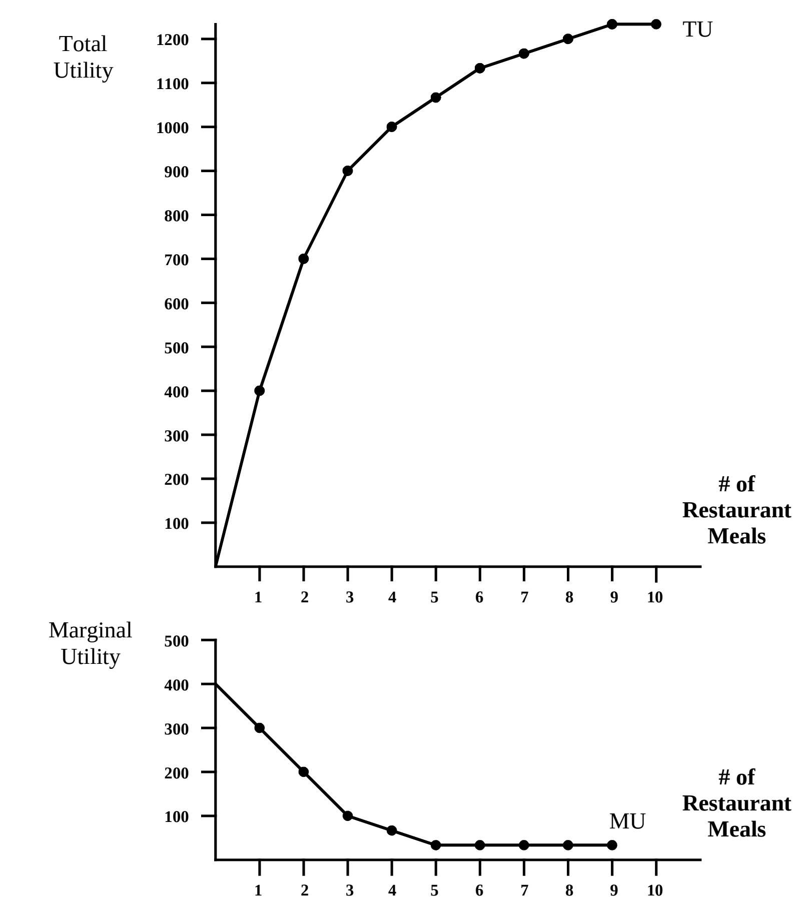


Figure 4-4
Total Utility and Marginal Utility Models for Professor Pete's Restaurant Meal Consumption

quantities 0 – 10 are shown in the second column. In the third column, we introduce information on Professor Pete's preferences.

We represent preferences in the model with the concept of total utility. Total utility (TU) is a representation of well being. The higher the total utility the higher the well being. In our model, a rational consumer will try to maximize his or her total utility. Note that as Professor Pete consumes more and more restaurant meals his total utility from restaurant meals rises. Consuming 0 meals gives 0 total utility. Consuming 1 meal yields a TU of 400, 2 meals 700, and so on, until Professor Pete receives a TU of 1,240 from 10 restaurant meals.

The column labeled Marginal Utility gives us additional information about Professor Pete's preferences for restaurant meals. You will find that the word "marginal" is a very important one in economics. The term marginal is always related to some kind of change, such as changing the amount of something you consume by buying more or less of it. Anything marginal in economics relates to change and change is the result of choice. That's why the term marginal is so important.

Marginal utility (MU) is the change in total utility achieved from a change in consumption of a good. In this case, we will give MU information for increments of one unit of the good. We can define marginal utility more formally as follows:

$$MU = dTU/dQ,$$

where the symbol d means "change in." So when Professor Pete is consuming 0 restaurant meals, he can increase his TU to 400 by consuming one meal, his first one. His TU rises from 0 to 400 with this action. So at a quantity of 0 his MU is 400. Likewise, at a quantity of 1, and a TU of 400, he can increase his TU to 700 by consuming one more restaurant meal. The change in TU by taking this action is 300. Therefore, we say at a quantity of 1 his MU is equal to 300. If we calculate the change in TU from one consumption level to the next, we can fill out the entire Marginal Utility column in Figure 4-3.

A very important concept appears in the numbers of the marginal utility column of Figure 4-3. As the example is constructed, marginal utility at all levels of consumption is positive, but as the quantity of restaurant meals consumed increases the magnitude of the marginal utility decreases. Each additional meal consumed adds to total utility, but each additional meal adds less additional utility. This is the second diminishing returns concept we've seen so far. The first was the law of increasing cost associated with the production possibility frontier. We call this one diminishing marginal utility.

We will make use of the right most column of Figure 4-3 below. Let's just define it here. The numbers in this column are the result of dividing the marginal utility by the price of an additional unit of consumption, in this case the marginal utility of restaurant meals divided by the price of restaurant meals. The units of this number, change in total utility per additional dollar spent, indicate how much additional utility we can achieve by spending more on the good. It's a "bang for the additional buck" concept. By comparing the "bang for additional buck" across goods, we will be able to see what combination of restaurant meals and rounds of golf maximizes Professor Pete's total utility.

In Figure 4-4, we represent the information of Figure 4-3 with graphical models. The model at the top of the figure is a graph of Professor Pete's total utility from consuming restaurant meals. Note that the more restaurant meals Professor Pete consumes, the higher is his total utility. This is another way of representing our fundamental economic assumption that more is preferred to less.

But note the slope of the TU function in Figure 4-4. The slope decreases as the quantity of restaurant meals increases. But what is this slope? Remember that the slope of a line on a graph is the "rise over the run," a change in a magnitude on the vertical axis divided by the corresponding change in the variable on the horizontal axis. But since the vertical axis is total utility, and the horizontal axis is quantity of restaurant meals, the slope of the TU curve is marginal utility. Remember that above we defined marginal utility as $MU = dTU/dQ$.

The lower portion of Figure 4-4 simply graphs the marginal utility numbers of Figure 4-3. As we have constructed the example, over the range of consumption we are considering, Professor Pete's MU is positive, but declining as the quantity of restaurant meals increases.

It is possible for marginal utility to become negative in some situations. I'll spare you the ugly details, but suffice it to say that this might be called the "Economics of Throwing-Up." Think of someone eating hotdogs in a one-hour period. After 20 or 25 hotdogs total utility might actually peak and begin to fall. A falling TU curve means the MU is negative. More hotdog consumption means less utility, not more. Certainly, no rational economics student would consume at this level.

Point on Budget Line	Q	Total Utility (TU)	Marginal Utility (MU)	MU/$
F	0	0	4,000	200
	1	4,000	3,000	150
E	2	7,000	2,000	100
	3	9,000	1,000	50
D	4	10,000	800	40
	5	10,800	600	30
C	6	11,400	400	20
	7	11,800	300	15
B	8	12,100	200	10
	9	12,300	100	5
A	10	12,400		

Figure 4-5
Total Utility, Marginal Utility, and Marginal Utility per Additional Dollars Spent on Golf Rounds

Figures 4-5 and 4-6 contain the same type of information about Professor Pete's preferences, but now for rounds of golf. Two differences appear, however. Note in Figure 4-5 that the letters corresponding to points on the budget line start at F and go up the alphabet to A, reflecting the appropriate point on Professor Pete's budget line. Note also that the total utility information for golfing has much higher TU numbers for golf, as compared to restaurant meals. Each round of golf has 10 times the total utility to Professor Pete than a restaurant meal has. For example, at one round of golf the TU for golf is 4,000, where the TU of one restaurant meal is 400. The TU of two rounds of golf is 7,000, where the TU of 2 restaurant meals is 700, and so on. Of course, this changes the marginal utility and marginal utility per additional dollar numbers as well.

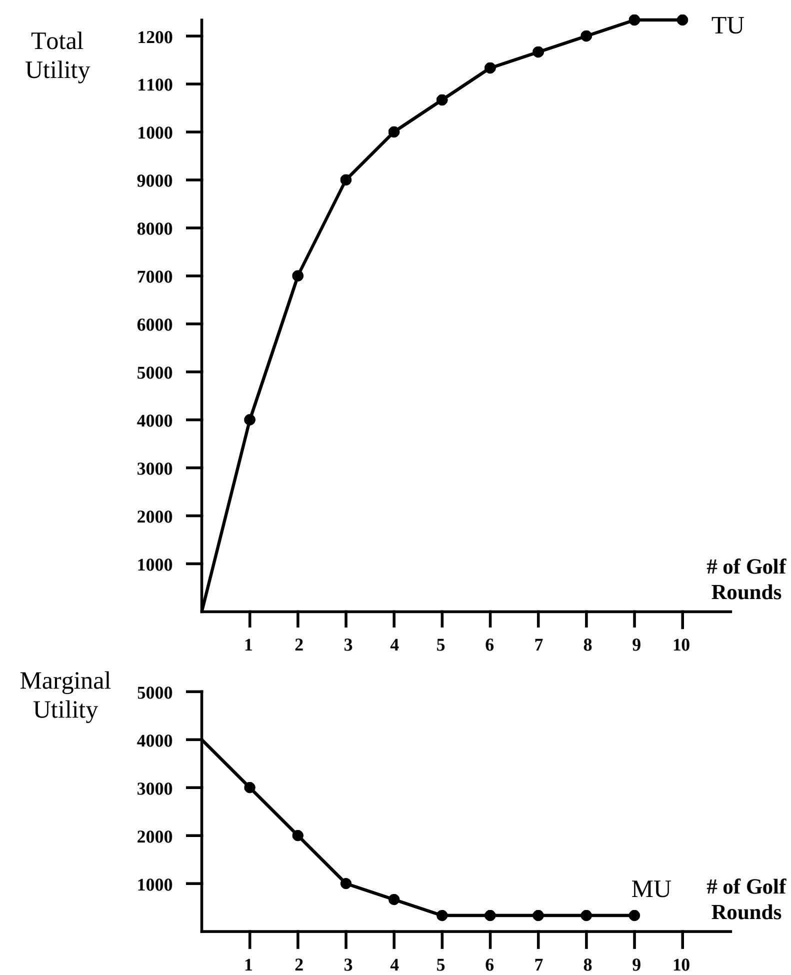


Figure 4-6
Total Utility and Marginal Utility Models for Professor Pete's Golf Round Consumption

Constrained Utility Maximization with the Equimarginal Principle

Now we are ready for the exciting part of our Professor Pete story. This is modeling his utility-maximizing choice. First, let's look at the terms in the title to this section. This is a problem in constrained utility maximization. Professor Pete, assumed here to be a rational individual, wants to make his utility as large as possible. He does not have free rein, however, in choosing restaurant meals and rounds of golf to maximize his utility. He has to live within his budget constraint. He only has $200 with which to maximize his utility.
The second term, equimarginal principle, is one of the most important, and unfortunately, one of the most tricky and difficult ideas in economics. The prefix "equi" means equal. The equimarginal principle means that in order to maximize utility Professor Pete has to have some marginal magnitude equal to another marginal magnitude. We will see as we progress through microeconomics that the condition where marginal something equals marginal something else is very common. But let's return to the problem of maximizing Professor Pete's utility.

Let's start with a simple question. Suppose you have $1 to spend on either of two goods X or Y. You can't have both. Suppose that by spending $1 on good X you could increase your total utility by 500 and that by spending $1 on good Y you could increase your total utility by 50. What would you buy? It's clear that you would buy good X and increase your total utility by 500. Spending the $1 on good Y would only increase your utility by 50. The change in your total utility, marginal utility, per additional dollar spent on good X would be 500. The marginal utility per additional dollar spent on good Y would be 50. Would it not be rational for you as a consumer to always choose an action with the highest marginal bang for the additional dollar spent on a good? This is the principle we will now apply to Professor Pete's utility maximization problem.

Figure 4-7 is simply the top part of Figure 4-1 with one additional piece of information. At each of the selected points A – F (and now also the points in between the points labeled with letters), we show in parentheses the marginal utility per dollar gained from purchasing and consuming another unit of each good. For example, look at point F, where Professor Pete would be consuming 10 restaurant meals and no rounds of golf. The first number in the parenthesis to the right of point F, 200, comes from the MU/$ column in Figure 4-5. At the point of 0 consumption of golf, consuming one additional golf round would increase total utility by 4000. But this additional round of golf costs $20. The number 200 is just 4000/20. The term NA appears for additional restaurant meals, as an additional meal is not possible in this example.

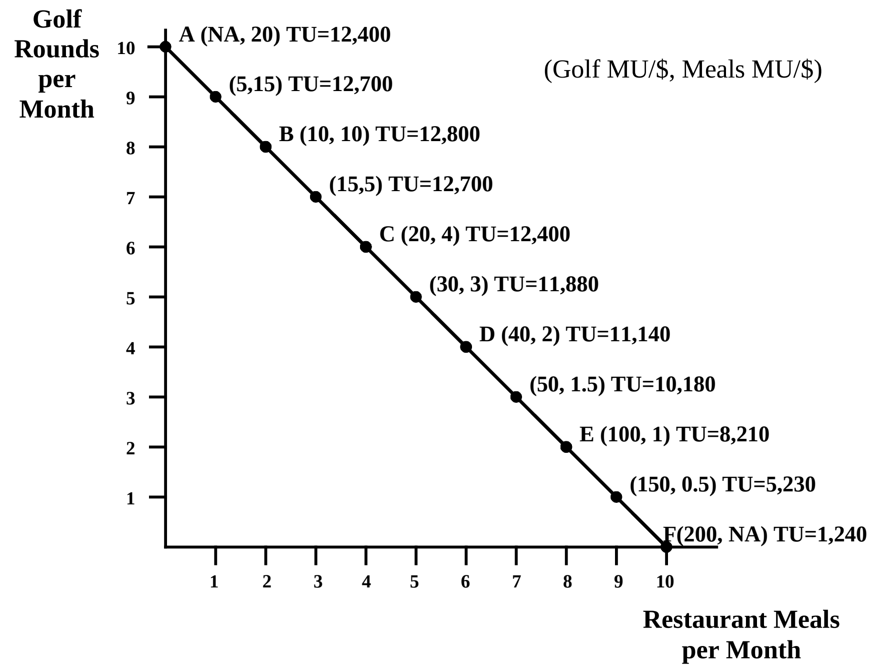


Figure 4-7
Professor Pete's Utility Maximizing Consumption of Restaurant meals and Golf Rounds

Although we have defined the terms inside the parenthesis at point F, our story is more instructive if we start at point E. If we look at Figure 4-5, we see that Professor Pete can increase his total utility by 2,000 by buying and consuming an additional round of golf. His MU for golf is 2,000 at this point. But he has to spend $20 for this round of golf. His MU per additional dollar spent is 100. We see this number in the rightmost column of Figure 4-5 at the quantity of 2 rounds, and inside the parenthesis next to point E on Figure 4-7. If we do the same thing for restaurant meals at point E, we see in figure 4-3 that the marginal utility per additional dollar spent on restaurant meals is only 1. This comes from dividing the MU of going from 8 to 9 restaurant meals, 20, by the $20 it costs for another meal.

Now comes the crucial question. Is Professor Pete maximizing his total utility? Let's apply the same "bang for the buck" reasoning we did for the good A and good B case above. By spending $20 more on golf, Professor Pete receives 100 TU per additional dollar. By spending $20 more on a restaurant meal he would receive 1 TU per additional dollar. Which is the better deal? Which gives more bang for the buck? Obviously golf. But Professor Pete needs $20 for another round of golf. Where can he get it? By spending less on restaurant meals. If we do a similar analysis at points D and C in Figure 4-7 (and the points in between if you like), we come to the same conclusion. It's better to spend additional money on golf, with the money coming from less spent on restaurant meals.

Finally, we do the analysis at point B, where Professor Pete would purchase and consume 8 rounds of golf per month and 2 restaurant meals. Note that the MU per additional dollar spent on each good is equal at point B. No case can be made for spending more on either good. Professor Pete's total utility is maximized at point B where the equimarginal principle is satisfied. The marginal utility per additional dollar spent on each good is the same.

Let's assume that we have what economists call an additive utility function, i.e., total utility from consuming some combination of goods is simply the sum of the total utilities from each utility function. The magnitudes also appear to the right of the points in Figure 4-7. We also have the total utility for the points between the points labeled with letters A – F. We can see that total utility from both goods is maximized at point B at a value of 12,800, and, of course, where the marginal utility per additional dollar spent on the goods is equal.

Note that the relative magnitude of golf rounds is much higher than restaurant meals for Professor Pete to maximize his utility. We can use our intuition to understand why this is the case. The price of the two goods is the same, but the utility from golfing is much higher. Professor Pete has to consume much more golf in order to drive down the marginal utility to a level that equalizes the marginal bang for the additional buck spent on restaurant meals.

Other Applications of the Equimarginal Principle: Equalizing Bang for the Buck at the Margin

We will return to other applications of the budget line model below, but while the equimarginal principle and consumer choice is fresh in our minds, it's worthwhile to see how the principle can be applied to other resource allocation situations. This will improve our understanding of both the theory of the consumer and resource allocation in general.

In general, the equimarginal principle can be extended to other allocation and choice examples with use of a simple pattern or template. In allocating a given amount of something across different choices in order to maximize something, one should allocate the resource such that the marginal (WHATEVER ONE IS MAXIMIZING) per (ADDITIONAL UNIT OF WHAT ONE IS ALLOCATING) must be equal across (WHATEVER THE CHOICE IS ABOUT). At first blush, this template is a little confusing, but with practice it will lead to an understanding of the equimarginal principle.

Let's apply the template to the case of the utility-maximizing consumer we studied above. We are maximizing utility. We are allocating dollars from a fixed budget. We are choosing how much of each good to buy. In this case the equimarginal principle is that a consumer should buy goods such that the marginal UTILITY, per ADDITIONAL DOLLAR SPENT ON EACH GOOD, is equal across GOODS.

Suppose you are a campaign manager in a statewide election campaign. You want to maximize the number of votes your candidate receives in the election. You have a fixed budget to allocate across the many voting districts in the state to achieve your goal. Let's follow the equimarginal template again. As a rational campaign manager, you would want to spend your campaign funds such that the marginal VOTES, per ADDITIONAL DOLLAR SPENT IN EACH VOTING DISTRICT, is equal across all VOTING DISTRICTS.

Suppose you were a conservation ecologist wishing to use a given amount of funds to preserve the maximum amount of an endangered species. You have several different habitat zones in which to try to save species. Again, follow the equimarginal template. As a species-maximizing ecologist you would want to allocate your budget such that the marginal NUMBER OF THE SPECIES SAVED, per ADDITIONAL DOLLAR SPENT IN EACH HABITAT ZONE, is equal across all HABITAT ZONES.

Finally, in all these cases, you should convince yourself that the principle works by asking the following question. What if the marginal something per additional unit of resource is NOT equal across whatever you are allocating across? Do a little numerical example. Suppose in voting district A we get 50 votes per additional dollar spent in the district. In district B, the bang for the buck is less at the margin, say 30 votes per additional dollar spent. Are you maximizing votes? The answer is no. By taking a dollar of spending away from district B you will lose 30 votes. But if you spend that dollar in district A, you get 50 more votes. That's pretty good deal, down 30 and up 50 for a net gain of 20 votes. The rational campaign manager will transfer spending at a net 20 votes per dollar as long as possible. Unfortunately, the net gains from such a transfer will not continue forever, as we probably have diminishing returns in campaign spending. As we spend more in district A, the marginal votes per dollar will decline. And, as we spend fewer dollars in district B, the marginal votes per additional dollar will increase. Somewhere between 50 and 30, say at 42 votes per additional dollar, the marginal votes per additional dollar will be equal and we will realize that no further reallocation of spending is warranted. We will have satisfied the equimarginal principle and our number of votes will be maximized.

Normal and Inferior Goods and the Budget Line Model

In Chapter 3, we learned that a change in the income of buyers affects the demand in a market. We also defined normal and inferior goods based on the direction of the demand change when income changes. Normal goods are those that a consumer buys more (less) of when income rises (falls), and inferior goods are just the opposite. As income rises (falls), a consumer buys less of an inferior good. We can use our budget line model to show this distinction.

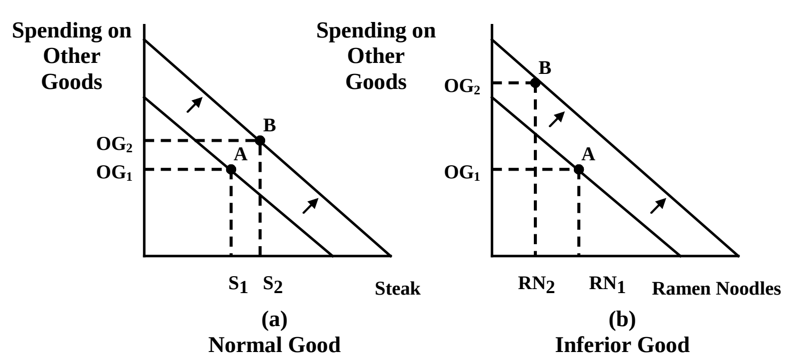


Figure 4-8
Normal and Inferior Goods and the Budget Line Model

The two panels of Figure 4-8 show the effect of an increase in income on a normal good and an inferior good, steak and Ramen noodles, respectively. The budget line models in Figure 4-8 have been modified slightly from the type we used for Professor Pete earlier in this chapter. Note that the quantities of the goods steak and Ramen noodles appear on their respective horizontal axes in the familiar way. But now on the vertical axis of the graph we have a composite good that we call spending on all other goods. Everything else about the budget line model is the same. The slope of the budget line represents the relative price of the good on the horizontal axis, and its position still represents the amount of income, which is the consumer's budget in this case.

Point A in both panels of Figure 4-8 represents the initial point on the budget line. In panel (a), the consumer purchases S_1 steak and spends OG_1 on other goods. The model also shows an increase in income with the budget line shifting outward. In panel (a), the case of a normal good, the consumer moves to point B and buys S_2 steak and spends OG_2 on other goods. As $S_2 > S_1$, steak is a normal good for this consumer. We say for this consumer, because a normal good for one consumer might be an inferior one for another consumer. We'll revisit this idea below. In panel (b) of Figure 4-8, we have a similar increase in income, represented again by a shift out of the budget line. In this case, however, $RN_2 < RN_1$. As income increases, the consumer buys less Ramen noodles. Ramen is an inferior good in this case.

The words "normal" and "inferior" are somewhat unfortunate choices to describe what happens to a consumer's consumption of a good when income changes. There may be nothing inferior about an inferior good. Consider a high-income person who leases a Lexus every year. Now suppose her income increases. She might now not lease a Lexus at all, and switch her automobile consumption to a Bentley, a Ferrari, or a Rolls Royce. Her income went up and her consumption of Lexus automobiles went down. A Lexus is an inferior good in this case, but there is nothing inferior about a Lexus as an automobile. Sometimes even jargon that appears reasonable at the outset can become confusing or misleading upon further consideration.

The Income-equivalent of a Price Change

The modified budget line model introduced in the previous section allows us to examine an important economic concept, the benefit to a consumer from a price change. Because it is impossible to measure utility directly, in economics we often use an equivalent income change as a measure of whether a consumer is better or worse off from an action or event. Of course, the benefit of an increase or decrease in income is easy to capture with this method. It's simply the amount of the income change. But what if something other than a change in income happens to the consumer? What if the price of a good changes?

Computer equipment is a good example to show this concept, as its price has fallen and continues to fall dramatically in the real world. How might we conceptualize the benefit to a consumer from a fall in the price of computer equipment? Figure 4-9 is a budget line model representing a consumer's decision regarding computer equipment and other goods. The consumer starts at point A on the initial budget line, purchasing CE_1 computer equipment and spending OG_1 on other goods. The price of computer equipment now falls, represented by a rotation of the budget line. Note that the budget line still intersects the other goods axis at the same point, as we assume here that income and the prices of other goods have not changed.

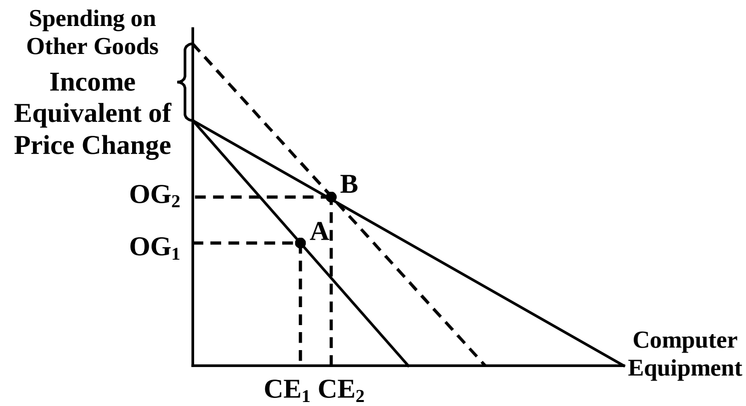


Figure 4-9
The Income Equivalent of a Price Change

We assume in this case that computer equipment is a normal good, so, after the fall in price, the consumer chooses to be at point B consuming CE_2 computer equipment and spending OG_2 on other goods. The consumer is obviously better off, but how much? Suppose we draw another budget line through point B, parallel to the initial budget line. This is the dotted budget line in Figure 4-9. Because we draw it parallel to the initial budget line, this new one represents the original price of computer equipment. Remember that the slope of the budget line represents the price ratio. This new budget line has the same slope as the original one, but it is farther out. Income is higher for the dotted budget line, and it is enough higher that the consumer could buy the combination of computer equipment and other goods with this income, but at the original price of computer equipment.

The distance between the two budget lines on the vertical axis represents this increase in income. Remember that the vertical intercept on this budget line model is the amount the consumer spends on other goods if she consumes no computer equipment at all. This is the same as her income. This distance is one measure of the income-equivalent of a price decrease. It represents the amount of income we would have to give the consumer to allow her to have the new consumption combination, represented by point B, without the decrease in price.

Note that we say this is one measure of the benefit of a price decrease. Unfortunately, there are several others, as well, but studying these would take us too far afield for an introduction to consumer theory, and would require development of more detailed and difficult models. But the point we learned here is a good one. A price change has an effect that is like giving the consumer a change in income, and if we can measure this "income effect," we can have an idea of the change in well being from a price change.

The Budget Line Model, In-kind Transfers and Black Markets

Our modified budget line model can yield insights into another very important real world phenomenon, the increase in well being derived from in-kind transfers. Let's first define what they are, and give some examples. A transfer is an economic term for a gift. If we give someone a gift of money, we economists call that a money or monetary transfer. Money a college student receives from home is a money transfer. A "care package" of chocolate chip cookies from Mom is an in-kind transfer. A check from the boss as a holiday bonus is a money transfer. A holiday turkey is an in-kind transfer. A welfare check is a money transfer. Food stamps represent an in-kind transfer.

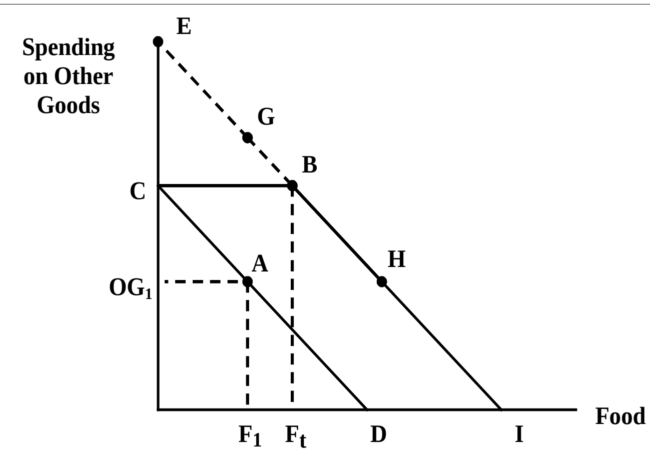


Figure 4-10
In-Kind Transfers and Consumer Choice

We represent an in-kind transfer in our budget line model by a modified shift of the budget line. Such a shift is shown in Figure 4-10. In this model the consumer is choosing between spending a given budget on food or other goods. The initial budget line is line CAD. Let's assume this consume chooses point A, consuming F_1 food and spending OG_1 on other goods. Suppose this is a poor person and we collectively as a government, or through some private charity, decide to give this person some food. Represent this amount of food by F_t on the food axis. Because the consumer can buy food in addition to the transfer, the budget line now has a horizontal portion and a kink in it. The new budget line is represented by the points CBHI.

Note that the dotted line EGB is not part of the budget line with the in-kind transfer. In order for this section to be part of the budget line, we would have had to give the individual money, not food. If we had given him additional money represented by the line segment CE, the consumer would have the dotted portion as part of his budget line.

Now comes the interesting part. If the consumer's preferences are such that he receives the in-kind food transfer and then buys additional food, the in-kind transfer has the same effect as an increase in money income. Note that the consumer in Figure 4-10 could get to point H with budget line CBHI or budget line EGBHI. On the other hand, suppose that the consumer didn't want all the food given in the transfer. Suppose that point G is preferred to point B or any other point on the segment of the budget line BHI. The consumer might want to have more alcohol and tobacco and less food. But point G is not attainable under an in-kind transfer system.

Economists always point out that if our goal is to increase the well being of the poor, we ought to give them money, and let them choose what to spend it on. To an economist, what the individual chooses is up to the individual. As we mentioned in Chapter 1, economists are reluctant to judge the appropriateness of individual choices. But in the real world, it's not just the preferences of the receiver that are taken into account. Givers of in-kind transfers might prefer that the receivers of them not buy alcohol and tobacco products with their gifts, so they restrict the receiver's choice.

Sometimes, however, judgmental givers of in-kind transfers are thwarted in the end. Markets have a tendency to emerge when willing buyers and sellers can be made better off. The U.S. government has a large in-kind transfer program called the food stamp program, where the poor can receive vouchers redeemable only for food. If "point G" individuals exist, i.e., if a receiver of food stamps would rather have more of other goods and not as much food as given in the transfer, economists would predict that a black market in food stamps would emerge. By selling food stamps, even at a discount, food stamp recipients would be able to get to a point above the segment of the in-kind transfer budget line CB, if such a point were preferred to one on the segment BHI.

The Budget Line Model and the Unanticipated Effect of a Subsidy

Another way to help a poor or otherwise-favored consumer is to subsidize the price of something they buy. Suppose we are concerned about substandard housing lived in by poor people. Suppose we start a subsidized housing program where the government pays a certain percentage of the monthly rent of a poor person, thereby allowing him to acquire a greater amount of housing. Such a case is shown in Figure 4-11.

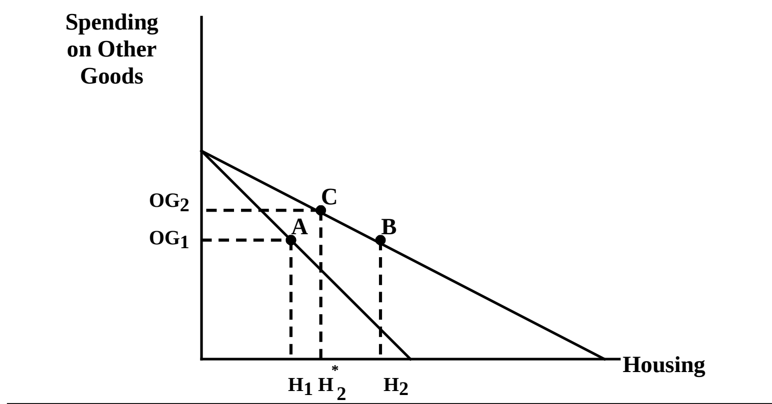


Figure 4-11
The Unanticipated Effect of Subsidies

Our consumer in Figure 4-11 is on his initial budget line at point A consuming H_1 housing and spending OG_1 on other goods. By subsidizing the price of housing, we rotate the budget line out the housing axis. Because we hold the consumer's income constant in this example, the budget line remains anchored at the same point on the other goods axis.

Now, imagine housing for the poor advocates doing a study to evaluate the effectiveness of the housing subsidy program on the quality and quantity of housing consumed by the poor. At the beginning of the study, housing subsidy advocates might be expecting to find an increase in housing consumed to H_2 in our model. Imagine the surprise when the program evaluation is delivered and a typical poor person has increased his consumption of housing only to H_2*. The housing program increased the quantity of housing purchased very little.

Was the program a failure? Only if the goal was to increase housing a large amount. But someone with economics training would understand that consumers have some flexibility under this program. Nothing requires them to spend the same amount on other goods as they did before the price of housing was subsidized.

The reduction in the price of housing had an income effect, like the one we showed in Figure 4-9, and this particular consumer chose to spend a large proportion of that increase in spending power on other goods. The surprising, unanticipated effect is neither surprising nor unanticipated to someone with knowledge of the theory of the consumer.

Deriving an Individual Demand Curve from the Budget Line Model

In a budget line model we represent a change in the price of a good with a rotation of the budget line. The budget line rotates out along the axis of a good when the price of the good falls, and rotates in when the price rises. Different points chosen by the consumer on different budget lines (ones with different prices) indicate the quantity the individual wishes to purchase at that price, income held constant. This is the quantity demanded at that price. We can read quantity demanded off the axis of the budget line model, but the price is hidden in the slope of the budget line itself. We have the information for a demand curve, but not enough axes on the graph.

Figure 4-12 shows how we can derive an individual's demand curve for a good from information in the budget line model. In this case, Professor Joe has a budget of $300 to allocate between two goods, Golf rounds and wine. If Professor Joe consumes no wine, he could buy 15 rounds of golf at $20 a round. His budget

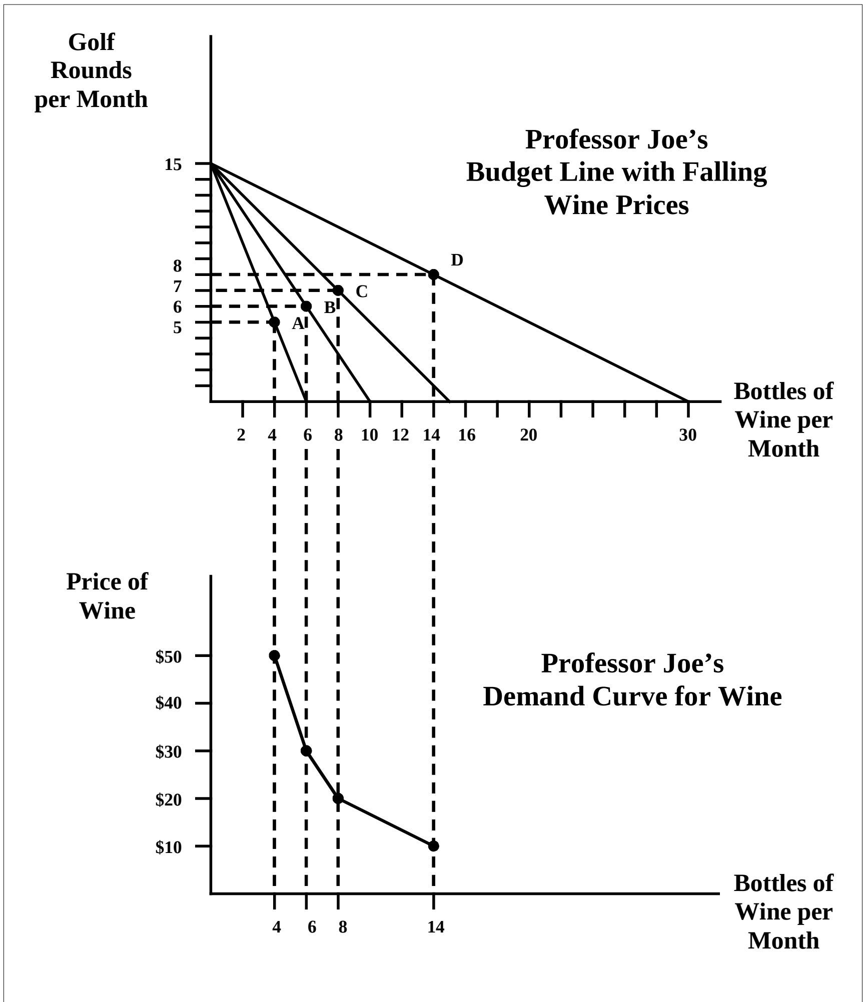


Figure 4-12
Deriving an Individual Demand Curve from the Budget Line Model

line intersects the golf rounds axis at 15 rounds. The four budget lines in Figure 4-12 represent four different wine prices, $50, $30, $20, and $10. As the price of wine falls, the budget line rotates out the wine axis in the usual fashion. Note that the intersection on the budget line with the wine axis is determined by dividing the budget of $300 by $50, $30, $20, and $10 respectively.

The points A, B, C, and D on the different budget lines represent the choices we have assumed for Professor Joe, and from these we can read the amount of wine demanded off the horizontal wine axis of the graph. The graph below the budget line graph is a demand curve. It has the same horizontal axis as the budget line model above it. Note that we plot the price of a bottle of wine (represented in the slope of the budget line) on the vertical axis of the lower graph. From these price and quantity combinations, we construct the relationship between price and quantity demanded, Professor Joe's demand curve.

From Individual Demand Curves to the Market Demand Curve

In this chapter, we moved behind the market demand curve to find its foundations in a theory of individual choice, the budget line model and the equimarginal principle. In the last section, we derived an individual's demand curve from the budget line model. In this section, we come full circle and show that the market demand curve we studied in Chapter 3 is merely an aggregation, a summing up, an addition of individual demand curves.

In the top portion of figure 4-13, we see three demand curves for bread for individuals Bill, Sam and Mary. We can imagine that we derived these demand curves for each individual from their respective budget line models, as we did for Professor Joe in the last section. The bottom portion of Figure 4-13 is the market demand curve derived by a horizontal summation of the three individual demand curves. Let's explore this horizontal summation.

When we add curves horizontally, we start with a particular value on the vertical axis, in this case the price axis, and add together the three corresponding quantities demanded from the individual demand curves. For example, at a price of $5 a loaf none of our individuals wants to buy bread. But at a price of $4, Bill demands 1 loaf, while the price is still too high for Sam and Mary. We add together (horizontally) the three

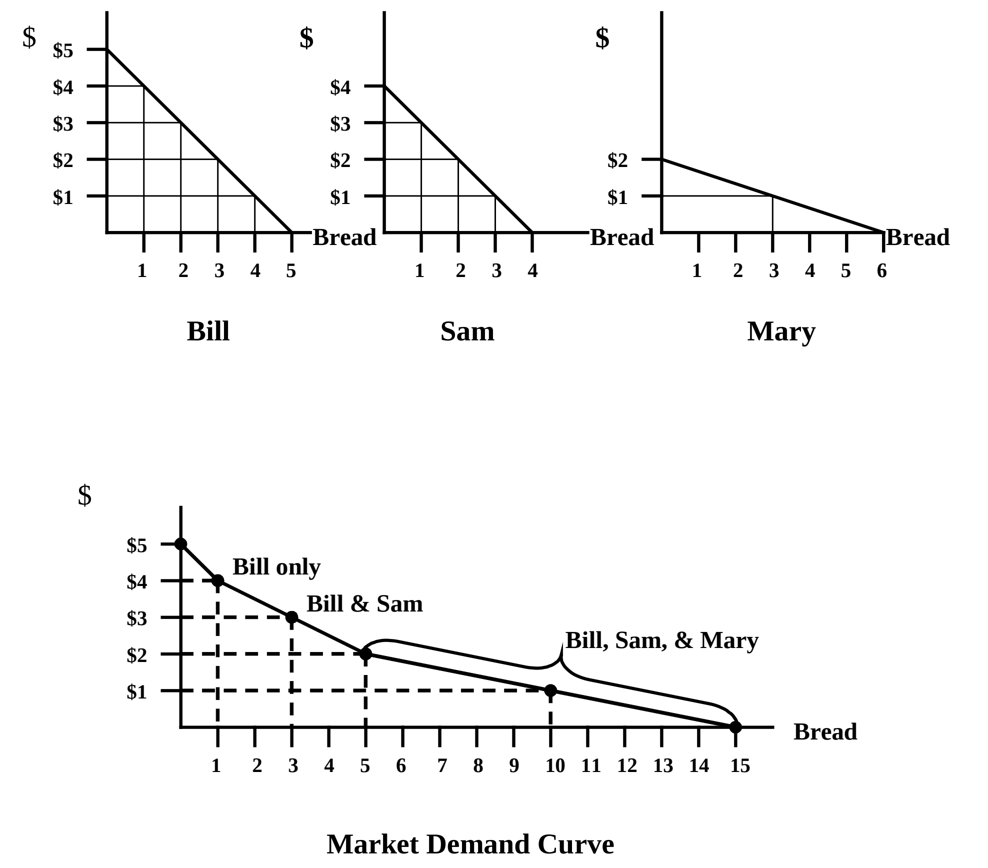


Figure 4-13
From Individual to Market Demand

quantities demanded at the price of $4 (1 + 0 + 0) and put that point on the market demand curve. We then follow the same procedure for other prices of bread. At a price of $3, Bill increases his quantity demanded to 2 loaves, Sam wants to buy 1 loaf, and Mary is still out of the market. By lowering the price, and summing horizontally at each price, we trace out the market demand curve for Bill, Sam, and Mary. Note that if bread has a price of $0, the quantity demanded would be 15 loaves (5 + 4 + 6). At different points or segments along the market demand curve in Figure 4-13, we have labels noting who among our three individuals is in the market wanting to buy at that price.

One implications of this aggregation is that the market demand curve has three distinct segments, each with a slope that represents an averaging of the slopes of the demand curves of individuals buying at prices over the segment. Another implication is more important. Note that at every price, Bill, Sam, and Mary buy just the amount they want, i.e., each is on their individual demand curve at that same price. Later we will use this observation and combine it with the additional argument that the individual demand curve is an individual marginal benefit or marginal value relationship. If the individual demand curve is an individual marginal benefit function, it means that each individual buys the good up to the point where their marginal benefit equals the price. This, again, is standard marginal analysis, if we consider that price is the marginal cost of another unit of the good to the individual demander. The amazing conclusion here is that if all face the same prices, everyone has the same marginal benefit from consuming the good, even though tastes and preferences and other individual characteristics give rise to greatly different individual demand curves. We'll return to this idea in Chapter 8 when we consider the normative characteristics of different kinds of markets.

Chapter 5: Marginal Analysis and a Model of a Business: Production, Cost, Revenue and Profit

Introduction

We defined economic systems in Chapter 1 according to the way people in them own and allocate resources, and learned that in market capitalism private individuals do the owning and the allocating. In Chapter 3, we learned that the interaction of buyers and sellers in markets could be captured with a model of a market, the supply and demand model. In Chapter 4, we went behind the demand curve and studied the most micro aspects of microeconomics, individual consumer choice. We represented this choice behavior with models such as the utility function and the budget line model.

In this chapter, we again do some disaggregation. We go behind the supply curve to another fundamental component of market capitalism, the private business. Formation of businesses is one way people own and allocate resources. Of course, in order to examine and begin to understand the role of business behavior in markets and market capitalism we need to construct a model of a business. This is the task here. We begin in the next section with a model of a business's production process, the production function.

The Production Function

We begin our discussion of the theory of the business with the production function, a model of the relationship between the business's use of resources and its output. We use the term "function" here in its mathematical sense. With different amounts of resources, the business produces different amounts of output. For simplicity let's assume that we have just two inputs, labor and capital. We can represent this relationship between resources and output in functional notation as follows:

$$Q = F(L,K)$$

where Q represents output, L represents labor, and K represents capital. The function F(L,K) indicates that for a particular combination of L and K the firm produces a particular amount of output Q. Just how much Q we obtain for different amounts of L and K depends on the specification of the function F.

In the top panel of Figure 5-1, we assume one specification of a production function and represent it with a graph. Because figure 5-1 is a two-dimensional graph, we have only two axes at our disposal. We could draw a production function in three dimensions, but let's not do that. This is an economics book, not one on drawing. In Figure 5-1, we place labor on the horizontal axis and output on the vertical axis. But what about the amount of capital in the production process? Similar to the procedure for constructing demand curves and supply curves when we have more variables than axes on the graph, we again must hold something constant. Here we choose to hold the amount of capital constant at the level K_1. This is why the production function in Figure 5-1 is labeled $Q = F(L,K_1)$, where capital is fixed at the level K_1. If we change the amount of capital used in our model of the firm's production, we would have to draw a different production function, but let's not do that right now.

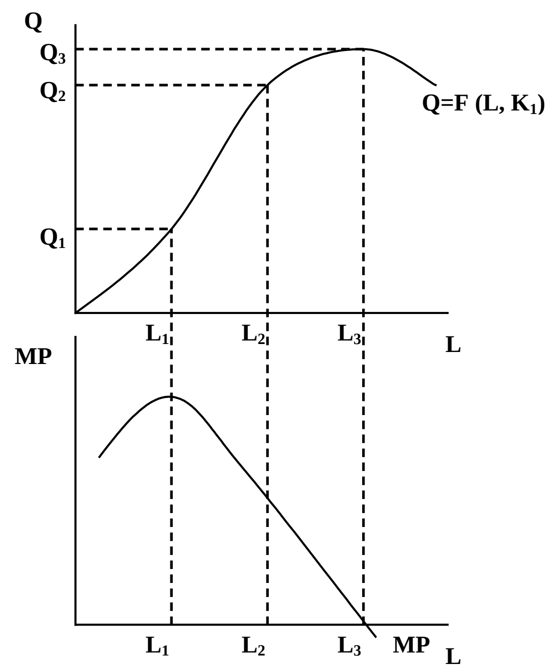


Figure 5-1
The Production Function and Diminishing Marginal Productivity

The shape of the production function in Figure 5-1 shows some important aspects of the production process. First, the general slope of the production function is positive. The more labor, given a fixed amount of capital, the more output the firm produces. When the firm employs L_1 labor, our model shows it produces Q_1 output. If it uses L_2 labor, it produces Q_2 output, and so on. Note that if the firm used more labor than L_3, output would actually start to decline. We will return to this part of the production function below.

While the slope of the production function is generally upward sloping, note that the slope is not constant. The slope of the production function with variable labor and fixed capital is called the marginal product of labor. Yes here is the familiar economics term "marginal" again. Remember in Chapter 4 we defined the slope of the total utility curve as marginal utility. If you think of the production function as representing "total product," it's easy to see that we are following the same procedure here. The slope of the production (total product) function is marginal product. The marginal product of labor is the additional amount of output obtained when we employ an additional amount of labor.

In the early stages of the production process, up to $L = L_3$ in figure 5-1, the slope of the production function is increasing. This represents increasing marginal product of labor. This is not uncommon in production processes. Recently, I rented an hydraulic wood splitter to split my winter's supply of firewood. I can run the splitter and fetch the unsplit rounds of wood by myself, but with the help of my friend Larry the process is much more efficient. By adding an additional worker, together we can split more than twice the amount of wood that I could split by myself in the same time period. The marginal product of wood splitting, given an amount of capital (one hydraulic wood splitter) is increasing from one to two workers.

After some amount of labor, represented by $L = L_1$ in Figure 5-1, the slope of the production function, while still positive, starts to fall. I expect this to be the case in wood splitting as well, but I've never been able to get enough friends over to help split wood and test my theory. This is the much more interesting portion of the production function from an economist's perspective, because it represents another form of the familiar diminishing returns concept. In this case the principle is called eventually diminishing marginal productivity. If we keep other resources fixed and increase the employment of just one, the marginal product of that resource, while positive, will begin to decrease, and will continue to decrease as we add more and more of that resource.

The principle of (eventually) diminishing marginal productivity is shown in the bottom panel of Figure 5-1. Here, distance on the vertical axis is marginal product, not total product. Note that the marginal product of labor rises until labor reaches the point L_1. This is exactly the point where the production function (the total product function) reaches an inflection point in the upper panel of Figure 5-1. And we can remember from algebra class or finite math that an inflection point is where the slope of a line stops falling and starts rising, or vice versa. From L_1 to L_2 and up to L_3, additional labor produces additional output, but the marginal product of labor falls with each additional amount of labor. After $L = L_3$, the marginal product of labor becomes negative and additional labor causes a decrease in the amount of output produced.

The next time you are in a fast food restaurant, one with a counter with a couple of cash registers, and a fixed amount of floor space where workers (let's call them "Blue Hats") take, fry up, and retrieve orders, think about the production function and the diminishing marginal productivity of labor. Imagine what would happen if you kept adding Blue Hats to the fixed work area. More Blue Hats means more output up to a point, but eventually, after a period of diminishing marginal productivity, so many Blue Hats occupy the work area that no one can move. Total output of fast food actually starts to fall as Blue Hats are crammed cheek by jowl into the work area. We will concentrate on the portion of the production function between L_1 and L_3, as no business would operate where marginal product is increasing or where marginal product is negative. A business will always want to add workers in the former case, and employ fewer in the latter.

From Production to Cost, Revenue, and Profit

While the production function provides a foundation for our model of a business, further development of the model requires a shift of attention to the cost of production. Our first task is to relate production to cost, and this begins in the first four columns of Figure 5-2, a numerical model of a business.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
L	K	Q	MP	P	TR	MR	TFC	W	TVC	TC	MC	AFC	AVC	ATC	Profit	Marginal Profit of Labor
0	1	0		$20	$0		$200	$120	0	$200		NA	NA	NA	$-200	
			20			$20					$6.00					$280
1	1	20		$20	$400		$200	$120	$120	$320		$10.00	$6.00	$16.00	$80	
			30			$20					$4.00					$480
2	1	50		$20	$1,000		$200	$120	$240	$440		$4.00	$4.80	$8.80	$560	
			25			$20					$4.80					$380
3	1	75		$20	$1,500		$200	$120	$360	$560		$2.67	$4.80	$7.47	$940	
			20			$20					$6.00					$280
4	1	95		$20	$1,900		$200	$120	$480	$680		$2.11	$5.05	$7.16	$1,220	
			15			$20					$8.00					$180
5	1	110		$20	$2,200		$200	$120	$600	$800		$1.81	$5.40	$7.27	$1,400	
			10			$20					$12.00					$80
6	1	120		$20	$2,400		$200	$120	$720	$920		$1.67	$6.00	$7.67	$1,480	
			5			$20					$24.00					$-20
7	1	125		$20	$2,500		$200	$120	$840	$1040		$1.60	$6.72	$8.32	$1,460	
			2			$20					$60.00					$-80
8	1	127		$20	$2,540		$200	$120	$960	$1160		$1.58	$7.50	$9.13	$1,380	

Figure 5-2
Production, Revenue, Cost and Profit: A Numerical Example

Figure 5-2 contains an almost mind-numbing amount of numbers taken in its entirety, but if we study it column by column and understand the relationships between the columns, we will have a good start in building and understanding a model of a business. The first four columns of Figure 5-2 represent a production function. Note that labor is the variable resource ranging from 1 to 8 workers in column 1. Let's further define the flow of the labor resource as workers per day. In column 2, we see that capital is fixed at a level of 1. We might think of capital as one shop, one store, or one production facility. Column 3 shows units of output per day produced by the different amounts of labor and the fixed capital. Note that as the business adds more labor to the production process it produces more output, as we would expect.

The example has also been constructed to show eventually diminishing marginal productivity of labor. We see the marginal product of labor in the fourth column. When the firm employs its first worker output per day rises from 0 to 20 units. The change in output is 20. The change in labor is 1. The marginal product is the change in output divided by the change in labor, so the marginal product is 20. Adding the second worker increases output per day to 50 units. Here the marginal product increases to 30. But as shown with the remaining entries in column 4, each additional worker adds a positive yet lower amount of additional output, i.e., the marginal product of labor eventually declines.

The numerical model of a business in Figure 5-2 is abstract by design, but it might help to imagine a real world production situation. Try to think of a business example that might fit these numbers. A take-out pizza shop works pretty well in my mind, one with fixed oven and pizza assembly capacity. A small shirt (or other garment) factory might work as well.

Columns 5, 6, and 7 of Figure 5-2 introduce the revenue side of our numerical model of a business. Here we assume that the price of the product is $20, and further, that the business can sell all it wants to at that price. Later, in Chapter 7, when we study something called "market structure," we will define this kind of pricing situation to be a characteristic of the market structure "perfect competition." Here, it just makes our example easier, because we don't have to worry about pricing decisions.

Column 6 shows total revenue (TR) for the business. Each number in column 6 is the product of the corresponding numbers in columns 3 and 5. Total revenue is the product of price and quantity.

In column 7 we have another important "marginal" concept, marginal revenue (MR). As is the case for anything marginal in economics, we define this one also as a change in one thing divided by a change in another. This time it's the change in total revenue divided by the change in quantity. Note that when output rises from 0 to 20, revenue goes from 0 to $400, and the change in revenue ($400) divided by the change in quantity (20) equals $20. When another worker is hired, and output rises from 20 to 50 units, total revenue increases from $400 to $1,000, a change in revenue of $600. The corresponding change in quantity is 30 units, so the marginal revenue is again $20 ($600 divided by 30). Note that marginal revenue is constant and equal to price at all levels of output in this example. This is not true in general, only when a business can sell all it wants to at a given price.

Columns 8 – 15 in Figure 5-2 represent the various components of the cost side of the business. In column 8, we see the most boring cost of all, total fixed cost (TFC). Fixed cost exists because we have a fixed resource in the production function, capital in this example. We assume here that the cost of the fixed capital is $200 per day. It might help to imagine our pizza shop or small shirt factory renting production facilities for $200 per day. Imagine that the business has a one-year lease on the production facilities that it cannot get out of, i.e., even if the business produces nothing it has TFC of $200 per day.

Column 9 shows the daily wage our firm has to pay each worker, assumed here to remain constant over the entire range of employment possibilities. Column 10 shows total variable cost (TVC), which is calculated as the product of the entry in column 1, the amount of the variable resource used, and the entry in column 9, the wage you have to pay the variable resource. As we would expect, TVC rises as we hire more labor and increase output.

Total cost (TC) appears in Column 11 of Figure 5-2. TC is the sum of TFC and TVC. You can check this by adding the number in column 10 to the number in column 8. Note that TC differs from TVC by a constant amount. For example, with 3 workers per day and 75 units of output, the difference between TC and TVC is $200. Note that this relationship also holds for any other level of labor input and corresponding output. This should not be surprising, however, as the $200 difference between TC and TVC is none other than TFC, at all levels of output.

In column 12, we again have something important, marginal cost (MC). Again, we have a standard "marginal" definition, the change in one thing divided by the change in something else. Here the definition is the change in TC divided by the change in Q. Check the numbers. When the business uses its first worker, output rises from 0 to 20 units per day. TC rises from \$200 to \$320, a change in TC of \$120. Because marginal cost is defined as the change in TC divided by the change in quantity, the marginal cost is \$6 (\$320 divided by 20 units of output). If the business hires an additional worker, output increases from 20 to 50 units per day, a change in output equal to 30 units. TC rises from \$320 to \$440, a change of \$120. The marginal cost is now \$4 (\$120 divided by 30).

It is no accident that the marginal cost fell with this change in production. Remember that at lower levels of production the marginal product of the variable resource often increases as we use more of it (see column 4). Yet hiring an additional worker costs the same amount, \$120. If the addition output from an additional worker increases, and we can hire him or her for the same wage, the cost of the additional output (marginal cost) has to fall.

This relationship between marginal product and marginal cost is an important one. It merits consideration because in this relationship we see the fundamental relationship between production and cost. If marginal product is rising, marginal cost is falling. If marginal product is falling, marginal cost is rising. Compare columns 4 and 12 of Figure 5-2 and try to reason through why this important relationship exists. Rising marginal cost of production is the result of diminishing marginal productivity of the variable resource in production.

The first twelve columns of Figure 5-2 are just about all we need for our model of the business. Columns 13-15, however, have some usefulness for other concepts we will develop later. I mention this because in economics we approach concepts that contain the word average with a little bit of caution, if not outright hostility. This is because economics is about choice, choices are actions, and actions are modeled with marginal concepts, e.g., marginal utility, marginal revenue, and marginal cost. That said, we can proceed to a discussion of average fixed cost (AFC), average variable cost (AVC), and average total cost (ATC).

As we see in column 13, AFC starts high and falls as output increases. We can think of the fixed cost being spread over larger and larger levels of output. AFC is never equal to 0, but it approaches 0 as output gets very large. In contrast, AVC falls at low levels of output, reaches a minimum, and then rises. In column 15, we see that ATC, often called unit cost, is the sum of AFC and AVC. The fact that ATC is the sum of the other two average costs, AFC and AVC, will help us understand why it initially falls then rises. Because at low levels of output both AFC and AVC are falling, ATC must fall as well. While AFC continues to fall, AVC eventually starts to rise. Eventually, the rising AVC outweighs the falling AFC, and the sum of the two, ATC, bottoms out and starts to rise as well.

The mass of production and cost data in Figure 5-2 finally comes together in column 16, Profit. We define profit as total revenue minus total cost. Note that profit is equal to -$200 at zero output, the business's cost that cannot be avoided even by producing nothing. But profit soon begins to rise as output increases. Note in column 16 that profit reaches a maximum at 120 units per day. At higher levels of output, profit is lower than at 120 units.

We save the most important lesson of Figure 5-2 for last, the rules a firm should follow to maximize profit. We have two rules here, one involving a choice about how much of the variable resource labor to employ, and the other about the choice of how much output to produce. Because we have only one variable resource in this case, if we satisfy one of the rules, we satisfy the other as well. Both rules involve marginal analysis.

If the business modeled in Figure 5-2 wishes to maximize profit, it should employ more of the variable resource labor if the change in profit from employing another unit of the resource is positive, i.e., if employing more labor raises profit, do it. If employing more labor doesn't raise profit, don't do it. Let's apply this rule to the production and cost data in Figure 5-2. First, let's define another new term, the marginal profit of labor, the change in profit divided by the change in the amount of labor employed. This division is easy in Figure 5-2, because we're adding only one worker at a time. This means we are dividing by 1 each time. The marginal profit of labor appears in column 17 of the table. Note that marginal profit is positive at levels of output up to and including the sixth unit of labor. If the manager of the business asks the question "Should I hire a sixth worker per day?" the answer is yes because by doing so profit will rise by the amount of the marginal profit, \$80. But once the sixth worker per day is hired, it doesn't pay to hire the seventh because doing so will decrease profit by \$20 per day. After six workers, marginal profit is negative.

The other rule for maximizing profit is about the output decision. If the business produces more output it will increase it revenue. That's good. Likewise, if it increases its output, it increases its cost. That's bad. What, then, is the profit-maximizing level of output. Our rule is that the business should produce more output as long as its marginal revenue is greater than its marginal cost. If marginal revenue is greater than marginal cost, the business will add more to the good thing, revenue, than it adds to the bad thing, cost, by producing more. This will increase the difference between the good thing and the bad thing, profit. Go down the output column (#3) of Figure 5-2 and check the relationship between marginal revenue and marginal cost. At 20 units of output MR = \$20 and MC = \$4. Because MR > MC, the firm would not be maximizing profit at 20 units per day. It should produce more. The next production level in our "lumpy" production function is 50 units of output. Here, MR again equals \$20, and marginal cost equals \$4.80. Notice that at 110 units of output (produced with 5 workers of labor) MR = \$20 and MC = \$12. Moving to the next level of output will increase profit. But at this next level, 120 units (produced with 6 workers per day), MR = \$20 and MC = \$24. Because MC>MR, by producing the next higher level of output the business would be adding more to total cost than it would add to total revenue. If we add more to the bad thing than we add to the good thing, the difference, profit, will fall.

Graphical Models of Total and Marginal Costs

The numerical model of a business's production and cost in Figure 5-2 is useful and instructive, but only up to a point. Numerical examples are too cumbersome when we want to use production and cost information to examine issues such as price and output determination, profit and loss, product differentiation, shut-down decisions, long-run adjustments, and the normative implications of market structure. In this section, we begin the construction of a graphical model of a business by constructing graphical models of the various costs mentioned above.

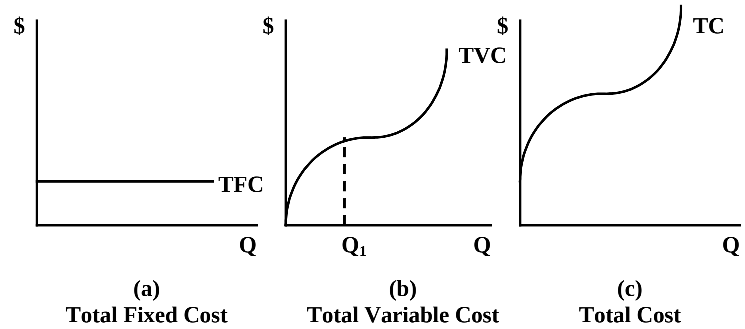


Figure 5-3
Total Fixed Cost, Total Variable Cost, and Total Cost Curves

In the three panels of Figure 5-3, we see models of the three total costs discussed above, Total Fixed Cost (TFC), Total Variable Cost (TVC) and Total Cost (TC). Note that the horizontal line in panel (a) indicates that TFC constant. As output expands, nothing happens to TFC. Panel (b) shows the characteristics of TVC. Note that at an output of 0 TVC is also 0, because to produce 0 output we hire no variable resources. Because we don't pay any variable resources at an output of 0, we have no variable cost. The shape of the TVC curve corresponds to the movement of the numbers of Figure 5-2. While TVC rises as output increases at all levels of output, it first rises at a decreasing rate, then at an increasing rate. We can see this in the slope of the TVC curve. The slope is always positive, but at lower levels of output the slope is falling as output increases. Eventually, the TVC curve has an inflection point at Q_1 and the slope of the TVC curve begins to rise. This shape of the TVC curve is the result of increasing marginal productivity of the variable resource in the early stages of the production process, and then eventually diminishing marginal productivity.

In panel (c) of Figure 5-3, we see the TC curve, the sum of the TFC and TVC curves. Note that the shape of the TC curve is exactly the same as the TVC. It's just displaced upward by the amount of fixed cost. Unlike TVC, TC is not 0 when output is equal to 0.

We have noted already on many occasions that any marginal relationship in economics is simply the change in some total magnitude divided by the change in the variable causing the total magnitude to change. For example, see the discussion in the previous section on marginal product, marginal revenue, marginal cost and marginal profit. On two occasions, we have described and graphed a marginal relationship as the slope of a total curve, e.g., marginal utility as the slope of the total utility curve (Figure 4-4), and marginal product of labor as the slope of the total product curve of labor (Figure 5-1).

In Figure 5-4, we see the graphical relationship between total cost and marginal cost. The slope of a curve at any point is represented by a tangent to that curve, so at output level Q_1 in the upper panel of Figure 5-4, marginal cost is the "rise over the run" at that point, the slope of the tangent as shown. Note that the slope of the TC curve, marginal cost, is falling up to the inflection point at Q_2, then increases beyond that level of output.

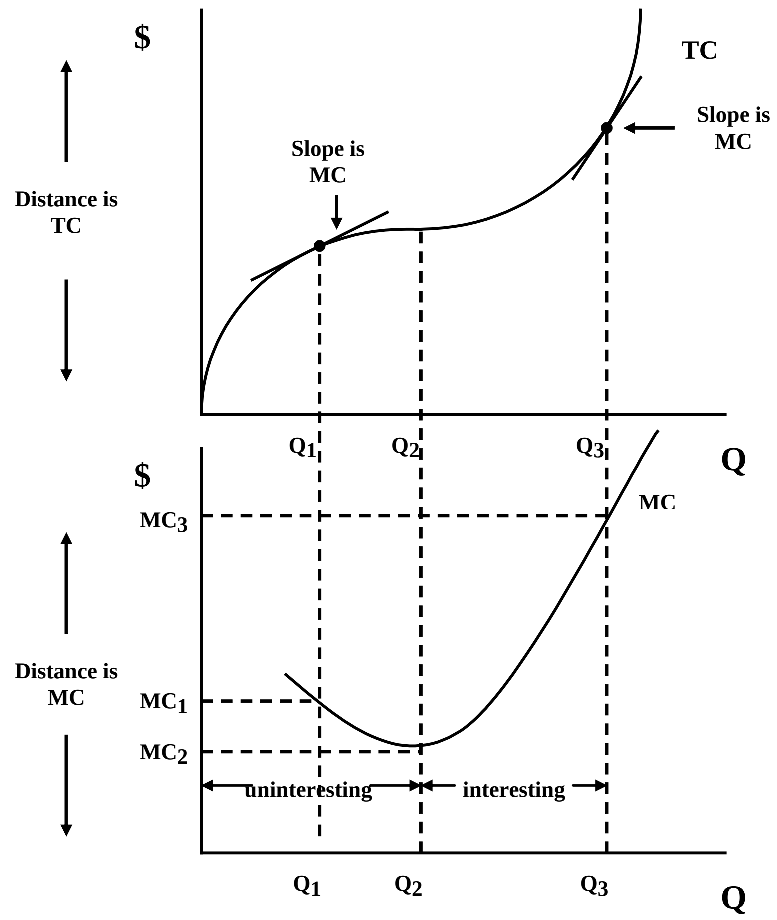


Figure 5-4
Total Cost and Marginal Cost Curves

In contrast with the top panel where we represent total cost as a distance up the vertical axis of the graph and show marginal cost as the slope of the total cost curve, in the bottom panel we represent marginal cost as a distance up the vertical axis. Note that from output level 0 to Q_2, the marginal cost is falling in the bottom panel of Figure 5-4. This corresponds to the falling slope of the total cost curve over this range of output in the upper panel. After Q_2, marginal cost rises as output increases, reflecting diminishing marginal productivity of the variable resource. The range of output 0 – Q_1 is also labeled as uninteresting in Figure 5-4. From a technical or engineering perspective this range perhaps is interesting, but from an economic perspective it is not, as no profit maximizing firm would ever choose a level of output where marginal cost is falling.

Graphical Models of the Average Cost Curves of a Business

While total and marginal cost are the most important costs to a business, and therefore our most important cost models, it is sometimes useful to consider average costs as well. In Figure 5-5, we show the relationship between TFC and average fixed cost (AFC), and introduce a general method for deriving the shape of an average cost curve from its corresponding total cost curve.

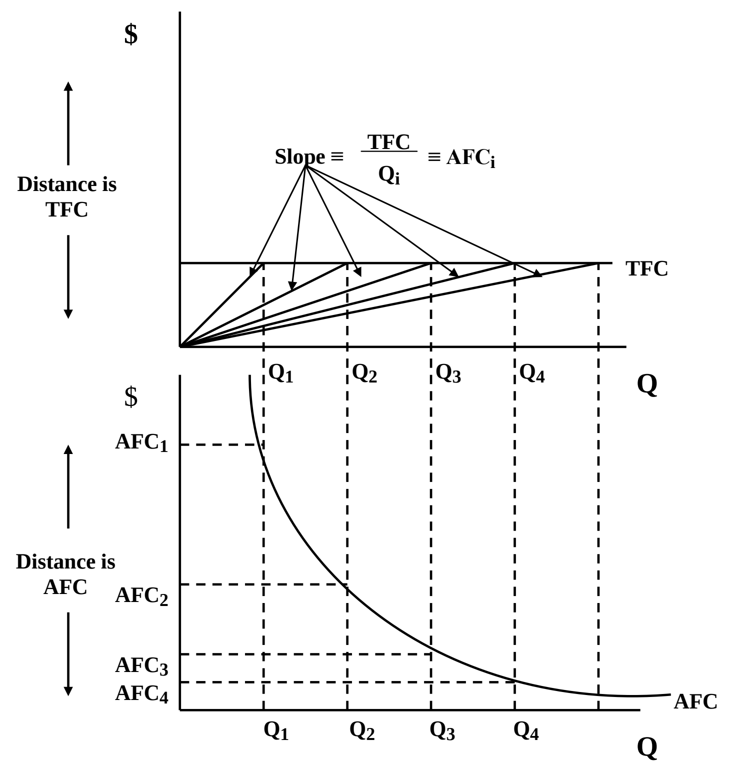


Figure 5-5
Deriving the Average Fixed Cost Curve from the Total Fixed Cost Curve

The TFC curve of the upper panel of Figure 5-5 is nothing new, a horizontal line representing the same amount of total fixed cost regardless of output level. The new additions to the figure are the upward-sloping straight lines from the origin to points on the TFC curve at output levels $Q_1 – Q_5$. The important information conveyed by each of these straight lines is found in their slope. For example, the slope of the line from the origin to the point on the TFC curve corresponding to output level Q_1 is the vertical distance from the horizontal axis to the TFC curve, divided by the quantity Q_1. Because the vertical distance is none other than TFC, the slope of this straight line from the origin is TFC/Q_1, the average fixed cost at output Q_1. Likewise, the slope of the lines from the origin to points on the TFC curve corresponding to output levels $Q_2 – Q_5$ represent the AFC at those levels of output.

Note that the slope of the straight line from the origin to the point on the TFC curve falls as output increases. This should not come as a surprise, as we saw in the numerical example in Figure 5-2 that AFC falls as output increases. In the bottom panel of Figure 5-5, we create a model of the firm’s AFC, the AFC curve.

Here, AFC is represented as a distance on the vertical axis of the graph, not a slope. As output increases, the AFC curve falls and approaches the horizontal axis of the graph asymptotically.

Now that we have mastered the general method of deriving average cost curves from a total cost curve, we can apply the method in a slightly more challenging context, deriving the AVC and ATC curves. These models appear as Figures 5-6 and 5-7, respectively. The slopes of the straight lines from the origin in the top panels of these figures represent AVC and ATC, just as the slope of the line from the origin to a point on the TFC curve represented AFC above. But note one difference in Figures 5-6 and 5-7. After the level of output labeled Q_3 in the top panel of each figure, the slope of the straight line from the origin to the point on the total variable cost or total cost curve rises. This means that AVC and ATC rise after that point. This is easy to see in the bottom panels of the figures, as the AVC and ATC curves are U-shaped, falling up to output level Q_3 and then rising. Please note, however, the amount of output represented by the notation Q_3 in Figure 5-7 is larger than the amount represented by output level Q_3 in Figure 5-6. ATC falls longer than AVC before it reaches a minimum and starts to rise. We can go beyond the mechanics of this result and bolster our intuition if we realize that ATC curve must continue to fall after the AVC curve has begun to rise, because the ATC curve includes average fixed cost. Because ATC is the sum of AFC and AVC, it takes a little more output for the rising AVC curve to overtake the effect of the falling AFC curve and make the ATC curve start to rise.

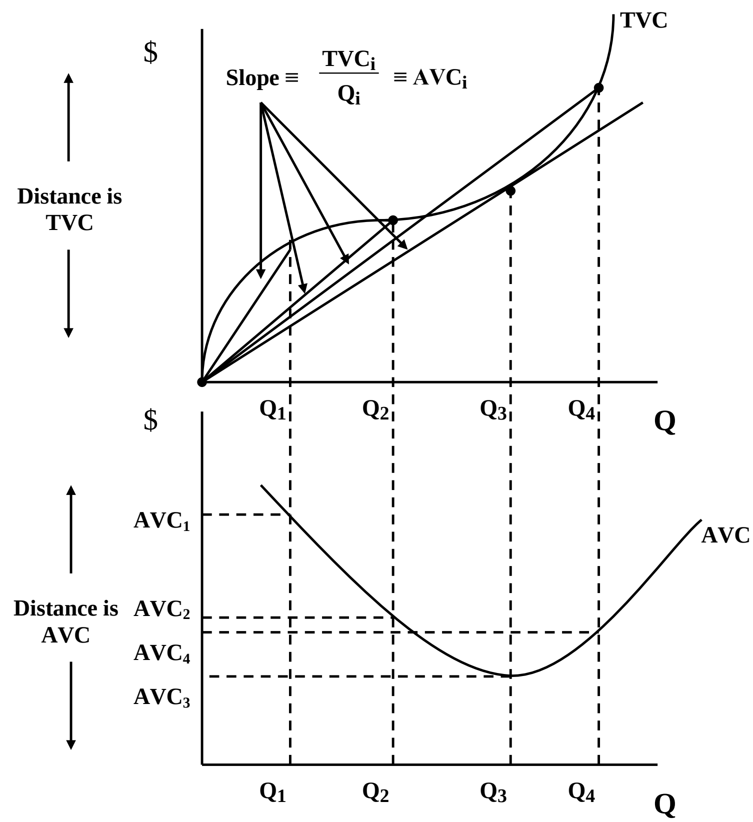


Figure 5-6
Deriving the Average Variable Cost Curve from the Total Variable Cost Curve

We can see one final important principle in Figure 5-7, the relationship between marginal cost and average total cost. At output level Q_3, ATC reaches a minimum. We show this in the upper panel by the fact that the slope of the straight line from the origin can be no lower than it is at output level Q_3. As noted above, beyond output level Q_3, the slope of the straight line from the origin increases, e.g., see the slope at Q_4. This is because at Q_3 the straight line from the origin is just tangent to the TC curve. With a slope any lower, it would not touch the TC curve. But what is the meaning of the slope of a tangent to the TC curve? It's none other than the marginal cost at that output. We can put these two facts together and conclude that MC equals ATC at the minimum of ATC. This realization will make our construction of the short-run cost model of a business much easier. We turn to this in the next section.

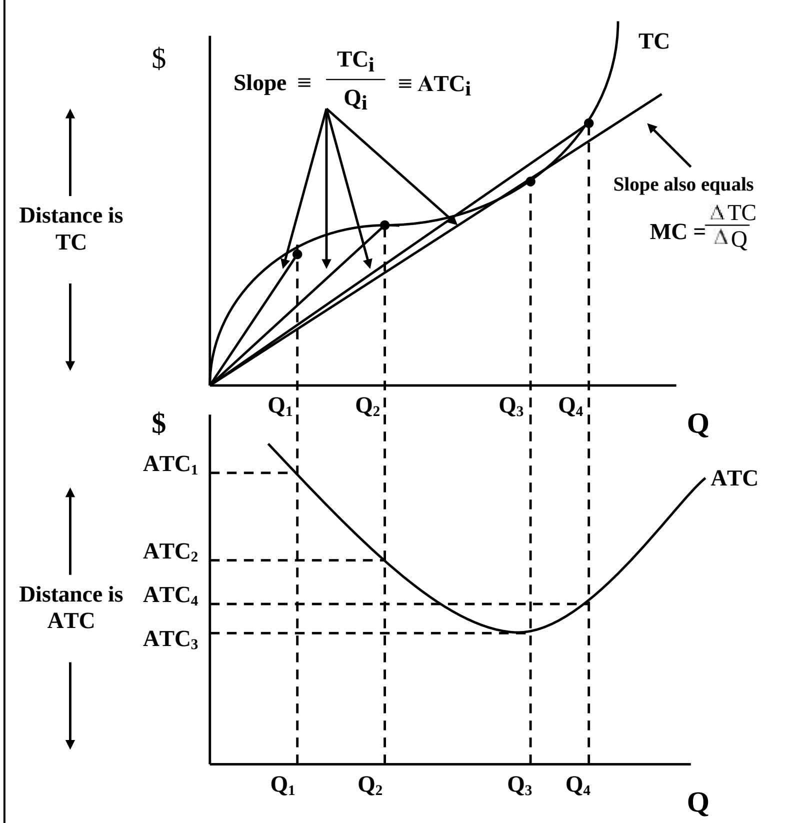


Figure 5-7
Deriving the Average Total Cost Curve from the Total Cost Curve

The Short-run Cost Model of A Business

In Figure 5-8, we collect the two average cost curves, AVC and ATC, and the MC curve to construct the cost model of a

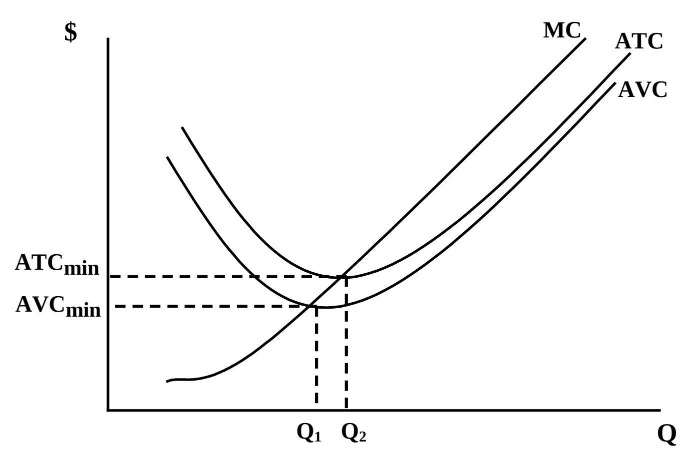


Figure 5-8
The Short Run Cost Model of a Business

business. For simplicity, we have now dispensed with the uninteresting falling portion of the MC curve. Two relationships are important in Figure 5-8. First, note that as output increases the AVC and ATC curves become closer and closer to each other. As output increases, AFC becomes smaller and smaller and therefore the difference between ATC and AVC becomes smaller and smaller. At low levels of output, AFC is high, so the difference between AVC and ATC is large. Second, MC crosses the AVC and the ATC curves at the minimum of each. We explained the relationship between ATC and MC above. The explanation for the AVC and MC relationship is identical. The slope of the TVC curve is also MC, and the minimum slope of a straight line from the origin to a point on the TVC curve also occurs at a tangent to the curve, a tangent that again represents marginal cost.

Grade Point Averages, Batting Averages, & the Relationship between Average Cost & Marginal Cost

The technical analysis in the preceding section is a good way to explain the relationship between marginal cost and average cost. A noneconomic example might also help to solidify our understanding. Think of your accumulated grade point average (g.p.a.). You can calculate it by multiplying the points for A, B, C, D, and F grades you receive (4, 3, 2, 1, 0 points) by the credit hours for each grade and then dividing by your total number of credit hours. Suppose you are a “B student.” You have a 3.0 g.p.a. Suppose you have a good semester and your g.p.a. for that semester is a 3.5. What will happen to your accumulated g.p.a.? It will go up. This is because your incremental g.p.a. is higher than you average g.p.a. If you have a bad semester and receive a 2.5 g.p.a., your accumulated (average) g.p.a. would fall. This is true of every average and marginal relationship. If the marginal is greater than the average, the average will rise. If the marginal is less than the average, the average will fall.

Let’s apply this principle to a baseball or softball batting average, calculated as the number of hits divided by the number of official “at bats.” Suppose someone is a “300” hitter. This means that out of every 10 official at bats, on average the player gets 3 hits. Suppose our player has a good day at the plate and goes four for four. What happens to her batting average. It goes up. Likewise, a bad day, where she bats zero for four, will result in a decline in her batting average. On the good day, the marginal performance is better than the average performance, so the average increases. On the bad day, the marginal performance is less than the average performance, so the average falls. Again, if the marginal is greater than the average, the average rises, and if it is lower, the average falls.

Marginal cost and average cost are no exception to this general rule about marginals and averages. If marginal cost is below average cost, average cost will fall. If marginal cost is above the average cost, average cost will rise. If marginal cost is equal to average cost, average cost will neither rise nor fall. This occurs only at the minimum of average cost.

The Revenue Model of a Business with No Market Pricing Power

Just as cost is the bad result of business activity, revenue is the good result. When a business sells its output, it receives revenue. Revenue is the product of price and quantity sold. To make things easier in our introduction to the revenue model of a business, we make a simplifying assumption. We assume initially that the business has no market pricing power. In Chapter 7, we will relax this assumption, as market pricing power is a key determinant of market structure.

With no market pricing power, the business takes the price as given. We call this type of business a price-taker. Two characteristics define a price-taking business. The first characteristic is size. For the price-taking business price is determined in the marketplace through the forces of supply and demand, and our business is such a small component of the total market supply that its output decisions have no effect on it. The business can sell as much as it wants without driving the price down, and if the business closes, the price in the market will not change.

The second characteristic of a price-taking business is the type of product it sells. For a business to have no market pricing power it must produce a homogeneous product, i.e., its product is just like that produced by many other businesses in the market. If the business tries to exert market pricing power by raising price above that charged by others in the market, no one will buy from it, preferring instead to buy the identical (homogeneous) product from one of the many other suppliers. Agricultural businesses are often price-takers. One grower's wheat is pretty much the same as another's. One grower's peas and lentils are much like his neighbor's peas and lentils. Farmers have little control over price. They discover the price on the daily farm report while listening to the radio eating a midday meal. They have no market pricing power.

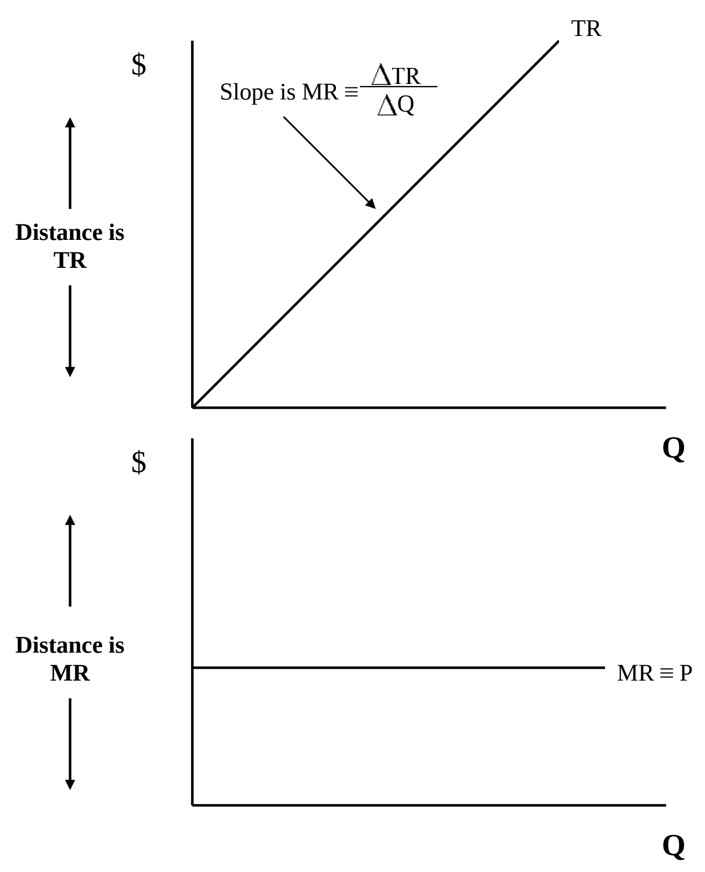


Figure 5-9
Total Revenue and Marginal Revenue Curves for a Business with no Market Pricing Power

Figure 5-9 shows the model for total and marginal revenue for a business with no market pricing power. The total revenue curve, shown in the upper panel, starts at the origin and rises at a constant slope. Because it can sell all it wants to at the market price, each additional unit sold increases revenue by that market price. Because the slope of any total curve is a marginal curve, the slope of the total revenue curve is marginal revenue, the change in total revenue divided by the change in quantity sold. In the bottom panel, we graph marginal revenue as a distance on the vertical axis. For the price-taking business, marginal revenue is identical to price. Price is constant, so marginal revenue is shown as a horizontal line on the graph.

Combining Cost & Revenue Models of a Business to Model Profit Maximization & Loss Minimization

In the three panels of Figure 5-10, we show different ways of modeling the profit maximizing output decisions of the price-taking business. Because the business is a price-taker, it has no pricing decision. What is the level of output that maximizes the profit of a business? First, we define profit as the difference between total revenue and total cost, and graph this relationship in the top panel of Figure 5-10. Note that the shape of the TC curve is like that in Figure 5-4, and the shape of the TR curve is like that in Figure 5-9

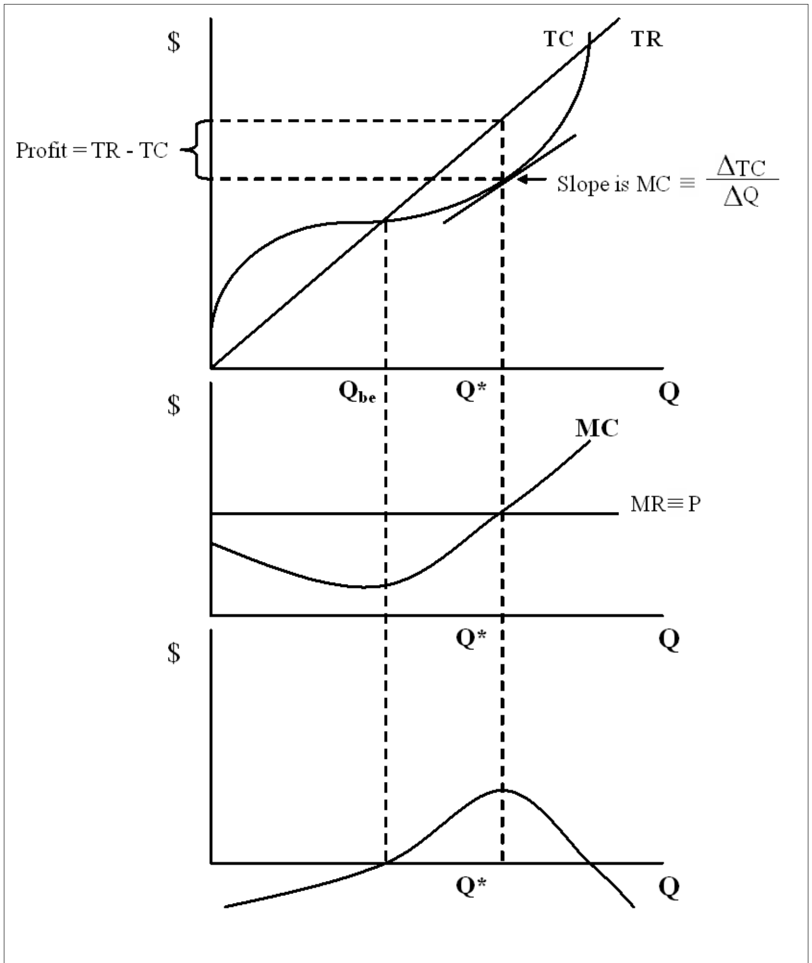


Figure 5-10
Three Models of Profit Maximization

Because both the total cost and total revenue models have dollars on the vertical axis and quantity on the horizontal axis, we can put both curves on the same graph. When we do this, we can define the vertical distance between the TR and TC curves as profit. Note the output level Q_{be}, the break even level of output. At this level of output the business's revenue just equals its cost. At levels of output below Q_{be}, the business would have negative profit. At output levels greater than Q_{be}, the vertical distance between TR and TC becomes positive and begins to grow. At output level Q*, profit is maximized. If the business expands output beyond Q*, profit, while positive, becomes smaller, eventually turning negative again when TC exceeds TR.

Note in the top panel that the maximum profit for the business occurs where the slope of the TR curve is equal to the slope of the TC curve. This is no accident, and represents the most important principle in our model of a profit maximizing business. In order to maximize profit, a business must choose an output where MR=MC.

The middle panel of Figure 5-10 shows this important marginal principle. Here, the MR curve is a horizontal line representing the fact that we assume no market pricing power. MR is constant and equal to the market price. We also graph the marginal cost of the business in the middle panel. Note that the MC falls (the uninteresting portion of the curve) and then rises. The intersection of the MR and MC curves in the middle panel occurs at the same output level, Q*, where the slopes of the TR and TC curves are equal in the top panel.

Finally, in the bottom panel of Figure 5-10 we have a graph of profit itself, the difference between TR and TC in the top panel. Initial profit is negative. Then it rises after Q_{be} as TR begins to exceed TC. Profit reaches a maximum at Q*, then begins to fall until it eventually becomes negative again.

At this time, it might be instructive to return to the numerical example in Figure 5-2. You should be able to see that the abstract modeling of the output choice for a profit maximizing business parallels that in our numerical example. The profit maximizing rule we used in the numerical example was that the business should increase output as long as marginal profit was positive (MR>MC). We see the same rule in the abstract model, but because we have nice continuous functions in this model, unlike the more lumpy production process of the numerical example, we can modify the rule to the precise MR=MC. Without a nice smooth, continuous production process the profit maximizing rule has to be modified slightly, as we might not be able to achieve an output where MR precisely equals MC.

An alternative model of the business and the profit maximizing output decision appears as Figure 5-11. This model is the same as the one in the middle panel of Figure 5-10, except for the addition of the ATC curve. The profit maximizing level of output is again shown as Q*, where MR=MC. At Q* level of output, the ATC, or cost per unit, is represented by the vertical distance 0A (or Q*D). Total cost can be determined as the product of cost per unit (ATC) and the number of units of output. In Figure 5-11, this is the area of the rectangle 0ADQ*. Because total revenue is the product of price and quantity, and price is the vertical distance 0AB, total revenue is represented as the area of the rectangle 0ABCDQ*. The difference between the TR rectangle and the TC rectangle is the rectangle ABCD, labeled profit. Should the business be in the situation where price is less than ATC, we have the result shown in Figure 5-12. Here the total revenue rectangle 0ADQ* is less than the total cost rectangle 0ABCDQ*, resulting in the loss rectangle ABCD. While the business is earning a loss in this case, because it chooses the level of output where MR=MC, it is earning the lowest possible loss, the lowest possible negative profit. MR=MC is the rule for both profit maximization and loss minimization in the short run.

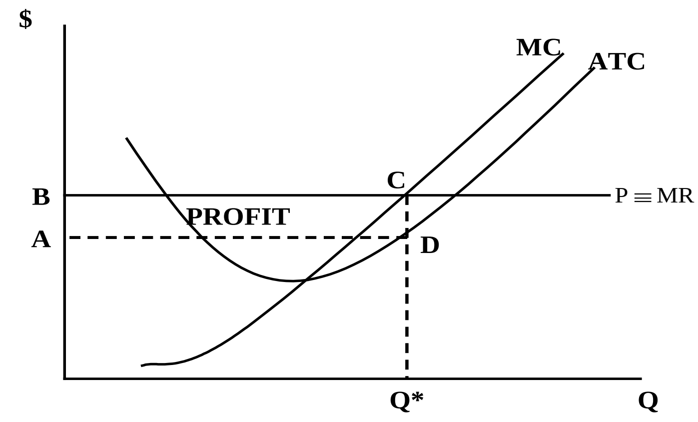


Figure 5-11
Profit Maximization in the Short Run Model of a Business without Market Pricing Power

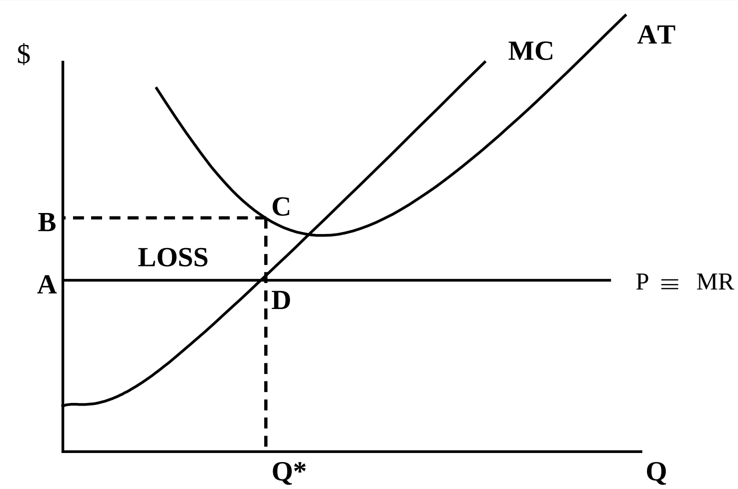


Figure 5-12
Loss Minimization in the Short Run Model of a Business without Market Pricing Power

A Model of the Short Run Shutdown Decision

The model of a business earning a short run loss, Figure 5-12, needs one major qualification we did not discuss above. Perhaps it would be better for the business to shut down and produce zero output. This would limit the loss to the business's total fixed cost. We have two possible approaches for modeling this situation. The first, using a total cost and total revenue approach is shown in Figure 5-13. Consider first the total revenue represented by the curve TR_1, and the TC and TVC curves. We show the loss minimizing output level Q^*_1 at the output where the slope of the TR_1 curve is equal to the slope of the TC curve. In this case the business should operate in the short run. Its loss is equal to the vertical distance ab. If the business shut down its loss would equal its TFC, the distance ac. Note that ac>ab. As long as the business receives revenue that is greater than total variable cost, it covers its variable cost with something left over. Its loss is less than its TFC.

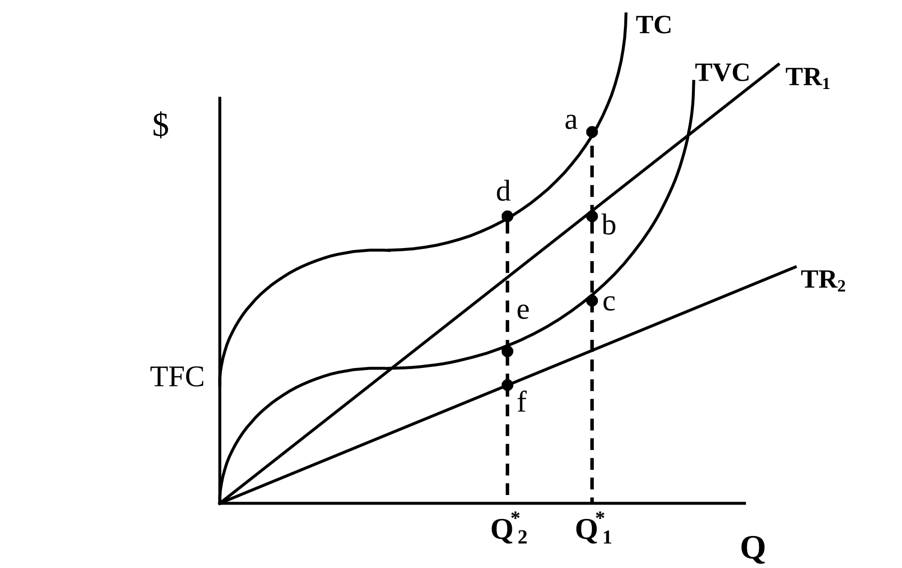


Figure 5-13
The Shutdown Decision: The Total Revenue and Total Variable Cost Approach

Now consider the shutdown decision if total revenue is represented by the total revenue curve TR_2. With a different total revenue curve we have a new possible loss-minimizing output level Q^*_2, where again MR=MC. This is only a possible (not the best) loss-minimizing output choice, however, because in this case it would be better for the business to shut down, even in the short run. If the business stays in operation, the best it can do is earn a loss equal to the vertical distance df. This is greater than the TFC, vertical distance de. In this case, the business should shut down and limit its loss to its total fixed cost. By staying in operation, it would incur its total fixed cost plus an additional loss that results from revenue not covering total variable cost.

An alternative model of the shutdown decision involves consideration of the ATC, AVC, and price curves in the model. This is shown in Figure 5-14. Facing a price of P_1, the business will minimize its loss with an output of Q^*_1, where MR_1=MC, earning a loss equal to the rectangle P_1edc. Note that P>AVC in this case. Price is high enough to cover the variable cost per unit by an amount equal to the vertical distance bc. The business's loss is reduced this vertical distance bc times the output level Q^*_1. Average fixed cost is the distance bd, the difference between ATC and AVC, so TFC is represented in the model by bd times the output Q^*_1.

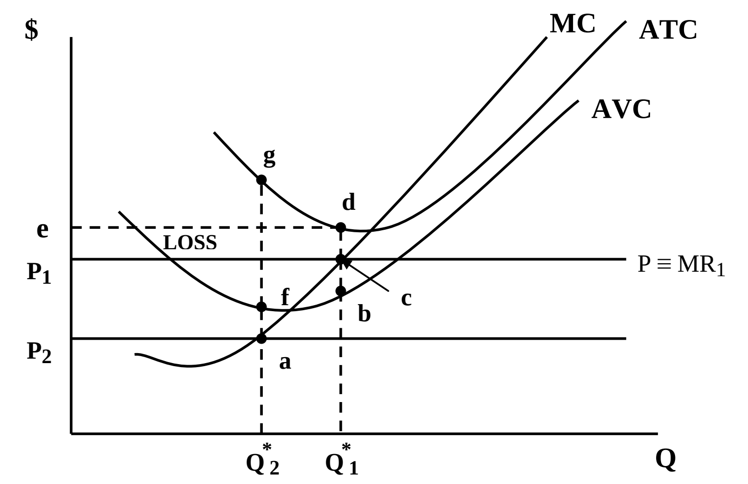


Figure 5-14
The Shutdown Decision: The Price and Average Variable Cost Approach

If price is P_1, the loss is obviously lower by operating and producing Q^*_1. If the price is at a level of P_2, however, Q^*_2 will not be the loss minimizing level of output. In this case, price is less than average variable cost by an amount equal to the vertical distance af. Production at this level of output will result in a loss equal to the vertical distance ag times the level of output Q^*_2. This is greater than the TFC, represented by the distance fg times Q^*_2. The rule in this case, while analytically a little complex, is actually straightforward. The business should stay in business in the short run, to minimize its loss, as long as price is greater than average variable cost.

The equivalence of the two approaches to the shutdown decision is also easy to see. Our first rule was TR>TVC. But TR is simply PQ, so the rule is PQ>TVC. But if we divide both sides of this inequality by Q, we have the equivalent rule P>TVC/Q or P>AVC.

Long Run Considerations

When we relax the assumption that the business has some fixed resources we move from the short run to the long run. In this case, all resources are variable. And, if all resources are variable, all costs are variable. No fixed cost exists in the long run. While the intricacies of a model of a business in the long run are well beyond the technical scope of our introduction to the theory of the business, one principle is worth noting, the idea of economies of scale.

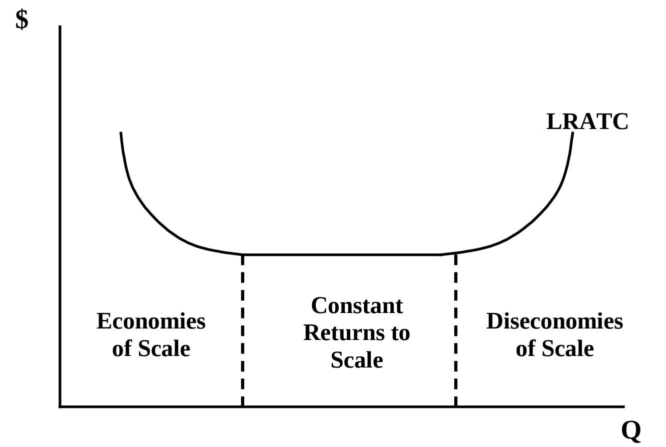


Figure 5-15
Scale Economies and Long Run Average Total Cost

In Figure 5-15 we see a U-shaped curve representing long run average total cost, labeled LRATC. Again all resources are variable. With the LRATC curve we can define the meaning of economies and diseconomies of scale. If LRATC falls as output increases, as the scale of the business increases, we say we are experiencing economies of scale. The wide range of output with a flat LRATC is what economists and production managers call constant returns to scale. Eventually, when LRATC increases, we say the firm is experiencing diseconomies of scale. The curve in Figure 5-15 is sometimes called the "bathtub curve," reflecting an empirical regularity of a wide range of output where the business has constant returns to scale.

Chapter 6: Marginal Analysis and a Theory of a Business: Resource Demand

Introduction

We saw in Figure 5-1 at the beginning of the last chapter a numerical example showing that any business faces two fundamental, related decisions, how much output to produce and how much of each resource to use in producing it. In Chapter 5, we focused on the output decision of the business. In this chapter, we focus on the resource decision by developing a model of resource demand.

To simplify the exposition, we again assume that the business has one variable resource, labor, and one fixed resource, capital. We also assume that the business has market power in neither the product market nor the resource market. This implies that the business is a "price taker" in the product market and a "wage taker" in the labor market, i.e., it takes both the product price and labor wage as given.

A Numerical Example Showing the Marginal Revenue Product & Marginal Resource Cost of Labor, a Variable Resource

We begin with the production function information shown in Figure 5-1. This is the same numerical representation of the production function that we used in Chapter 5. The first three columns of Figure 6-1 contain the production function data. The business can hire from 0 to 8 workers per day that can work with 1 unit of fixed capital. Column 3 shows the output obtained from the different levels of resource use.

1	2	3	4	5	6	7	8	9
L	K	Q	MP	P	TR	MR	MRP	W
0	1	0		$20	0			$120
			20			$20	$400	
1	1	20		$20	$400			$120
			30			$20	$600	
2	1	50		$20	$1,000			$120
			25			$20	$500	
3	1	75		$20	$1,500			$120
			20			$20	$400	
4	1	95		$20	$1,900			$120
			15			$20	$300	
5	1	110		$20	$2,200			$120
			10			$20	$200	
6	1	120		$20	$2,400			$120
			5			$20	$100	
7	1	125		$20	$2,500			$120
			2			$20	$40	
8	1	127		$20	$2,540			$120

Figure 6-1
Marginal Product, Marginal Revenue, and Marginal Revenue Product: A Numerical Example

In column 4, we see the important information on marginal product, the change in output produced by hiring one additional unit of labor. Because of the principle of diminishing marginal productivity, the marginal product of labor begins to decline eventually, in this case, after hiring the second worker. Column 5 shows the constant price of $20 per unit of output, and column 6 shows total revenue, the product of price and quantity produced, the number in column 5 times the number in column 3. Because price is constant in this example, marginal revenue is the same as price, $20. Remember that marginal revenue is defined as the change in revenue divided by the change in quantity produced.

In column 8, we have a new concept, the marginal revenue product. Marginal revenue product is calculated by multiplying marginal product and marginal revenue. Hiring an additional worker will increase output by the marginal product of labor. But how much additional revenue will this additional output bring to the business? If we multiply the additional output by the additional revenue created by this output we get the answer.

The marginal revenue product is the additional revenue brought in by hiring an additional unit of the variable resource, labor in this case. In order to get additional revenue, the business must produce more output. The marginal product gives how much more the additional labor produces. But revenue doesn't come from mere production alone. The business has to sell the additional output, and how much it can sell the additional output for, the price in this case, determines marginal revenue. The product of marginal product and marginal revenue is marginal revenue product.

Note that marginal revenue product rises initially in column 8, but then, like marginal product, it falls as additional units of the variable resource are hired. The declining marginal revenue product in this case is totally due to the principle of diminishing marginal productivity. It has to be solely the result of diminishing marginal productivity, because marginal revenue remains constant in this example as additional labor is employed.

We need one additional piece of information before we can determine the profit-maximizing level of resource use. Revenue is only half of the profit picture for a business. The marginal revenue product of labor tells us only what additional labor generates in revenue. But labor is not free. We have to buy it in the labor market and pay the going wage. In deciding how much of a variable resource to hire, the business must compare the marginal revenue product with the marginal resource cost, the daily wage in this example. The constant marginal resource cost (wage) is shown in column 9 of Figure 6-1.

To determine the right amount of the variable resource to hire, the business must compare the marginal revenue product to the marginal resource cost (wage). Using standard marginal analysis reasoning, if additional labor brings the business more revenue than it has to pay out in wages, profit of the business will increase, and increasing profit is the goal of the business. As long as marginal revenue product is greater than marginal resource cost the business should hire more labor. In the example in Figure 6-1, the profit maximizing business will hire 6 workers per day. Up to 6 workers per day the marginal revenue product exceeds the daily wage. Hiring the seventh worker will increase revenue by $100. That's good. But to get that additional revenue from hiring the additional worker, the business would have to pay $120 a day. Paying $120 to make $100 is a losing proposition.

A Model of Marginal Product and Marginal Revenue Product

The two panels of Figure 6-2 are models of the marginal product (MP) and marginal revenue product (MRP) of labor For simplicity, we have ignored the upward sloping part of the functions. In both panels, the line on the graph is downward sloping because the marginal product of labor declines as employment of labor increases.

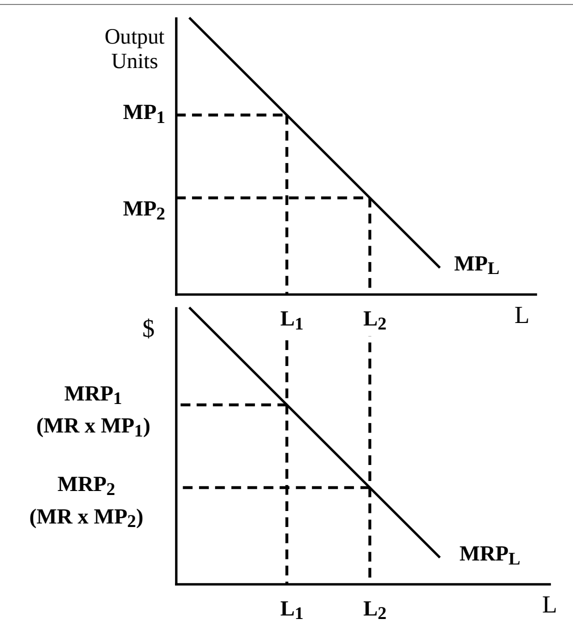


Figure 6-2
A Model of Marginal Product and Marginal Revenue Product of Labor

The horizontal axes of both graphs are the same, the amount of labor employed. The vertical axis of the upper graph is in output units and measures marginal product. In the lower graph, we multiply marginal product by marginal revenue to obtain marginal revenue product, which is measured in $ units on the vertical axis.

The Effect of Capital on the Marginal Product and Marginal Revenue Product

In Figure 6-2, we constructed the MP and MRP curves for a given level of capital, the fixed resource. In Figure 6-3, we examine the effect of a change in the amount of capital in the production process. In each of the panels, we have curves drawn for two different levels of Capital, K_1 and K_2, where $K_2 > K_1$. Note that with more capital in the production process labor becomes more productive. With K_1 capital and the amount of labor L_1, the marginal product and the marginal revenue product are MP_1 and MRP_1, respectively. But at this same amount of labor, and more capital, the MP and MRP curves shift up, and the marginal product and marginal revenue product are now MP_1' and MRP_1'.

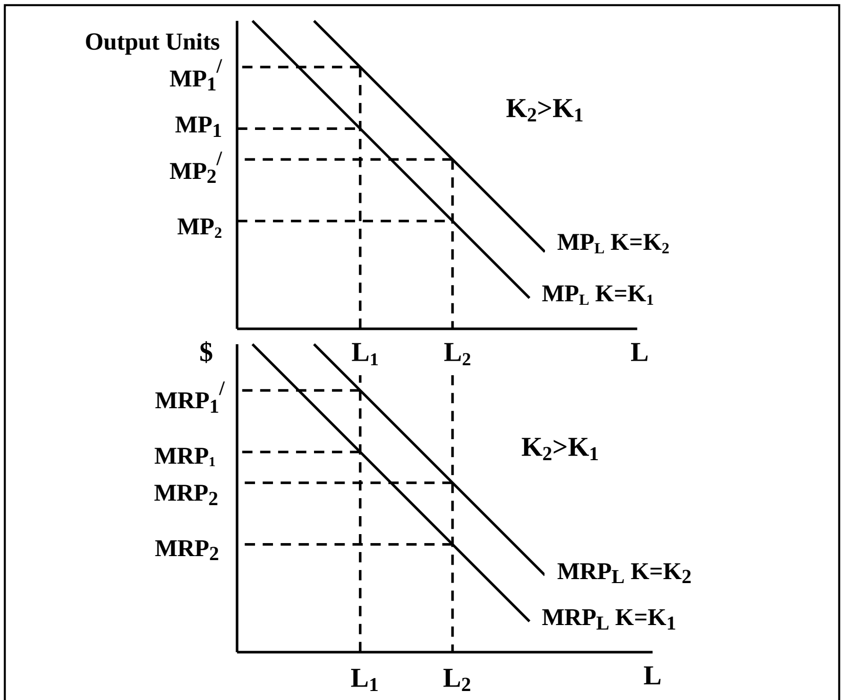


Figure 6-3
Marginal Product and Marginal Revenue Product of Labor with Different Amounts of Capital

While the models in Figure 6-3 are abstract, it is not hard to imagine this type of effect on the marginal product of labor in the real world. Consider a simple production process, moving a pile of rocks 100 yards, with a wheelbarrow as the fixed capital. Additional labor will definitely increase the amount of rocks hauled. But with a pickup truck as capital, additional labor will move a much higher additional amount of rocks in a unit of time.

The Effect of Product Price on Marginal Revenue Product

A different amount of capital is not the only change that can affect the marginal revenue product of a variable resource. A change in product price also changes the marginal revenue product. This is shown in Figure 6-4. In both panels of Figure 6-4, we assume the business has the same amount of capital $K = K_1$. Note that in the bottom panel the marginal revenue product curve shifts up with an increase in price from P_1 to P_2. In the bottom panel, with labor at L_1, the increased price increases the marginal revenue product from MRP_1 to MRP_1'. Likewise, with labor at L_2, the increase in price increases the marginal revenue product from MRP_2 to MRP_2'. On the vertical axis, we show in parentheses under the MRP notation the method of calculating the MRP.

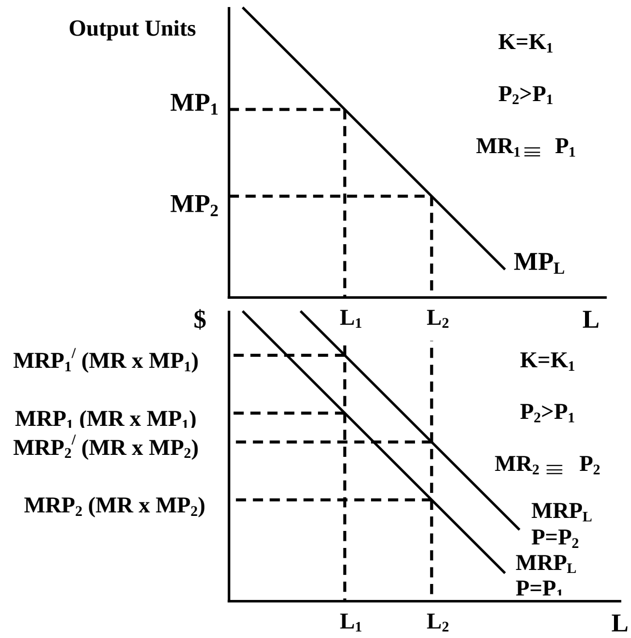


Figure 6-4
Marginal Product and Marginal Revenue Product of Labor with the Same Amount of Capital and Different Output Prices

Nothing happens to the physical aspects of the production process when the price of the product changes. This is why there is no shift in the marginal product curve in the upper panel of the Figure 6-4. The physical productivity of labor is not enhanced with an increase in product price, just the revenue productivity.

While the model in Figure 6-4 is definitely abstract, it is not difficult to think of real world examples of product price increasing the revenue productivity of a variable resource. Consider a manufacturing process for electrical components. While a great deal of capital is involved in this production process, much labor is involved as well, e.g., sorting, testing, and packaging the product. More labor, even with a fixed amount of capital will increase output. The marginal product of labor is positive, as is the marginal revenue product. Suppose the price of electrical components increases. While the marginal product of labor does not change, the marginal revenue product of labor increases, because the additional output produced by the additional labor can be sold by the business at a higher price.

Marginal Analysis and the Profit Maximizing Employment of a Variable Resource

Up to this point in the chapter we have concentrated on what additional labor can do for a business in the form of more output and more revenue from selling that output. But hiring labor, or any variable resource, is not just a matter of how much revenue a certain amount of labor can generate. Labor is not free. It has to be purchased in the marketplace. Figure 6-5 is a model of this employment decision.

On the horizontal axis, we have the amount of labor, and on the vertical axis we can read the marginal revenue product and the wage paid to labor. The marginal revenue product curve is downward sloping, reflecting diminishing marginal productivity. The marginal resource cost (MRC) is represented in the model by a horizontal line equal to the wage of labor. The constant wage is due to the assumption that the neither the business nor workers have any market power. The wage is established in the labor market. The business is a "wage taker." It can hire as much labor as it wants without having to raise the wage to entice more workers to work for it.

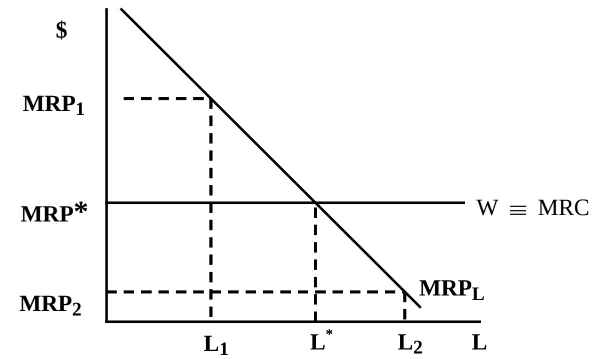


Figure 6-5
Marginal Analysis and the Profit Maximizing Amount of Labor

The marginal analysis we employ in the model in Figure 6-5 is the same as the analysis we applied in the numerical example at the beginning of the chapter. The marginal analysis reasoning is that the business should hire additional labor as long as the additional labor brings in more revenue than it creates in cost. The additional cost, of course, comes from the additional wage payments to the new labor.

At all levels of labor less than L* in Figure 6-5, e.g. $L = L_1$, $MRP_L > W$. If we ask whether hiring additional labor will bring in more revenue than it generates in cost, the answer will be yes at points to the left of L*. At L*, where $MRP_L = W$, adding additional labor will not add more to revenue that cost. At the amount of labor $L = L_2$ $MRP_L < W$. Here, hiring more labor would add more to wage cost than it would generate in revenue. Certainly, the business would not want to hire more labor in this case. In fact, when $MRP_L < W$ it would pay the business to fire some workers. Marginal revenue product is positive, so firing workers will result in less output and less revenue from selling that output. But firing workers will reduce wage cost as well, and because MRP_L<W, the reduction in wage cost will exceed the reduction in revenue. Profit will increase.

Again, while analysis in this section is admittedly abstract, the model represents a very sound business principle that can be applied in the real world. As long as hiring additional labor, or any variable resource, adds more to revenue than it does to cost, doing so will earn higher profits for the business. If firing workers lowers wage cost more than it lowers the revenue from the reduced output and sales, profit of the business will increase. Thinking at the margin in resource decisions has immense practical importance.

Marginal Revenue Product and the Demand for Labor

In Figure 6-6, we extend the model of Figure 6-5 to show that the marginal revenue product curve is really a model of the demand curve for a variable resource. Using marginal analysis and three different wage rates facing the business, we see in Figure 6-6 that as the wage falls the business will want to hire more labor, and if the wage increases it will want to hire less. This relationship is none other than a resource demand relationship. For this reason, we have added the additional notation d_L to the marginal revenue product curve in Figure 6-6. We use a small d to signify that that this is a demand curve by a business, not the demand curve in the market as a whole.

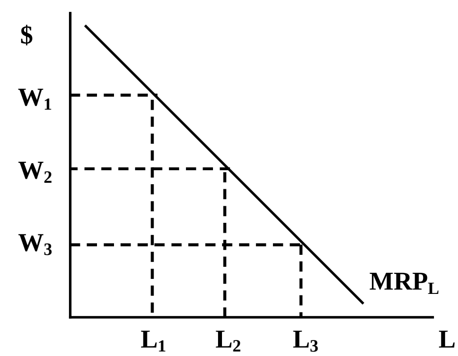


Figure 6-6
Marginal Revenue Product and the Demand for Labor

Figure 6-7 shows the effect of different amounts of capital and different product prices on the business' demand for labor. Each of these changes shifts the labor demand curve. In the top panel, we see the business hiring L_1* labor when K = K_1. If capital increases, the demand curve shifts up and to the right and the business chooses to increase its employment of labor to L_3*. Likewise, at a lower level of capital the labor demand curve shifts down and to the left and the business chooses to employ less labor at the given wage, e.g., L_2*

In the bottom panel of figure 6-7, we show the effect of product price change on the business' demand for labor. At $P = P_1$, the firm chooses to hire L_1* labor. With an increase in product price, the business' demand curve for labor shifts up and to the right and it chooses to employ more labor, L_3* in the model. Likewise, when product price falls, the labor demand curve shifts down and to the left and the business chooses to hire less labor at the given wage.

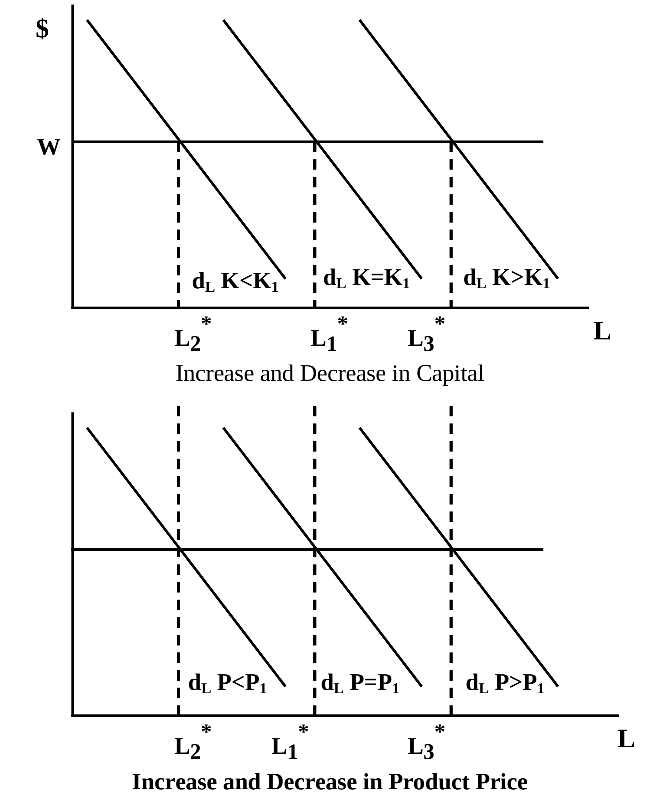


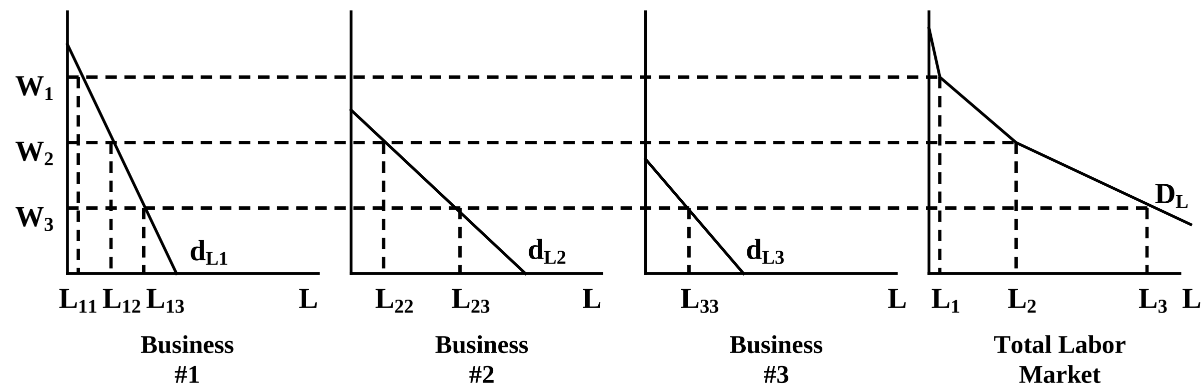
Figure 6-7
Changes in Capital and Product Price and a Business' Demand for Labor

Aggregating Individual Labor Demands to Market Labor Demand

We can add together the individual labor demand curves for each business to obtain a market demand curve for labor. This is a horizontal summation just like that we performed when we were aggregating individual demand curves in Chapter 4 and individual business product supply curves in Chapter 5.

In Figure 6-8, we have models of the labor demand for three businesses and the market demand for labor. At $W = W_1$ only Business #1 is demanding this type of labor. The notation L_{11} means the amount of labor demanded by business #1 at $W = W_1$. L_{12} is the amount of labor demanded by Business #1 at $W = W_2$. L_{22} is the amount of labor demanded by Business #2 at $W = W_2$, and so on. The market demand for labor in the rightmost graph has only one subscript, e.g., L_1 at $W = W_1$. At $W=W_2$, both Business #1 and Business #2 want to purchase labor. So we add together their individual demand to form the quantity of labor demanded in the market, $L_{12} + L_{22} = L_2$. Finally, at the lower wage $L = L_3$, all three businesses demand this type of labor, so the horizontal summation yields a new point on the labor market demand curve, $L_3 = L_{13} + L_{23} + L_{33}$.

Figure 6-8
From Business Labor Demand to Market Labor Demand



A final interesting point here is that if all businesses face the same wage, they will all hire labor up to the point where the marginal revenue product of labor is equal to the common wage. Because each business faces the same wage, each has the same marginal revenue product at the profit maximizing level. Because we assume no business has any market power in the product market, their marginal revenue product is equal to price times marginal product. This is really the value of the additional output that can be produced by adding labor in any of the businesses. This principle has normative significance for resource allocation. If all businesses are price takers in their respective product markets, and all are wage takers in the labor market, the value of the marginal product (same as MRP in this case) of labor will be the same for all firms. This means that an economy could not grind any more value out of its labor resource by reallocating it among businesses. Profit maximization on the part of businesses, and a lack of market power in both product and resource markets, leads to a situation that maximizes the value of the output from that resource. We'll return to this principle again in Chapter 8 when we study normative economics in more detail.

The Time Value of Money, Future Value and Present Value

To this point in the chapter, we have examined marginal analysis and the demand for a variable resource in the short run. Remember that the short run is a time period in which one or more resources are fixed. In our discussion, we assumed capital was the fixed resource. But in the long run, a business must determine the amount of capital as well, which in turn leads to the demand for capital goods. When making long run decisions, time becomes an explicit component in the analysis, especially the time value of money.

The time value of money is a direct result of the productivity of saving in the economy. If we forego consumption today, this saving can be invested in productive opportunities that will yield, on average, a positive rate of return over time. As time progresses, we can earn a return not only on our initial savings, but also on the return itself. This is the common notion of compound interest, which leads to the related concepts present value and future value.

The relationship of present value, compound interest and future value is shown in Figure 6-9. Suppose at time period 0 we begin with a sum of money PV_0. If we can invest this money in a productive opportunity, perhaps a bank account, with a percent rate of return of r per period, at the end of one period we will have earned rPV_0 in interest. Let's also assume that our initial sum is also preserved in our investment so that at the end of the first period we still have our initial sum of money PV_0 plus our interest return rPV_0. Let's call this amount at the beginning of the next period the future value of a sum invested for one period, FV_1, such that $PV_0 + rPV_0 = FV_1$. Factoring out the PV_0 term yields the equation in the Future Value column of Figure 6-9.

Future Value	**Present Value**
FV	**PV**
$\mathbf{PV_0\,(1+r) = FV_1}$	$\mathbf{PV_0 = \dfrac{FV_1}{(1+r)}}$

Figure 6-9
The Relationship of Future Value and Present Value of a Lump Sum: The One-Period Case

The idea of future value is very easy to understand, because most of us have experienced the idea of earning interest in our lives. Indeed, one of the aspects of good parenting is teaching children about the benefits of thrift, one of which is the ability to earn interest from our ability to save, to be compensated for our ability to delay gratification by foregoing consumption. By contrast, the notion of present value is more difficult to understand, even though it is based on exactly the same concept, an interest rate of return.

If we divide both sides of the equation in the Future Value column of Figure 6-9 by (1 + r) we obtain the equation in the Present Value column. The prevent value of some amount FV_1, received one period in the future, has a present value of $FV_1/(1 + r)$ one period earlier. In my experience teaching the idea of present value, this simple algebraic maneuver initially loses more than half of the students of present value.

Because we assume the interest rate is positive, $(1 + r) > 1$, when we divide a future value by $(1 + r)$ we get a smaller number. This is why present value is often called discounted present value. The interest rate r is called the discount rate and $1/(1 + r)$ a discount factor. If we multiply a future value by a discount factor, the equivalent of dividing it by 1 + the discount rate, we obtain the present value of a lump sum to be received one period hence.

This process of discounting actually has a very intuitive explanation. Suppose I propose that you give me \$1,000 and I will return \$1,000 to you in a year. You won't do this because I will have the \$1,000 now, and could put it in the bank at some positive interest rate and have \$1,000(1 + r) in a year. I would earn the interest on the money, not you. \$1,000 a year from now has a present value today that is less than \$1,000. Because interest rates are positive we forego interest when we delay receipts. With positive interest rates, there is an opportunity cost of waiting, because if you have a sum of money earlier, you can put it in some productive opportunity and earn a positive rate of return.

$$\left.\begin{array}{l} \mathbf{PV_0\ (1+r) = FV_1} \\ \mathbf{FV_1\ (1+r) = FV_2} \\ \mathbf{FV_2\ (1+r) = FV_3} \\ \mathbf{FV_3\ (1+r) = FV_4} \end{array}\right\} \quad \textbf{four one-period interest returns}$$

$$\mathbf{\underbrace{\underbrace{\underbrace{PV_0\ (1+r)}_{FV_1}\ (1+r)}_{FV_2}\ (1+r)}_{FV_3}\ (1+r) = FV_4}$$

$$\mathbf{PV_0\ (1+r)^4 = FV_4}$$

Figure 6-10
The Relationship of Future Value and Present Value of a Lump Sum: The Multi-Period Case

So far we have examined present value and future value only in the context of one period investments. Multiperiod examples are a simple extrapolation from the one-period case. This is shown in Figure 6-10. At the top of the figure we see a collection of four one-period present value and future value equations. If we substitute judiciously, however, we can turn the four separate one-period equations into one four-period equation. Just substitute $PV_0(1 + r)$ for FV_1 in the second equation and we get $PV_0(1 + r)(1 + r) = FV_2$. Substitute this for FV_2 in the third equation and we get

$PV_0(1 + r)(1 + r)(1 + r) = FV_3$, the third equation, and so on. These substitutions are summarized in the middle portion of Figure 6-10, with brackets indicating each round of compound interest. At the bottom of the figure, we simply collapse the four (1 + r) terms multiplied together into the expression $(1 + r)^4$.

In Figure 6-11, we generalize the equations for present and future value to a multiperiod case involving t periods. Note again that future value involves compound interest for t periods. The future value of a lump sum t periods in the future is larger than the present value, with the difference in magnitude a function of the rate of return, r, and the number of periods. Likewise, the present value of a lump sum future value received t periods in the future is less than that future value. Because of the foregone interest caused by waiting for the lump sum, a future value must be discounted. The magnitude of the discounting depends on the foregone rate of return, r, also called the discount rate in this case, and the number of time periods. The greater the discount rate, the lower the present value. The greater the number of time periods, the lower the present value.

$$\mathbf{PV_0\ (1+r)^t = FV_t}$$

or

$$\mathbf{PV_0 = \frac{FV_t}{(1+r)^t}}$$

Figure 6-11
General Equations Relating Present Value and Future Value of a Lump Sum Over t Periods

The final step in understanding the relationship between present value and future value is made by considering the present value of several future values that occur at different time periods in the future. Economists often call this a “stream” or a “flow” of future returns. In order to compute the present value of a stream of future returns, we only need to apply the present value (discounting) operation several times, one for each of the distinct time periods in the future. In each time period, we apply a different discount factor due to the different length of time we have to wait for the lump sum of that period. The present value of this stream of returns is simply the sum of all these individual lump sum future values discounted back to the present with the appropriate discount factor, $1/(1 + r)^t$. This relationship is shown in Figure 6-12.

This present value relationship is one of the most useful relationships in business and economics. Often the relationship is called a capitalization formula, i.e., it generates a lump sum capital value of a stream of future returns discounted at some discount rate r. If you wish to value corporate stocks and bonds, figure out the value of an apartment building, build a dam or buy energy-efficient appliances, among many other actions, knowledge of the present value relationship is indispensable.

$$PV_0 = \frac{FV_1}{(1+r)} + \frac{FV_2}{(1+r)^2} + \ldots \frac{FV_t}{(1+r)^t}$$

or

$$PV_0 = \sum_t \frac{FV_t}{(1+r)^t}$$

Figure 6-12
The Present Value of the Sum of Many Future Values Occurring in Different Time Periods

Net Present Value, Internal Rate of Return, and the Demand for Capital

Now that we have an understanding of how to figure out the value today of a stream of future returns, present value, we can now approach a theory of a business' demand for capital. We need the understanding of present value because the nature of the capital resource is that we produce it today so that it can produce for us in the future. The demand for capital is very much forward looking, and involves a comparison of the present value of future returns with the cost of capital.

Two useful concepts in understanding the demand for capital are net present value and internal rate of return. The top two equations in Figure 6-13 define net present value (NPV) of an investment in capital. The (1 + r) terms in the denominators of the equations are the now familiar discount factors discussed in the previous section. The R terms in the numerators are net returns, positive cash flows minus negative cash flows, from an investment in capital. The NPV of this investment is simply the discounted present value the future stream of net returns. When we calculate NPV, we use a particular value of the

$$NPV = \frac{R_1}{(1+r)} + \frac{R_2}{(1+r)^2} + \ldots \frac{R_t}{(1+r)^t}$$

$$NPV = \sum_t \frac{R_t}{(1+r)^t}$$

$IRR \equiv$ value of r, such that:

$NPV = 0$

Figure 6-13
Definition of Net Present Value and the Internal Rate of Return

discount rate r. Determining the proper discount rate for a particular business is beyond the scope of our efforts here, but is an important topic in finance.

Related to the NPV is another concept called the internal rate of return (IRR). As noted in Figure 6-13, the IRR is the discount rate in the NPV relationship that makes the NPV equal to zero. The relationship of the NPV and the IRR is shown graphically in Figure 6-14. Assume we have a capital project with net cash flows such that the NPV falls as the discount rate rises. This would be a project that has either relatively balanced positive net cash flows over time, or one that has negative net cash flows concentrated in early periods. These would be the typical capital projects, with negative returns in the first period when the capital good is purchased, and positive periodic returns over time in the future. If, however, large negative cash flows occurred in the distant future, it is possible that a higher discount rate would lead to an increase in NPV, as these negative returns would be reduced in the overall summation due to a larger discount factor. In the case assumed here, as the discount rate rises the NPV falls. The discount rate indicated by the intersection of the declining NPV line and the horizontal axis is the IRR.

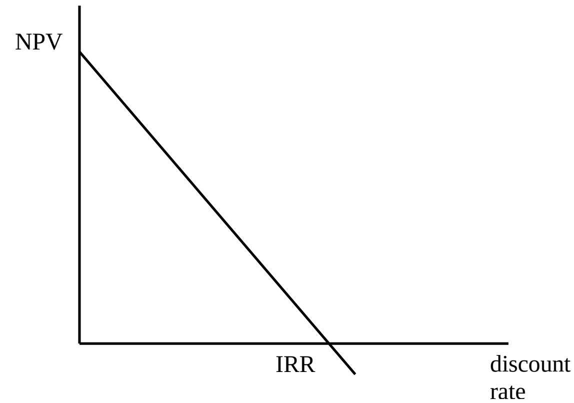


Figure 6-14
Relationship Between Net Present Value and the Internal Rate of Return

The IRR measures the percent rate of return on a project. A business would prefer a capital project with higher IRR to projects with lower ones. If we have a stream of future net cash flows that requires a high discount rate to drive the present value of that stream of cash flows to zero, the stream of net cash flows must be a good one. If the future stream of net cash flows requires only a low discount rate to drive the NPV to zero, that stream of net cash flows is less good.

If the business had a large and infinitely divisible number of capital projects, it could rank them from highest to lowest. As the business employed more and more capital it would adopt projects with lower and lower internal rates of return. Such a relationship is represented by the downward sloping IRR_K line in Figure

6-15. How much capital should the business employ? The answer to this question depends not only on the return from capital, but also on the cost of capital, with the cost of capital defined as the opportunity cost to the business of investing in capital projects. An important topic in finance is determining for a business the magnitude of this cost of capital. The cost of capital, in turn, depends on how the capital projects are financed, e.g., through retained earnings, or issuing new equity or debt. If we know the proportion of each in a business' financing mix, we can compute a weighted average cost of capital for the business. Again, how this is actually done is beyond the scope of our efforts here, but an important topic in finance.

Let's just assume here that the business has a constant marginal cost of capital indicated by the horizontal line in Figure 6-15. The profit maximizing business will want to employ marginal analysis in its demand for capital. As long as the marginal rate of return from additional capital investment, the IRR_K, is greater than the marginal cost of capital, profit and the value of the business is increased by employing additional capital. In Figure 6-15, this occurs at K* where the IRR_K is just equal to the marginal cost of capital. An amount of capital less that K* "leaves money on the table" so to speak. Capital investment beyond K* costs more than it returns to the business.

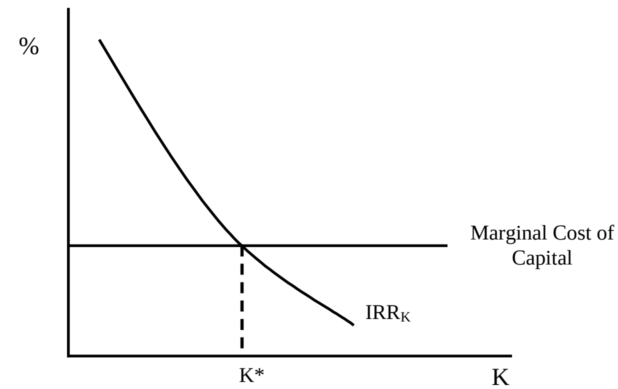


Figure 6-15
Internal Rate of Return, the Cost of Capital, and the Profit Maximizing Amount of Capital

Finally, the IRR_K curve plays the role of a demand curve for capital just as the MRP_L curve is a demand curve for labor. A higher marginal cost of capital, other things held constant, will lead to a lower demand for capital. A lower marginal cost of capital will lead the business to choose a higher amount of capital. This is a demand curve for capital.

Chapter 7: Market Structure: Perfect and Imperfect Competition

Market Power, Market Structure and Perfect and Imperfect Competition

In this chapter, we classify businesses according to their market power. Market power comes in a short run and a long run form. A business has market power in the short run if it has pricing power, the ability to raise its price and not have its sales go to zero. In the long run, a business has market power if it can continue to earn greater than normal profit. Characteristics of the market as well as characteristics of the product combine to determine the degree of market power. The different market and product characteristics that determine a

MARKET STRUCTURE

Market and Product Characteristics	Perfect Competition	Monopolistic Competition	Oligopoly	Monopoly
Number of Sellers	Many	More than a few, less than many	Few	One
Product Characteristics	Homogeneous	Differentiated	Homogeneous or Differentiated	Only
Entry to Market	Easy	Easy	Difficult (“Barriers to Entry”)	Very Difficult (“Barriers to Entry”)
Pricing Power	None	Some	Some, depending on rivals	High
Profit in Long Run	No	No	Possible	Possible

Figure 7-1
A Market Structure Taxonomy

business’s market power are summarized in Figure 7-1. The characteristics run down the first three rows of the table and the four main market structures run across the columns. The bottom two rows of the table indicate whether the business in that particular market structure has pricing power in the short run and/or the power to earn greater than normal profit in the long run.

In the first market structure column, we see perfect competition. In this market structure, we find a large number of businesses each producing a homogeneous (standardized) product. This combination yields an absence of pricing power. Suppose a perfectly competitive business tries to raise its price. Buyers won't buy from this business, as they can buy the identical product at a large number of other businesses. Businesses in perfect competition are price takers, not price setters. Many agricultural markets have a large number of sellers and homogeneous products. Producers of computer memory chips are in a similar situation. If Farmer Brown asks a wheat buyer to pay him a premium for his wheat, the buyer says no and buys from Farmer Smith next door. With no pricing power, farmers listen to the farm report during their noon meal to get information on prices, rather than engage in lengthy pricing strategy meetings.

Perfect competition is a perfectly miserable business to be in from the standpoint of a seller. Not only do you have no pricing power, but if you are fortunate enough to receive a price greater than you average total cost and earn greater than normal profit, easy entry to the market will eliminate the profit in the long run. Note that we have classified perfect competition as having easy entry.

For our second market structure let's jump to the rightmost column of Figure 7-1. Here, we find the opposite of perfect competition, Monopoly. A key to understanding monopoly is to recognize the term mon in its name. A monarchy is rule by one, usually a king or queen. Here, we have one seller who supplies the entire market, and, by default, it has a unique product. This is the epitome of pricing power, as the monopolist sets the price that will get buyers to buy the amount that it wants them to buy, the amount that maximizes the monopolist's profit. In contrast with the price-taking behavior of the perfectly competitive business, the monopolist is a price setter.

Furthermore, a monopolist enjoys barriers to entry to the market. It can earn greater than normal profit, and when this signal travels to the marketplace, no one can follow the signal and enter the market. A monopoly can earn greater than normal profit even in the long run. From the point of view of owners of the business, a monopoly is the kind of business you want to have. You have control of your price and the possibility of persistent greater than normal profits.

Barriers to entry can come in many forms. Legal barriers are common. Patents give producers exclusive rights to sell a product for a certain amount of time. Many pharmaceutical companies have exclusive rights to sell a drug. The company sells at a high monopoly price until the patent runs out, and then generic drugs enter the market and drive the price down. Economies of scale often lead to conditions where one business can supply the entire market at a lower cost than many smaller ones. Local retail electricity and natural gas companies resemble this special form of monopoly called natural monopoly. We'll have more to say about natural monopoly in the next chapter when we consider public policy alternatives for dealing with monopoly.

Barriers to entry can occur even if more than one business supplies the market, conferring a degree of long run market power on sellers in the market. Many cities limit the number of taxicabs on the streets. In addition to some degree of consumer protection, licensing often gives long run market power to businesses in a market. Contractors, hairstylists, and cable television companies are just three among many examples of businesses who get market power from licensing.

The term monopolistic competition is a mouthful to say. This seemingly paradoxical market structure is a combination of the characteristics of competition and monopoly. Businesses in this type of market share with perfect competition the characteristic of a large number of sellers, although we often think of the number of sellers in monopolistic competition as being lower than that in perfect competition. The key difference between perfect and monopolistic competition is seen in the product characteristics row of Figure 7-1. In monopolistic competition, sellers have differentiated products. This product differentiation allows a level of product preference and loyalty to occur, and gives the seller some pricing power. The market for fast food is a good example here. Some people just like McDonalds better than Burger King, and vice versa. These and other fast food establishments can raise their prices above those of others in the market and still retain some sales, because their product is just a little different and worth the extra price in the eyes of some consumers. This pricing power is the characteristic that monopolistic competition shares with monopoly.

In the long run, however, the ease of entry to the marketplace makes monopolistic competition more similar to perfect competition than to monopoly. And, with easy entry, comes the inability to earn greater than normal profit in the long run. Consider retail coffee shops such as espresso carts and drive-in coffee stands. Each business has a product that is differentiated from its rivals in the market, yet if the word gets out that owners of coffee carts or stands are making high profits, the good times will be fleeting. Entry will drive price down to meet costs in the long run.

Finally, we have the very interesting market structure, oligopoly. The key to understanding this market structure is again in the prefix, oli. Where mon appears in monarchy, oli appears at the beginning of the word oligarchy, rule by a few. In oligopoly, we have selling by a few. The product can be either homogeneous or differentiated. Markets for steel and aluminum and dimension lumber are oligopolies characterized by homogeneous products. Soft drink companies, airlines and tobacco producers are also few in number, but have differentiated products.

Sellers in oligopoly have some pricing power, but their degree of pricing power depends on the actions of their rivals in the marketplace. Think of airline pricing strategy. If one airline reduces its price on a route, the success of that price reduction depends greatly on whether other airlines maintain prices on the route or drop theirs as well. Simultaneous increases in price by all of the businesses in an oligopolistic market will have generally more success than if just one business raises price. This suggests that joint price determination on the part of members of an oligopoly, something called collusion, is something to watch for in this market structure.

We often think entry to oligopolies is difficult, due to large capital demands to start such a large business. But this need not be the case. Entry of low-cost carriers to the airline industry is happening worldwide. The recent success of Ryanair in Great Britain and JetBlue in the United States are but two examples.

Another key result of fewness in oligopoly is that interdependence is a major factor. This is why many economists who study oligopoly use a body of theoretical tools called game theory to examine this market structure. Indeed, John Nash, made famous in the movie *A Beautiful Mind*, won the Nobel Prize in Economics for his pioneering work in this area.

Price and Marginal Revenue for a Business in the Market Structure Perfect Competition

Now that we have seen the broad overview of market structure it's time to examine each form in some analytical detail. We begin with figure 7-2, a model of the revenue side of a business in the market structure perfect competition. This is not the first time we've seen part of this figure. In Chapter 5, we learned that marginal revenue for a business without pricing power is identical to price. In Figure 7-2, we see that this price is determined in the overall market for the good in question according to the forces of supply and demand. The perfectly competitive business takes the price as given. Because there are a large number of sellers in the market, its output is very small as a proportion of total output in the market. It can sell all it wants without shifting the market supply curve enough to change the price. Because the perfectly competitive business sells a product exactly like that of other businesses in the market, it will sell nothing if it attempts to raise its price. Note that we have a capital Q for the quantity supplied and demanded in the market as a whole, and a lower case q for the output of the individual business.

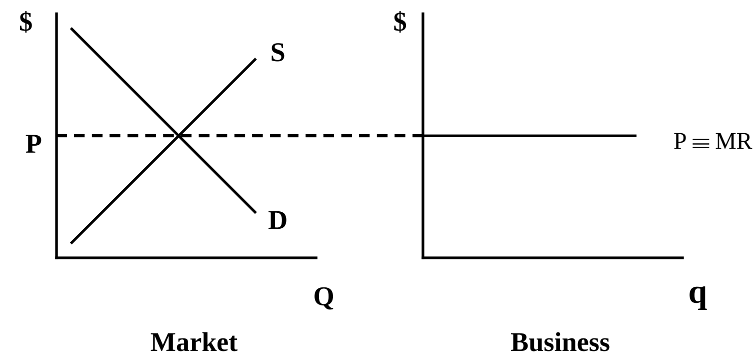


Figure 7-2
Price and Marginal Revenue for the Perfectly Competitive Business

A Model of the Perfectly Competitive Business

If we combine the rightmost model of Figure 7-2 with the short run cost model of a business from Chapter 5, we obtain a short run model of a business in the market structure perfect competition. This is shown in Figure 7-3. Note again that the assumption of profit maximization leads the business in the model to choose to produce output level q*, where price, identical to marginal revenue in this case, equals marginal cost. At output level q*, total revenue is represented as the area of a rectangle with sides representing price and quantity. This is rectangle Oabcdq*. We can also represent total cost in this model as the product of average total cost

(cost per unit) and quantity. This is rectangle Oadq*. Note that the difference between the area of the revenue rectangle and the cost rectangle is the area abcd, which represents short run profit for the business.

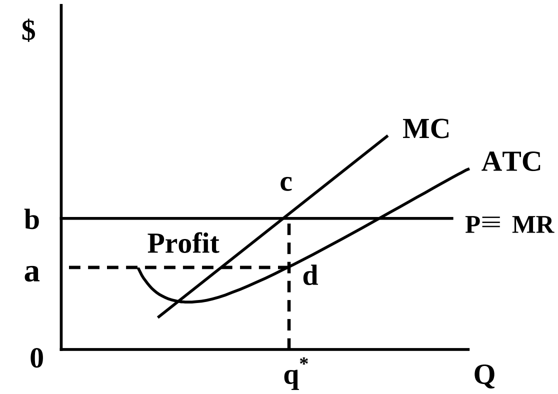


Figure 7-3
A Model of the Business in Perfect Competition in the Short Run

The business modeled in Figure 7-3 and others like it in the market are enjoying favorable circumstances, most of which are beyond the control of the individual business. Price just happens to be above average total cost. We will see below, when we consider the implications of easy entry to the market, that these favorable circumstances are unlikely to last.

The Supply Curve of a Business in Perfect Competition

We now possess enough analytical hardware to define a supply curve for the business in perfect competition. In Figure 7-4, we have added the average variable cost curve to the model. This is necessary because of the shutdown consideration we discussed in Chapter 5. Even if it is earning a loss because price is below average total cost, the business will not shut down in the short run, as long a price is greater than average variable cost. In this case, it is making its loss smaller than the fixed cost it would incur if it shut down. P > AVC means the business is covering its variable cost with something left over. This amount in excess of variable cost reduces the loss. The first part of the

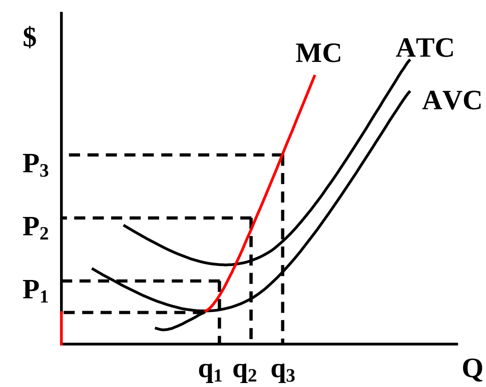


Figure 7-4
Marginal Cost and the Supply Curve for a Business in Perfect Competition

business's supply curve is a vertical red portion of the vertical axis. We'll later ignore this inactive portion of the supply curve, but for now it is instructive. Because profit maximization leads the business to choose the level of output where price equals marginal cost, the portion of the marginal cost curve lying above average variable cost is the active portion of its supply curve. This is also shown in red in Figure 7-4.

From Individual to Market Supply

Each of the businesses in a perfectly competitive market has a supply curve that is identical to its marginal cost curve. In Figure 7-5, we see that the market supply curve can be constructed as an aggregation of the supply (marginal cost) curves of all the businesses in the market. While we show this aggregation for only three businesses in Figure 7-5, we could imagine that many of each of these three exists, so we can be consistent with the requirement of a large number of sellers. The notation in Figure 7-5 is a little cumbersome. Because all businesses face the same price we only need a one-part subscript on the price axis, e.g., P_1, P_2, and P_3. In contrast, note that we have a two-part subscript for magnitudes on the quantity axes of the models in the graphs. q_{11} is the amount supplied by business #1 at the price P_1. Business #2 supplies q_{21} at the price P_1. Business #3

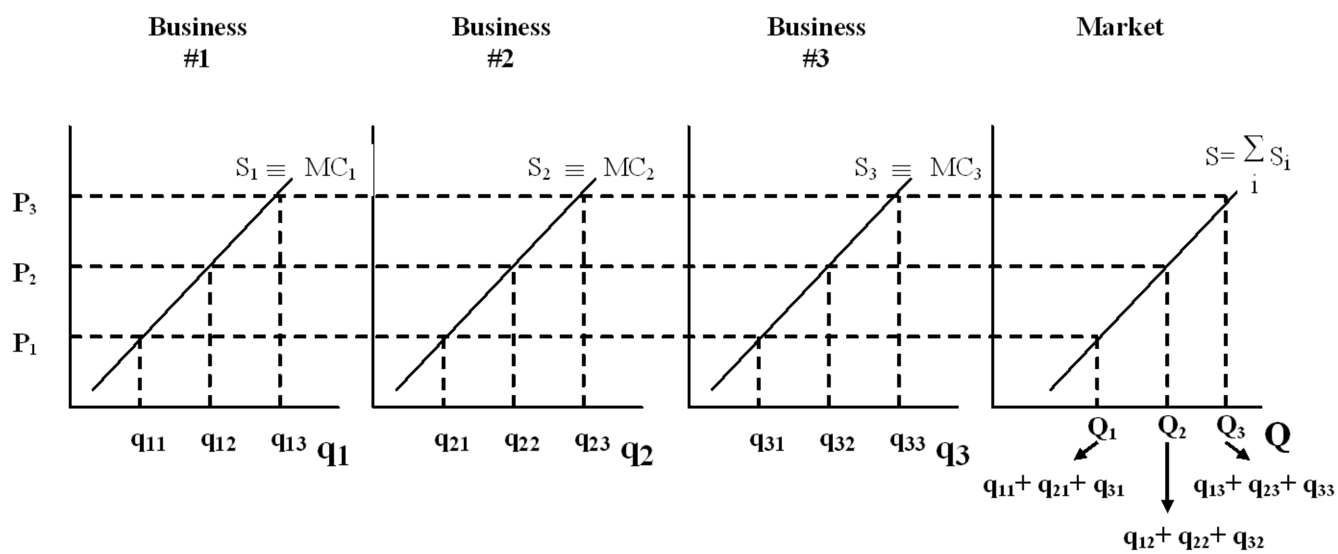


Figure 7-5
From the Individual Business Supply Curve to the Market Supply Curve

supplies q_{31}. The amounts supplied at the price P_2 by Businesses #1, #2, and #3, are q_{12}, q_{22}, and q_{32}, respectively. Similar notation applies to the quantities supplied at the price P_3.

The process of constructing the market supply curve from the individual supply curves in Figure 7-5 is exactly like the process we used in Chapter 4 when we constructed a market demand curve as a "horizontal summation" of individual demand curves. We see in the rightmost graph of Figure 7-5 that the market quantity supplied at a particular price is the sum of the amounts supplied by all the businesses at that price, i.e., $Q_1 = q_{11} + q_{21} + q_{31}$, and so on.

Another very important principle emerges from Figure 7-5. Note that if all businesses face the same price, each will choose a level of output where price equals marginal cost. Profit maximizing output decisions, when combined with a market requiring price taking behavior, lead to the equality of marginal cost across all businesses in the market. So, for any price read off a supply curve in a model of a competitive market, we also know what the marginal cost is in the market. In this sense, the market supply curve can be considered to be a social marginal cost curve. From it we can determine the marginal cost at any level of quantity supplied in the market.

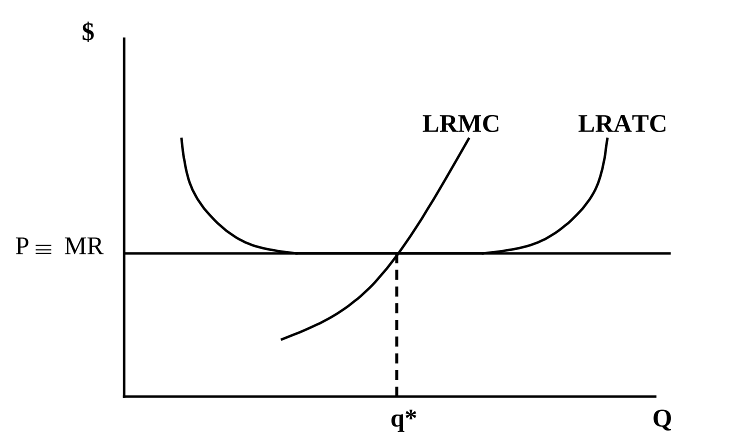


Figure 7-6
Scale Economies and Long Run Average Total Cost

Long Run Adjustment and Equilibrium in a Perfectly Competitive Market

Finally, we end our discussion of the theory of perfect competition by considering long run adjustments. We do this with the model shown in Figure 7-6. Here, we have introduced the notion of long run average total cost, LRATC, and long run marginal cost, LRMC. While there are analytically interesting relationships between SRATC and LRATC and short run marginal cost and LRMC, they are not necessary for the depiction of long run adjustment in a perfectly competitive market.

The long run adjustments are the result of the positive profit signal in the short run. If businesses are earning greater than normal profits as indicated in Figure 7-3, both existing businesses and new businesses will expand production in the market. New businesses now can adjust their previously fixed resources to new conditions. Easy entry allows new businesses to enter. Both actions have the effect of increasing supply in the market, and with a given demand this will drive down price. How far will price fall? It will fall enough so that no positive profit signal exists. When businesses in the market are earning just normal profits, earning a return just equal to the opportunity cost of capital and owner labor invested in the business, expansion of existing businesses and entry of new ones will cease. This is the long run equilibrium in a competitive market shown in Figure 7-6.

Price and Marginal Revenue for the Price Setting Business

The revenue side provides the main analytical difference between the model of a business in a perfectly competitive market and one in either monopoly or monopolistic competition. In these two forms of imperfect competition a business has pricing power. A monopoly derives its pricing power from its position as the sole producer in the market. The business in monopolistic competition gets its pricing power from a differentiated product. Each business faces a downward sloping demand curve. When a business faces a downward sloping demand curve, marginal revenue is no longer equal to price. The relationship between demand, price and marginal revenue for a business with pricing power is shown in Figure 7-7. Note that we have constructed this revenue model with a marginal revenue curve below the demand curve at all levels of output. Let's see why this is the case.

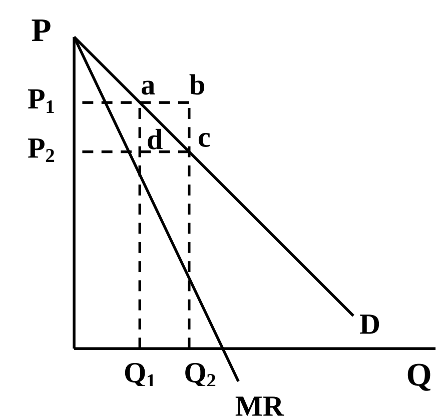


Figure 7-7
Demand and Marginal Revenue for a Business with Market Pricing Power

Suppose the price setting business is charging a price of P_1 for its product, with quantity demanded Q_1. If the business were a price taker it could sell a higher level of output, say Q_2, at the same price, P_1. Additional revenue from the increased output could be represented by the revenue rectangle

Q_1dabcQ_2. Of course, as we explained in the last section, marginal revenue is the same as price in this case. But the business in Figure 7-7 is a price setter, not a price taker. In order to get its customers to buy the quantity Q_2, it must lower the price to P_2. It gains the revenue rectangle Q_1dcQ_2 from this action. Note that this is less than the price taker's revenue gain by the area abcd, so we already can make the argument that MR<P for the price setter. But the case is even stronger.

Customers were purchasing Q_1 at a price of P_1 before the price decrease. Now, even though they are willing to pay P_1 for the quantity Q_1, they can now purchase at the lower price P_2. We are assuming here that everyone in the market pays the same price, i.e., no price discrimination exists. The revenue rectangle P_1adP_2 represents more reduced revenue due to the fact that the business had to lower the price to sell more. At Q_1 marginal revenue is less than price by the amount of these two revenue reductions.

Another point helps in the understanding of this tricky relationship. Note that the vertical distance between the demand curve and the marginal revenue curve becomes greater as the level of output increases. At higher levels of output, the business lowers the price on a larger amount for which consumers were paying a higher price before the price reduction, thereby increasing the difference between price and marginal revenue. Note that price and marginal revenue are equal where the demand curve and the marginal revenue curve meet on the vertical axis at an output level of 0. At zero output, lowering the price does not cause a reduction in revenue from customers who were paying a higher price before, because they were not buying any before.

Marginal Revenue and the Price Elasticity of Demand

We can now relate two important concepts for the business, marginal revenue and the price elasticity of demand. In Chapter 3, in Figure 3-30 to be precise, we showed the relationship between the price elasticity of demand and different points on a demand curve. By examining gain and loss revenue rectangles, we were able to conclude that demand is elastic in the upper part of a demand curve, at points with a high price and low quantity demanded. Likewise, at low price and high quantity demanded we have low price elasticity of demand. In the middle range, price elasticity of demand is unitary.

These elasticity ranges are reproduced in Figure 7-8, but here we also show the marginal revenue curve and the sign of marginal revenue in different ranges of the demand curve. First look at the point where the marginal revenue curve crosses the horizontal axis, and the point on the demand curve represented by P_2 and Q_2. Marginal revenue equals zero at this point. Note that this corresponds to the point on the demand curve where the price elasticity of demand is equal to 1. This is no accident. When the business lowers price and revenue does not change the price elasticity of demand equals 1. The percent decline in price is exactly equal to the percent increase in quantity demanded. But because quantity demanded increases and revenue stays the same, this means marginal revenue must be zero.

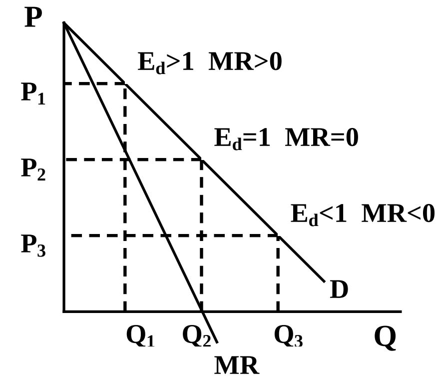


Figure 7-8
Demand, Marginal Revenue and Elasticity of Demand

Now look at the point on the demand curve at P_1. The price elasticity of demand at this point is greater than one. Demand is price elastic. We know that when demand is elastic lowering price will increase revenue, as the quantity demanded increases by a larger percentage than the percentage fall in price. Because the change in quantity is positive, and the change in revenue is positive, the quotient of those two magnitudes, marginal revenue, is also positive. This is why the marginal revenue curve is in positive territory to the left of Q_2.

At the lower price P_3, with quantity demanded Q_3, we are at a point on the demand curve where demand is price inelastic. We know from Chapter 3 that if we lower price with inelastic demand, revenue will decrease, i.e., as quantity demanded increases with a decrease in price and revenue falls, the quotient of these two magnitudes, marginal revenue, must be negative. This is why the marginal revenue curve is in negative territory to the right of Q_2.

Models of Businesses in the Market Structures Monopoly and Monopolistic Competition

If we combine our standard cost model of Chapter 5 with the revenue model of the price setting business, we obtain the short run models of businesses in both the market structure monopoly and monopolistic competition. This is shown in Figure 7-9. Here the business has a two-part decision. First it must choose the level of output that maximizes profit. This is shown by the quantity Q*, where MR = MC. At the intersection of MR and MC we must be careful not to look immediately horizontally over to the vertical axis. Rather we need to ask the question, if the business wishes to sell Q* in the market, what price should it charge. We find this price on the demand curve and label it P*.

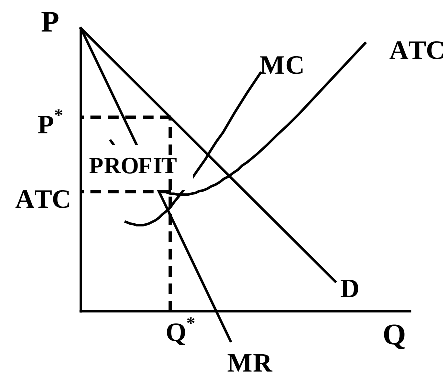


Figure 7-9
Profit Maximizing Price and Quantity for a Business with Pricing Power: Monopoly and Monopolistic Competition

We show the business under a set of favorable circumstance is Figure 7-9. Price is greater than average total cost, so the business earns a profit, represented by the rectangle (ATC)P*ab. This can occur in the short run in both monopoly and monopolistic competition. For the monopoly business, however, because of barriers to entry, greater than normal profit can occur in the long run as well. But, with easy entry to the market, profit will dissipate in monopolistic competition in the long run.

Figure 7-10 shows the long run situation for a business in monopolistic competition. Entrants to the market, with their own differentiated products that are substitutes for other products in the market, take away some of the

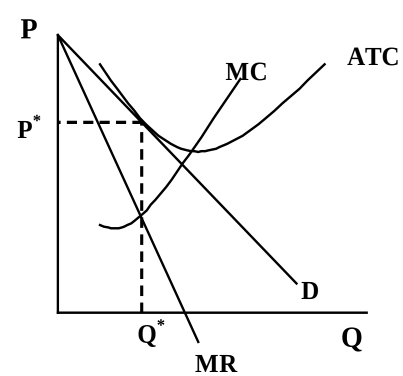


Figure 7-10
Long Run Zero Profit Equilibrium in Monopolistic Competition

demand for the product of the business in question. As long as greater than normal profit exists for businesses in the market, entry will continue to occur. Only when P=ATC will entry cease. An interesting model result occurs where the demand curve facing the monopolistically competitive business is just tangent to the demand curve at the level of output Q*, where marginal revenue equals marginal cost.

A Model of a Sellers Cartel

We mentioned above that perfect competition is a miserable market structure from the point of view of the producer. No pricing power and easy entry limit market power in both the short run and the long run. This is why businesses in perfect competition often dream of banding together to form a joint monopoly, a cartel.

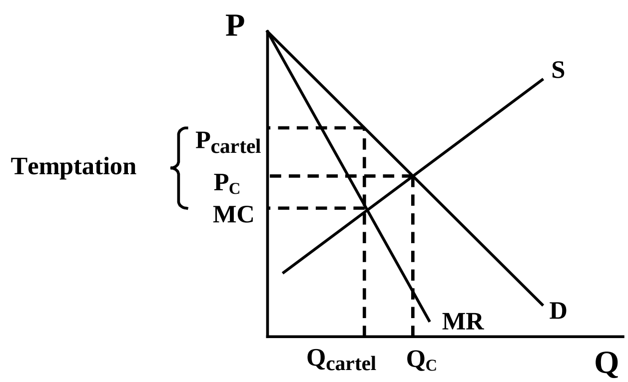


Figure 7-11
A Model of a Sellers Cartel

A model of a sellers cartel appears in Figure 7-11. We begin with the model of a perfectly competitive market, where the supply curve intersects the demand curve at quantity Q_c, giving a market price P_c. Note that the supply curve is the summation of the marginal cost curves of all businesses in the market. In order to form an effective cartel, these businesses have to act collectively like a monopoly. They have to restrict output, and if they can do this price will rise. The profit maximizing cartel quantity and price is found just how a monopoly finds the profit maximizing quantity and price, where marginal revenue equals marginal cost.

Production quotas less than the competitive level are needed to get each producer to reduce output. Equal percent cutbacks would be just one method of reducing output. The problem with the joint monopoly solution for the cartel is also illustrated in Figure 7-11. At the joint profit maximizing cartel output Q_{cartel}, price is greater than marginal cost for each member of the cartel. If the individual cartel member is small in relation to the total output in the market, which we assume here, and if this member assumes that everyone else in the cartel will maintain its production cutback, the cartel price becomes this individual member's marginal revenue. And if MR > MC, we know what action will lead to even greater profit, sell more. We label this distance between the cartel price and marginal cost as "Temptation" in Figure 7-11.

Unfortunately for our individual cartel member, she is not the only one tempted by the potential profit differential. If every cartel member assumes that every other member will maintain its output restriction, all will increase output, price will fall, and the cartel will fail. The search for higher profit, the force that initiated the cartel action, brings down the cartel in the end. If cartels are to succeed, it's better to start them from an oligopoly market structure, than from perfect competition. With just a few sellers, policing the action of individual members is easier, and the threat of being apprehended might outweigh the temptation to cheat on your cartel colleagues.

Chapter 8: Normative Economics and the Case For and Against Collective Action

Normative Economics and Collective Choice Criteria

Up to this point we have attempted to study positive rather than normative economics. This is the scientific aspect of economics where we try to explain the allocation of resources in different economic systems, especially how it's done in market capitalism. When we practice positive economics we refrain from judging whether this allocation is good or bad, or in need of improvement. We simply study how people in economies answer the fundamental economic questions of what to produce, how to produce it, and to whom the output of the economic process accrues.

It's positive economics where economists have both an absolute and comparative advantage over other disciplines. Unfortunately, explaining "what is" satisfies neither the consumers of economics nor its producers. "What should be" is an important issue, one that economists have not been able to resist. We turn now away from the descriptive side of economics to the proscriptive side, normative economics.

In the standard economics of western market capitalist countries, we begin the discussion of normative economics by looking first at the outcome provided in the market. Perhaps this is where we can stop our inquiry. Maybe the allocation found in the market is just fine. But, if it is not just fine, if the market fails to produce the right allocation of resources, somehow defined, we then look to collective (government) action as a potential way to improve on the market allocation. The failure of the market is then a justification for government action. In deciding whether the market allocation is just fine, or whether we need government action, economists use collective choice criteria.

Before we begin the discussion of collective choice criteria, it is worthwhile to qualify this market vs. government distinction. First, we can't really have a market capitalist economic system without having a government to get it going. We need, at a minimum, a way to establish and enforce property rights. It would be possible for businesses to have their own armed guards in every store and their own posses of contract enforcers, but this private enforcement of property rights would be very costly. A rule of law may not be essential for some anarchic form of market capitalism, but it sure makes market capitalism a lot easier and cheaper.

Second, our market capitalist bias is hardly disguised here. If we were predisposed to command socialism, we might approach normative economic questions in just the opposite way. We might check to see if the centrally planned government allocation of resources is just fine, and if it is, according to some criterion, then there might be no need to allow market forces to enter the resource allocation picture. But, if the government allocation is not OK, there might be a justification for market allocation to improve on the failure of government. Here, the resource allocation failure of government is a justification for market allocation.

When mainstream economists in market capitalist environments evaluate market allocations and consider collective choice criteria they usually focus on human welfare. All collective choice criteria don't have this emphasis. A Green economic collective choice criterion might see a justification for the government to change a market allocation in order to improve "environmental integrity" or foster "natural sustainability" irrespective of the effects on human welfare. Likewise, a theocratic economic system might make collective choice decisions based on principles of a holy book. In the Middle Ages, for example, Thomas Aquinas and his colleagues justified economic outcomes according to how they stacked up against biblical dogma.

The Pareto Collective Choice Criterion

We see the influence of human welfare in normative economics when we examine our first candidate for a collective choice criterion. This one is named for early 20th century economist Vilfredo Pareto. The Pareto criterion says that we should alter an allocation of resources if someone can be made better off without making someone else worse off. An allocation where no one can be made better off without making someone else worse off is said to be Pareto Optimal.

Figure 8-1 helps to shed light on the Pareto collective choice criterion and its companion concept, Pareto optimality. Figure 8-1 contains a utility possibility frontier for an economy of two individuals, Mary and John. Mary's utility is represented on the vertical axis and John's on the horizontal. The downward sloping line looks like a production possibility frontier, but, in this case, it's utility on each of the axes, not the level of production of a good. The tradeoff relationship from the production possibility frontier is preserved in the utility possibility frontier, however. In this economy, we would like to have unlimited utility for each individual, but scarcity constrains us to the utility possibility frontier. Points A and B represent two points on the utility possibility frontier. If the economy moves from point A to point B, Mary's utility falls from M_1 to M_2, and John's utility rises from J_1 to J_2. The reverse holds for movements from point B to point A. Any movement along the utility possibility frontier means that if someone is made better off, someone else is made worse off. Economist call points along the utility possibility frontier Pareto optimal, i.e., no one can be made better off without making someone else worse off. Movements along a utility possibility frontier do not satisfy the Pareto criterion.

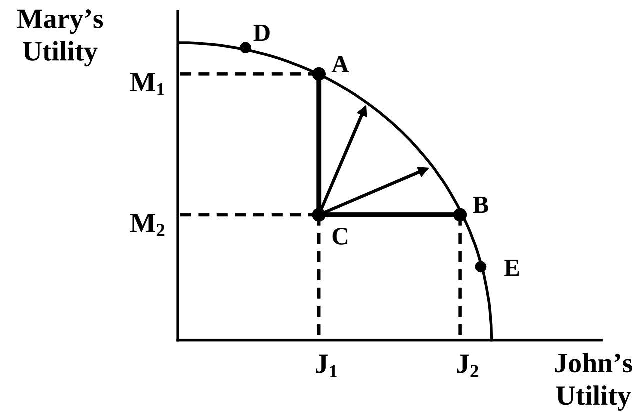


Figure 8-1
A Utility Possibility Frontier, Pareto Optimality and Pareto Improvements

Some movements from points interior to the utility possibility frontier to points on it do satisfy the Pareto criterion. For example movements from point C toward any point on the utility possibility frontier inside the L-shaped line ACB satisfy the Pareto criterion. A movement from point C to a point such as D or E does not satisfy it, however, as one of the individuals would be made worse off in such a move. Economists often call changes that satisfy the Pareto collective choice criterion Pareto improvements or Pareto moves.

The Pareto collective choice criterion would seem to be noncontroversial. Who, but an envious person, could argue with making someone better off if no one else is harmed? It's not envy, however, that makes the Pareto collective choice criterion relatively unused in practice. The problem is that it is too conservative. It's hard to think of collective actions that harm no one. Economists would be quick to point out that establishing a market where none existed before might come close. As voluntary exchange is mutually beneficial, both sides of a trade are made better off. Market transactions, provided that they involve no negative effects on third parties, result in Pareto improvements.

The Social Net Benefit Collective Choice Criterion

Suppose we were willing to accept a government action to improve a market resource allocation that involved large gains for some members of an economy and only small losses for others. In this case, the gains to the gainers are greater than the losses to the losers. Enough net gains exist so that we could take some of the gains from the gainers, compensate the losers, and leave some remaining gains with the gainers. If this compensation were actually accomplished, the action would be a Pareto improvement satisfying the Pareto collective choice criterion. But what if the compensation is not made? Social net gain or net benefit occurs, i.e., the gains to the gainers are greater than the losses to the losers. This is the net benefit collective choice criterion, i.e., we will allow government action to change a market allocation of resources if it leads to an increase in net benefit. This criterion is also called the benefit-cost criterion or the Kaldor-Hicks criterion after the two British economists who invented it, Nicholas Kaldor and John Hicks.

	Satisfies Pareto Criterion and Net Benefit Criterion	**Satisfies Only Net Benefit Criterion**
State 1	John’s NB=50 Mary’s NB=100 John’s + Mary’s NB=150	John’s NB=50 Mary’s NB=100 John’s + Mary’s NB=150
State 2	John’s NB=60 Mary’s NB=110 John’s + Mary’s NB=170	John’s NB=80 Mary’s NB=90 John’s + Mary’s NB=170

Figure 8-2
A Numerical Example Comparing and Contrasting the Pareto and Net Benefit Collective Choice Criteria

Figure 8-2 compares and contrasts the Pareto and net benefit collective choice criteria. Two states of the world exist, State 1 and State 2, shown in the two rows of the figure. In State 1 the sum of John’s and Mary’s net benefit equals 150, and in State 2 this sum equals 170. Clearly in an aggregate sense State 2 is better than State 1. But could we justify government action that moved the economy from State 1 to State 2? In column 1 of the Figure we see a case where the net benefit of both individuals increases by the move, John from 50 to 60, and Mary from 100 to 110. This move satisfies both collective choice criteria, as is indicated at the top of the column. In the rightmost column, however, we see a case where the change in the aggregate net benefit is the same, from 150 to 170, but now John's net benefit increases from 50 to 80 and Mary's decreases from 100 to 90. A government action that accomplishes this result would not satisfy the Pareto criterion, but it would satisfy the net benefit criterion.

A Model of Social Net Benefit Maximization

While any move that increases social net benefit satisfies the social net benefit collective choice criterion, perfect information about the effects of actions following this criterion would eventually lead to maximization of net benefit. A model of social.net benefit maximization appears as Figure 8-3. We should also note that this model shows marginal rules for the maximization of the social net benefit of just about anything.

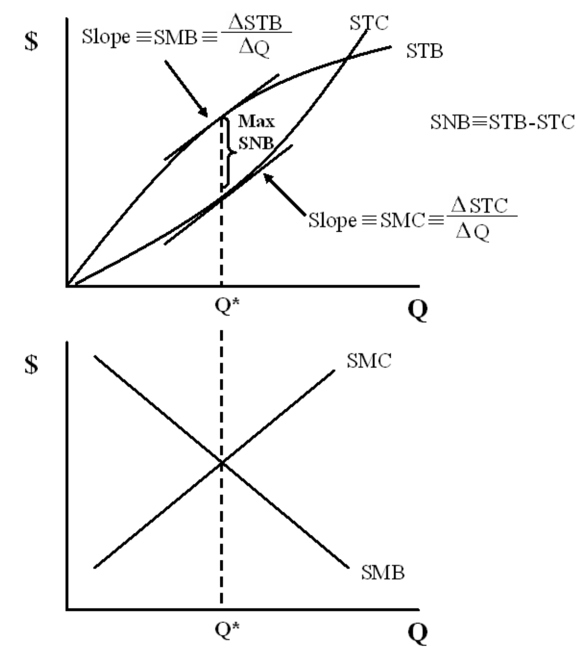


Figure 8-3
A General Model of Social Net Benefit Maximization

Two curves appear in the top graph of Figure 8-3, one showing Social Total Benefit (STB) and the other Social Total Cost (STC). The term social here is used as a synonym for the term collective so that we can be in line with more common economics language. The STB curve rises throughout its range, but each additional unit of some action Q creates less of an increase in STB than a previous change in Q. This is the familiar diminishing returns principle applied in a social context. The change in STB divided by a change in Q is defined as Social Marginal Benefit (SMB). As is the case for any total and marginal relationship, SMB is represented as the slope of the STB curve. This relationship is exactly like the relationship between total revenue and marginal revenue we saw in our models of a business in Chapter 5.

Likewise, the generally upward-sloping STC curve has as its slope the Social Marginal Cost (SMC). We assume this slope increases as Q increases, much as the total cost curve for a business we studied in Chapter 5. This too reflects a diminishing returns concept. The more of something you do, the more difficult it becomes to do an additional amount of it.

The vertical difference between the STB and STC curves in Figure 8-3 represents Social Net Benefit (SNB). As indicated in the top portion of the figure, the maximum distance between STB and STC occurs at Q = Q* where the slopes of the two curves are equal, or in words, maximum social net benefit occurs when social marginal benefit equals social marginal cost. The lower graph of Figure 8-3 shows the net benefit maximization result with marginal curves rather than total curves. The SMB curve starts out high at low levels of Q and then falls as Q increases, following the decline of the slope of the STB curve in the graph above it. Likewise, the SMC curve starts out low at low levels of Q and rises as Q increases. The level of Q that maximizes social net benefit occurs at the intersection of the SMB and SMC curves. The marginal analysis here exactly parallels that of profit maximization in Chapter 5, where a businesses profit is maximized at a level of output where marginal revenue equals marginal cost.

An Example of Social Net Benefit Maximization: Optimal Water Quality

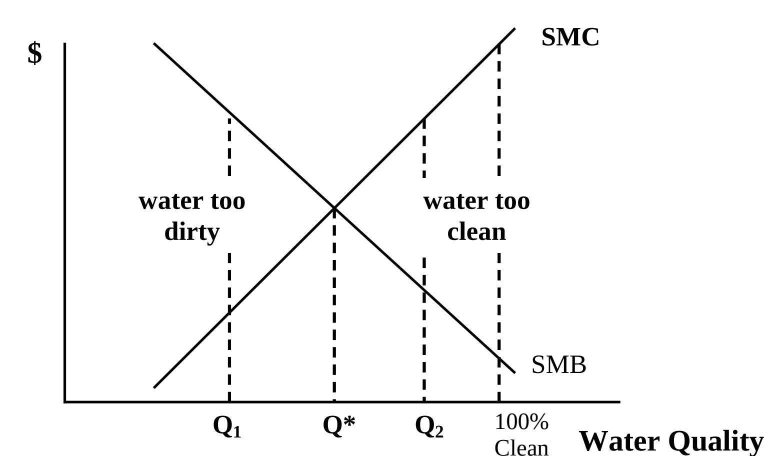


Figure 8-4
The Optimal Level of Water Quality

In Figure 8-4, we apply the social net benefit maximization model to the issue of optimal water quality. Two curves represent the economic choice, a downward-sloping SMB curve and an upward-sloping SMC curve. These are the most likely general representations of the behavior of marginal benefit and marginal cost curves, but either a constant (flat) SMB or a constant SMC would work as well. With increasing water quality, Q increasing from left to right on the horizontal axis, we see that Total Social Benefit increases (MSB is positive), but each additional improvement in water quality provides less of an increase than previous unit of improvement. The social marginal cost is increasing because it becomes more and more difficult to improve water quality as the quality improves. We do the easy, less costly things first. Going from 99% to 100% clean water increases Social Total Cost more that going from, say 10% to 11%. Again this is a diminishing returns concept. Another way of thinking about rising SMC is to think of the process of wringing water out of a washcloth. The more water is wrung out the harder it is to wring the same amount out with the next twist. Improving water quality follows the same principle.

As is the case in most economic optimization problems, the optimal amount of the activity does not occur at the 100% level. The social net benefit maximizing amount of water quality is at Q* where SMB = SMC. At any level of water quality less than Q*, say at Q_1, economics is a friend of the environment. At Q_1 SMB > SMC. This means that additional improvements in water quality (Q moving to the right on the horizontal axis) will add more to social total benefit than it does to social total cost. In that case, social net benefit would increase. Social net benefit increases up to Q*.

At a water quality level greater than Q*, the economist is the environmentalist's opponent. For example, at a level of water quality indicated by Q_2, SMB < SMC. More water quality is certainly not warranted, as this would add more to social total cost than it would add to social total benefit. This would make social net benefit decline. Here the social net benefit collective choice criterion would say that we should make the water dirtier. By doing so, by reducing water quality, we will lose social total benefit. This is because social marginal benefit is positive and reducing Q will reduce STB. But, by reducing water quality, we reduce social total cost more than we reduce social total benefit, and social net benefit increases. We lose some benefit from water quality, but we forego less other alternatives, such as golf and sewage disposal.

The Geometry of Deadweight Loss

In order to show better the change in social net benefit from changes in resource allocation we need to develop a few additional analytical concepts. We will represent shortfalls from maximum social net benefit as areas on SMB and SMC graphs. We'll call these magnitudes deadweight loss, which when eliminated result in increases in social net benefit. We begin in Figures 8-5 and 8-6 by showing how social total benefit can be represented as an area under a SMB curve.

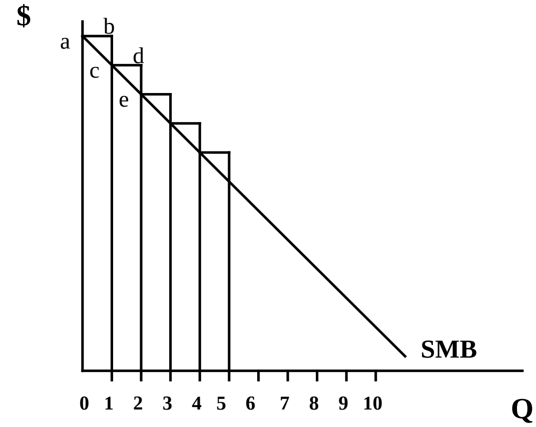


Figure 8-5
Showing Social Total Benefit as the Sum of Social Marginal Benefits

In Figure 8-5, we show a standard downward-sloping SMB curve. Let's focus on a one-unit movement from a quantity of zero to the quantity 1. Because SMB is defined as the change in STB divided by the change in quantity, and the change in quantity is equal to 1, we can represent approximately the addition to STB as the first vertical rectangle 0abc1. For this first unit of quantity the SMB and the STB are equal. If we generate another unit of quantity, moving from 1 to 2 we can represent approximately the additional STB by the vertical rectangle 1cde2. Now, however, the STB is the sum of the two vertical rectangles 0abc1 and 1cde2.

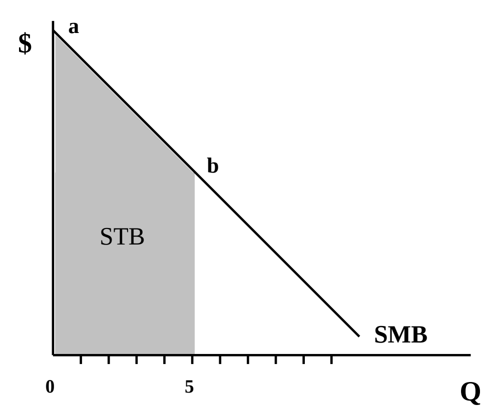


Figure 8-6
Social Total Benefit as the Area Under a Social Marginal Benefit Curve

As we increase quantity up to 5, we add additional STB approximately equal to the vertical rectangles shown in Figure 8-5, and the STB at each quantity is the sum of these vertical change in STB rectangles. The lumpiness of our quantities makes our summation of rectangles more than the area under the SMB curve, but as changes in quantity become small relative to the total quantity, the little triangles like abc and cde become vanishingly small. If we ignore the magnitude of the small triangles, we can represent STB as the area under the SMB curve up to the quantity in question. If you remember your calculus, you will recognize this STB magnitude as the integral of the SMB function from 0 to the quantity in question. In Figure 8-6, we show the STB as the upward sloping shaded trapezoid 0ab5, the area under the SMB curve from 0 up to a quantity of 5.

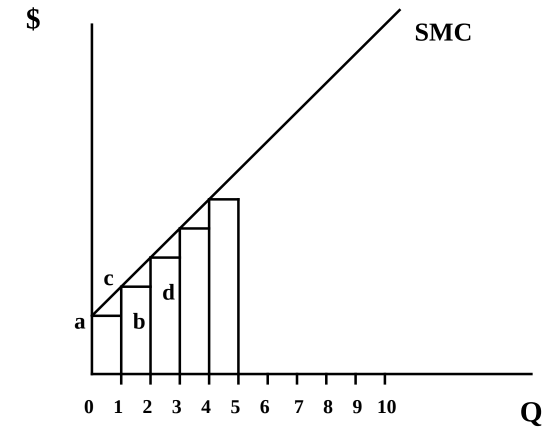


Figure 8-7
Showing Social Total Cost as the Sum of Social Marginal Cost

In a similar manner, we can represent STC as the sum of a number of SMC rectangles. This is shown in Figure 8-7. At a quantity of 0, the SMC is shown as the intercept on the vertical axis. We can approximate the SMC of moving from 0 to a quantity of 1 by the vertical rectangle 0ab1. Moving from a quantity of 1 to a quantity of 2 adds another rectangle of incremental social cost equal to the area of the rectangle 1bcd2, and so on. Again, ignoring the small triangles from the lumpy quantities, we show in Figure 8-8 that STC is the area under the SMC curve. At a quantity of 5, this is the shaded trapezoid area 0ab5.

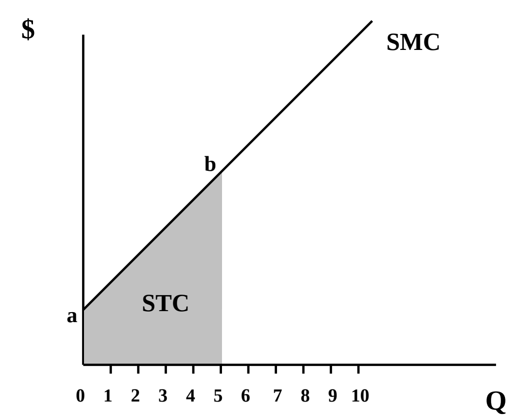


Figure 8-8
Social Total Cost as the Area Under a Social Marginal Cost Curve

Putting the SMB and SMC curves back together we can show in Figure 8-9 an area equal to social net benefit, SNB. At the optimal level of quantity Q*, where SMB=SMC, the STB is a lightly-shaded area under the SMB curve between 0 and Q*. The STC is the area under the SMC curve between 0 and Q*. The darkly-shaded area is the result of partially overlapping STB and STC cost areas. If we subtract STC from STB we get SNB, which is the lightly-shaded area above the SMC curve and below the SMB curve from 0 to Q*. This area is also labeled SNB in

Figure 8-9. The important principle in the figure is that this SNB area is as large as it can be, i.e., we have shown that resource allocation resulting in the quantity Q*, where SMB=SMC, maximizes social net benefit.

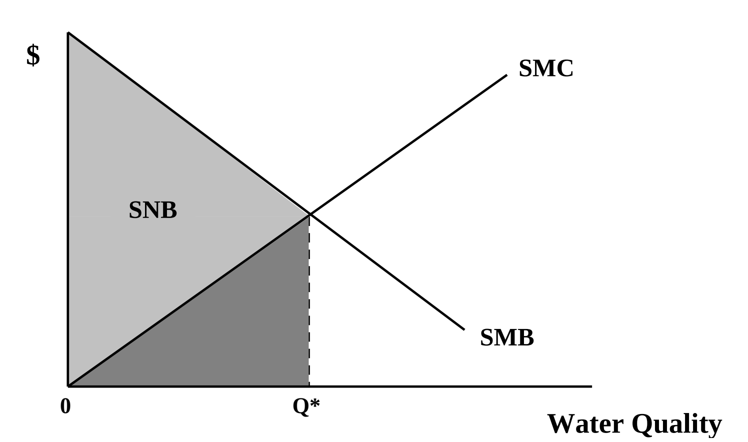


Figure 8-9
Showing Social Net Benefit with a Model of Social Marginal Benefit and Social Marginal Cost

By choosing quantities other than Q*, we can see that SNB is not maximized. This is shown in Figure 8-10. In the top graph of the figure, we have a quantity Q_1 which is less than Q*. The area under the SMB curve between 0 and Q_1, the area bounded by the line segments 0abceQ_1, shows SNB. The STC is the area under the SMC curve between 0 and Q_1, the area bounded by line segments 0aeQ_1. In this case, unrealized social net benefit exists. If we reallocated resources and increased the quantity from Q_1 to Q* , STB would increase by the area Q_1ecdQ*. Likewise, STC would increase by the area Q_1edQ*. The surplus of additional STB over the additional STC is the shaded area triangle ecd. This triangle is labeled D.W.L. for deadweight loss to society. By producing Q_1 rather than Q*, people in the economy, in the aggregate, suffer a loss of net benefit. The allocation of resources fails to maximize SNB.

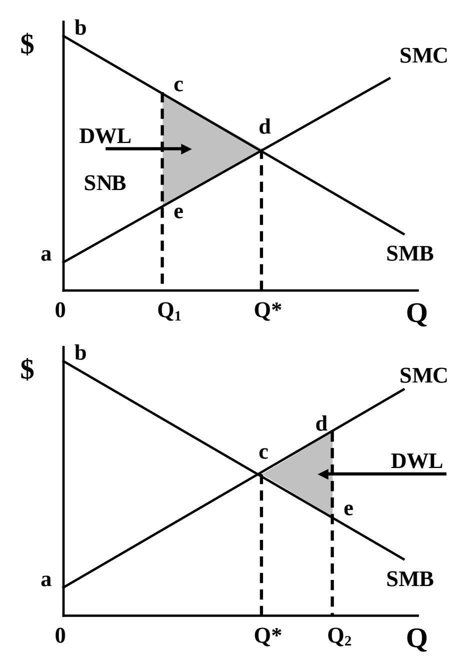


Figure 8-10
Too Little, Too Much and Deadweight Loss

In the bottom graph of Figure 8-10, we see the result of too much quantity at Q_2. The area under the SMB curve from 0 to Q_2, the area bounded by line segments 0abceQ_2, shows SNB. STC is the area under the SMC curve between 0 and Q_2, the area bounded by line segments

$0acdeQ_2$. Here, in contrast with the case shown in the upper graph, STC > STB by the amount of the horizontal-lined deadweight loss triangle cde. People in the economy receive benefit from a movement from Q^* to Q_2. It's just that the cost of such a movement exceeds that benefit. If the economy is at a resource allocation yielding Q_2, a movement from Q_2 back to Q^* reduces STC more than it reduces STB, eliminating the deadweight loss.

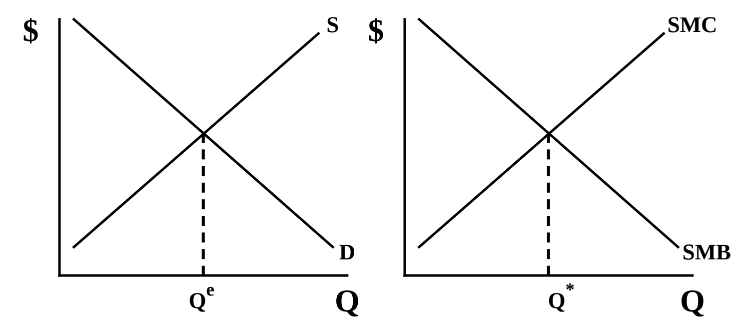


Figure 8-11
Two Suspiciously Similar Models

Perfect Competition and Social Net Benefit

In Figure 8-11, we see two very similar graphs. In fact, if we removed the labels of the lines on the graphs, the two models would be indistinguishable. The graph on the left is the standard supply and demand model. The graph on the right is the standard SMB and SMC model. In each, the intersection of the lines on the graph has an important meaning, market equilibrium in one case and social net benefit maximization in the other. Imagine what the implications would be if we could interpret a market demand curve as a SMB curve and the market supply curve as a SMC curve. Then the equilibrium quantity would be the quantity that maximized social net benefit. This would be an important normative result, indeed. This is the task to which we now turn.

We have already done some of this work. At the end of Chapter 4, we showed that the market demand curve is a horizontal summation of the demand curves of individual consumers. We also hinted that these individual demand curves could be interpreted as marginal benefit curves for the individual. We examine this latter issue further in Figure 8-12. One way to imagine that the individual's demand curve is her marginal benefit curve for the good in question is to read

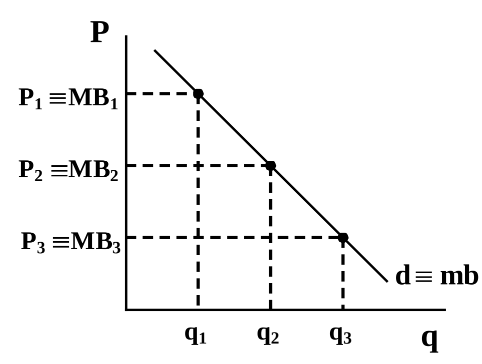


Figure 8-12
An Individual's Demand Curve as a Marginal Benefit Curve

the demand curve "backwards." Usually with demand curves we ask how much the individual wants to buy at a given price. In Figure 8-12, we would say that this individual wants to buy a quantity q_1 at price P_1, a quantity q_2 at the price P_2, and so on. Alternatively, we could ask what is the maximum price the individual would be willing to pay for the quantity q_1. Note that the maximum willingness to pay for the quantity q_2 is less than the amount she is willing to pay for q_1. This is because of diminishing marginal utility. As the individual buys more and more of the good, the additional utility from another unit falls, and the individual is willing to pay less for that additional unit. If the individual buys q_1 for a price of P_2, the individual gets a good deal, as she was willing to pay more for all the preceding units, for example, P_1 for the quantity q_1. If an individual can buy a unit of the good for less than she values that unit, she will do it.

Suppose the individual in our example is given the opportunity to by q_1 quantity at a prince of P_2. She would certainly do this, as the last unit up to q_1 is valued at $P_1>P_2$. That's what reading the demand curve backwards means. But what if the individual after having been offered q_1 at the price of P_2 is given another opportunity, to buy more at the price of P_2. She will do this up to the point on her demand curve, the quantity q_2. The individual demander purchases up to the point where the value of another unit, its marginal value, is just equal to the price she has to pay. Let's just rename this marginal value marginal benefit. The individual demand curve is an individual marginal benefit curve.

As we noted in Chapter 4, this provides an amazing normative insight. If every buyer faces the same price, each will have the same marginal benefit from the good. All marginal benefits are equal. Of course, all individual marginal benefit curves are not the same, because all individual demand curves are not the same, but when we face the same price, all buyers get to a point on their own demand curve where marginal benefit is equal to price. This is good from a normative standpoint, for if someone had a higher marginal benefit than another did, we could generate more social net benefit by allocating more of the good to the individual with the higher marginal benefit. Social net benefit maximization requires that every buyer has the same marginal benefit, and a market system where all face the same prices brings this about.

We see in Figure 8-13 that the market demand curve is the sum of all individual demand curves. And because individual market demand curves are individual marginal benefit curves, it's the sum of them as well. This means that any point on a market demand curve tells us, on the vertical axis, the (identical for all buyers) marginal benefit of another unit of the good. A curve that does this is a social marginal benefit curve, as is indicated in the figure.

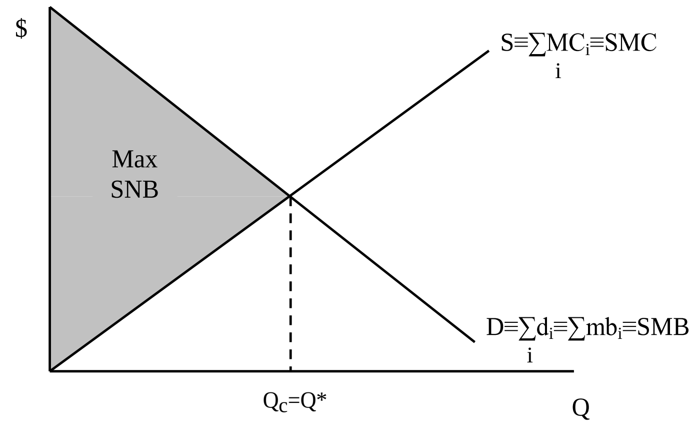


Figure 8-13
Perfect Competition and Maximization of Social Net Benefit

In Chapter 7, we determined that the market supply curve is the horizontal sum of all the marginal cost curves of the businesses in the competitive market. If all businesses face the same product price, they will choose an output where that price equals marginal cost, in order to maximize profit. Facing the same price means that each business has the same marginal cost. Not all marginal cost curves are the same, but each business chooses a point on its marginal cost curve where the marginal cost at that point is the same as the marginal cost of every other business in the market. Again, the normative implications are powerful. In order to maximize social net benefit, it is necessary that our resource allocation have all goods produced at the lowest possible cost. If one business had a marginal cost lower than another did, we should allocate more production to that business, and less to other businesses, until marginal costs are equalized. Ensuring equal marginal cost among businesses would be a difficult task in a command socialist economic system, but in market capitalism it happens automatically if all businesses face the same prices and maximize profit.

Now armed with the realization that the demand curve is a SMB curve and that a supply curve is a SMC curve, we can conclude that the equilibrium quantity in the competitive market, Qc, is also is the quantity Q*, where SMB=SMC and social net benefit is maximized. In Figure 8-13, the shaded area below the demand curve and above the supply curve represents maximum social net benefit. Any quantity in the market either less than or greater than the equilibrium, market-clearing quantity results in a deadweight loss and market failure.

Monopoly and Market Failure

We now can demonstrate why monopoly results in market failure. The monopolist's choice of quantity, while maximizing its profit, results in market failure for the economy as a whole. Social net benefit is not maximized. Look at Figure 8-14. As the monopolist is the only producer in the market, its marginal cost curve is the social marginal cost curve. In this market, the quantity Q* is the allocation of resources that maximizes social net benefit, where SMB=SMC. The monopolist, however, is interested in maximizing profit, and this occurs not where SMB=SMC, but where the monopolist's marginal revenue equals its marginal cost, at Q_m. We find the social marginal cost from that point on the marginal cost curve. Social marginal benefit, identical to price in this case, we find from the demand curve. At the price P_m, all consumers are buying up to the point where their individual marginal benefit equals that price.

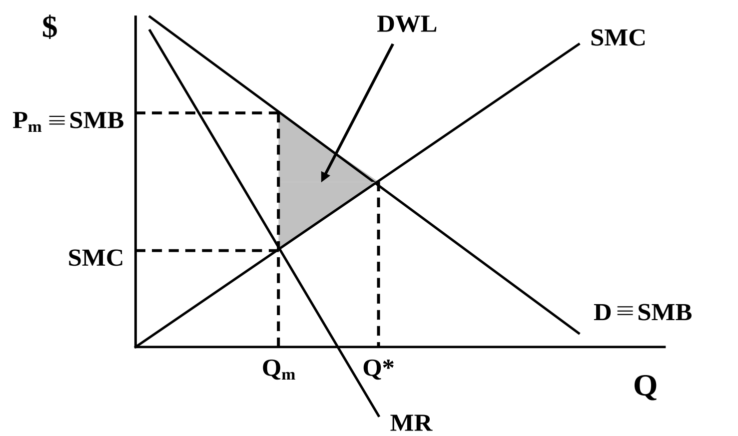


Figure 8-14
Monopoly and Market Failure

Note that SMB>SMC when the market is monopolized. Suppose you were a benevolent social net benefit maximizing economic czar for this economy. Your underlings tell you that SMB>SMC. What would you do? You would order that more should be produced because buyers of the product value another unit more than the opportunity cost of that additional unit. Monopoly results in a deadweight loss indicated by the shaded triangle in Figure 8-14.

The economist's problem with monopoly is much different than a normal person's problem with monopoly. A normal person might worry that the monopoly is "gouging" consumers with higher prices and making profit that is too high. "Gouging" is not a word in the economist's lexicon. After all, some very nice grandmothers own stock in monopolies and do quite well. No, the economist's problem is one of resource allocation. With monopoly, we allocate too few resources to producing this good. Net benefit to society is forgone. Economists leave it up to others to sort out the equity issues involved in monopoly.

Natural Monopoly and Public Utility Regulation

Natural monopoly is a special case of the monopoly market structure. The term "natural" comes from the salient characteristic of natural monopoly, falling long-run average cost over the relevant range of output. With falling average cost, it's natural to have one business produce all the output in the industry. In a practical sense, the first business in the market will likely have a cost advantage due to its scale of operations that will act as a deterrent for new entrants to the market. Furthermore, as average cost is falling, we would not want the output of the industry to be produced by a number of smaller businesses, because these smaller businesses would each be producing at a higher average cost and the sum of their costs would be greater than the cost of the single monopoly firm.

From a normative point of view, however, the standard normative problem exists with natural monopoly. It is still a monopoly, and if we allow a monopoly to maximize profit, it will do what all monopolies do, restrict output and raise price. The dilemma posed by falling long-run average cost and monopoly behavior is shown in Figure 8-15. Note that LRAC is falling as it crosses the demand curve, i.e., it's falling over the relevant range of output. Note also that LRMC is below LRAC, the necessary condition we learned in Chapter 5 for LRAC to fall. If left to its own profit maximizing behavior, the monopoly will choose to produce output Q_M and charge price P_M, the price and quantity combination where marginal revenue equals marginal cost. Of course, this results in market failure, as discussed in the last section. In order to maximize social net benefit, we must have a level of output where price equals marginal cost, indicated in Figure 8-15 as the price labeled P=MC.

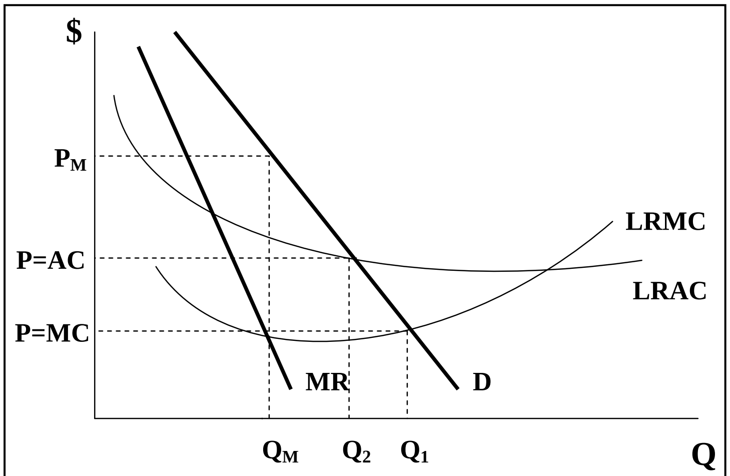


Figure 8-15
Regulation of Natural Monopoly

A common collective procedure to deal with natural monopoly is for the government to set up a regulatory mechanism to force the natural monopoly to produce more output at a lower price. Electricity, natural gas and telephone service are industries with the characteristics of natural monopoly, and for which we have something called public utility regulation. Think about where you live. There is usually only one electric company, one natural gas provider, and one land-line local service telephone company. These companies are usually subject to regulation by a state government agency called the public utility commission or the public service commission or an agency with a similar name.

If the regulatory agency allows the natural monopoly to charge a single price that maximizes social net benefit, that price would be that where P=MC. Note in Figure 8-15 the interesting problem created by this policy. At the price P=MC, consumers will purchase Q_1 level of output. Unfortunately, this socially optimal level of output results in a case where price is below average cost. The business incurs a loss. It will either go out of business, or the government will have to subsidize its losses. Sometimes it is possible to charge declining block rates for natural monopoly products. This means that consumers pay higher prices for lower levels of service or output and lower prices for higher levels of service. This could allow the business to bring in enough revenue to cover its costs and still have price equal to marginal cost on the marginal unit. For example, an electricity customer might pay 8 cents per kilowatt-hour (kwh) of electricity consumption for the first 1000 kwh and 6 cents for anything over 1000 kwh. If the marginal cost is 6 cents, the correct normative price could be achieved and the extra revenue on the first 1000 kwh could help solve the revenue problem.

Another alternative is to abandon marginal cost pricing altogether and adopt a fair rate of return policy. Here the regulator allows the natural monopoly to earn a normal or fair rate of return. Because a normal return to capital in the business is part of economic cost, this normal return is already included in the LRAC curve of Figure 8-15. At the price P=AC, consumers would purchase Q_2 of the natural monopolist's output, and the business would earn a normal profit. Note that at output Q_2, however, we incur a deadweight loss, because P > MC.

The entire field of regulatory economics is a fascinating one, but one well beyond the scope of our introductory inquiry. We must remember that the cost of the regulatory apparatus itself must be considered in our normative calculations (even though it provides good jobs for economists). And imagine what you might have a tendency to do as a business, if you where allowed a certain percent rate of return on your capital investment. Might you have a tendency to have a lot of capital? Check out the electric company's trucks. Are they new, or are they old beaters? And is it possible that regulators might not be as tough as they could be, given the opportunity to cross over the line at some time and work for the regulated in the industry? The field of regulatory economics has interesting questions, indeed.

Collective Goods and Market Failure

Another form of market failure comes in the form of the existence of collective goods. A collective good, sometimes called in economics a public good, is best defined by relating it to the kinds of goods we have been considering to this point, private goods. Figure 8-16 summarizes this relationship.

	Pure Private Goods	Pure Collective Goods	Impure Collective Goods	Impure Private Goods
Consumption Characteristic	Rival	Nonrival	Rival	Nonrival
Exchange Characteristic	Excludable	Nonexcludable	Nonexcludable	Excludable
Examples	hamburger pair of socks	national defense species existence value	common grazing ocean fishing	satellite TV uncongested movie digital music
Market Provision	Yes	No	Not Likely	Yes, but too little

Figure 8-16
Characteristics of Collective and Private Goods

Two general characteristics of goods help us separate them into two pure forms and two impure forms. One characteristic is how a good is consumed. The other is how it is exchanged. In the first column of Figure 8-16, we find pure private goods, exclusively our subject up to this point. Pure private goods are rival in consumption. This means that if one person consumes a good another person cannot consume it. If I eat a carrot, you can’t eat the same carrot. If I wear a pair of socks, you cannot wear that pair of socks, at least at the same time. If I own a share of stock in Microsoft, you cannot own that same share. If I consume a kilowatt-hour of electricity at my house, you cannot consume the same kilowatt-hour of electricity at your house. If I have a ticket to a seat at the NCAA Final Four, you cannot sit in the same seat. Each of these goods is rival in consumption.

Pure private goods are also excludable in exchange. If I don't pay for a hamburger or a pair of socks in a store, I can't have them unless I steal them. The ability to exclude opens up the possibility for a market to provide a good. Without the ability to exclude and charge people for what you produce, it's difficult to have a business that earns even a normal profit. Markets can produce the right amount of private goods.

In the second column of Figure 8-16, we see the characteristics of pure collective goods. A pure collective good is nonrival in consumption. Once the good is provided, we can add additional consumers of the good at no cost. I can enjoy the benefits of the good without detracting from your enjoyment of the benefits of it as well. National defense is the classic collective good in terms of consumption. Once a missile defense system is provided in the U.S., I can be protected with that system without limiting your ability to be protected as well. Knowledge of the existence of a wildlife species also is nonrival in consumption. If you derive utility from knowledge that gray whales swim in the ocean, your neighbor can derive the same kind of benefit without detracting from yours.

Pure collective goods are unlikely to be provided at a level that maximizes social net benefit. In fact, the market might not provide pure collective goods at all. Once the good is provided, self-interested consumers will have an incentive to be "free riders" and consume the benefits of national defense or species existence while relying on someone else to pay. Of course, if no one pays, if everyone is a free rider, the good will not be produced.

The final two columns combine a rich mixture of consumption and exchange characteristics to form impure forms of collective and private goods. Sometimes economists refer to these goods as mixed goods, but we're going to reserve that term for some interesting cases we will discuss later in this chapter. The column labeled impure collective good has the consumption characteristic of a private good, rival in consumption, and the collective characteristic in exchange. We could actually call this an impure private good, but the inability to exclude is a key in precluding market provision of it. A situation known as common property is often involved here. Think of grazing on the open range, fishing in ocean waters, drawing oil or water from a common pool underground. It could be difficult for a private entity to exclude people from grazing, fishing and pumping. This would certainly be the case if the private entity didn't own the property, and even if it did, exclusion might be prohibitively costly. Assignment of private property rights to the resource in question would at minimum be a necessary condition for market provision. Assignment of such rights could be practically difficult, however. It would be very difficult to sell the oceans, as no one owns them to begin with, and selling the public lands in the United States would be politically difficult for even the most conservative national, state and local governments.

The final column we label impure private good. It is possible to exclude, so a private business could earn a profit by charging for the good, but the good is nonrival in consumption. Once the good is provided, additional consumers can be added at no additional cost. Satellite TV is a good example of this type of good. I can consume a satellite TV signal without detracting from your ability to do so at the same time. Yet the TV company can scramble the signal until I pay my bill. A movie theater is another example. Movies are easily excludable, but if the theater is not filled to capacity additional consumers could be added at little additional cost. Computer software and digital music and video are other examples. Music files can be downloaded from one computer to another without detracting from anyone listening to the music. Only the initial sale of the original disc or downloadable file is excludable in the marketplace. Further excludability usually relies on the illegality of unauthorized pirated use, which of course involves collective action.

So in the case of pure and impure collective goods, and perhaps in the case of impure private goods as well, the market is unlikely to allocate resources in a way that maximizes social net benefits. The existence of characteristics of collectiveness, especially nonexcludability, results in market failure.

Mixed Goods and Market Failure: Private Goods with Collective External Cost or Benefit

To this point we have assumed that all the opportunity costs involved in producing a good and all the benefits enjoyed from consuming it are paid for and received by participants in the marketplace. All who bear costs, who suffer opportunities foregone, are more than compensated for their losses. We've seen that workers are paid wages and salaries. Raw materials suppliers are paid for their products. Investors in a business receive interest, dividends or capital gains. Likewise, to this point all the benefit of consuming a good accrues to the person buying the good. A consumer purchases a good up to the point where their marginal benefit equals the price he has to pay. In this section, we relax this assumption and see the implication this has for market failure. We begin with external cost.

Economic activity does not always result in opportunity costs internal to the market. Sometimes we have external cost, also called a negative externality. When external cost exists market failure can result. Think of a production process that involves waste gasses disposed of up a tall smokestack into the air, or liquid waste running out a pipe into a stream, river, lake or ocean. Lest we be too quick to point a finger at businesses as the generators of external cost, think of the exhaust gasses emerging from your tailpipe when you drive your car, or the smoke from your fireplace or woodstove chimney.

External cost often occurs from the disposal of residuals from production or consumption. If we have someplace to dump the residuals that is free to us, we are likely to do it. Unfortunately, these residuals can cause harm to others without any compensation to them. It's the lack of compensation that makes this cost an external cost. It's a one-way transaction. But external cost is cost nonetheless. And if external cost is not paid by the generator of it, we'll likely have too much of the activity.

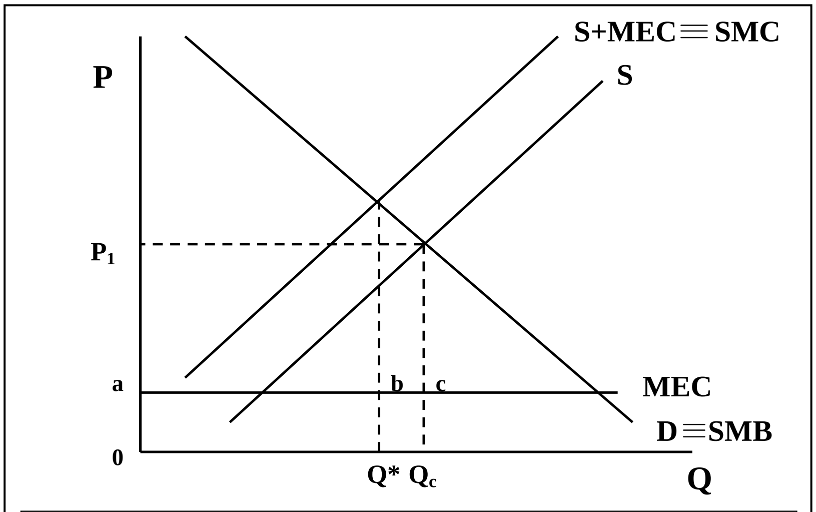


Figure 8-17
External Cost and Market Failure

Figure 8-17 is a modification of a standard supply and demand model. As before, the demand curve is the social marginal benefit curve, but note the supply curve is not the social marginal cost curve in this case. The supply curve still comes from the summation of the marginal cost curves of the individual businesses in the market, but these marginal costs do not include external cost. The horizontal line labeled MEC represents marginal external cost. Every unit of output Q generates the vertical distance 0a in external cost. For simplicity, we assume MEC is constant. It certainly would not have to be so.

We can obtain a SMC curve in Figure 8-17 by adding the MEC curve vertically to the supply curve. Think of the supply curve as being displaced upward by the amount of the external cost. The new social marginal cost curve, S + MEC intersects the SMB curve to the left of the market quantity Q_c. Q* is less than Q_c. When external cost is unaccounted for, the market fails by generating too much of it. We also see in Figure 8-17 the deadweight loss triangle created when the market produces Q_c instead of Q*.

Note that at the social net benefit maximizing level of output Q*, the external cost is not eliminated, just reduced. Originally in market equilibrium the total external cost is the area under the MEC curve bounded by the line segments $0abcQ_c$. Moving to the optimal level of output in the market merely reduces the external cost by the amount of the rectangle Q^*bcQ_c. Reducing the external cost more than this lowers benefit more than the reduction of internal and external cost. We'd have to give up too much of the net benefit of the good to receive the additional improvement in, say, air or water quality.

We often call the type of good shown in Figure 8-17 a mixed good. Because the good itself is rival and excludable, the market can provide it. But the external cost is most likely collective in nature. For example, think of water pollution caused by a production process. This is a collective bad, the reduction of which is a collective good. If the water is cleaned up for me, if I can now swim, fish and recreate on the water, this doesn't preclude you from consuming the benefits of improved water quality as well. But suppose consumers of river or lake services were asked to pay for cleaning up the water. Because exclusion from the benefits of cleaner water is not possible, free riders will emerge.

Cases do exist, however, where establishment of rights to pollute or to be free from pollution can result in market solutions that reduce the negative externality to the optimal level. In this case, no market failure would exist. This case was originally recognized in 1960 by Professor Ronald Coase of the University of Chicago, a body of work for which he was later honored with the Nobel Memorial Prize in Economics. This interesting analysis is beyond the scope of our introductory inquiry into mixed goods. For our purposes here, we will assume that external benefit and external cost have collective characteristics.

Now let's consider a mixed good with external benefit rather than external cost. Economic actions often generate benefit to others for which they receive no compensation. For example, if I increase my level of education, I certainly benefit from it, but others might benefit from my education as well. If I can read, you can develop a print advertisement to encourage me to buy your product or service. A politician can send me campaign literature. If I'm a scientist, I might find a cure for breast cancer or AIDS, or, as an economist, I might find the way to end poverty. In all these cases, by acquiring education, I provide benefit to myself, but also to others.

Unfortunately, in a human sense people have a tendency to undertake activities based on the benefit that will accrue to themselves, not to others. I get education mainly to improve my life. I get a flu shot to protect myself from the flu, not to protect others from getting the flu from me. But from a social net benefit perspective, we collectively should not ignore external benefit. A benefit is a benefit and external ones should be counted too.

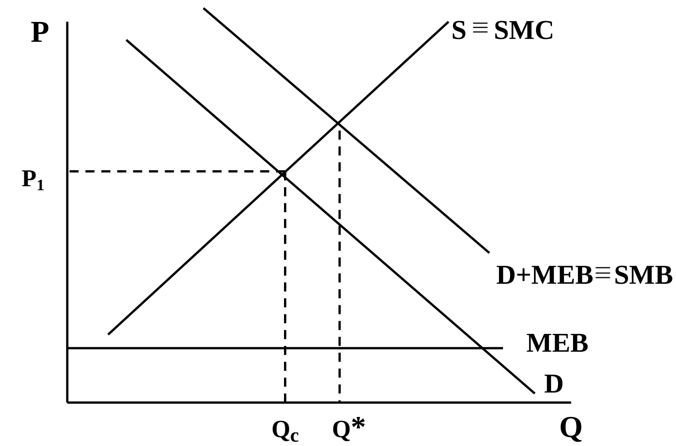


Figure 8-18
External Benefit and Market Failure

Figure 8-18 shows how we can include external benefit into our consideration of market resource allocation and market failure. This figure looks much like the preceding figure, Figure 8-17 about external cost. But here it's the demand curve that is not the social marginal benefit curve, where it was in the external cost case. The supply curve is the SMC curve because we assume we have no external cost. The horizontal line in Figure 8-18 is marginal external benefit, a measure, assumed constant here, of the amount of external benefit generated by each unit of quantity produced in the market. The demand curve represents only the private marginal benefit. We have to add to the demand curve the marginal external benefit to arrive at the SMB curve.

After performing this vertical summation, we see that the intersection of the SMB and SMC curves occurs at a point to the right of the market-determined quantity Q_c. In contrast with the external cost case where Q* was less than Q_c, with the existence of external benefit the market fails to generate the social net benefit maximizing amount of the good or service, i.e., $Q^* > Q_c$. Market failure exists.

Government Failure

Until the 1960s in economics, the collective choice story was nearly over once we noted the various forms of market failure. Market failure was a justification for government action to correct it, and we assumed that the government simply came in and moved the market from a resource allocation that failed to maximize social net benefit to one that did. As the study of the economics of collective choice evolved in later decades, economists came to realize that when people took the civil service exam, won an election to public office, or became administrator of a government agency, they didn't automatically become candidates for sainthood.

In real life, government officials and employees are just like other people. They have preferences and tend to act in there own interest. Often these interests coincide with a mission of a government agency. The government does not necessarily act to follow any collective choice criterion set up by crazy economists. The government does what it does to satisfy the political preferences of the constituents it serves, at least in a democratic government. This does not necessarily move a market to an allocation of resources that maximizes social net benefit. If the government does not move the economy in the direction of maximizing social net benefit, it fails the social net benefit collective choice criterion. We have in this case government failure.

We can demonstrate one type of government failure with an extension of Figure 8-17, the case of a private good the production of which generates external cost. Suppose in response to external cost in the form of air pollution, the government establishes a department of air quality. Establishing such a department will gain votes in the next election from those who have strong preferences for improved air quality. Suppose the first administrator of the department is a zealous fighter of air pollution. We'll call this the case of the over-zealous regulator, and show it in Figure 8-19.

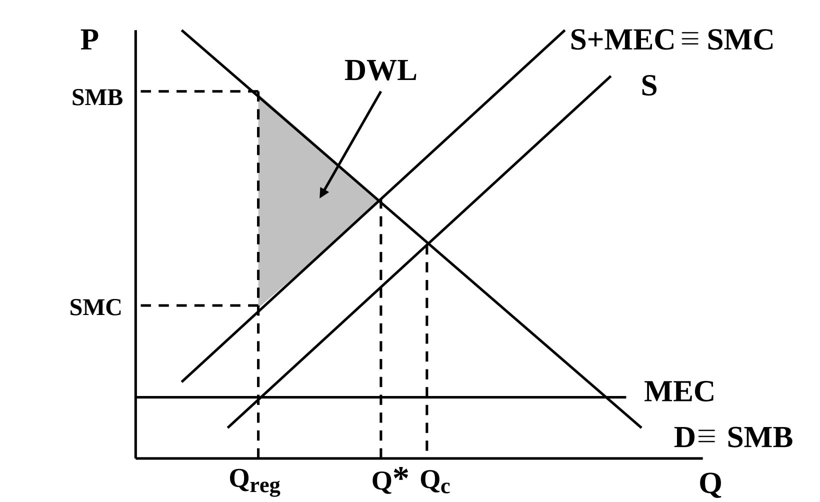


Figure 8-19
Government Failure from an Overzealous Regulator

Most of Figure 8-19 we've seen before in Figure 8-17. Actions in the market with collective external cost lead to market failure where $Q_c > Q^*$. But now the department of air quality undertakes a regulatory policy that results in reduction of industry output to Q_{reg}, with the related reduction in the external cost of air pollution. Unfortunately, from a social net benefit perspective, the government policy to reduce air pollution is too extreme, and results in a larger deadweight loss than we had with the market failure.

Government failure need not arise just from attempts to reduce market failure. Government failure can occur simply from a political process driven by political forces different from the economist's collective choice criteria. A common case of government failure is government provision of private goods. Often the government steps in to produce private goods when private participants in the market recognize that demand for the good is insufficient to generate production of it. Government produces many private goods, e.g., facilities for golf, tennis, and swimming, and libraries, electricity, and water. Each of these is a private good. It is rival in consumption and excludable. Try golfing without paying greens fees, or consuming electricity or water without paying your electric or water bill. Even library services are excludable. Now some of these goods might be mixed goods and have external benefits with collective characteristics, and this might warrant some government participation to encourage more consumption, but probably not to the extent of producing the good.

The government produces these goods due to political not economic pressure. Interest groups with political power gain favor with elected officials, and convince elected officials to coerce people to pay for these goods with tax dollars. This lowers the cost to people who want to consume the goods, and forces people to pay for them who could care less if they are provided. By definition, government production of private goods is government failure.

Another form of government failure arises because incentives within government are not aligned to minimize cost of producing a good. If at the end of a fiscal year a government agency has been successful in accomplishing its mission while spending less than its budget, the agency is unlikely to return unspent funds to the legislature. Legislatures will certainly reduce the agency's budget the next year. Rather, government agencies engage in the practice of end-of-year feeding frenzies where they spend "end-of-year money" on any product that can be delivered in time. In contrast, in the private marketplace owners of businesses get to keep what's left over at the end of the year in the form of profit. An incentive exists to control costs. Controlling cost is not the government's strong suit.

Likewise, promotion through the administrative ranks in government often requires the supervision of some number of workers before one can be promoted to a higher position where she can supervise a larger number of workers, and earn a higher salary. Size of the workforce is to be maximized, not kept to the most efficient least-cost level. Again the incentive is not to control cost.

Market Failure, Government Failure, and the Economist's Collective Choice Dilemma

Market failure is a justification for collective action to correct it. Government failure is an argument to live with market resource allocation. We are caught on the horns of a collective choice dilemma. While we can't offer any universal solutions to this dilemma, economists have a couple of practical suggestions that might lower the severity of the dilemma. These are service contracting and voucher systems. Service contracting is the practice of soliciting bids from private sector providers to produce the collective good we wish the government to provide. Voucher systems give purchasing power to demanders of goods so that these demanders can buy the collective goods in the marketplace where they are produced by private businesses. We'll discuss each of these in more detail below, but first let's look at the overriding principle leading to them.

First, let's eliminate one form of government failure by assuming that the government gets involved only in the provision of collective goods. Service contracting and voucher systems could also be used in government provision of private goods as well, but we will not consider that case here. We will be using two words that appear to be quite similar, but in the context here are very different. These words are provision and production. By provision we mean raising the funds to pay for the good in question. Usually, the government does this through taxation, a coercive form of finance. By production we mean the actual production of the good, the building of buildings, employment of labor, purchasing supplies, etc.

Service contracting and voucher systems, while different, are similar in that they attempt to separate provision from production of collective goods. In each case, we take the best of both worlds, the government and the market. Collective goods are underprovided by private businesses because of the inability to exclude consumers from the benefits of them. As we have seen, free riding emerges. The government, however, has coercive powers to make people pay. In the U.S., we grant these powers to the government in federal and state constitutions. Government is the only legitimate coercive authority in the otherwise free marketplace. Government excels in coercing people to pay.

On the other hand, government does not excel in assessing the preferences in the market and producing products people want at minimum cost. This is the province of the market, through the forces of supply and demand. Even if we use the government to force payment, to undertake provision for the good, there is no need for the government to produce it. Let's look at service contracting.

Take garbage collection as an example. Garbage collection is a mixed good. Much of the benefit of garbage collection accrues to the household having its garbage collected. But external benefits with the characteristic of a collective good describe garbage collection, too. Leaving garbage collection to the forces of supply and demand would result in a lot of roadside and canyon dumping, generating external cost. Market failure calls for government involvement. The government has the power to coerce all households to pay for garbage collection, either through taxes or user fees, but there is no need for the government to purchase and maintain garbage trucks, manage landfills and incineration facilities, and hire garbage collectors. The government can solicit bids from private sector providers and choose the low cost provider that meets required specifications. Likewise, government need not build it's own roads, or run its own fire or police departments, or prisons and other correctional facilities. These can be contracted out to the lowest bidder.

Voucher systems differ from service contracting because with voucher systems consumers choose the providers. The government does not choose a service provider. Rather the government raises funds through coercing people to pay for something with taxes, and then gives consumers of the good a voucher that is "almost as good as money." The voucher is not quite money, because it can be used to purchase only the good it's intended for.

The food stamp program is a good example of a widely used voucher system. Provision for the feeding of the poor can be considered a collective good. I can consume knowledge that the poor are provided for without detracting from your knowledge of this as well. Likewise, once the lot of the poor is improved, no one can be excluded from consuming the knowledge that it has happened. A "Robin Hood" government can tax the rich and feed the poor. But does the government have to have farms, factories, and distribution systems to provide food for the poor. The government can give a poor person a voucher, a food stamp in this case, that the poor person can redeem for food at any private sector business where food is sold. The business returns the voucher to the government for reimbursement in money. Voucher-like systems are also used in healthcare and housing for the poor, and have been proposed and used in limited settings for education.

The Distribution of Income

One of the most common actions of government involves redistribution of the fruits of the economic process. Some of this redistribution is done horizontally, i.e., people with the same level of economic well being, usually measured in terms of income, are favored by the government at the expense of others. The other form of redistribution, is the "Robin Hood" form. This is vertical redistribution, where income, and to a lesser extent wealth, is take from the more well off and given to the less well off. In this section, we address how economists attempt to measure this distribution of well being that leads to redistribution actions by governments.

In an age of standardized testing we are all familiar with percentiles. If one scores in the 99th percentile on college entrance exams, one is more likely to gain admission to college than if one scores in the 9th percentile. Test percentiles are 1% increments of performance ranging from 0 – 99%. In assessing the distribution of income in an economy, we use a more aggregated percentile measure, usually a quintile.

Where a percentile is a 1% increment, a quintile is a 20% increment. It takes 100 percentiles to make a whole, but only five quintiles. In calculating an economy's distribution of income, we determine the percent of the entire income in the economy received by each quintile in the income distribution. A perfectly equal distribution of income would mean that each quintile, from the poorest to the richest, receives 20% of the economy's income. A perfectly unequal distribution would involve one person, or household, or family receiving all the income.

Figure 8-20 shows a way we can represent the distribution of income graphically with a Lorenz curve. On the horizontal axis, we represent cumulatively the percentage of some group arranged by general level of income received, with quintile markings. In this figure, we use households as the group, but the same approach could be applied to families or individuals as well. Along the horizontal axis we start with the poorest household, then add the next poorest and finally we reach 20% of households, the poorest quintile. We then add the next poorest (or richest, if you like) 20% to arrive at 40% of the households, and so on, until we reach a cumulative value of 100% of the households. On the vertical axis we measure the percent of income received by each percent of households, again with 20% demarcations.

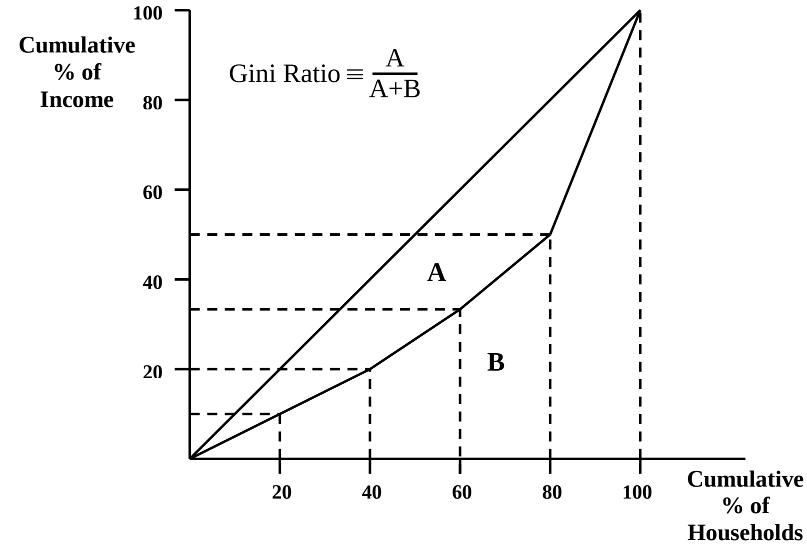


Figure 8-20
The Lorenz Curve and Gini Ratio

The 45-degree line is the line of perfect income equality. If this were the Lorenz curve, it would mean that 1% of the households received 1% of the income, 20% received 20%, and so on. The actual Lorenz curve pictured in Figure 8-20 curves away from the 45 degree line and shows an economy with an unequal distribution of income. Try to interpret the Lorenz curve here. Note that in this hypothetical economy the poorest 20% of the households receive only 10% of the income. The next richest 20% of the households receive another 10% of the income, so the cumulative total for the poorest 40% of households is 20%. In this economy, 60% of the households receive about 35% of the income. It takes the cumulative income of the bottom four quintiles to account for about 50% of income received. This means that the richest 20% of households receive 50% of the income. Note of course that 100% of the households must receive 100% of the income, so the Lorenz curve returns to the 45-degree line at a value of 100%.

The more the Lorenz curve diverges from the 45 degree line the more unequal is the distribution of income. A convenient way of summarizing this result is also shown in Figure 8-20 by the Gini ratio. If we label the area between the 45-degree line and the Lorenz curve as area A, and as area B the area between the Lorenz curve and the L-shaped line made up of the horizontal axis and the vertical line from 100% of households up to the 45 degree line, we can measure divergence from income equality by the ratio A/(A+B). Perfect income equality would be represented by a Gini ratio of 0, and perfect inequality by a ratio of 1.

How we measure income affects the distribution of it in an economy. For example, if we measure income before government transfers such as welfare payments the distribution will be more unequal than if we include transfer payments, assuming, of course, that transfers go from the richer to the poorer groups. Likewise, if we measure the distribution of income before taxes, it will be more unequal than if we measure it after taxes, because in most economies taxes as a percent of income rise with income. Finally, if we were able to measure a household's or individual's income over the entire lifetime, the income distribution would be more equal, because individuals usually start off with lower incomes and receive more income as they age, up to a point, where income then declines with advancing age. As is always the case in comparing economic magnitudes across different economies, it is important to determine that the distribution of income is measured the same way in each economy.

Progressive, Proportional and Regressive Taxes

In the last section, we saw that the distribution of income could be affected by how we treated taxes, for if tax paid as a proportion of income rises with income an after-tax distribution will be more equal than a before-tax distribution. In this section, we examine this issue of the relationship of the amount of tax paid to the level of income. Economists call this determining whether a tax is progressive, proportional, or regressive.

The three general types of taxes are summarized in Figure 8-21. In the first column of the table, we see the three types of taxes. In the second column we find information about the average tax rate, the tax paid divided by income. The average tax rate for a progressive tax rises with income. An example of a progressive tax is the federal income tax. State income taxes are also progressive, but usually less so than the federal tax system.

Proportional taxes have average tax rates that are constant for all levels of income. It's hard to find proportional taxes in practice. The closest we come is the payroll tax. In the U.S., this tax is assessed on wages, salaries of employees and incomes of self-employed persons. It's made up of the social security tax, which funds a transfer program from the working to the elderly and disabled. Because there is an upper limit on the social security tax (around $90,000 in 2004), it is only proportional up to that point. No limit exists on the Medicare tax in the U.S., a tax which funds a transfer program for healthcare for the elderly. Sometimes in discussions of tax reform and tax policy, the idea of a "flat tax" is discussed. A "flat tax," conceptually, is an income tax where everyone pays the same tax rate. Those who argue for a flat tax are arguing for moving from a progressive tax to a more proportional income tax.

Type of Tax	**Average Tax Rate (Tax/Income)**	**Example**
Progressive	Rises as income rises	Federal income tax
Proportional	Constant as income rises	Payroll tax is close
Regressive	Falls as Income rises	Sales tax

Figure 8-21
Progressive, Proportional and Regressive Taxes

Finally, we have the regressive tax. This is a "reverse Robin Hood" tax, where the richer you are the less tax you pay as a proportion of your income. The average tax rate falls as income increases. The best example of a regressive tax is a sales tax. While everyone in the same jurisdiction pays the same sales tax rate, the level of taxable purchases as a percent of income declines as income rises. This means that the amount of sales tax you pay falls as your income increases. Think of food in a state that has a sales tax on food. Food expenditures as a percent of income fall as income rises. Bill Gates and Warren Buffet have incomes much larger than the average individual, but their food expenditures (even allowing for higher quality) don't rise proportionately with their incomes. Higher income people save more of their incomes. This avoids the sales tax. And the expenditures of higher income households on things like accounting, legal and brokerage services escape the sales tax.

As might be expected, because equity and fairness of a tax system is an important political issue, reforming the sales tax is often a topic of discussion. The State of Washington, for example, has no sales tax on food, in an effort to make the sales tax less regressive. In many state legislatures, we often find discussions of expanding the sales tax base, what the tax rate is multiplied by, in order to include expenditures more likely made by households higher up in the income distribution. Also, discussions of national tax reform often include consideration of a national sales tax or its equivalent, a value added tax. How to handle the regressivity of this type of tax reform is always a companion issue.

Chapter 9: Introduction to Macroeconomic Definitions and Issues

From Microeconomics to Macroeconomics

We can detect the essence of macroeconomics in the root of the word itself, macro. Macro means big. Up to this point we have studied microeconomics. Micro means small. We also associate the words aggregated or added-up with macroeconomics. In macroeconomics, we add up a bunch of small things to get a big thing. We aggregate microeconomic magnitudes to obtain macroeconomic magnitudes.

In microeconomics, we studied markets for individual goods and services and developed the supply and demand model to explain how much consumers and producers buy and sell and at what price. We even went smaller, searching behind the market supply and demand curves for the behavior of individual businesses and consumers. In macroeconomics we ignore individual producers and consumers and specific product markets and add up the amount produced in all markets. We seek a level of output for the entire economy. That's a big number. We also study the growth and decline in national output over time, the ups and downs described by something we will call the business cycle.

Rather than study price determination for individual goods and services, in macroeconomics we study and attempt to measure changes in the average of all prices in the economy. We will call this change in the average price level inflation, and will develop theories and look at empirical data in an attempt to explain the causes of inflation.

In microeconomics, we were interested not only in the supply and demand for products, but for resources as well. We saw that one particular resource, labor, might experience chronic excess supply in a labor market if a wage floor were implemented, i.e., a minimum wage. In macroeconomics, we study employment and unemployment in the total economy, not in particular markets, and calculate magnitudes such as the national unemployment rate.

In microeconomics, we studied the production function for a business and how characteristics of this production function, such as the marginal product of labor, were related to cost and profit in a business. In macroeconomics, we focus on the average productivity of all labor in the economy, and relate changes in aggregate labor productivity to changes in aggregate levels of economic well being over time.

In studying macroeconomics, it is important to realize that much is hidden behind the aggregation and averaging of macroeconomic numbers. Output in one market can be declining at the same time aggregate output is increasing. A price of one good may be declining even as the average price level rises rapidly. The national unemployment rate may be low, but this may be of little consolation to an individual who has lost his or her job. The economy may be growing in total but particular geographical areas of the country can be in severe decline at the same time. Much is lost in aggregation, but, as we will see, much is gained as well.

Definition of Gross Domestic Product

When we attempt to add up the amount of everything produced in the economy we immediately run into a problem. What units should we use? Kilograms doesn't seem to be a good choice, as lead, bricks and rocks would account for a large portion of national output, and pillows, diamonds, and visits to the doctor for not much at all. We solve this problem by adding up the value of goods and services in the economy. We'll soon see that this solution to the adding up problem creates problems of its own, but first let's give a more precise definition of our macroeconomic measure of output.

Gross domestic product (GDP) is the total market value of all final goods and services produced by resources within the borders of a country in a given time period, usually a year. We can go a long way toward understanding GDP just by carefully examining each component of the definition.

The term gross should be understood as the opposite of net. In fact, another macroeconomic output measure exists, net domestic product. Net domestic product subtracts from GDP production to replace worn-out capital. In our gross measure of output we include production of capital goods used to replace the depreciation of capital. We have to produce some capital goods just to keep the same amount of capital we had before some of it wore out.

The term domestic has the same meaning as it does in other areas of economics, within the country in question, and by this we mean produced by resources inside the borders of the country. U.S. residents need not own these resources. The output of a Japanese automobile company producing cars in the U.S. is part of the GDP of the U.S., not Japan. For many years we measured national output with a measure called gross national product, which added up all production from a country's resources no matter where these resources produced in the world. The U.S. output of a Japanese automobile company would be part of Japan's gross national product, but part of U.S. GDP.

As we mentioned above, the term market value is a key ingredient of the GDP definition. We use it as a common denominator to add things up, but we don't add up the value of all market transactions, just transactions involving final goods and services. This distinction is best made with the use of an example. Consider the simplified case depicted in Figure 9-1, which shows the different stages of production for the final consumer good, a bag of potato chips.

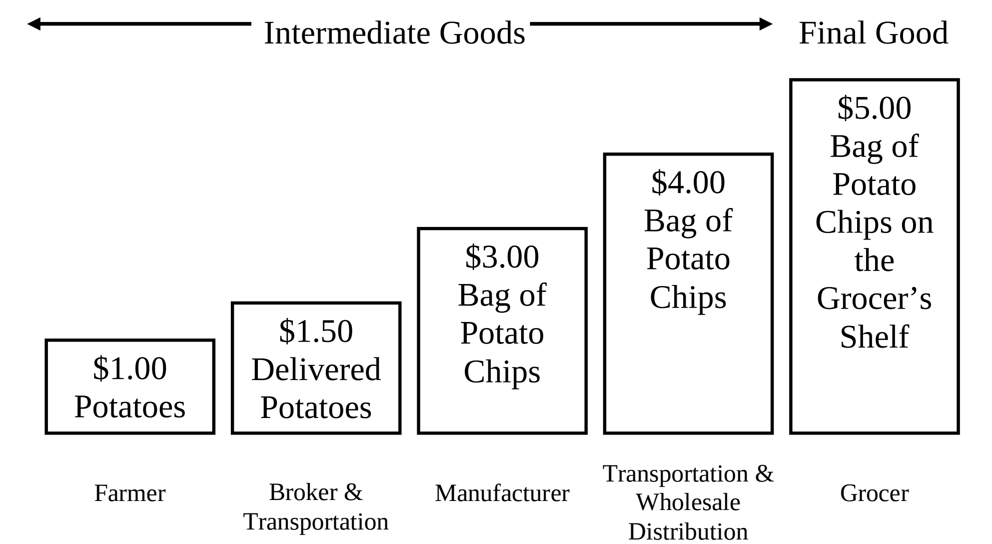


Figure 9-1
The Stages of Production, Intermediate Goods, Value Added, and Final Sales

At the left of the diagram we see the initial stage of production, a farmer growing potatoes. Suppose that every bag of potato chips contains $1.00 worth of potatoes. Assume the potato farmer sells the potatoes at the farm to an intermediary we will call a broker for $1. The farmer had created the initial $1 of value that will eventually become a bag of potato chips.

After the market transaction of buying $1 of potatoes, the broker performs some marketing and transportation services, and delivers the potatoes to a potato chip manufacturer, receiving $1.50 for the delivered potatoes. This is another market transaction where a market value could be recorded. One dollar of this value was created not by the broker, however, but by the farmer. The broker added just fifty cents to the value of the product.

Assume the manufacturer produces the bag of chips and sells it to a wholesaler for $3. This $3 in value has not been created entirely by the manufacturer, as the farmer and the broker have previously created $1.50 in value. The value added at the manufacturer's stage of production is only $1.50.

The wholesaler performs some storage and transportation delivery services and sells the bag of chips to the retail grocer, who sells the bag of chips to a consumer for $5. This is finally (pun intended) the final purchase. We add this $5 to the economy's GDP. We do not, however, add the $1 the farmer received, the $1.50 the broker received, and the $3 and $4 received by the manufacturer and wholesaler, respectively. This would be counting the value added by the farmer 5 times, the broker 4 times, the manufacturer 3 times, and so on. GDP is related to sales of the final good not intermediate goods.

Note, however, that if we added up the value added at each stage of the production process we would obtain the same amount we determined from the final sale, i.e., $1 + $.50 + $1.50 + $1 + $1 = $5. We'll examine different ways to measure GDP in the next section.

Before turning to the ways of measuring GDP, one additional component of the definition needs attention, the time period. We measure GDP per unit of time, usually a year. This means that GDP is an economic flow measure. It has units of time attached to it. A magnitude that does not have a time dimension we call a stock variable in economics. This does not mean a share of common stock, a piece of ownership in a corporation, but rather a hunk or mass of something. Your net worth (assets minus liabilities) is a stock variable. Your income (dollars per year or per month or per hour) is a flow variable. Consider a physical analogy. Water flowing in a stream is measured in cubic feet per second. Water stored behind a dam, a stock variable, is measured in cubic feet, or acre-feet, the amount of water it takes to cover an acre one foot deep with water. Acre-feet has no time dimension. Because GDP has a time dimension, it is a flow variable.

Three Ways to Measure GDP

In addition to summing valued added at each stage of the production process for all goods and services in the economy, the measurement method described in the last section, we have two additional ways to measure GDP. The three methods are shown in Figure 9-2.

GDP≡Consumption + Investment+Government +Net Exports
Spending

GDP≡Sum of all Income + Depreciation + Sales Taxes

GDP≡Sum of all Value Added

Figure 9-2
Three Ways to Measure GDP

The method most like the definition of GDP is on the top line of Figure 9-2. Here, we see GDP defined as the sum of spending on all final goods and services. Consumption is the largest of the final spending categories, accounting for more than 2/3 of GDP. These are purchases made by consumers on final goods. While consumption is the largest component of GDP, it is also the most stable, i.e., it doesn't change a lot from one period to the next. This is good from a macroeconomic standpoint, for if consumption were volatile, the aggregate economy would be in for a very bumpy ride indeed.

The second category of final spending comes mainly from businesses through their purchase of capital goods. Investment in economics means something a little different from the meaning of the term in common usage. Investment is not the purchase of stocks and bonds. Economists would call this financial investment, and it is related to investment through the financing of capital goods purchases. But, in economics, investment is defined as the purchase of final capital goods. Households also contribute to investment spending through their purchase of new residential housing, residential investment. We also include in investment the buildup of unsold inventory by businesses, even though this unplanned spending is not on capital goods as usually defined. Compared with consumption spending investment is a much smaller component of GDP, about 16%, but what it lacks in size it makes up for in volatility. Changes in investment are often responsible for fluctuations in macroeconomic activity.

Government spending is an odd category of final spending. Of course, the government buys many different goods, from paper and paper clips to aircraft carriers and fighter jets. We also include government spending on wages and salaries of government employees in GDP. Just think of the government purchasing the consulting services of the employees it hires. One important thing to remember about government spending is that the entire government budget is not counted in GDP. A large and growing component of government activity in all economies is transfer payments. A transfer payment is not a purchase of a good or service, but rather a transfer of spending power from one American to another. When the receiver of a transfer payment, such as a welfare check or farm subsidy, spends the transfer, it is then counted in GDP, but not when the transfer is made.

Finally, we have the effect of the foreign sector on GDP, net exports. The category net exports is exports minus imports. Exports, the sale of goods and services to foreigners, are counted as part of our GDP, because we produce the goods and services in the U.S. An import, the purchase of a foreign good or service by an American, is not part of our GDP, because the good or service is produced abroad. We have to subtract this part of spending from GDP.

When we spend on final goods and services of any kind someone has to provide resources to the production processes involved in producing those goods. Final sales provide income in the form of value added at the final stage of production, but intermediate sales involve goods produced by other resources. The value added at each stage of production is made up of payments to resource owners. This is why we have another way of measuring GDP, the sum of all income payments in the economy. That which is purchased must have been produced, and when it was produced it generated income to the producers, the resource owners. This is shown on the second line of figure 9-2. In order to make things add up correctly, we have to add sales taxes and depreciation to the sum of income payments to arrive at GDP, but this is a minor adjustment. For our purposes here, it is fine to think that national output equals national income. As we saw in Chapter 1 in the circular flow model, output and income are just two sides of the same macroeconomic coin.

Real GDP and the Business Cycle

Because we measure GDP in monetary terms, we are able to add together very different goods and services to determine it. But this monetary aggregation presents a major problem. What if the monetary unit we use to measure GDP changes in value over time? Let's consider a simple example. Suppose from one year to the next the physical quantity of all goods and services is exactly the same, but the price of every good and service rises by 10%. GDP in the second year will be 10% greater that it was in the previous year. Obviously, national output is no greater in the second year than it was in the first. Later in this chapter, we'll look in detail at this change in the average price level, something economists call inflation.

Because of the problem of a changing price level we have developed an additional GDP concept, real GDP. Real GDP is GDP corrected for a change in the average price level, corrected for inflation. Real GDP is a more meaningful measure of national output than is nominal GDP. The term nominal means not corrected for inflation.

We use real GDP to track the performance of an economy over time. If real GDP grows, we have economic growth. If real GDP falls, we have economic decline. Economists call the periodic rise and fall of real GDP in the economy the business cycle. A generic business cycle is shown in Figure 9-3.

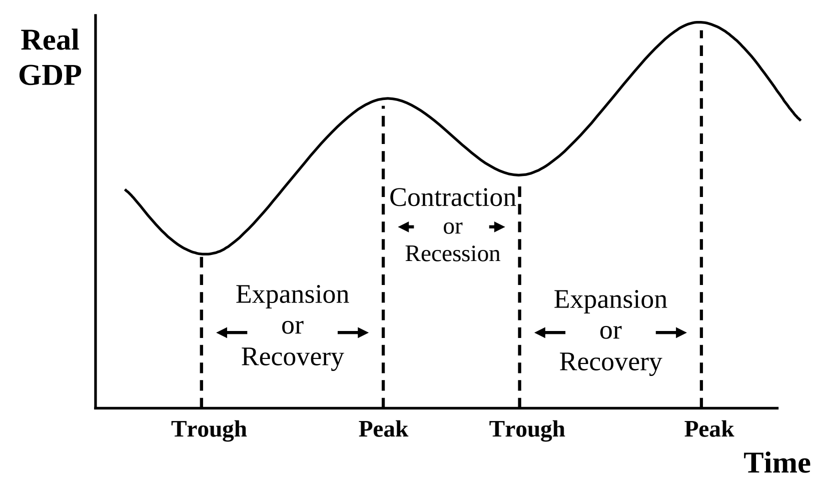


Figure 9-3
The Business Cycle

Peaks and troughs and expansions and contractions define a business cycle. After a period of decline in real GDP the economy reaches a trough where real GDP starts to rise again. The period of rising real GDP is called an expansion or sometimes in the early stages a recovery. The economy expands until it reaches a peak, then it declines through a contraction until it reaches the next trough, and so on.

We call a contraction that lasts long enough a recession. A private sector organization called the National Bureau of Economic Research (NBER) officially certifies recessions. The NBER has a group of macroeconomists on a committee called the business cycle dating committee. This committee looks at various macroeconomic measures such as real GDP, industrial production, employment and income and determines if economic activity has declined enough and over a long enough period to be officially declared a recession. The committee picks the months where peaks and troughs occur. The business and economic press define a recession by a rule of thumb measure, two consecutive quarters of declining real GDP. While the NBER disavows this definition, most recessions satisfy the popular criterion. The most recent recession, however, certified by the NBER to have occurred between March and November 2001, did not have consecutive quarters of declining real GDP. For that reason, this recession remains a little controversial.

Gross Domestic Product and National Well Being

Although it is difficult to believe after reading press accounts of rising or falling real GDP, where every bit of growth is good and the smallest decline a reason for worry, movements in real GDP leave a lot to be desired as a measure of national well being. Real GDP is a flawed measure of national well being. We summarize the reasons for this in Figure 9-4.

Correct for Inflation

Divide by Population

Subtract Value of External Cost

Add Value of External Benefit

Subtract Defensive Expenditures

Add Value of Leisure

Add Value of Home Production

Add Value of Underground Economy

Figure 9-4
Possible Corrections to GDP to Improve it as a Measure of National Well Being

Once we have corrected GDP for a changing price level, we commonly divide it by population to compute GDP per capita. If there are more mouths for a certain amount of real GDP to feed, so to speak, there is less GDP for each mouth. This is the most common measure used to compare different economies, each economy's real GDP per capita. We often call this measure the "standard of living."

But even real GDP per capita is flawed as a measure of national well being. The aggregate measure says nothing about aspects of economic activity that affect well being but that are not captured in market transactions. For example, the value of external cost is not subtracted from GDP. If growth in real output also creates air and water pollution, the related human costs are not subtracted. In fact, if pollution causes more doctor visits and hospital visits, pollution can actually cause GDP to rise. Of course, to be fair, we would want to add the value of external benefits to GDP as well. Perhaps higher material production and consumption raises even the poorest members of an economy from abject poverty, and this might reduce a number of social ills.

Certain defensive expenditures are included in GDP, and while they add to well being they would not have the same value as GDP from the production of more fine dining. For example, a rise in crime could increase the production of guns and burglar alarms in one economy and add to its GDP, while in another economy the increase could come from the production of holy books and Christmas presents for grandmothers. We don't account for this compositional difference in measuring GDP.

One major valuable activity not captured in GDP accounting is nonmarket leisure. If I take a wonderful walk in the woods or along the seashore, this is not included in GDP. Of course, to the extent that I use equipment in my leisure activities, the purchase of it is included in GDP, but, unless I pay admission to a park, a trail, a ski area, a golf course, or a fishing or hunting site, the activity itself has no GDP value.

Home production adds much value to a person's life, but it's not included in GDP. If you clean your apartment, GDP remains the same. If you hire Maids-R-Us to clean your apartment, GDP rises. As more and more adults in a household enter the labor market over time, childcare in the home declines, and GDP rises by the amount of market daycare expenditures. If everyone in a household works, no one wants to come home and cook. Meals eaten out increase. The value of preparing meals in the home does not enter GDP, while a burger at McDonalds does. So, it's possible that increases in GDP from a more fully employed household labor force, exactly what has been happening for decades in the U.S., might overstate the value of this increased production a great deal.

Finally, GDP does not include goods and services in the "underground economy." The underground economy involves transactions which escape official record keeping. Traffic in illegal goods is one example. A hidden transaction to avoid the taxman is another. Even though many underground activities such as gambling and internet pornography have risen above the legal surface in recent years, illegal drugs, prostitution (outside Nevada) and tax-related unreported income still understate GDP. Because the magnitude of the underground economy differs across countries, consideration of it is important when making cross-national comparisons of GDP.

Inflation and Three Types of Averaging

We've alluded to the effect of a rising average price level above when discussing the difference between GDP and real GDP, but in this section we'll be more precise. We begin with a definition. Inflation is a rise is the average price level, usually measured as a percent change on an annual basis. Alternatively, inflation is a decline in the purchasing power of money. After inflation, it takes more money to buy a given collection of goods and services. The word average in the definition comes from a three-part average, one across goods and services, another across individuals and a third across geography.

To begin, think of the goods and services you buy in a month. Economists sometimes call this your market basket, even though it's hard to (or at least amusing to) think of your hair stylist or insurance agent peeking up at you from the confines of a shopping cart. Each of the items in your monthly market basket has a price and a quantity. Now suppose that the next month you buy the same amount of everything, but the price of every item has changed. What has happened to the average price level of your purchases?

A simple way of measuring this change in your average price level would be to add up all the percent price changes and divide by the number of prices. This would be an average of all the price changes from one month to the next. But it wouldn't be a good measure of monthly inflation for you. Some price changes are more important than others are. For example, you probably spend more on housing than you do on movies. If your apartment rent went up 10% and movie prices fell 10%, in terms of your purchasing power these price changes would not be offsetting. Rather than a simple average, you would want to weight the price changes by their importance in your expenditure budget. You'd want give more weight to the rent than to the price of a movie. We'll return to this idea of weighting below when we discuss the consumer price index (CPI), our most common way of measuring inflation, but let's look at another kind of averaging involved in calculating the rate of inflation in an economy.

In macroeconomics, we want to measure the percent change in the average price level in the whole economy. You might be concerned only with your own average price level and what is happening to it, but, in macroeconomics, we need to aggregate over many individuals. So, in calculating a national economy-wide inflation rate, we have to average over different goods and over different people. And, because these people don't all live in the same place, and price changes differ spatially, we have to average over geography as well. Even though each individual American experiences a unique rate of inflation determined by his or her spending patterns and geographical location, macroeconomics requires averaging over all of them.

Measuring Inflation with a Price Index

Usually we measure inflation as a percent change in a price index. Before proceeding to a discussion of the most commonly used price indexes we need to explain the meaning of an index and an index number. An index number is simply the ratio of some number in one period compared to a number in a base period. This definition, along with an example is summarized in Figure 9-5. In the top line of the figure, we see the definition of an index number, but as is the custom we multiply the ratio by 100. This is simply a convention. The definition of an index number would be fine without the multiplication by 100, as long as we are consistent and knowledgeable of what we are doing.

Definition of an Index Number	$\frac{\text{Value of a Measure in Current Period}}{\text{Value of a Measure in Base Period}}$ X 100
Example of an Index Number: The Weight Index (October 2004=100)	$\frac{\text{Weight in 1/05}}{\text{Weight in 10/04}}$ X 100 $\frac{228}{220}$ X 100 =103.6

Figure 9-5
A Definition and Example of an Index Number

The second row of Figure 9-5 contains a specific example of the creation of an index number. Suppose you wanted to beat yourself up and start a new year with a bout of self-loathing and resolution about weight gain and weight loss. We could summarize holiday weight activity with a Weight Index. Let's choose October 2004 as the base month. We can choose any period as the base period. There is nothing mysterious about choice of a base period. It's arbitrary. Suppose our hypothetical holiday reveler weighs in October and then again after the holidays in January. In October, our portly friend weighed 220 pounds. After a few too many cookies and glasses of holiday cheer (or perhaps just the right amount) our friend weighed 228 pounds. The January value for our holiday monthly weight index reads 103.6. This is a pure number with no units. The value means that since the base month, our friend's weight had risen by 3.6% [(103.6-100)/100]. Note that the value of an index number in the base period is always 100, as any number divide by itself is 1, and 1 time 100 equals 100. Should our friend diet and lose weight in February, the index number would fall, and if he gained weight, the number would rise. Suppose our friend gains another 5 pounds in a stressful January. In February, he weighs 233 pounds. The February index would be (233/220)x100, a value of 105.9. We can calculate the percent monthly weight gain from January to February as the percent change in the weight index, (105.9-103.6)/103.6 x 100 = 2.2%. We also know by just looking at the February index number that his weight has risen 5.9% since October, the base month.

Calculating the rate of inflation is no different than calculating a percent weight gain. We just do it with a different index, a price index. The Bureau of Labor Statistics in the U.S. Department of Labor publishes the most well known price index in the U.S., the Consumer Price Index (CPI). Every month the BLS sends hundreds of its employees into 87 urban areas across the country to collect information on about 80,000 prices. It also periodically uses expenditure survey information about how an average urban consumer spends his or her income, and uses this information to develop the market basket it monitors and to formulate weights for the various prices in this market basket.

Definition of the Consumer Price Index

$$\frac{\text{Cost of Market Basket in Current Period}}{\text{Cost of Market Basket in Base Period}} \text{ X 100}$$

Definition of an Annual Rate of Inflation

$$\frac{\left[\text{CPI in Current Month}\right] - \left[\text{CPI in Same Month Last Year}\right]}{\left[\text{CPI in Same Month Last Year}\right]} \text{ X 100}$$

Example of an Inflation Rate Calculation for February 2005

$$\frac{191.8\text{-}186.2}{186.2} \text{ X 100 =3.0\%}$$

Figure 9-6
Calculating the Rate of Inflation with the Consumer Price Index

In the top line of Figure 9-6, we see the definition of the CPI. Note that this index number is conceptually identical to our weight index number we described above, a number in the current period divided by the corresponding number in the base period. The base period in the current CPI is an average of the months in the years 1982-1984. In this case, the numbers represent the cost of the market basket of goods and services in the two periods for all urban workers.

In the middle portion of Figure 9-6 we show how we calculate the rate of inflation as the percent change in the CPI, just as we used the numbers in the weight index to compute a month to month percent weight gain. Here we are calculating the annual rate of inflation as the percent change in the CPI that has occurred between the current month and the same month one year ago.

The bottom portion of Figure 9-6 shows an actual calculation for February 2005. In February 2005, the value of the CPI index was 191.8. In February 2004, the comparable number was 186.2. As show in the figure, the percent change in the CPI over this one-year period was 3%. It is important to note that we cannot read the rate of inflation directly from the CPI. Rather, we have to calculate the percent change in the CPI to find the inflation rate.

<u>CPI Shortcomings</u>	Amount Inflation <u>Rate Overstated</u>
Misses Some Quality Change	.2%
Long Lag to Include New Goods	.3%
Fails to Recognize Substitution	.3%
Fails to Recognize Increased Purchases at Discount Stores	.2%
Total Overstatement	1.0%

Source: Panel of economists picked by the Senate Finance Committee. September, 1995. Reported in David Wessel. "Consumer Price Index May Overstate Increase in Living Costs, Panel Reports." *Wall Street Journal*. September 18, 1995.

Figure 9-7
Reasons the CPI Overstates Inflation

The CPI and Overstating the Inflation Rate

In 1995, the Finance Committee of the U.S. Senate convened a panel of expert economists to address the problem of measuring inflation accurately. The panel concluded that the CPI overstates inflation by about 1 percent. The reasons for this overstatement, and the panel's estimate of the amount of the overstatement appear in Figure 9-7.

While the economists and statisticians at the B.L.S. attempt to ferret out quality change in the market basket they use for the CPI, they miss some changes in quality. If a good increases in quality its price might rise because of it. It would not be correct to count this increase in price as an actual increase in price. A good with a different quality is really a different good. Suppose the safety features of cars improve from one year to the next. Airbags, antilock breaks, and fog lights require resources to produce them. The value of these quality changes will increase the price of the car, but this does not contribute to inflation. If this price increase is counted in the CPI, inflation is overstated.

When new products come on the market they are usually priced higher than they are later in their product life cycle. Early adopters are willing to pay higher prices for new products, so businesses price them high initially to earn higher profits. Unfortunately, it takes awhile before the B.L.S. includes new goods in its market basket. This delay misses the early high price of the product and the subsequent fall in price. Missing the fall in price of new goods causes an upward bias to CPI inflation.

In Chapter 3, we spent a great deal of effort studying the law of demand, the idea that consumers buy less of a good when its price rises. Because the CPI comes from a fixed market basket, this is tantamount to assuming that consumers consume the same amount of the good when prices increase. This overstates CPI inflation.

Finally, the CPI is slow to incorporate changes in the places or businesses where consumers shop. Even after Wall Mart opens in an area, there is a time over which the B.L.S. still assumes consumers are shopping at mom and pop hardware stores paying higher prices. This, too, overstates inflation.

Other measures of inflation exist. One broader measure is the GDP deflator, with which economists compare nominal and real GDP in a year and use the difference to calculate how much inflation occurred in that year. A subcomponent of the GDP deflator accounts for differences in the nominal and real values of personal consumption expenditures, and from this we can compute a personal consumption expenditure (PCE) deflator, an alternative to the CPI. Inflation measured with the PCE deflator is the number preferred by Alan Geenspan, Chairman of the Federal Reserve Board, and a frequent critic of the CPI as a way to measure inflation.

Of course, all price indexes have flaws, but it is suggestive to compare changes in the CPI and the PCE deflator over time. Using average annual data, according to the PCE deflator the average price level rose 107%

between 1980 and 2004. According to the CPI, the price level rose 129%. This is a big difference, averaging about .9% per year.

Perhaps Mr. Greenspan knows something here. In the next chapter, we will study the U.S. Federal Reserve System (Fed) and actions they take to control inflation. If the Fed's job is to control inflation, we can imagine that they are concerned with measuring it properly. Another of Mr. Greenspan's concerns is that we use inflation as measured by the CPI to adjust social security and other government benefits annually. We also use the CPI to adjust annually the standard deduction, the personal exemption, and the tax brackets in the federal income tax system. If we measure inflation inaccurately, we adjust for inflation incorrectly. This adjusting for inflation is the topic of the next section.

Converting Nominal to Real Magnitudes

As mentioned above, we use inflation calculations based on price indexes to correct economic magnitudes for the effect of inflation. We use the term nominal to apply to any monetary magnitude that is uncorrected for inflation, and the word real to refer to magnitudes that have been corrected. In Figure 9-8, we see how to perform this adjustment.

How to Convert a Nominal Magnitude to a Real Magnitude

$$\text{Real Magnitude (in Base Year \$)} = \frac{\text{Nominal Magnitude}}{\text{Price Index}} \times 100$$

Example of Converting from a Nominal to Real Magnitude

$$\text{Real Annual Income} = \frac{\text{Nominal Magnitude}}{\text{Current CPI}} \times 100$$

$$\text{Real Income in 1982-1984 \$} = \frac{\$100{,}000}{188.9^*} \times 100 = \$52{,}938$$

*CPI in 2004

Figure 9-8
Converting Nominal to Real Magnitudes

As shown in the top part of the figure, to convert any nominal magnitude to a real one we simply divide the nominal magnitude by a price index and multiply by 100. This will create a dollar magnitude in dollars of base period purchasing power. If we have had inflation over the period from the base year to the year in question, we will be dividing by a price index number greater than 100. For years after the base year we will be deflating the nominal magnitude down to the smaller real magnitude. Should we wish to convert a nominal magnitude that occurred earlier in time than the base year, we would be dividing by a price index number that is less than 100. This means we would be inflating the nominal magnitude to obtain a real value in dollars of base year purchasing power.

The bottom two lines show an important conversion from nominal to real values. Suppose we wish to see if there has been an increase in an individual's income over time. We know the individual's current annual income is $100,000, but we have had inflation every year since the base year. What income in 1982-84, the average base year for the CPI, would be equivalent to $100,000 earned in 2004? If we divide the current nominal annual income of $100,000 by the price index for 2004, and multiply the number by 100 we see that it takes $100,000 in 2004 to buy what $52,938 bought between 1982-84. We say that $52,938 is the individual's 2004 annual income in constant 1982-84 dollars.

If the individual had received an income of $40,000 in 1983, we would conclude that she had experienced a real increase in income between 1983 and 2004. Had her income been $60,000 in 1983, she would have experienced a decline in real income. Of course, this assumes we are measuring inflation correctly with the price index we use.

Rebasing a Price Index

Sometimes it makes sense to change the base year of a price index so that when we use it to adjust nominal to real values the base year occurs at a more meaningful time. Because we all are alive and spending today, we know the purchasing power of a dollar today. Many college students reading this were yet to be born in 1982-1984, the current base year for the CPI. It often makes a time series of economic values easier to understand if the are converted to values of real current purchasing power. This procedure is shown in Figure 9-9.

	Price Index (year 2 is base year)	**Price Index** (year 5 is base year)
year 1	90	$\frac{90}{140}$ X 100 = 64.3
year 2	100	$\frac{100}{140}$ X 100 = 71.4
year 3	110	$\frac{110}{140}$ X 100 = 78.6
year 4	120	$\frac{120}{140}$ X 100 = 85.7
year 5	140	$\frac{140}{140}$ X 100 = 100

Figure 9-9
Rebasing a Price Index

In Figure 9-9, we have three columns, the year of the number in the time series, a hypothetical price index with year 2 as the base (note its value equals 1000), and the rightmost column, the same price index with the most recent year, year 5, as the base year. Rebasing is accomplished simply by dividing all price index numbers by the original index number in the year you want to be the base year. Because we want year 5 to be the base year we divide EVERY price index number in the series by 140 and multiply the result by 100. Note that this makes the rebased index number in year 5 equal to 100 and all preceding index numbers less than 100. This is because the original series indicated inflation over the period, i.e., an ascending price index over time.

Inflation and Nominal and Real Interest Rates

In order to understand interest rates we must know something about inflation. An interest rate is an amount a lender requires a borrower to pay for use of money over a given time period. Because lenders understand the erosive power inflation has on the purchasing power of money, they add an inflation premium onto a real rate of interest to arrive at a nominal interest rate. This relationship is shown in the top equation of Figure 9-10. The nominal interest rate, the rate we might see in a financial market, is the sum of a real interest rate and the rate of inflation. If inflation rises, or is expected to rise, interest rates in the market will rise as well.

Real Interest Rate + Rate of Inflation = Nominal Interest Rate

OR

Real Interest Rate = Nominal Interest Rate – Rate of Inflation

Figure 9-10
Inflation and Interest Rates

The real interest rate is much more stable than the nominal interest rate. The real interest rate is determined by the interaction of factors such as individuals' propensity to save and businesses' perception of productive opportunities for capital. While these factors change, they usually change slowly. Inflation, on the other hand can rise and fall dramatically in short time periods, and when it does it carries nominal interest rates along with it.

An example might help to explain this relationship. Suppose in a world of no inflation you are willing to lend me $1,000 for a year, if I pay the principal back at that time and also pay you an interest payment of

$50, an annual interest rate of 5%. Now suppose that the inflation rate in the economy is 5%, and that you still wish to obtain a 5% real return on your loan to me. If you leave the interest rate at 5%, at the end of a year you would receive $1,050 from me, but over the year the purchasing power of that money would have declined by 5% due to inflation. You would have received a real return of 0%. To protect yourself from the erosive power of inflation on the purchasing power of money, you could add 5% on to the interest rate and charge me 10% interest. After a year you would have received a 10% nominal return, lost 5% on the purchasing power of money, for a net real return of 5%.

We also see in Figure 9-10 that we can rearrange the original equation such that the real rate is the nominal rate minus the rate of inflation. This formulation is more useful ex post, when we wish to determine the real rate of return received. Here, we would subtract the actual rate of inflation that occurred from our nominal return. When we are forward looking, however, the original formulation makes more sense. Here we add an expected rate of future inflation onto a desired real rate in order to arrive at a nominal rate that includes an inflation premium.

Inflation has been moderate in the late 1990s and early 2000s in the U.S. This is one reason interest rates have been at levels not seen in 40 years in the U.S. If the inflation rate increases, nominal interest rates will rise. One of the reasons that lenders on home loans introduced variable or adjustable rate mortgages in the 1980s is that they had been burned by unexpected high inflation and fixed long term rates. Suppose you had a fixed rate mortgage in 1980 that you got in 1970 at a rate of say 7%. When inflation went to double digits in the late 1970s, the real interest rate on than mortgage was negative, good for you the borrower, but terrible for your lender.

Inflation and the Stock Market

Because inflation affects nominal interest rates, and because inflation can lead the central bank to undertake actions to control it that affect the real interest rate in the short run, and because interest rates are used in computing the present value of a future stream of a business earnings, inflation can affect the value of businesses. If inflation affects the value of businesses, it affects the price of shares of stock issued by these businesses. Inflation affects the stock market negatively.

This somewhat confusing and convoluted relationship might be made a little clearer by looking at the equation at the top of Figure 9-11. This is a discounted present value equation like the ones we studied in Chapter 6. The value of any asset depends on a future stream of returns. For example, if the asset is an apartment building, the return in any year is the rent paid by renters less the cost to manage and maintain the building. If the business is a fast food restaurant, the stream of returns is composed of revenues from food sales minus the cost of producing these meals in each of the years in the future. We could just sum up all the future earnings, but, as we learned in Chapter 6, the present value of a dollar of earnings farther in the future is less than that of a dollar of earnings received nearer in time. Because of the opportunity cost of waiting, we discount the future, and because the opportunity cost of waiting is due to foregone interest, the interest rate is a key ingredient in calculating present value.

$$\textbf{Value of an Asset} = \sum_{t=0}^{T} \frac{\text{Earnings}_t}{(1+i)^t}$$

Two Reasons Inflation Hurts Stock Market Performance

- i = real rate + inflation rate
- inflation rate↑→ i↑
- i↑, with same earnings →Value of Assets↓
- inflation→Central Bank will↑real rate in short run, and
- real rate↑→ i↑, so with same earnings value of Asset↓

Figure 9-11
Inflation and the Stock Market

Note that the interest rate i appears in the denominator in the term $(1 + i)^t$. If the interest rate rises, we discount the earnings in any future period more, and, if future earnings are discounted more, the value of the assent declines. If this asset is a corporation, participants in the stock market sell the stock and its price falls. If this happens for all companies, the total stock market declines.

This mechanism either assumes that earnings are fixed in nominal terms or that money illusion exists with respect to company earnings. Money illusion means the act of confusing nominal and real values. If investors think earnings won't rise with inflation, but that inflation will affect interest rates, we get an increase in the denominators of the present value formula but not in the numerators. But if inflation rises, prices must be rising. So, some form of money illusion must exist in the minds of stock market investors.

Another affect of inflation on the stock market is less direct. Central banks such as the U.S. Federal Reserve System, which we will study in the next chapter, have as one of their main goals a stable price level. If inflation rises, it is more likely that the central bank will undertake policies to control inflation. As we will see in the next chapter, these policies can lead to an increase in the real interest rate in the short run and an increase in nominal rates in any case.

Country	Time Period	Annual Inflation Rate (in percent)
Former Yugoslavia[a]	1992	15, 201
Zaire[a]	1992	3,860
Brazil[a]	1992	1,038
Argentina[b]	annual average for 1980-87	346
Mexico[b]	annual average for 1980-87	73
Israel[c]	September 1984	900
Bolivia[d]	some periods in 1985	116,000

[a] International Monetary Fund, *World Economic Outlook,* May 1993 as reported in *Time*, August 2, 1993

[b] International Monetary Fund, *International Financial Statistic*, various issues.

[c] *Desseret News*, October 16, 1984

[d] Sonia L. Nazario, "When Inflation Rate Is 116,000% Prices Change by the Hour," *Wall Street Journal*, February 7, 1985.

Figure 9-12
Hyperinflation Around the World

Hyperinflation Around the World

We live in a country with modest inflation. Others are not so fortunate. Figure 9-12 is a compilation of different rapid rates of inflation from around the world. Here we see annual inflation rates ranging from 73% – 116,000 %. That's rapid inflation by anyone's standards. One lesson to be learned from these episodes of hyperinflation is that inflation cannot be the cost of living. If the cost of living rises over 1,000% everyone in the country would die that very year. Actually, people tend to live with hyperinflation, by indexing much of the economy and incurring the avoidance and distributional costs we discussed earlier in the chapter.

Another lesson to be learned from Figure 9-12 is that these are not the most politically stable countries in the world, and this gives us a hint at the causes of hyperinflation. Hyperinflation occurs when a government in trouble tries to lessen its political problems by printing money and spending it. If it does this at a rapid rate, demand rises so rapidly that prices rise dramatically. We'll have more to say about money creation and hyperinflation in the next chapter.

The Labor Force, Employment, Unemployment and the Unemployment Rate

A normal person might think that employment and unemployment would be easily defined, but nothing is easy in economics. The U.S. Department of Labor, again through its Bureau of Labor Statistics (BLS), goes through a rather elaborate procedure to determine the level of employment, unemployment and the unemployment rate. This process is summarized in Figure 9-13.

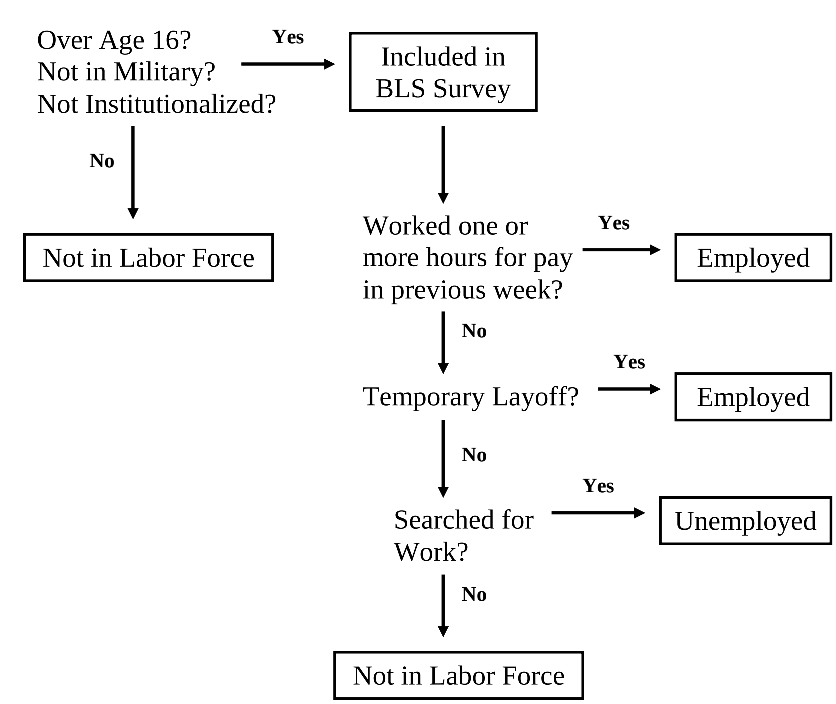


Figure 9-13
Measuring Employment Status

This figure is designed around several questions that the BLS needs to answer in its monthly surveys of U.S. households. The three questions in the upper left of Figure 9-13 are really definitional. The broadest labor economics category, except for population, is the adult, civilian, noninstitutionalized labor force. If a person is under the age of 16, or in the military, or in an institution such as a prison or mental hospital, he or she is neither employed nor unemployed, but merely not in the labor force. We will see another way not to be in the labor force in just a moment, but first let's look at the series of questions on the right side of Figure 9-13.

Individuals are randomly selected to answer the questions posed by the BLS. The first question is whether the individual worked one of more hours for pay in the previous week. If the answer is yes, that person is employed, and the interview is over. If the answer is no, the BLS asks if they are temporarily laid off. If yes, that too leads to a classification as employed. If the individual didn't work the previous week and is not temporarily laid off, the BLS asks if they searched for work. If the answer is yes, they searched for work unsuccessfully, they are classified as unemployed. But if the answer is no, if they didn't look for work, they are not unemployed, but simply not in the labor force. Unemployment is different than merely not having a job. To be unemployed you must not have a job and also wish to have one and be engaged in searching for one. Many people are voluntarily unemployed. They may be involved full time in education, or home production such as childcare, or retirement, or just merely the pursuit of activities that might be characterized as laziness and sloth.

Once the BLS has an estimate of the number of unemployed individuals and estimate of the labor force, the sum of the unemployed and the employed, they can calculate the unemployment rate. The definition of the unemployment rate appears in Figure 9-14. The unemployment rate is simply the percent calculated by dividing the number of unemployed by the labor force and multiplying by 100.

$$\text{Unemployment Rate} \equiv \frac{\text{Number of Unemployed}}{\text{Labor Force}} \text{ X } 100$$

Figure 9-14
Definition of the Unemployment Rate

While the national unemployment rate is an important macroeconomic number, we must remember that this number, like all numbers in macroeconomics, is very much aggregated and averaged. The national unemployment rate might be, say, 5-6%, yet in certain parts of the country it might be double or triple that rate. The unemployment rate for white adult males is lower than the average unemployment rate. The rate for blacks is higher than the average, and the rate for black teenagers is higher still. As always, a single macroeconomic number camouflages a great deal of socio-demographic variation.

We've seen that the definition of unemployment is based on a requirement of job search. Suppose the economy is in a large and lengthy recession. Some people have looked for work unsuccessfully for a long time. They may give up their job search, feeling that no prospects exist. Why search if no jobs exist? Economists call these discouraged workers, and the existence of them understates the true rate of unemployment. At the depth of the Great Depression in the U.S. in 1933, 25% of the labor force was unemployed. But there were also a large number of discouraged workers during this period, so the very high rate of unemployment then was likely understated.

In the next few chapters, especially when we talk about macroeconomic policy to influence the overall level of economic activity in the economy, we'll be using the term full employment. When an economist says full employment he or she does not mean zero unemployment. Figure 9-15 explains this idea.

Frictional Unemployment	Positive at Full Employment
Structural Unemployment	
Cyclical Unemployment	Zero at Full Employment

Figure 9-15
Three Types of Unemployment and a Definition of Full Employment

At any point in time some workers are quitting their jobs to look for better ones. They are by definition unemployed if their job search lasts long enough to go a week without employment. This kind of unemployment we call frictional unemployment. Like friction in the physical world it's always with us and while we sometimes try to reduce it with oil and other agents, we can't do much about it. In fact, in economics, frictional employment is actually good. We want workers to specialize in activities in which they have a comparative advantage, and sometimes it takes a little time unemployed to find that best job.

The second kind of unemployment we allow at full employment is structural unemployment. The economy is always undergoing structural change. New goods and services replace old ones. New production processes emerge that use fewer or different kinds of workers. Workers are displaced by structural change, and when this happens workers require additional training in order to find employment in the newly changed economy. Some structural unemployment is always with us, so we don't count it against the employment record at full employment.

Finally, cyclical unemployment arises from the contraction part of the business cycle, hence its name. Economists worry about cyclical unemployment, and this is the unemployment that we require to equal zero if we are to define the economy as experiencing full employment. Cyclical unemployment is that unemployment that we attempt to reduce through the use of macroeconomic policy, a topic we will address in detail in the next few chapters.

Chapter 10: Money, the Federal Reserve System, and Monetary Policy

Introduction

Money and misunderstanding inhabit the same room in the house of a normal person. Perhaps no other economic concept is subject to more myth and mystery than money is. We'll dispel many myths about money in this chapter and lift the veil of mystery so that you will understand what money is, why we have it, where it comes from, and how a central bank manipulates the amount of it in an economy in order to affect key macroeconomic variables, such as inflation, unemployment and economic growth. We will also examine a very old central bank function, one that has been used extensively in recent times. This is the "lender of last resort" function, used to address financial system crises.

The Functions of Money

Unit of Value (unit of accounting)

Store of Wealth (asset value)

Means of Payment (medium of exchange)

Figure 10-1
Defining Money by Its Functions

We define money by the valuable functions it performs in an economy. These are summarized in Figure 10-1. First, money makes our economic lives easier by serving as a unit of value, a unit of accounting. Think of how difficult life would be without the ability to converse in terms of money. Suppose you wanted to tell someone your annual income or hourly wage. Without money what would you say? I earn enough to rent an apartment, buy a car, eat, clothe and entertain myself. But how much is that? Without money we would spend a great deal of time making lists to convey the information easily delivered by use of a unit of accounting, money. Without money we would have difficulty talking about wages, prices, profit, income, wealth, and many other important economic magnitudes.

Money also makes our lives easier by serving as a store of wealth, an asset. Wealth is stored purchasing power that we acquire mainly by not consuming, so that the part of income we don't consume accumulates. We can hold wealth in various types of assets, such as stocks, bonds, real estate, cows, cars, antique furniture, paintings and baseball card collections. Money separates itself from this crowd of assets because we can spend it easily, which brings us to the last and most important function of money.

Money serves as a means of payment, or what economists call a medium of exchange. Money makes our lives easier because with it we can avoid the high cost of a barter economy. In a barter economy, we must have a double coincidence of wants. For an exchange to take place, you must want what someone else has to trade and they must want what you have to trade. Transaction costs associated with barter, even in the computer age, are extremely high. Money alleviates the need for barter and greases the gears of exchange. If I want to sell economics knowledge for a living, I don't have to find a poultry farmer to give a lecture to if I want to make an omelet or roast a chicken. That saves me a lot of time and effort.

Some Important Characteristics of Money

In common parlance, we might ask someone, "How much money do you make?" But with this question we really mean how much income do you earn. We don't imagine the person down in the basement running off a fresh batch of "Benjamins." Money and income are related, but distinct. Because money is a unit of accounting we use it to describe income, but money is not income. We emphasize this fact in Figure 10-2 as the first important characteristic of money. Money is a stock, not a flow.

Money is a Stock, not a Flow.

Money is the Most Liquid Asset.

To have Value Money must be Limited.

Figure 10-2
Three Important Characteristics of Money

Flow magnitudes in both the natural and social sciences have a time dimension in the units we use to define them. If we want to know the amount of water flowing in a stream, we determine the cubic feet per second flowing past a particular point on the bank. Likewise, in economics, we denominate income in dollars per year, per month, per week, or per hour. Income has a time dimension. But if you think about the bills in your wallet and the coins in your pocket, part of our economy's money, you don't think in a time dimension. You might have $22 in your wallet, not $22 per hour. When a variable has no time dimension, when it is merely a hunk of something, we call it a stock. This is not a share of stock in a corporation, but just a magnitude without a time dimension. We will refer to the amount of money in an economy as the stock of money. Again, money is a stock, not a flow.

In the last section, we defined money as a store of wealth, an asset. That's not all. Money is the most liquid asset. This doesn't mean that you have to carry money around in a jar. In economics and finance, the term liquidity means the ease with which an asset can be converted into money. Because money is, well, already money, it's the most liquid asset. You can spend a liquid asset. You can't spend a house, an acre of land, or a share of corporate stock. First, you must convert these relatively illiquid assets into money. Liquidity allows spending, so it's valuable. We don't receive the value of liquidity for nothing, however. Money, the most liquid asset, usually earns the lowest rate of return, receives the lowest interest rate.

The third important characteristic of money shown in Figure 10-2 relates to many of the topics we will address in the remainder of this chapter. Money derives its value from its functions, a unit of accounting, a store of wealth, and a medium of exchange. A certain amount of money tells us the worth of a good or service, the value of an asset, and how much of a good or service we can obtain in exchange for a given amount of money. If money were ubiquitous and plentiful, it would not buy very much and a unit of it would have little value. Just imagine walking out your door and finding $100 bills one foot deep on the ground as far as you could see. A $100 bill would be worth less than the modern U.S. penny, for which I personally will not risk the possibility of a back injury by bending over to pick one up from the sidewalk. To have value, we must limit the amount of money in the economy. We make counterfeiting a crime and establish central banks to limit the supply of money.

Commodity Money: Shells, Beads, Cows, Gold, Silver

Fiat Money: Coins (with little or no commodity value), Banknotes, Central Bank Notes

Bank Money: Checkable Deposits, Electronic Payments

E-Money: Debit Cards, Smart Cards, e-cash

Figure 10-3
Types and Examples of Money

Types of Money

In Figure 10-3, we show some different types of money and examples of these types. First, we have commodity money. Commodity money is a commodity in its own right. Precious metals have served as commodity money throughout human history. Gold and silver coins have always had an allure as money. Commodity money gives us a feeling that our money is worth something, but as we'll see in a moment, money does not need to have value as a commodity to perform its valuable functions.

Commodity money has several disadvantages as a type of money. If cows were money, money might die. A dead cow is very bad money. Even if a cow doesn't die, it has to be fed and milked, an opportunity cost of cow money. And if a cow serves as money, we can't eat the cow, so we incur an opportunity cost there, too. Gold and silver coins have their own unique problems, as well. If the value of the metal in a coin rises sufficiently above the value of the purchasing power of the coin itself, a rational individual would melt down the coin and sell the metal to obtain even more purchasing power. But if everyone converts money to metal, our valuable medium of exchange function vanishes. Likewise, precious metal money creates incentives for debasement, the practice of taking just a little bit of the metal from the coin. One could clip off a small piece, file a few shavings, or heat up the coin to the point a drop of liquid metal might separate from the coin. After a few rounds of clipping, filing, and dripping the coin ceases to look like a coin. No one will accept a sufficiently debased coin, so it stops serving as a medium of exchange.

We can solve the physical debasement problem of commodity money by creating paper bills and backing them with commodities such as precious metals. We have had gold certificates and silver certificates in the past in the U.S. Economy, but long ago we decided that gold and silver were better suited for jewelry and tooth repair than used as money. Think about the process of using gold as backing for paper money. We employ miners at great monetary and external cost to dig deep in the ground and extract gold ore. We extract the gold from the ore at great out-of-pocket and external cost and produce gold ingots. We place these on a truck or train and haul them to Fort Knox, Kentucky, where we dig another hole in the ground and place the gold back in it. At best this is costly, at worst very silly.

Problems with and the opportunity cost of commodity money led humans to invent something called fiat money, the second type listed in Figure 10-3. Fiat money is not Italian money that won't start on a cold morning. Fiat money arises from government fiat, by decree. Something is money because the government or central bank says it is. Most of the currency and coin we have in the U.S. today is fiat money. Look at the bills in your wallet. They will say a lot of nice things on them about God and the United States of America, and are signed by a Secretary of the Treasury and the Treasurer of the United States, but you won't find anything guaranteeing you the right to exchange the bill for a certain amount of gold or silver. They say Federal Reserve Note, which means the Federal Reserve System creates them by fiat. We have much more to say about the Federal Reserve System below. Please note that paper fiat money is not worth the paper it is printed on. In fact, paper fiat money is worth much more, that which you can obtain by exchanging the note in the marketplace for goods and services. As long as money is accepted for payment, it has value.

In the U.S., the Department of the Treasury creates fiat money in the form of coin. Dimes and quarters have a silver appearance, but they contain no silver. Silver is too valuable to use as fiat money. If you look at the side of a dime or quarter, you will see a definite dull color indicating some other metal. Coins are not worth the metal contained in them. Just as paper bills, they are worth much more, the amount they can purchase in the marketplace.

To economists and to consumers as well, even fiat money is often too costly as a means of payment. We risk losing bills or having them stolen. We have to go to the bank or an ATM to get them. Because of this, bank money evolved. Bank money exists in an account at a bank on which you can write checks or make electronic payments. If someone will agree to take your check as payment, or to receive an electronic funds transfer from your account to his or hers, bank money is a very low-cost form of money.

Finally, we have the newest form of payment, electronic money. Examples of these are debit cards, smart cards with prepaid balances, or forms such as e-cash on the Internet. We must distinguish these forms of purchasing power from credit cards. Credit cards are what their name implies, a source of instant credit or debt. E-money is not debt, but rather an immediate reduction in a bank balance or reduction in a prepaid balance on some other form of card or account.

The Stock of Money

M1		**Currency**
	+	Traveler's Checks
	+	Demand Deposits (checks or electronically accessible)
	+	Other Checkable or Electronically Accessible Deposits
M2		**M1**
	+	Time Deposits and Repurchase Agreements (small denomination)
	+	Savings Deposits
	+	Money Market Mutual Fund Shares (noninstitutional)
M3		**M2**
	+	Time Deposits and Repurchase Agreements (large denomination)
	+	Money Market Mutual Fund Shares (institutional)
	+	Eurodollars

Figure 10-4
Monetary Aggregates and the Stock of Money

We measure the amount of money in an economy with one of three monetary aggregates indicated by ascending numbers 1, 2, and 3. A monetary aggregate is the sum of different types of money in an economy. These monetary aggregates are defined in Figure 10-4. The most spendable form of the money stock is M1, the sum of currency, traveler's checks, demand deposits and other checkable or electronically accessible deposits.

Currency is paper money and coin. We have currency as part of the money stock because consumers and businesses want it for convenience. Checks are not a good means of payment for small transactions, and for transactions where the seller does not know the buyer. Often we see a sign in an establishment that reads, "No out-of-town checks!" In these cases, currency is king. Currency is also the preferred means of payment for transactions where the buyer and seller don't wish to leave a record of the transaction, e.g., the purchase of illegal goods and transactions designed to avoid payment of taxes.

Traveler's checks have purchasing power almost as good as cash. You purchase them for a fee at a financial institution with another means of payment and then use them for spending in retail outlets. They have the advantage of insurance against loss or theft. If someone steals your cash while you are traveling you are out of luck, but lost traveler's checks can be refunded by the issuing organization.

If you have money in the form of a "demand deposit," you can demand that the financial institution holding your deposit do what you tell them to do with it. For example, if you write a check and the receiver deposits the check, your bank is required to honor it, provided you have sufficient funds in your account. You can also arrange to have automatic or discretionary electronic funds transfers from your demand deposit account.

A couple of decades ago, our discussion of M1 would end at this point, as demand deposits at banks were the only form of checkable deposits. But financial deregulation since the 1980s has resulted in other forms of checkable deposits at many types of financial institutions. You might have an interest-earning savings deposit at a savings institution or a credit union upon which you can write a negotiable order of withdrawal, which looks very much like a check, but technically is not. Because of the ability to use these instruments as a means of payment we include them in the definition of M1.

When we move down in Figure 10-4, we keep the previous monetary aggregate and add additional near-liquid assets to it. M2 consists of M1 plus other near-liquid stores of wealth. Time deposits are savings vehicles where the saver agrees to leave the principal on deposit for a period of time. Repurchase agreements, mostly made by the Fed, involve the temporary purchase of an asset, with an agreement to sell it back in a period of time. A savings account contains a legal provision that allows the financial institution to require a waiting period before a depositor can withdraw funds. This rarely-used feature distinguishes a savings account from a checking account (demand deposit). M2 also contains money market mutual fund shares not owned by institutions. A money market mutual fund allows an individual to have access to financial instruments with higher interest rates, by pooling his or her assets with others in a fund managed by a business for a fee.

When we move to M3, we move farther from the function of money as a medium of exchange and toward the interest-bearing store of value function. Here we add large denomination time deposits and repurchase agreements, institutional money market mutual fund shares, and an asset called Eurodollars, U.S. dollars deposited in foreign banks outside the U.S. or in foreign branches of U.S. banks. It is easy to see that these latter components added to arrive at M3 are far from means of payment indeed.

Money Creation in a Fractional Reserve Banking System

When you learn for the first time that money has little or no value as a commodity, or that no valuable commodity backs it, you might become a slight bit uneasy. If you have these concerns, you might want to make sure you are sitting down when you read this section, because the story now becomes even scarier.

The checkable deposits we include in all definitions of the stock of money are not even "backed" by currency and coin. If we all went to our financial institutions today and wrote checks for cash, we would have a scramble not seen in the U.S. since the early 1930s. Banks do not have currency reserves backing their deposits. This is because we have in the U.S. a fractional reserve banking system. We show the main characteristics of this kind of banking system in Figure 10-5.

$$\text{Reserves} = \text{reserve requirement} \times \text{Deposits}$$

or

$$\frac{1}{\text{Reserve requirement}} \times \text{Reserves} = \text{Deposits}$$

or

$$\frac{1}{\text{Reserve requirement}} \times \Delta\text{Reserves} = \Delta\text{Deposits}$$

Figure 10-5
Deposits, Reserves, and Money Creation in a Fractional Reserve Banking System

As the name implies, banks in a fractional reserve banking system have to have reserves that are only a fraction of their deposits. This fact is summarized in the first equation at the top of Figure 10-5. A banking system does not have to be a fractional reserve system. We could have 100% banking, where the reserve requirement is equal to 1, but that is not how our banking system evolved. These reserves are not even held in the form of currency in a bank's vault, but rather are held as a deposit with the central bank, the Federal Reserve System in the U.S.

The second equation in Figure 10-5 is simply a manipulation of the first. If we divide both sides of the first equation by the reserve requirement, we obtain a very useful form of the equation. The deposits of a banking system, a main component of the stock of money, can be determined as the reciprocal of the reserve requirement times the reserves of the banking system. If we apply a change operator to both sides of the equation, it is easy to see how the money supply can change in a banking system. The money supply changes when the reserves of the banking system change, and because the reserve requirement is less than 1, the reciprocal of the reserve requirement is greater than 1. Any change in the reserves of the banking system results in a multiple expansion of deposits, a major component of the money supply or money stock.

We will return below to this important feature of a banking system. Money is created by a banking system, not a printing press. The government or central bank prints enough currency and coin to satisfy the currency and coin preferences of the public, but the money supply can be changed by the central bank if it changes the reserves of the banking system. Altering the money supply in this way is what we will call monetary policy. Before we turn to this discussion, we need to elaborate on the functions of central banks and the characteristics of the U.S. central bank, the Federal Reserve System (Fed).

Central Banks

U.S. Federal Reserve System
Bank of England
Bank of Japan
European Central Bank
Banco Central do Brasil

Figure 10-6
Central Banks Around the World

In Figure 10-6, we see some examples of central banks around the world. In some countries, a central bank is part of the government, where the leaders of a central bank can be removed by the government at any time. This level of political influence on a central bank would lead to a lack of independence of it. But often, as is the case in the United States, the central bank is a separate entity with some level of independence from the government and the political pressures that come from it. Economists have found that the degree of central bank independence is inversely related to the average rate of inflation over time. A central bank's independence insulates its activities from the desire of government officials to solve political problems through the creation of money.

Holds reserves of banks
Lends to banks
Issues currency
Supervises and regulates banks
Provides Deposit Insurance
Conducts Monetary Policy
Lender of Last Resort

Figure 10-7
Functions of a Central Bank

We summarize the major functions of a central bank in Figure 10-7. A central bank is a bank for banks. It holds banks' required reserves in an account at the central bank. It can also lend to a bank, should the bank have a temporary deficiency in required reserves.

We have already discussed the third function of a central bank listed in Figure 10-7. A central bank issues fiat money in the form of currency, Federal Reserve Notes in the case of the U.S. central bank, pound notes in the case of the Bank of England, and Euros from the European Central Bank.

A central bank also performs supervisory and regulatory functions in the banking system. These functions can either be concentrated in the central bank, or, as in the United States, distributed among various central bank and other government agencies. In the U.S., some central bank regulatory functions are performed by the U.S. Treasury through the office of the Comptroller of the Currency.

Central banks often provide deposit insurance to the banking system. Deposit insurance guarantees a depositor's bank balance, usually up to a maximum limit, should the bank fail. One purpose of deposit insurance is to provide a disincentive for "bank runs," a condition where depositors lose confidence in a bank's solvency and run to the bank to withdraw their balances in the form of currency. As we have already seen, with a fractional reserve banking system this could cause a bank to close its doors in a hurry. In the U.S., this important central bank function is not provided by the central bank itself, but rather by a separate government-sponsored agency, the Federal Deposit Insurance Corporation (FDIC). The FDIC also performs regulatory functions in the banking system.

By far, the most important central bank function is monetary policy. Monetary policy, which we will discuss at length below, involves changing the supply of money in an economy in order to affect interest rates. By influencing interest rates in an economy, a central bank can affect interest-sensitive spending and through this effect on spending affect the growth of real GDP, the unemployment rate, and the rate of inflation.

Finally, a central bank plays the role of "lender of last resort" in an economy. This function helps to stabilize the economy by making it less susceptible to financial panics that might spiral out of control. Sometimes participants in financial markets collectively engage in behavior that could cause severe disruption of an economy. Often these financial crises involve a preponderance of asset selling activity and a rapid decline in the liquidity of portfolios. The central bank can avoid the severe affects of such financial panics by extending a willingness to make loans not only to banks but also to other financial institutions and even to nonfinancial corporations.

The Federal Reserve System

Main Component of U.S. Central Bank

Independent of U.S. Government

Run by Board of Governors appointed by president for 14 year terms

One governor appointed by the president, serves a 4-year term that overlaps administrations.

The Federal Open Market Committee (FOMC) Conducts Monetary Policy.

A regionalized central bank

Figure 10-8
Major Characteristics of the Fed

We see the main characteristics of the Federal Reserve System (Fed) summarized in Figure 10-8. While the Fed has other relatives in its U.S. central bank family, namely parts of the U.S. Treasury and the FDIC, it remains the dominant family member. The Fed was created by an act of congress in 1913, but in spite of its governmental origins it remains very much independent of both the congress and the president. This was by design. The authors of the Federal Reserve Act recognized the problems that could befall an economy if its political branch had control of the money supply. The temptation to solve political problems by creating money is simply too great, and when government officials create money to finance their spending inflation or even hyperinflation usually results.

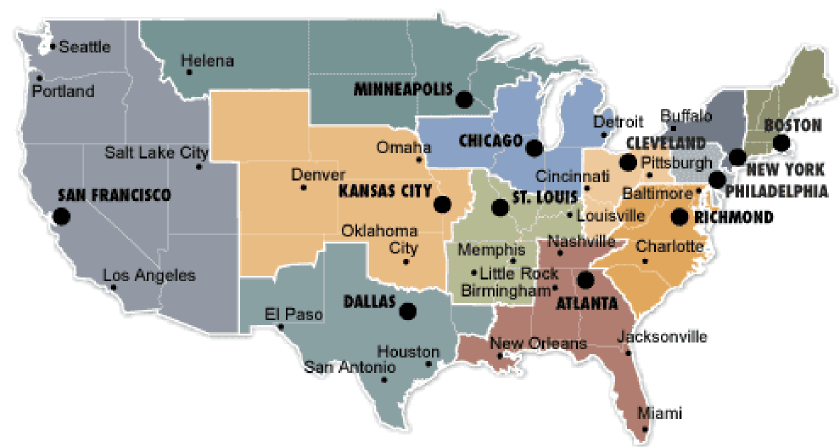
As noted in Figure 10-8, seven governors of the Federal Reserve Board run the Fed, each appointed by the president subject to the advice and consent of U.S. Senate. These governors serve for a maximum of fourteen years, though most return to the private sector long before their term runs its course. Once one has served on the Federal Reserve Board, many types of businesses find one’s expertise and inside knowledge extremely valuable, and then the opportunity cost of serving on the Board of Governors becomes great. Governors of the Fed are appointed to long terms for the same reason that judges are, to insulate them from political pressure.

The president names one of the Fed’s governors to be chairman of the Federal Reserve Board. This appointment, however, comes not at the beginning of a president’s term, as is the case with cabinet members, but rather in the middle of the term. Therefore, the term of the chairman often overlaps presidential administrations. This further insulates monetary policy from presidential political control.

In the conduct of monetary policy, the Board of Governors’ power is diluted somewhat with the addition of additional individuals to a decision making body called the Federal Open Market Committee. Five presidents of the twelve regional Fed banks on a rotating basis vote on monetary policy decisions. This brings us to the final point in Figure 10-8. The Fed is a regionalized central bank.

When the Fed was founded in 1913 the U.S. economy was far less geographically and economically integrated than it is today, and banking was regionalized as well. National financial networks were far less developed, so financing of business activity relied much more on local and regional financial flows of funds, rather than national and global flows of funds that occur today. Many of the political issues of the day rested on a foundation of mistrust of the centralized control of government operation and business finance that existed in the nation's governmental and financial capitals, Washington, D.C. and Wall Street in New York City. Framers of the Federal Reserve Act regionalized the Fed to get the hinterland of America on the central bank train. Without regionalization, the train would never have left the station.

The twelve Federal Reserve Districts are pictured in Figure 10-9 along with the cities where the main regional bank and sub-district banks are located. In order to have a modicum of proportional representation, the districts had to be designed with approximate population equality. This is why the western districts, especially the 12th District (San Francisco), are so much larger. Because few lived in the West in 1913, the western districts had to be large to incorporate enough population to give some sense of population equality among districts. With the growth of population in the West, just the opposite is true today. The San Francisco Fed District has a much greater population than any other Fed district.



Source: http://www.frbsf.org/federalreserve/system/districts/index.html

Figure 10-9
Regional Configuration of the Fed

Figure 10-10 shows the population of the regional fed districts in the Fed's early days and in modern times. Note the large increase in the percent of population living in the San Francisco district, and the decline in the East, Northeast and Chicago. If democratic representation on the FOMC is important, then the power of the San Francisco, Atlanta, and Dallas districts has declined over time.

	% Population 1934	**% Population 1997**
Boston	6.3	4.4
New York	13.3	9.1
Philadelphia	6.1	4.4
Cleveland	9.2	6.1
Richmond	8.9	11.0
Atlanta	9.1	13.1
Chicago	15.0	11.9
St. Louis	7.8	4.8
Minneapolis	4.2	3.0
Kansas City	6.4	5.3
Dallas	5.8	7.7
San Francisco	8.0	19.3
	100%	100%

Source: Jon R. Miller and Ismail Genc. 2002. "A Regional Analysis of Federal Reserve Districts." *The Annals of Regional Science* 36: 123-38.

Figure 10-10
Population of Federal Reserve Districts 1934 and 1997

The Fed's Monetary Policy Tools

As we learned above, when the Fed conducts monetary policy it changes the reserves of the banking system, which in turn alters the money stock and interest rates. When interest rates change, consumers and businesses change the amount they borrow to consume and invest. This altered spending affects macroeconomic activity. This process is summarized in Figure 10-11.

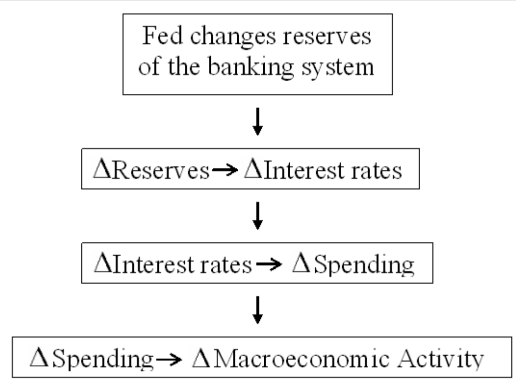


Figure 10-11
The Fed's Monetary Policy and Macroeconomic Activity

As carpenters use tools to construct houses, the Fed uses tools to build a monetary policy. Unlike carpenters, however, the Fed's toolbox contains some tools that remain unused almost all the time. We list the monetary policy tools of the Fed in Figure 10-12. Because we have a fractional reserve banking system the Fed could change the reserve requirement to alter the excess reserves of the banking system. By lowering the reserve requirement, the Fed instantly creates excess reserves. Because excess reserves don't earn a bank income, banks would eagerly make loans and buy other earning assets with these excess reserves. This eagerness would put downward pressure on interest rates and stimulate the economy. Just the reverse would happen if the Fed raised the reserve requirement.

Change reserve requirement

Change the discount rate

Open market operations

Figure 10-12
Main Monetary Policy Tools of the Fed

But the Fed never uses the reserve requirement tool in the construction of monetary policy. Building a monetary policy is like finish work in carpentry, e.g., the use of moldings, casings and cabinetry. By contrast, a reserve requirement change is the five-pound hammer of monetary policy, too blunt and powerful to use in fine tuning the level of reserves in the banking system. For this reason, reserve requirement changes remain largely unused in the Fed's toolbox.

The Fed also has the discount rate in its toolbox. This is the rate the Fed charges member banks that borrow from it. Banks don't like to borrow from the Fed. Borrowing from the Fed can elicit a response similar to that of a Father whose teenager has spent her weekly allowance by Wednesday. Saturday night is coming and the teenager needs a loan. If the teenager approaches Dad, she might get the loan, but, in addition, she will create added scrutiny of her social activities and her financial practices. Usually a lecture accompanies the loan. Instead, she calls a friend.

Just like a teenager in a liquidity crisis, a bank worries that "Father Fed" will act the same way, inquiring why the bank is short of reserves, and suggesting that a visit by a bank regulator might be warranted. Banks don't want that extra scrutiny, and the general stigma associated with borrowing from the Fed. Just as a teenager turns to an understanding friend for a temporary loan to get through the weekend, banks turn to friendly banks for short term loans to meet their required reserve positions. We call this market for short-term, interbank loans the federal funds market, and the interest rate on those loans the federal funds rate. These terms are enormously confusing, because the federal funds market has nothing to do with the federal government, and little to do with the Fed. But this is common terminology, so we'll just have to live with it.

The discount rate occasionally performs another function for the Fed, one indirectly related to monetary policy. If the Fed wants to announce a change in policy, it can change the discount rate as a signal of a policy change. In recent years, the Fed has gradually become more open and forthcoming about the direction of monetary policy, and has relied on the "announcement effect" from discount rate changes less and less.

Finally, we can pull the handiest tool out of the Fed's monetary policy toolbox, open market operations. When using this tool, the Fed enters the market for U.S. government debt either as a buyer or a seller. This is the "market" in open market operations. The term "open" means that anyone can buy or sell in this market, not just the Fed. Let's use the generic term government bonds to apply to the different kinds of government debt. We'll address the important issue of time to maturity of this debt below, and the different terms applied to debt with different terms to maturity.

The government issues debt for the same reason you do. Its outgo is greater than its income. If you buy, say, a 30-year bond issued by the government, you don't have to hold that bond for thirty years before you can get your money back. Well-organized bond markets allow you to sell your bond at any time. The price of your bond is not guaranteed, however. It fluctuates with the forces of supply and demand. Likewise, if you want to buy a government bond, you don't have to wait for the U.S. Treasury to issue new debt. Rather, you can buy existing bonds in the bond market and pay the price of the bond at the time.

When the Fed buys and sells bonds it's not trying to "play the market" and get rich, although the Fed earns a huge interest income on its government bond portfolio. When the Fed uses open market operations, it does so to change the reserves of the banking system and interest rates, in an attempt to influence interest-sensitive spending in the economy. If the Fed wants to increase the reserves of the banking system and the money stock that those reserves support, it buys bonds. When the Fed buys a bond in the bond market, it writes a check to the seller. When the seller of the bond deposits that check in its bank, excess reserves of the banking system increase. When the Fed sells a bond, the buyer writes a check to the Fed, and, when the check clears, reserves of the banking system decrease. How this activity affects interest rates is slightly more complicated, a topic we will address in the next few sections.

The Fed could buy long-term bonds, medium-term notes, or short-term bills. Bonds, notes and bills are government debt with different terms to maturity. When a debt instrument matures, the borrower, the government in this case, pays the lender back the initial amount that it borrowed. A bond might have a twenty or thirty year maturity. A Treasury note can have a maturity of 5 or 10 years. Short-term instruments of debt are called bills, often with maturities of thirteen weeks. These are sometimes called 3-month treasury bills. The Fed usually operates in the short-term market.

The Inverse Relationship Between Bond Prices and Interest Rates

We continue our discussion of open market operations with one of the most important topics in financial economics. This is the relationship between bond prices and interest rates. We need to understand this principle to understand the Fed's actions, but also just to manage our own portfolios with a good understanding of how bond markets operate. If you buy bonds at the wrong time, the result could be financially devastating. Let's first look at what we will call a long-term bond example.

By term we mean term to maturity as we discussed above. If a bond has a long term to maturity, we can approximate the rate of return on a bond with a simple relationship. This relationship appears in the top part of Figure 10-13. When a long-term bond is first issued, it has a legal contractual interest rate that matches interest rates in the economy for financial instruments of the same risk at that time. The borrower wants to closely estimate the going interest rate on the issue date so that the new bond can sell near its face value of $1,000. The face value is the amount the borrower, the government in this case, will pay back to the lender at the bond's maturity date. Because interest rate forecasts are never perfect bonds, rarely sell exactly for their face value.

Long Term Bond

$$\text{annual yield} = \frac{\text{annual payment received}}{\text{price of bond}}$$

Short Term Bond

$$\text{one-year interest rate} = \frac{\$1{,}000 - \text{price of bond}}{\text{price of bond}}$$

Figure 10-13
Two Ways to Understand the Relationship Between Bond Prices and Interest Rates

The face value of a bond serves another important purpose. By multiplying the face value times the contractual interest rate, we obtain the annual interest payment the borrower will pay the lender, the buyer of the bond. The actual rate of return on the bond is not the contractual rate, however. We can approximate this rate of return, called the annual yield on the bond, by dividing the annual payment by the price the buyer pays for the bond. For example, suppose the bond has a contractual rate of 5%. This means that the owner of the bond will receive $50 a year from the government ($1,000 x 5%). But the rate of return on the bond depends on the price the buyer pays for the bond. In the unlikely event that a bond sold for its face value, the yield would match the contractual rate, 5% ($50/$1,000). But suppose the buyer could buy the bond in the bond market for $500. The annual yield on the bond would now be 10% ($50/$500). If the buyer had to pay $2,000 for the bond the annual yield would fall to 2.5% ($50/$2,000). Note what happens in these examples. If the bond price rises, the yield on the bond falls. If the bond price falls, the yield on the bond rises. We have an inverse relationship between a bond's price and its yield. The yield on the bond is what financial investors consider the interest rate of return on the bond, the rate they compare to the returns on other possible financial investments. The inverse relationship between bond prices and bond yields is an inverse relationship between bond prices and interest rates, because a bond's yield is an interest rate.

Before we bring the Fed back into the picture, let's look at an important practical implication of understanding the inverse relationship between bond prices and interest rates. Suppose you buy a bond that has a combination of annual payment and bond price that results in a yield of 6%. This means that interest rates in the economy for financial instruments like your bond must also be yielding 6%. An important characteristic of financial instruments is risk. For financial instruments of the same risk as your bond, the market is yielding 6% on all of them. To understand this, just ask yourself what you would do if another bond with the same risk were yielding 8%. Which would you rather have, your bond yielding 6% or the other yielding 8%? Of course, you would want the bond with the 8% yield. A yield differential like this will last a very short time in modern financial markets, because people like you, holding bonds yielding 6% will sell them and purchase the bonds yielding 8%. But what happens to the bond prices? The price of the bond yielding 6% falls as it's supply increases in the bond market, and the price of the bond yielding 8% rises as the demand for it increases. With fixed annual payments on each bond, their yields will soon be the same. This is an example of what we might call the "law of one interest rate." Holding risk constant, yields on two bonds cannot differ. They must be the same.

Now suppose that your are again holding a bond yielding 6%. Then interest rates in the economy rise to 8%. Let's not worry about why rates rose. Just assume they did. How does your bond look now? Not very good. If you sold your bond in the market the price of the bond would be lower than the price you paid for it. You would have suffered what is called a capital loss in the bond market. When interest rates in the economy rise and you are holding long bonds, you are in trouble. But if you are holding long bonds and interest rates in the economy fall, your bond now looks good compared to others, and the price of your bond will rise. You can sell it for a capital gain. This is the way speculators operate in the bond market. Bond market speculators essentially bet on the course of interest rates. If rates fall, they earn capital gains. If rates rise, they earn capital losses. This is why interest rate forecasting is a popular pastime for participants in bond markets.

Now let's bring the Fed back into the picture. If the Fed buys bonds in the open market, bond prices rise because of the increase in the demand for bonds. With fixed annual payments, the yields on these bonds fall. If bond yields fall, the prices of other financial assets will adjust so that after correcting for risk, all returns are the same. So, buying bonds in open market operations lowers interest rates and stimulates interest-sensitive spending in the economy. This is why we call buying bonds by the Fed expansionary monetary policy. Likewise, if the Fed sells bonds, the increased supply drives down bond prices and bond yields rise. Other interest rates in the economy adjust accordingly. A rising interest rate curtails some interest-sensitive spending in the economy. This is why we call selling bonds by the Fed contractionary monetary policy.

Another way to see the inverse relationship between bond prices and interest rates is with a short-term bond example. This is shown in the bottom portion of Figure 10-13. Suppose we have a one-year bond. These bonds are often sold "at a discount," a price less than the face value of $1,000. But in one year the government will pay you back $1,000. The cash you receive from the bond is the difference between $1,000 and the price you pay for the bond. The one-year interest rate is this difference divided by the price of the bond. Suppose you pay $900 for the bond. Your return on the bond is 11.1% ($100/$900). But if you pay $950 for the bond, a higher price, your return is 5.3% ($50/$950). Again, we see the inverse relationship between bond prices and interest rates. If the Fed buys short-term government bonds, it alters bond prices and short term annual interest rates move inversely to these bond prices.

A Money Market Model and the Effect of Monetary Policy on Interest Rates

We can also capture the effect of monetary policy on interest rates by constructing a model of the "money market." The money market is not a market where individuals and businesses buy and sell money, as they would buy and sell wheat in a wheat market. In the money market, however, people and businesses demand money and the Fed supplies it. Let's start with the demand for money.

When we demand money we don't consume it like we would consume milk after demanding milk. We demand money to have it in our portfolio of assets, because it's the only asset we can spend. As we have mentioned earlier in this chapter, however, money yields a zero or low rate of return when we hold it. A bond will pay us an interest rate over time, but we can't spend a bond, and the transaction cost of selling bonds to obtain money every time we want to spend is high. We incur an opportunity cost when we hold wealth in the form of money, and we can approximate this opportunity cost with the interest rate. The higher the rate of interest in the economy, the lower is our quantity of money demanded. The lower the interest rate, the higher is the quantity of money demanded.

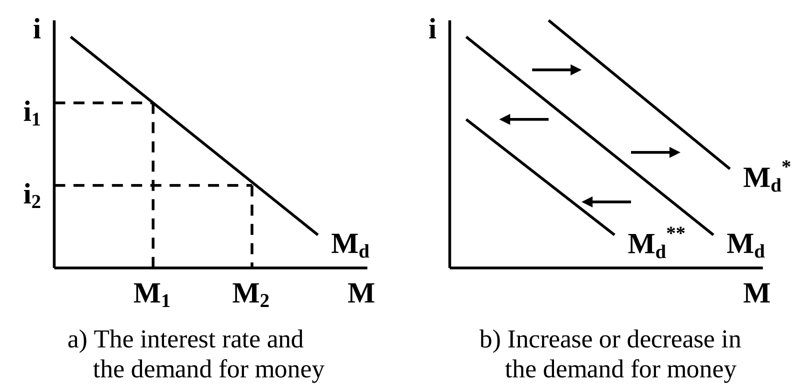


Figure 10-14
The Demand for Money

We show the relationship between the quantity of money demanded and the interest rate in panel a) of Figure 10-14. As is the case with any demand curve, we have an inverse relationship. If the interest rate is i_1, we demand M_1 for our portfolios. If the interest rate is lower, at say i_2, we want to hold M_2 money in our portfolios. Again, the demand for money is a demand for a particular type of asset, a liquid type of asset that we can spend. When the interest rate is high, we economize on our money balances. When the interest rate is low, we suffer a lower opportunity cost from holding money, so we hold more of it in our portfolios.

In order to draw a particular money demand curve, we have to hold other determinants of the demand for money constant. Three of the main other determinants of the demand for money are income, the price level, and cultural or institutional factors. If any of these change, the money demand curve will shift. Suppose the level of income in the economy increases. Because higher income leads to higher spending, and higher spending requires more money, an increase in income shifts the money demand curve to the right, from M_d to M_d* in panel b) of figure 10-14. A decrease in income shifts it to the left, from M_d to M_d**.

Likewise, an increase in the price level will require more money to support the spending at that price level. A decrease in the price level will require less money. Suppose the prices of all goods and services double. If you carried $7 in your pocket to buy a movie ticket before the increase in the price level, you now have to carry $14 in your pocket. Because your market basket of goods and services is now priced higher, you need to have more money in your portfolio of assets. The relationship between changes in the price level and changes in the demand for money will be very important in the next chapter when we develop an aggregate demand and supply model. We'll see that the effect of the price level on money demand and the effect this change in money demand has on the interest rate determines, in part, the slope of the aggregate demand curve.

Finally, institutional factors can affect the demand for money. Suppose on-line banking becomes more prevalent in an economy, and with it we have the possibility of making low cost transfers from an interest bearing account to a checking account. As the cost of asset transfers falls, we are more willing to economize on money balances. We don't need to leave as large a balance in a checking account earning no or little interest. Rather we could leave funds in interest bearing accounts and transfer funds to a checkable account right before we spend.

In Figure 10-15, we add the supply of money to the money market model. We make this a vertical line because we assume that the central bank controls the amount of money in the economy at some fixed level. At the point of intersection of the money demand and money supply curves, we find the equilibrium interest rate, i_e in Figure 10-15, the interest rate at which the amount of money demanded just equals the amount of money supplied by the central bank.

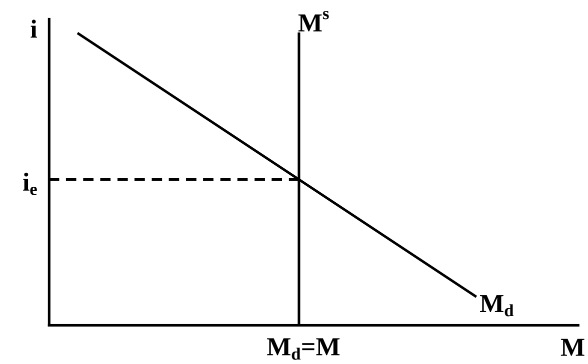


Figure 10-15
The Money Market Model

In panels a) and b) of Figure 10-16, we show the effect of an increase and decrease in the demand for money respectively. When the demand for money increases from M_d to M_d* in panel a), the previous interest rate i_1 is not high enough to equilibrate the money market. The excess demand for money drives up the interest rate to a level i_2, where the quantity of money demanded again equals the amount of money supplied. Because we learned about the inverse relationship between bond prices and interest rates earlier in the chapter, we can now tell a little dynamic story about how the interest rate rises from i_1 to i_2. Suppose only one asset other than money exists in financial markets, bonds. If people and businesses want more money in their portfolios, they must sell bonds. This drives down bond prices and raises the yield on bonds, and the yield on bonds is the interest rate in this simple economy.

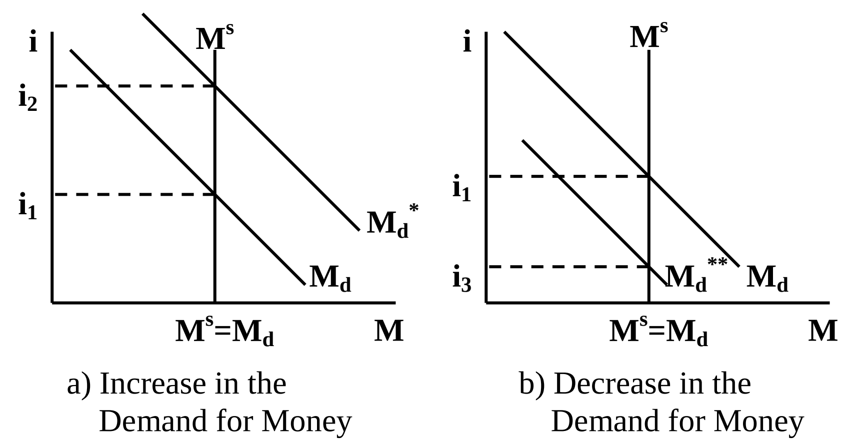


Figure 10-16
The Money Market and a Change in the Demand for Money

In panel b) of Figure 10-16, we show the effect of a decrease in the demand for money from M_d to M_d**. In this case we have a temporary excess supply of money in the money market, which puts downward pressure on the interest rate. Again, we can use the simplified financial market example to explain the dynamics of this adjustment. If people and businesses demand less money, they now have the ability to buy earning assets with their wealth. If they buy bonds, this increases the demand for bonds and drives up the price of bonds. As bond prices rise, bond yields fall, and a bond's yield is again the interest rate.

In Figure 10-16, we held the money supply constant and changed the demand for money. In Figure 10-17, we hold the demand for money constant and change the money supply. We see an increase in the money supply in panel a) of the figure. This represents expansionary monetary policy. Note that this characterization makes sense, as an increase in the supply of money by the central bank, from M^s to M^s* causes a decrease in the interest rate from i_1 to i_2. Likewise, we show contractionary monetary policy in panel b) of the figure with a shift to the left of the money supply curve. This reduction in the money supply leads to an increase in the interest rate from i_1 to i_3. Again, a bond prices and interest rates story lurks just beneath the surface in Figure 10-17. When the Fed increases the money supply, it buys bonds, driving up bond prices and driving down interest rates. When it sells bonds the money supply falls, and the increased supply of bonds drives down bond prices and drives up interest rates. The inverse relationship between bond prices and interest rates and the effect of an increase in the money supply in the money market are merely two sides of the same coin (pun intended).

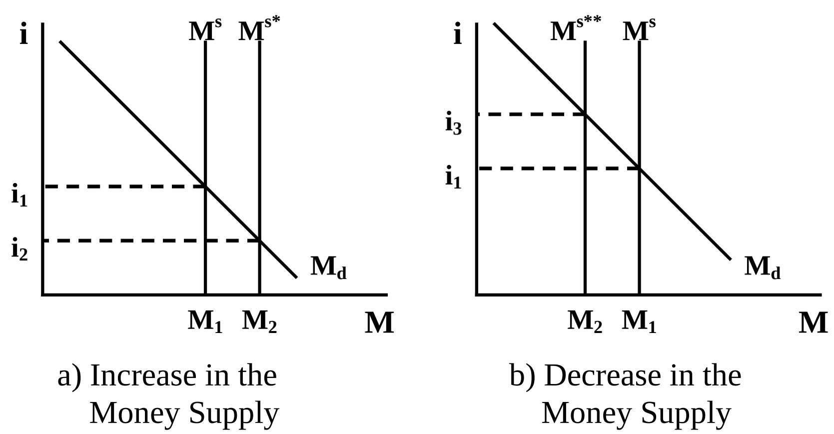


Figure 10-17
The Money Market and a Change in the Money Supply

Monetary Policy and the Federal Funds Rate

In addition to understanding the inverse relationship between bond prices and interest rates, and the money market model, we can understand the relative stimulus or restraint of the Fed's monetary policy by looking at movements in the federal funds rate. As we saw earlier in this chapter, this interest rate's name can cause misunderstanding. When a bank borrows from another bank, often for a time period as short as overnight, it pays the federal funds rate. Very little in this transaction is even remotely federal, and the only small link to anything federal is one to the Fed, not the federal government. The federal funds rate derives its name from the fact that banks are usually engaging in interbank financial transactions to meet reserve requirements. Also, the Fed can control this rate with a very high level of precision. A summary of the process of controlling the federal funds rate appears in Figure 10-18. The left (right) column pertains to expansionary (contractionary) monetary policy. When the Fed buys (sells) bonds excess reserves in the banking system increase (decrease). With more (less) excess reserves in the banking system lenders have less (more) bargaining power in setting the rate on interbank loans. Interbank loans become easier (harder) to obtain and the federal funds rate falls (rises). Other short-term interest rates in the economy adjust to the federal funds rate and when they do consumers and businesses adjust their spending accordingly. More spending, the result of an expansionary monetary policy of lowering interest rates stimulates the economy. Less spending constrains it.

Expansionary Monetary Policy	**Contractionary Monetary Policy**
Fed buys bonds	Fed sells bonds
↓	↓
Excess reserves in banking system increase	Excess reserves in banking system decrease
↓	↓
Interbank loans easier to obtain	Interbank loans harder to obtain
↓	↓
Federal funds rate falls	Federal funds rate rises

Figure 10-18
Open Market Operations and the Federal Funds Rate

Monetary Policy and the Interest Rate on Bank Loans

Yet another way to understand the effect of monetary policy is to understand the relationship between open market operations and the loans banks make. We see this process summarized in Figure 10-19. Again, the left (right) column refers to expansionary (contractionary) monetary policy. When the Fed buys (sells) bonds reserves of the banking system increase (decrease). With more (fewer) reserves banks increase (decrease) the supply of loans in financial markets. Loans are easier (harder) to obtain and the interest rate on them falls (rises).

Expansionary Monetary Policy	**Contractionary Monetary Policy**
Fed buys bonds	Fed sells bonds
↓	↓
Excess reserves in banking system increase	Excess reserves in banking system decrease
↓	↓
Banks make more loans (supply of loans rises)	Banks make fewer loans (supply of loans falls)
↓	↓
Interest rate on loans falls	Interest rate on loans rises

Figure 10-19
Open Market Operations and Interest Rate on Bank Loans

In Figure 10-20, we see a summary of the many different ways to think about the influence the Fed has through its monetary policy. Of course, all methods of understanding are related and all involve open market operations, the effect open market operations have on interest rates, and the effect changes in interest rates have on spending in the economy. The first of these four methods, described in detail above, is the inverse relationship between bond prices and interest rates, either with a long term bond example, or a short term bond example. We also can use the money market model, which highlights the link between changes in the money supply and changes in the interest rate. The federal funds market provides the third method. This one is very important, because the financial and economic press focuses almost entirely on

Inverse Relationship Between Bond Prices and Yields
- Long bond example
- Short bond example

The Money Market Model with Changes in Money Supply

Open Market Operations, Changes in Excess Reserves of the Banking System, and the Federal Funds Rate

Open Market Operations, Changes in Excess Reserves of the Banking System, and the Interest Rate on Bank Loans

Figure 10-20
A Summary of Ways to Understand the Relationship of Monetary Policy to Changes in Interest Rates

movements in the federal funds rate as an indicator of whether the Fed's policy is expansionary, neutral, or contractionary. And finally, we can understand the effect of open market operations by looking at the effect on reserves of the banking system and the effect changes in reserves have on bank loans and the interest rate on these loans. While a complete understanding would need mastery of all these channels of monetary policy, mastery of any will improve one's economic literacy enormously.

The Fed as a Lender of Last Resort: Managing the Financial Crisis

The lender of last resort function of central banking is not a new idea. Indeed, its origin is associated with Walter Bagehot, a 19th century banking scholar, who encouraged central banks to lend freely in a financial panic. A financial panic, now more commonly called a financial crisis, usually involves a large and sudden drop in liquidity in the financial system. We defined liquidity earlier in this chapter as the ease with which an asset can be sold for money. And the Fed does operate in this area, through open market operations and other actions. But a broader definition of liquidity is needed to understand the lender of last resort function in the modern economy, one with highly developed financial markets. This broader definition is about financial system liquidity, which is important for the stability of the financial system.

We can broaden the definition of liquid assets to include short term financial instruments, agreements to provide something in the near future for the ability to do something now. A short term loan is a simple example. A borrower gets to spend now and repay later. A loan could be thought of as liquid if the loan could be refinanced at the end of its (short) term. If consumers, businesses and governments can "roll over" short term debt when it is due, the system has more liquidity in it. Think of it this way. Suppose you were expecting someone to pay back a loan you made to them. Your asset is more liquid, and less risky, if the borrower has the ability borrow again to pay you off.

Modern life, especially in business and government, but also for consumers, involves short term debt. Businesses find it more profitable to borrow in order to handle fluctuations in payments and receipts than to self-finance by building up a larger cash reserve. Incurring short term debt is also a relatively low risk way of obtaining financial leverage, which can amplify the return on equity. In normal times, large corporations borrow regularly in something called the commercial paper market. A sudden inability to borrow in this market would require them not to make payments they had committed to make. Others waiting for one of these payments would be similarly disrupted. Financial crises can spread in a chain reaction across the financial system when more and more entities cannot honor payment commitments.

For about two and a half decades before 2008, the Fed performed the function of lender of last resort sparingly. In response to the potential collapse of Continental Illinois Bank in 1984, at that time one of the 10 largest banks in the country, the Fed made large loans to Continental Illinois not just to meet a temporary shortfall of required reserves, but rather to prevent the failure of this large financial institution, and the spillover effects this would have had on other institutions in the financial system. The Fed again "opened its discount window" to financial institutions during the rapid depreciation of Asian currencies in the Asian financial crisis of the late 1990s, and did so again immediately after the World Trade Center and Pentagon terrorist attacks in September 2001.

The Fed's use of its lender of last resort function changed dramatically in late 2007 and 2008, in response to several simultaneous destabilizing events in the financial system. In 2007 and 2008, the U.S. financial system came under severe stress. Here are just a few of these events, all of which the Fed was involved in or responding to, along with the U.S. Treasury, the FDIC, the Securities and Exchange Commission, and other regulatory agencies. The current severe and long recession began in December 2007. The stock market had its largest decline since the Great Depression of the 1930s. The U.S. mortgage market was in turmoil, when a speculative bubble burst in the U.S. housing market and it became clear that defaults on "subprime" mortgages, those to very risky borrowers, would make them and the assets in which they were bundled for sale, mortgage-backed securities, much less valuable. With assets plunging in value across the financial system, borrowers became much more risky and short term lending to them dried up.

In early 2007, The Federal Home Loan Mortgage Association (Freddie Mac), a quasi government agency designed to support home buying in the U.S., announced that it would no longer buy the most risky mortgages and mortgage-backed securities. In 2008, troubled mortgage lender Countrywide Financial had to be purchased by Bank of America, in lieu of going bankrupt. Likewise, investment bank Bear Stearns Companies, Inc. was purchased by JPMorgan Chase & Co. Both the Federal National Mortgage Association (Fannie Mae) and Freddie Mac were taken over by the federal government. Carlyle Capital Corporation failed to meet margin calls (demands for payment by lenders) on its mortgage bond fund. Well-known financial institution Merrill Lynch & Co. had to be purchased by Bank of America. Lehman Brothers Holdings went bankrupt. Washington Mutual Bank was acquired by JPMorgan Chase. The banking operations of Wachovia Corporation were acquired by Wells Fargo & Co. The large insurance company AIG Corporation teetered on the brink of bankruptcy. These were not normal financial and economic times, and the Fed responded with a massive program of lending and asset purchases. The Fed's comrade in arms, the U.S. Treasury, responded as well. We focus here on the Fed's actions.

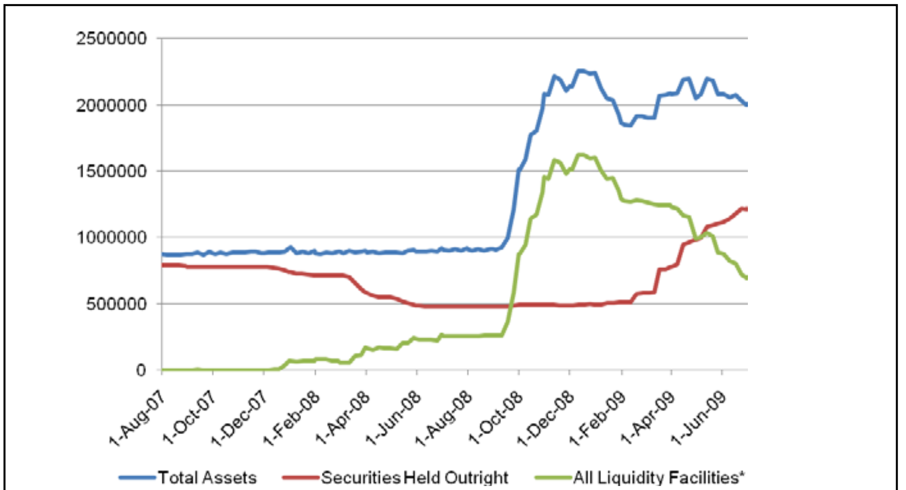


Figure 10-21
Federal Reserve Assets
08/01/07 – 07/01/09

Figure 10-21 shows changes in the total amount of assets and the composition of the assets in the Fed's portfolio in the period from August 2007 to July of 2009. Before December 2007, the Fed's assets consisted primarily of short term treasury bills. The flat Total Assets line (the top line in the figure, blue if you are viewing in color) looked that way into the indefinite past. In December, 2007, the composition of the Fed's portfolio of assets began to change. Through various kinds of short term financial transactions such as loans and asset swaps, the Fed injected higher quality assets onto the balance sheets of financial institutions. This increased the credit-worthiness of borrowers, because the quality of their assets increased. It also freed up lending as lenders had to be less defensive about their own portfolios.

The green line of Figure 10-21, the one that rises rapidly in September 2008, shows the unprecedented moves by the Fed to inject liquidity to the financial system through various types of what it calls "liquidity facilities." These efforts continue as this is written in July of 2009. In 2009, the Fed started to address the quality of assets in the financial system by buying directly mortgage-backed securities guaranteed by the now federal mortgage institutions Fannie Mae and Freddie Mac. This is why the red line, showing Securities Held Outright, starts to rise in this period.

The actions of the last two years by the Fed, and by the U.S. Treasury, the FDIC and others, have been well beyond the normal types of financial actions undertaken by these regulatory institutions, both in scale and in scope. At no other time in the history of central banking has a central bank and its government allies so fully used its lender of last resort function. Were the actions necessary? Have they been enough to stem a financial crisis? A cautious yes answer, at least on the part of the Fed, might be warranted at this time. But the jury is still out. The economy continues in a deepening recession. The financial system, while improved, still struggles under the burden of large amounts of mortgage debt with questionable value, at the same time the bottom of the housing price decline has not been reached. These are truly interesting times for central banking in the U.S. and around the world.

Chapter 11: The Aggregate Demand and Aggregate Supply Model

Introduction

In Chapter 3, we developed the supply and demand model and used it to explain many interesting and important microeconomic relationships. In this chapter we will develop a comparable model for macroeconomics, the aggregate demand and aggregate supply (ADAS) model. We will use this model to examine many of the macroeconomic relationships that are familiar to us.

We will use our macroeconomic "workhorse" model to look at different parts of the business cycle, recessions and expansions, indicated directly in the model by movements in real GDP and implicitly by changes in the unemployment rate. We will see that the model suggests that different phases of the business cycle have different implications for inflation and deflation. We will be able to see how changes in different components of aggregate demand affect the economy and how the Fed can counteract or exacerbate these effects through monetary policy. We will also look at the government's role in macroeconomic stabilization or destabilization through an examination of fiscal policy.

We can also visualize with the ADAS model several episodes in macroeconomic history, such as the Great Depression, the recent slump in Japan, the non-inflationary expansion of the late 1990s in the United States, and the recovery from the 2001 recession. But, before we can plunge into the refreshing waters of theory and history, we must learn to swim, and learning macroeconomic swimming is not easy. This could be the most challenging material we have encountered so far. We begin by learning the mechanics of the model.

Slopes and Shifts Again

Our method for developing the ADAS model parallels the method we used in developing the microeconomic supply and demand model. We must understand why the two main curves in the model slope the way they do. We also must understand what is held constant when we construct our models of aggregate demand and aggregate supply. As in microeconomic supply and demand analysis, these other things held constant will become important when we use the model, for we will capture changes in the economy through changes in these other things held constant.

Figure 11-1 provides a good start. First, look at the vertical and horizontal axes of the model. On the horizontal axis, we have the macroeconomic counterpart to quantity in the supply and demand model, real GDP. Rather than price on the vertical axis, we now have the average price level measured with a price index such as the CPI. Now look at the point in the model representing an economy with a price level of CPI_1 and a level of real GDP of $RGDP_1$. Points like these, different combinations of the price level and real GDP will soon emerge from the model as equilibrium combinations represented by the intersection of two curves. For now let's just think of point A as an initial point for an economy.

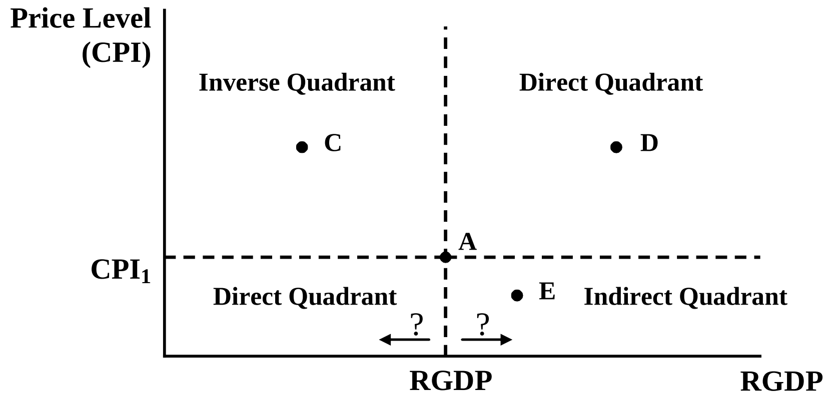


Figure 11-1
Searching for Slopes of the Aggregate Demand and Aggregate Supply Curves

Suppose we consider aggregate demand, the sum of consumption, investment, and government spending on final goods and services, and net exports. Remember, this was one of the ways we measured GDP in Chapter 9, but GDP and aggregate demand are usually just two sides of the same coin. Can we imagine a line in the graph in Figure 11-1 that represents how the price level and real GDP might be related through changes in aggregate demand? Might there be a corresponding level of real output that would be consistent with this new price level? And could we tell a theoretical story about how changes in the price level might cause changes in aggregate demand and how this change in spending affects real GDP? These are the questions at hand.

Suppose we experience a rise in the price level. Will aggregate demand fall leading to a decline in real GDP. If so, the aggregate demand curve would have points in the upper left inverse quadrant of Figure 11-1. The aggregate demand curve would show an inverse relationship between the price level and real GDP. But, if aggregate demand in the economy responded directly to an increase in the price level, a point in the upper right direct quadrant of the graph would represent the new position of the economy.

Likewise, with aggregate supply, would an increase in real GDP lead to an increase in costs per unit of output, and would businesses try to pass the increased costs on to consumers in the form of higher prices. If so, the aggregate supply relationship would show a movement from point A to a point like D in the model. But, if growing real GDP led to a lower price level, we would see a movement from point A to a point like E. We will determine these slopes shortly.

You may not have caught the subtle difference between the AD and AS questions in the last two paragraphs. When considering aggregate demand, we ran the direction of causality from the price level through aggregate demand to real GDP. When asking about the aggregate supply relationship we ran the direction of causality from real GDP to the price level. While subtle, these are important distinctions, so much so that we summarize them in Figure 11-2. When attempting to understand the relationship we call the AD curve, we go from changes in the price level to changes in real GDP. When attempting to understand the relationship we will call the AS curve, we go from changes in real GDP to changes in the price level. We do this because it makes learning the model much easier.

Aggregate Demand Curve

ΔPrice Level $\longrightarrow$ ΔReal GDP

Aggregate Supply Curve

ΔReal GDP $\longrightarrow$ ΔPrice Level

Figure 11-2
Direction of Causality in the AD and AS Curves

The Downward-Sloping Aggregate Demand (AD) Curve

In microeconomics, if price rises people buy less of a good, other things held constant. This explanation is simple, quick and transparent. One example of behavior drives the theory. The AD curve slopes downward for much more complicated reasons than the microeconomic demand curve does. We summarize these in Figure 11-3.

We emphasize two channels through which changes in the price level affect aggregate demand and real GDP. These are represented by two chains of economic events beginning with the demand for money and net exports. On the top line of the figure, we see that an increase in the price level increases the demand for money in the economy. As we concluded in the last chapter, with a higher price level, people and businesses need more money to enable the higher spending level caused by an increase in the price level. The next event in the chain involves the interest rate. With the money supply held constant by the Fed, the increase in money demand leads to an increase in the interest rate to reequilibrate the money market.

Demand for Money Reason	Price Level↑→ Demand for Money↑→ Interest Rate↑→ Interest Sensitive Spending↓ → Real GDP↓
	Price Level↓→ Demand for Money↓→ Interest Rate↓→ Interest Sensitive Spending↑ → Real GDP↑
Net Exports Reason	Price Level↑→ Relative Price of U.S. Goods↑→ Net Exports↓→ Real GDP↓
	Price Level↓→ Relative Price of U.S. Goods↓→ Net Export↑→ Real GDP↑

Figure 11-3
Reasons the Aggregate Demand Curve Slopes Downward

Continuing to read across the chain of events from left to right we see that an increase in the interest rate leads to a decline in interest-sensitive spending such as consumption and investment. Because these are two of its main components, aggregate demand falls. Businesses see this decline in demand and lower production to meet it. Real GDP falls.

So an increase in the price level works through the money market to raise the interest rate and reduce aggregate demand and the real GDP that responds to it. We have an inverse relationship, a downward-sloping AD curve. Likewise, reading across the chain of events in the second line we see that a decrease in the price level reduces the demand for money, thereby lowering the interest rate, increasing interest-sensitive components of aggregate demand leading to an increase in real GDP. Again, we have reasoned through a complex chain of events in the economy to derive a downward-sloping AD curve.

The bottom two lines of Figure 11-3 involve the effects of an increase or a decrease in the price level on net exports. Here the chain of events is more limited. If the price level in the U.S. rises and the price levels in all other countries remain the same, our goods are relatively less attractive on world markets. Foreigners will buy less in the U.S. and U.S. consumers and businesses will spend more abroad. Exports fall and imports rise. This decline in net exports is a direct reduction in aggregate demand. The decline in aggregate demand leads producers to respond to it by producing less. Real GDP falls. A fall in the U.S. price level has the opposite effect in the foreign sector. In both cases through the net exports channel, an inverse relationship exists between the price level and the level of real GDP that is consistent with it.

We capture these relationships in an abstract aggregate demand model like the one in Figure 11-4. At a price level of CPI_1, aggregate demand in the economy calls forth production of $RGDP_1$. A rise in the price level to CPI_2 works through chains of events in the money market and net export channels to result in a reduction in aggregate demand and the real GDP it supports. It is important to realize that what looks simple and analogous to the microeconomic demand model is really much more complex.

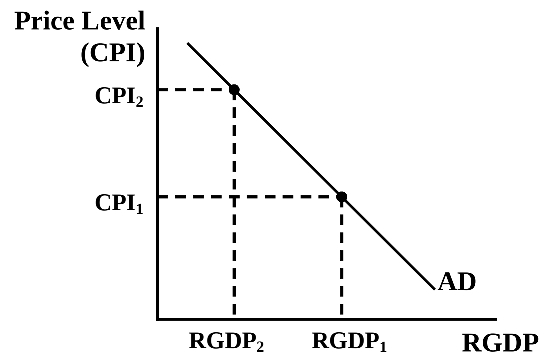


Figure 11-4
The Aggregate Demand Model

Compared to the microeconomic demand model the action in the AD model is more behind the scenes. The behavior is hidden. All we see are different points that relate different price levels and different levels of real GDP that correspond to them. When explaining the slope of the AD curve DO NOT say, “when the price level rises people buy less.” While they might buy less, it’s not a simple behavioral response to a rise in the price of a good. In fact, if all wages and prices in the economy rise in the same proportion, consumers could well buy the same amount of each good, as the relative prices of goods and the level of real income in the economy would not change.

Shifts in the AD Curve

In any graphical economic model involving more than two variables, something has to be held constant when we select two variables to put on a two-dimensional graph. A relationship represented by a line on an economics graph will shift if one of the other variables held constant no longer remains so. We have only to remember the microeconomic demand model to illustrate the effect of changes in other things (previously) held constant. When a determinant of demand changed the demand curve shifted. When a determinant of supply changed the supply curve shifted.

In Figure 11-5, we show the macroeconomic variables that cause shifts in the AD curve. In the column labeled AD Shifter, we use up and down arrows to indicate an increase or decrease in the aggregate demand shifter. In the Direction of Shift column, we note whether the AD curve shifts left or right. It is important here to think of the AD curve as shifting left or right, not up and down. This is consistent with our adopted direction of causality, from price level to real GDP. If aggregate demand increases for some reason other than a change in the price level, the level of real GDP corresponding to any price level will differ from that which occurred before the change in aggregate demand.

AD Shifter		Direction of Shift
Autonomous Consumption	↑	Right
Autonomous Consumption	↓	Left
Autonomous Investment	↑	Right
Autonomous Investment	↓	Left
Autonomous Net Exports	↑	Right
Autonomous Net Exports	↓	Left
Government Spending	↑	Right
Government Spending	↓	Left
Taxes	↑	Left
Taxes	↓	Right
Money Supply	↑	Right
Money Supply	↓	Left

Figure 11-5
Shifters of the Aggregate Demand Curve

Note that the first three pairs of shifts contain the word autonomous. This means that the changes occur for some reason other than a change in the price level. We have already seen that consumption changes when a change in the price level creates a change in the demand for money and a change in interest rates. But this is a movement along the AD curve caused by a price level change. But consumption spending can change for other reasons, too. This is what we mean by autonomous consumption. For example, consumers might begin to feel optimistic about the future of the economy. When consumers are optimistic they spend more. At any given price level aggregate demand, and real GDP increase. If consumer confidence drops, the reverse happens. Consumers become cautious and start saving for a "rainy day." An increase in autonomous consumption shifts the AD curve to the right as shown in panel a) of Figure 11-6. A decrease in autonomous consumption shifts it to the left, as in panel b).

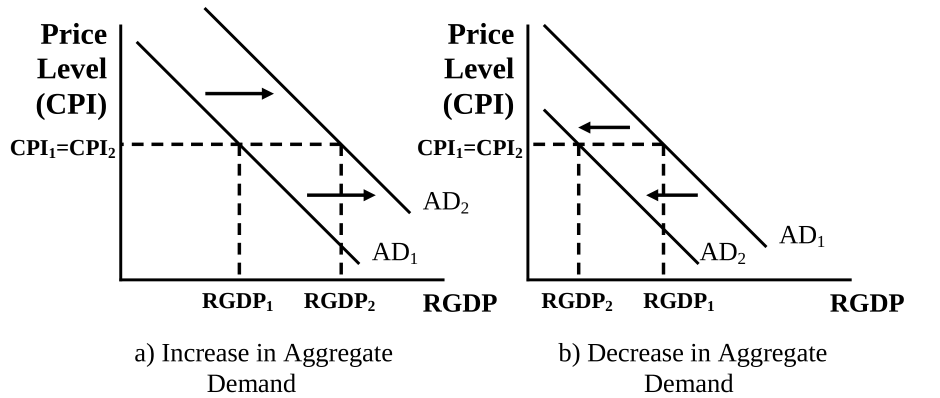


Figure 11-6
An Increase and Decrease in Aggregate Demand in the Aggregate Demand Model

Investment is also subject to changes that are not induced by changes in the price level. British economist John Maynard Keynes, whom we will discuss more in the next chapter wrote that investment depends mainly on the expectations of business owners, and these could be influenced, according to Keynes, even by the digestion of these owners, and what Keynes called the "animal spirits" of capitalism. If the future looks rosy, then investment can increase in anticipation of meeting future demand. If investors see bleakness in the future, investment might fall. Again, by investment we mean the purchase and production of capital goods. An autonomous increase in investment will shift the AD curve to the right, and a decrease will shift it to the left.

We have already seen that one of the reasons for movements along an AD curve comes from the foreign sector. But these movements require a change in the price level with other determinants of exports held constant. But suppose German consumers just wake up one morning with a strong demand for Ford pickup trucks. U.S. exports will increase and next exports will increase as well, even though no change in the price level caused this change. This autonomous change in net exports shifts the AD curve. If autonomous net exports rise, the AD curve shifts to the right. If autonomous net exports fall, the AD curve shifts to the left.

Just below the autonomous changes in consumption, investment and net exports in Figure 11-5 we find one direct and one indirect AD shifter. Because government spending is part of aggregate demand, if the government spends more on goods and services aggregate demand increases. Because we said nothing about a change in the price level starting this story we are not talking about a movement along the AD curve here. An increase in government spending will increase aggregate demand regardless of the price level. An increase in government spending shifts the AD curve to the right. A decease shifts it to the left.

Taxes affect aggregate demand mainly through their effect on disposable income of consumers. Disposable income is income after taxes. You can't spend the money the government takes out of your check each pay period, only the income you "take home." Reduced taxes increase disposable income and consumers spend a good portion of it. This is not caused by a change in the price level and the chain of events this sets in motion, so we don't handle tax changes by movements along the AD curve. Lower taxes increase aggregate demand and shifts the AD curve to the right. Higher taxes shift the AD curve to the left. Later in this chapter we will discuss the use of changes in government spending and taxation for the purposes of influencing economic activity in the economy. This is called fiscal policy.

Finally, actions of the Fed affect aggregate demand. Increasing the money supply lowers interest rates and stimulates interest-sensitive spending. Decreases in the money supply do the opposite. Increases in the money supply shift the AD curve to the right. Decreases shift it to the left.

The Upward-Sloping Aggregate Supply Curve

The aggregate supply curve has a closer relationship to the microeconomic supply curve than the aggregate demand curve does to the microeconomic demand curve, but we still have to resort to "chain of events" reasoning for an explanation. Two channels explain the upward slope of the AS curve, and both are related to changes in production cost and an assumption that businesses change prices when costs change. We see these two channels in Figure 11-7. As we saw in Figure 11-2, we now direct our reasoning from changes in real GDP to changes in the price level, just the opposite of what we did when we explained the downward slope of the AD curve.

First, let's examine the resource prices channel. Imagine an expansion in the economy, as indicated by the first change in our chain of events in the top line of Figure 11-7. As output expands, businesses purchase more resources to produce the higher level of real GDP. This increases demand in resource markets. As demand in these markets increases, wages and other resource prices rise. Rising resource prices causes average

Resource Prices

RGDP↑ → Resource Demand↑ → Resource Prices↑ → ATC↑ → Businesses Raise Prices → Price Level↑

RGDP↓ → Resource Demand↓ → Resource Prices↓ → ATC↓ → Businesses Lower Prices → Price Level↓

Diminishing Marginal Productivity

RGDP↑ → Resource Use↑ → Resource per Unit of Output↑ → ATC↑ → Businesses Raise Prices → Price Level↑

RGDP↓ → Resource Use↓ → Resource per Unit of Output↓ → ATC↓ → Businesses Lower Prices → Price Level↓

Figure 11-7
Reasons the Aggregate Supply Curve Slopes Upward

total cost (ATC) to increase. For the purposes of the ADAS model, we assume that businesses in general have some market power in the economy, and, when cost increases, they attempt to maintain profit by passing the increased cost on to consumers in the form of higher prices. As businesses raise prices, the price level increases and we've completed our first chain of events. An increase in real GDP leads to a rise in the price level in the economy. Real GDP and the price level are directly related. The AS curve slopes upward. If we follow the chain of events in the second line of the resource prices channel, a decline in real GDP, we find the same result, an upward sloping AS curve.

Diminishing marginal productivity also drives the effect of changes in real GDP on the price level. These chains of events appear in the bottom two lines of Figure 11-7. Because of diminishing marginal productivity caused by the existence of fixed resources in the short run, as real GDP rises and output expands and resource use increases, resource use per unit of output increases. This also increases production cost and businesses pass this increased cost on to consumers in the form of higher prices. When real GDP falls, the opposite occurs. Again, real GDP and the price level are directly related.

In our discussions in the remainder of this chapter, we will focus mostly on the resource prices channel. And because wages, salaries and other forms of employee compensation account for over 70% production cost, we will concentrate on the relationship between changes in real GDP and wages, and the important variable unit labor cost, labor cost per unit of output. In the real world, economists track unit labor cost in an economy closely, because unit labor cost affects the price level. Our model of aggregate supply appears as Figure 11-8. As real GDP rises from $RGDP_1$ to $RGDP_2$, resource prices rise, production cost increases, and the price level rises from CPI_1 to CPI_2. A decline in RGDP has the opposite effect.

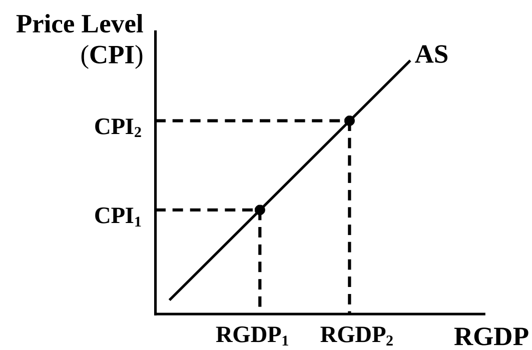


Figure 11-8
The Aggregate Supply Model

Shifts in the AS Curve

In order to construct the upward sloping AS curve, we must hold some other determinants of production cost constant, and, when these determinants of cost change, the AS curve will shift. We summarize AS shifters in Figure 11-9.

We will consider three general types of AS shifters here, supply shocks, price level expectations and productivity. These comprise the three chain of events groupings in Figure 11-9. A

	AS Shifter		Direction of Shift
Supply Shock	Autonomous Change in Resource Prices	↑	Up
	Autonomous Change in Resource Prices	↓	Down
Price Level Expectations	Price Level Expectations	↑	Up
	Price Level Expectations	↓	Down
Productivity	Autonomous Change in Productivity	↑	Down
	Autonomous Change in Productivity	↓	Up

Figure 11-9
Shifters of the Aggregate Supply Curve

supply shock occurs when a resource price rises for some reason other than an increase in resource demand caused by expansion in the economy. Adverse supply shocks in the form of rising oil prices have had an effect on aggregate supply at various times in the past three decades in the United States. First in the 1970s and again in 2004 and 2005, world oil prices rose dramatically. With autonomous increases in oil prices production cost rises and businesses raise prices to maintain profit margins. We show this with a shift up in the AS curve. When resource prices eventually decline, as oil prices did in the 1980s and 1990s, favorable supply shocks occur. We show this with a shift down in the AS curve. Both upward and downward shifts of the AS curve appear in Figure 11-10.

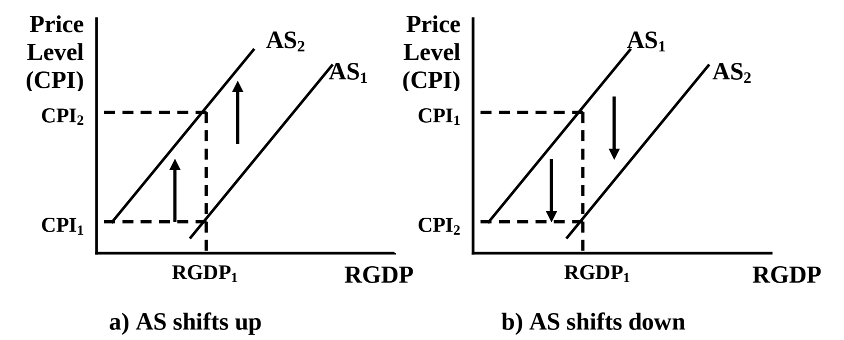


Figure 11-10
An Increase and Decrease in Aggregate Supply in the Aggregate Supply Model

Other more subtle supply shocks also shift the AS curve. An increase in union power in the economy can put upward pressure on wages and encourage businesses to raise prices. More environmental and occupational safety and health regulation can increase the cost of running a business, and, as cost rises, so do the prices businesses set. While not exactly a supply shock, the level of competition in markets, either domestic or foreign competition, can affect the pricing power of businesses in a way similar to a supply shock. With less (more) competition and more (less) market pricing power business will raise prices by a larger (smaller) margin over cost. This shifts the AS curve.

It’s important to think of the AS curve shifting up and down, rather than left and right. This helps to underscore the point that businesses attempt to increase their prices (and indirectly the price level) at any level of real GDP. We don’t have to have an expansion in the economy for businesses to raise prices when supply shocks occur.

In the middle section of Figure 11-9, we have the price level expectations channel. Workers are aware that changes in the price level can affect their real incomes if price level changes are not matched by increases in wages. This thinking has given rise to the cost of living adjustment (COLA) in wage contracts. If workers expect the price level to increase they will drive harder bargains in wage negotiations and businesses will be more likely to give them higher wages if they expect they can raise prices. So, if workers expect the price level to increase, it will affect wages, and wage change will be passed on in higher prices. The AS curve shifts up with an expectation of an increase in the price level and down with an expectation of a decrease. Again, this modification in business pricing does not depend on a change in real GDP. As before, we reserve changes in real GDP, and the price increases this creates, for movements along the AS curve, not shifts in it.

Finally, in the third collection of events we have productivity. Let's focus on labor productivity here, as labor is such a large component of production in an economy. Labor productivity can increase from having relatively more other resources in the production process, from education and job training, and from the mysterious forces of technological change. As labor productivity increases in an economy, unit labor cost falls for any given wage level. Declining unit labor cost has an effect on the AS curve similar to a favorable supply shock. Productivity increases in the late 1990s and early 2000s were partially responsible for the lack of necessity of businesses to raise prices even in a strong economic expansion. This, as we shall see, decreased the rate of increase in the price level in these periods.

The ADAS Model

Now that we have developed both sides of the ADAS model we can do the exciting part, put AD and AS curves together and begin to apply the model. The complete ADAS model appears as Figure 11-11. As in the microeconomic supply and demand model, the intersection of the two curves has special significance. At output level $RGDP_1$ and price level CPI_1, AD equals AS. This means that the economy is in an equilibrium of sorts, a macroeconomic equilibrium. For our purposes, this important condition means that the economy of our model will remain at this output and price level combination until something happens

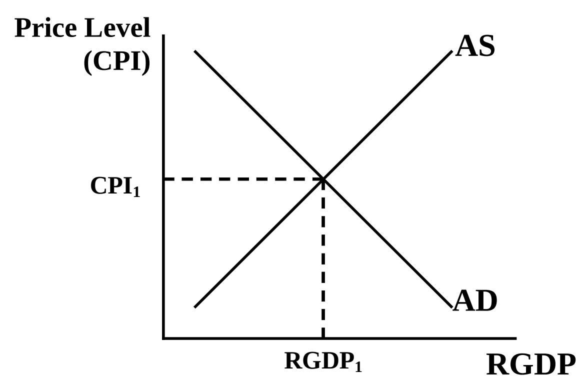


Figure 11-11
The Aggregate Demand and Supply Model

in the economy to displace it. When this happens we get action in the model. The curves shift and the equilibrium point moves.

Before moving on to applications of the ADAS model, we summarize in Figure 11-12 the similarities and differences between the microeconomic supply and demand model and the macroeconomic ADAS model. Similarities appear at the top of the figure. Both models have a price variable on the vertical axis and a quantity variable on the horizontal axis of the graph. Of course, the macroeconomics price and quantity variables are much more involved and complicated than price and quantity in a particular market. Both models have a downward-sloping demand relationship and an upward-sloping supply relationship. Again, the meaning of the relationships in the ADAS model is much more complicated and involved that in the supply and demand model. Single behavioral relationships dominate chain of events relationships in terms of simplicity. In both models, the intersection of the curves implies an equilibrium, the most important implication of which is that nothing will happen in the model unless one of the other things held constant no longer remains so. Which brings us to the final similarity. Using each model requires us to change at least one of the other things held constant.

Similarities Between Models

Price variable on vertical axis

Quantity variable on horizontal axis

Downward-sloping demand relationship

Upward-sloping supply relationship

Intersection of curves means equilibrium

Using the models involves changing other things held constant

Differences Between Models

S&D model shows behavioral relationships while ADAS shows chain of events relationship

Action is on the curves in S&D model while action is behind the scenes in ADAS model

Slope of curves, while generally the same, happens for vastly different reasons

Figure 11-12
Superficial Similarities and Fundamental Differences Between the S & D and ADAS Models

Differences between the models appear in the bottom portion of Figure 11-12. We alluded to the first difference above. In the microeconomic supply and demand model, straightforward behavioral relationships explain the curves. If price rises, buyers buy less, other things held constant. If price rises, sellers wish to sell more, end of story. As we have mentioned several times already, slopes of the AD and AS curves involve chain of event relationships. The second difference is actually a corollary of the first. While the general slopes of the demand and supply curves are the same in both models, they have those slopes for vastly different reasons. Finally, the action in the microeconomic supply and demand model is right on the surface, while the action in the ADAS model is behind the scenes.

Expansions, Contractions, Recessions, Recoveries, Booms and Corrections in the ADAS Model

We begin with some basic macroeconomic episodes or events and how we represent them in the ADAS model. We also name the episodes according to where they occur in the business cycle. The arrows in Figure 11-13 align movements in the model to the terms that apply to them. Price level movements are straightforward. An increase in the CPI registers inflation, a decrease deflation. We name movements in real output, in part, according to the initial level of real GDP preceding the change. Note the vertical line intersecting the horizontal axis at $RGDP_{FE}$. This is the level of real GDP that generates full employment in the labor market. Recall that this does not mean that the unemployment rate is zero. Rather, at full employment cyclical unemployment is equal to zero, while frictional and structural unemployment remain positive.

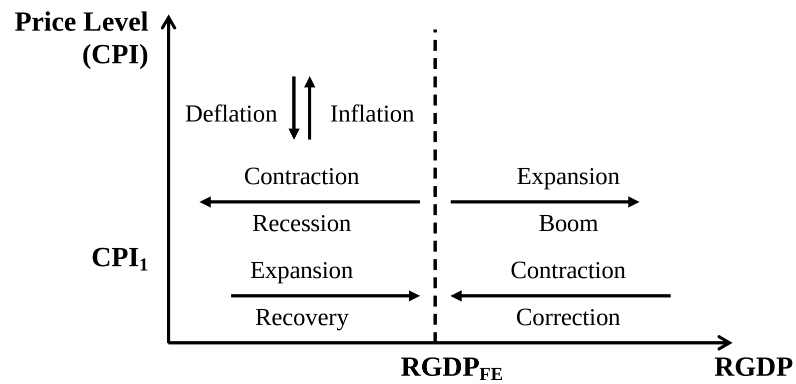


Figure 11-13
Full Employment, Booms and Recessions
Inflation or Deflation in the ADAS Model

When the intersection of the AD and AS curves moves to the left in the model, it signifies a decline in real GDP, a contraction in the economy. If this contraction lasts long enough, the NBER might even label it a recession. Most recessions begin at or near full employment and are caused by a decline in aggregate demand. However, the economy can contract if the level of real GDP has been above the full employment level. This would be a recession following a boom, something we'll call a correction. We'll define a boom as an expansion of real GDP beyond the full employment level. As we will see, an economy can move temporarily beyond full employment by grinding more and more work out of a fully employed labor force, but this situation will not last. If the economy expands after a recession, we call that a recovery.

Figure 11-14 shows a recession with deflation. Note that we start the model out at full employment, where the AD curve labeled AD_1 intersects the AS curve at $RGDP_1$ and CPI_1. We then assume a decline in aggregate demand from either autonomous reductions in consumption, investment, or net exports, or a reduction in the money supply or government spending, or an increase in taxes. The AD curve shifts to the left to AD_2. The new macroeconomic equilibrium occurs at $RGDP_2$ and CPI_2.

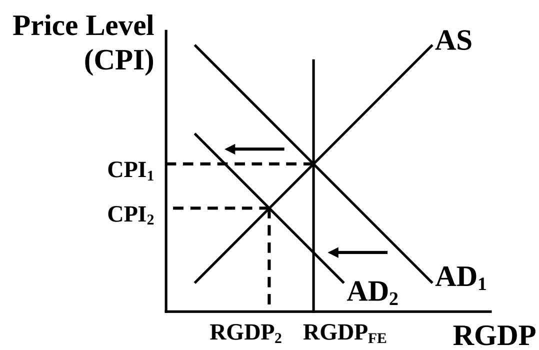


Figure 11-14
A Recession with Deflation in the ADAS Model

In the real world, deflations are uncommon. Recessions usually occur with disinflation, not deflation. Disinflation is a reduction in the rate of inflation, not and actual decline in the price level. This is a shortcoming of the ADAS model, but a macroeconomic model with inflation instead of the price level would involve dynamic analysis inappropriately difficult for an introduction to economics. That said, recessions with deflation do occur in the real world. The graphical story in Figure 11-14 depicts the period 1929-1933 in the U.S. very well. The Great Contraction or Great Depression, two names applied to this period, resulted in a large decline in the price level and a very deep recession. Also, in the 1990s, Japan experienced recession with deflation, although not as severe as in the Great Depression in the U.S.

In Figure 11-15, we see an expansion after a recession. This recovery begins at output level $RGDP_1$ and price level CPI_1,

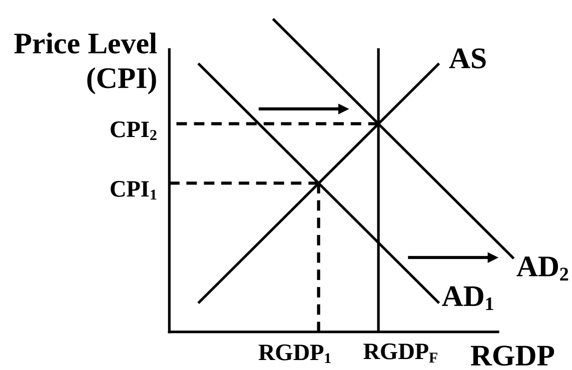


Figure 11-15
A Recovery with Inflation in the ADAS Model

where the AD curve AD_1 intersects the AS curve. Note that we are starting below the full employment level of real GDP. The unemployment rate rose during the contraction of the recession and falls again with the recovery. Now we show an increase in aggregate demand from AD_1 to AD_2. As the economy expands however, demand for resources rises, putting upward pressure on wages and other resource prices and production cost. With rising cost and increasing demand for their output, businesses raise prices. The CPI rises from CPI_1 to CPI_2. Recoveries from recessions involve renewed inflation.

In Figure 11-16, we show an expansion that moves into the boom stage. Here we start at full employment real GDP where the AD_1 intersects the AS curve at $RGDP_{FE}$ and CPI_1. The aggregate demand curve shifts to the right to AD_2 and real GDP and the price level increase to $RGDP_2$ and CPI_2, respectively. Later in the chapter, we will return to the boom, because we are telling only half the story here. A boom of this nature is temporary because it will set off a self-correcting mechanism in the economy that will lead to an end of the boom. But first, we need to examine macroeconomic policy in the context of the ADAS model.

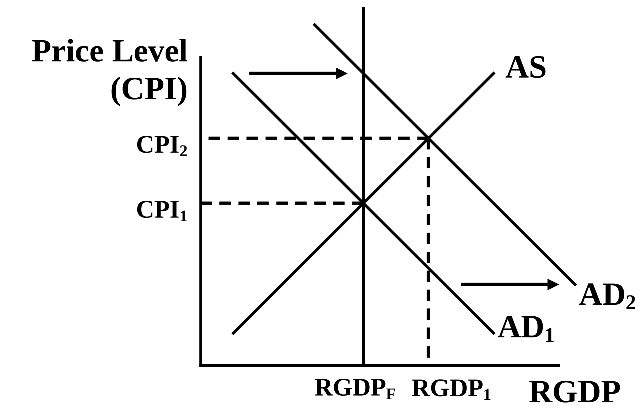


Figure 11-16
A Boom with Inflation in the ADAS Model

Macroeconomic Policy and the ADAS Model

Macroeconomic policy comes in two forms, monetary policy and fiscal policy. We summarize the essential characteristics of each in Figure 11-17. In the left column of the figure we see that the Fed undertakes

	Monetary Policy	**Fiscal Policy**
Undertaken by	Fed	Congress & President
Policy Instrument	Money Supply and Interest Rates	Government Spending and Taxation
Influence of Politics	Low	High
Ability to Time Correctly	Difficult	Very Difficult

Figure 11-17
Some Major Characteristics of Macroeconomic Policy

monetary policy. The Fed, or the comparable central bank in other economies, uses as a policy instrument changes in the money supply and interest rates. When interest rates change, aggregate demand changes. Monetary policy, especially when undertaken by an independent central bank, avoids political influence. Finally, we see that the ability to time monetary policy correctly is difficult. We will address this timing problem in more detail in the next chapter, but let's just recognize here that timing problems stem from the difficulty in recognizing where the economy is in the business cycle, where it's going, and where it will be after the time period that elapses between the implementation of a policy change and the effect of that change.

In the right column, we have fiscal policy, undertaken buy the Congress and the President. By changing government spending and taxes, the President and the Congress can change aggregate demand. We'll see in the next chapter that this process affects the magnitude of the deficit or surplus in the federal government's budget. Because fiscal policy involves government spending and taxation, and these are two of the major ways politics finds expression in government, the influence of politics on fiscal policy is high. Because fiscal policy must traverse the slippery slope of politics, it incurs an additional lag that monetary policy does not have. It takes a while for governmental officials to implement a fiscal policy. The Fed can implement monetary policy in the next hour. By contrast, fiscal policy requires bills entered into the congressional lawmaking process, committee meetings, floor discussions, House-Senate conference committee meetings and, finally, signature by the President. And while these debates are happening, the economy is not standing still.

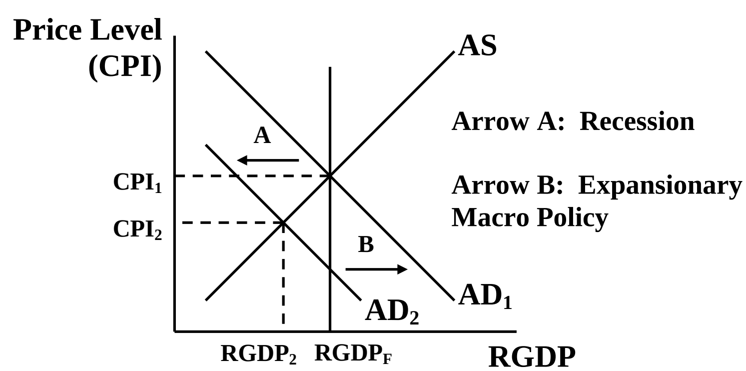


Figure 11-18
Expansionary Macroeconomic Policy to Combat a Recession

In Figure 11-18, we represent expansionary macroeconomic policy used to counteract a recession. Suppose the economy starts off in the model at full employment real GDP, $RGDP_{FE}$, and a price level CPI_1. Now suppose some component of aggregate demand declines and the economy slips into a recession. Arrow A in the figure indicates a recession, a leftward shift of the AD curve, perhaps a shift to AD_2. The Fed could undertake expansionary monetary policy, buying government bonds through open market operations, and shift the AD curve back to the right. Likewise, the Congress and the President could increase government spending or cut taxes to boost aggregate demand. Arrow B represents these shifts of the AD curve. A perfectly timed policy would be one designed for the increased spending induced by the policy to "kick in" at the same time the other components of aggregate demand are falling. In this case, we would see no leftward or rightward shift in the AD curve whatsoever. The stimulus of the macroeconomic policy might just precisely counteract the other shortfalls of aggregate demand.

In Figure 11-19, we see the result of contractionary monetary or fiscal policy to counteract a boom. Here, we see the economy either moving beyond the full employment level of real GDP, or at least threatening to do so. This boom effect is represented by Arrow A in the figure, a rightward shift of the AD curve, an increase in real GDP to $RGDP_2$ and an increase in the price level to CPI_2. If the Fed or the Congress and the President wait too long to implement their contractionary policy, they will have to induce a recession by a rightward shift in the AD curve from AD_2 back to AD_1. Arrow B represents this contractionary policy in Figure 11-19. If their timing is right, however, the boom itself, and the recession correction might be avoided. In that case, the economy remains at full employment with a stable price level, with perfectly calibrated and timed contractionary policy compensating for the upward pressure on aggregate demand caused by the impending boom.

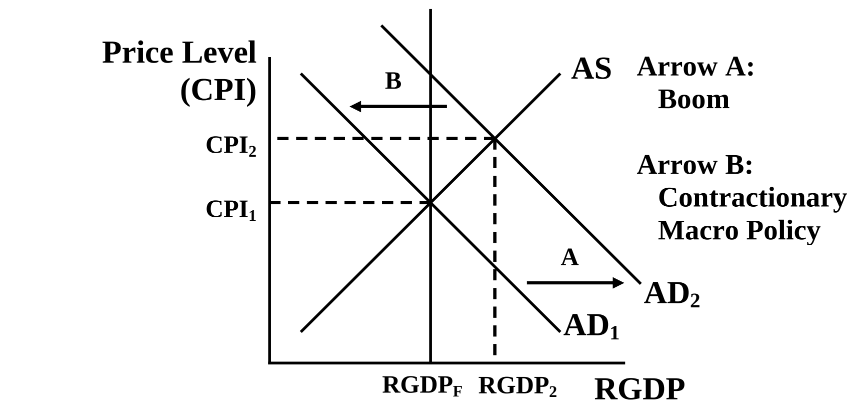


Figure 11-19
Contractionary Macroeconomic Policy to Combat a Boom

We have been emphasizing timing in the discussion of expansionary and contractionary policy in this section. We must not forget that the Fed and the government must also undertake the correct magnitude of the stimulus or restraint on aggregate demand. If either magnitude or timing is off, then macroeconomic policy is less stabilizing. If the timing and magnitude are very much wrong, macroeconomic policy might even be destabilizing. We'll examine this issue in the next chapter.

Supply Shocks and the ADAS Model

Every application of the ADAS model we've made to this point has involved a change in aggregate demand with aggregate supply remaining unchanged, i.e., we have shifted the AD curve, but have not shifted the AS curve.

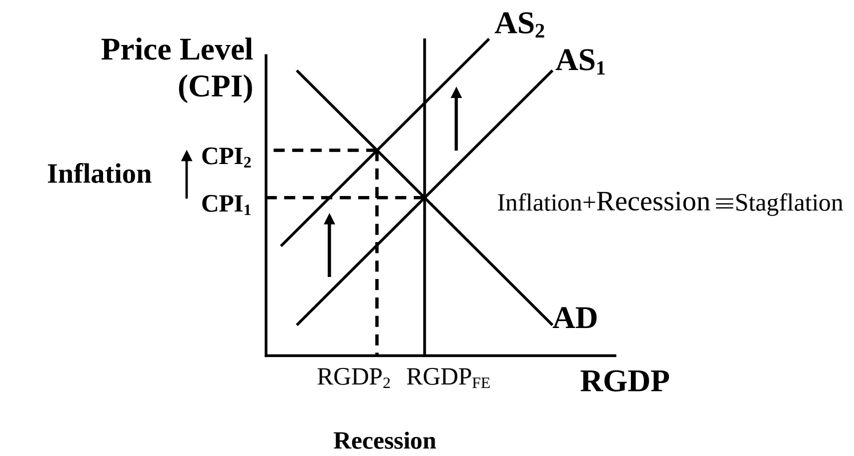


Figure 11-20
Stagflation from a Supply Shock

We now consider adverse supply shocks. Figure 11-20 shows the case of stagflation, the result of an adverse supply shock, such as a large increase in the price of oil.

Suppose the economy starts off at full employment real GDP and a price level indicated by CPI_1 in the model. Then suppose the price of oil rises. We represent this autonomous increase in an important resource price by a shift up of the AS curve. The economy enters a recession as real GDP falls to $RGDP_2$, and the price level rises to CPI_2. We have recession and inflation at the same time, which is often called stagflation.

Note that stagflation presents a difficult choice for macroeconomic policy. To fight the recession, the government or the Fed would have to undertake expansionary policy. This would increase real GDP, but it would exacerbate inflation by increasing the price level even more. If the policy makers choose to cool off the economy with contractionary policy, they can reduce the price level, but it makes the recession deeper.

The best example of the effect of a favorable supply shock is the case of noninflationary economic growth. Suppose that the economy starts out at $RGDP_1$ in Figure 11-21 and experiences an increase in aggregate demand. The AD curve shifts from AD_1 to AD_2. Without a favorable supply shock this expansion of the economy would result in an increase in the price level from CPI_1 to CPI_2. But suppose oil prices decline and labor productivity increases, which happened in the 1980s and the late 1990s. Now the AS curve shifts down to

AS_2 and real GDP expands further to $RGDP_2$*. We experience real economic growth and a stable price level. The macro economy doesn't get much better than this.

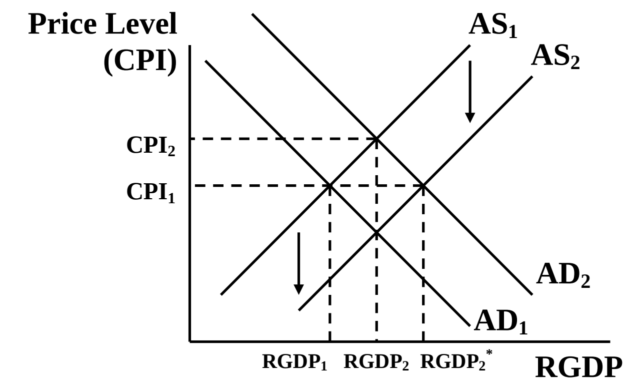


Figure 11-21
Favorable Supply Shocks and Noninflationary Economic Growth

The Self-Correcting Mechanism and the ADAS Model

Up to this point in this chapter we have shown the effect of macroeconomic policy used in order to stabilize the economy, by contracting the economy in the face of a boom and expanding it in the face of a recession. We now examine in the model whether the economy might correct itself in a boom and in a recession. We begin with the case of the boom.

In Figure 11-22, we start the economy off at full employment at $RGDP_{FE}$ and the price level CPI_1. The AD and AS curves intersect at point A in the model. Now suppose a boom ensues and neither the Fed nor the federal government makes any effort to counteract it. The AD curve shifts to AD_2 and intersects the AS curve at point B, at $RGDP_2$ and CPI_2. But this isn't the end of the story. As the economy experiences the boom and the initial inflation it is unlikely that workers and businesses will maintain the same price level expectations. If workers and businesses revise upward their price level expectations the AS curve will shift up, from AS_1 to AS_2, creating another round of inflation. This shift of the AS curve comes from the rise in wages in anticipation of more inflation, with businesses increasing prices to complete a self-fulfilling prophesy. If people expect an increase in inflation, that increase is likely to

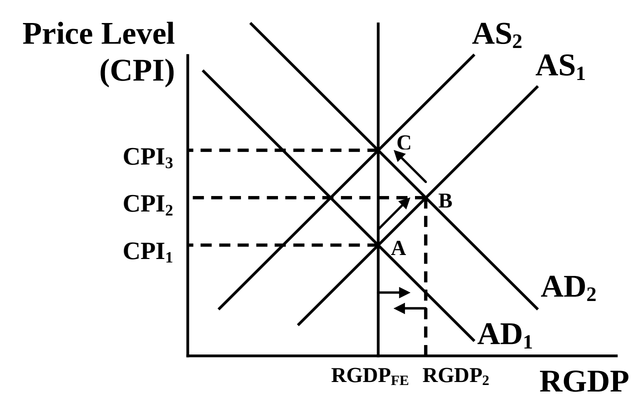


Figure 11-22
The Self-Correcting Mechanism and a Boom

occur. The movement from point B to C in the model also results in a boom-ending recession at the same time we are experiencing another increase in the price level. The unfortunate result of an unfettered boom is what makes especially the Fed so vigilant in keeping the economy from overheating. Better to nip the flower of a boom in the bud that have it eventually go to seed in a wage-price spiral and eventual recession.

In theory, a self-correcting mechanism exists in a recession as well. We show his in Figure 11-23. We again start at full employment where AS_1 intersects AD_1 at point A in the model. A downturn in aggregate demand shifts the AD curve from AD_1 to AD_2. A recession with deflation ensues, with real GDP falling to $RGDP_2$ and the price level falling to CPI_2.

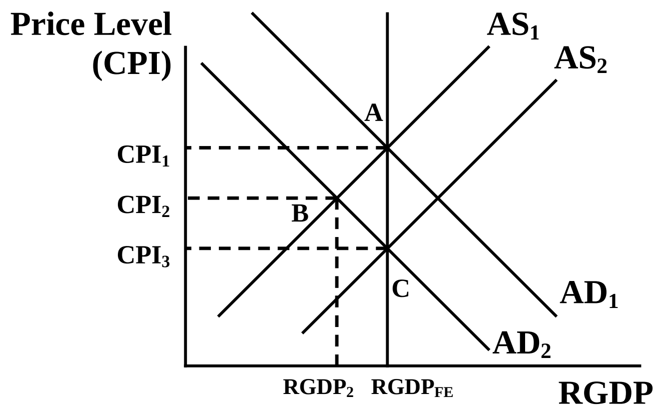


Figure 11-23
The Self-Correcting Mechanism and a Recession

But again, this is not the end of the story as people in the economy now revise their price level expectations downward. Wages fall accordingly, leading to a fall in production cost and a further round of deflation, which lowers money demand and interest rates and stimulates aggregate demand as the economy moves back to full employment.

One practical real world fact implies that the self-correcting mechanism probably won't work to counteract a recession. The self-correcting mechanism, in both the case of the boom and the case of the recession, requires wages to adjust to a change in price level expectations. This is not a problem in a boom, as wages adjust upward when people revise upward their expectations of the price level. But in a recession wages have to fall to make the self-correcting mechanism work. If workers resist reductions in their nominal wages the self-correcting mechanism will had a difficult time working.

Probably the main reason we don't see the self-correcting mechanism correcting recessions is because of political behavior. Macroeconomic policy to counteract recessions is "fun" to do. The Fed gets to stimulate the economy. They don't have to ruin the economic party buy taking away the punchbowl, or by taking the keg away from a kegger. Fighting recessions is also very popular fiscal policy. Members of Congress get to spend more on their constituents or lower their taxes, things elected officials like to do. So we almost never get to a point in modern economies where the self-correcting mechanism would even be allowed to operate. Expansionary fiscal and monetary policy is simply too popular.

Chapter 12: Politics and Macroeconomic Policy

Introduction

Macroeconomic policymakers work in the dirty air of politics, not the clean vacuum of a controlled economics experiment. What could be more political than fiscal policy, with its adjustment of government spending and taxes? On whom do we spend? Whom do we tax? Theodore White once wrote that the essence of American politics reflects three fundamental relationships, rich and poor, black and white, and war and peace. Certainly, these issues capture the debate on government spending and taxation.

As we will see in this chapter, government spending and taxing decisions allow not only expression of differing views on macroeconomic theory, but also an outlet for core political beliefs and philosophies. Even monetary policy, undertaken in the U.S. by an independent Fed, requires that the President appoint members of the Board of Governors and the Chairman of the Fed, of course with the advice and consent of the Senate. Politics reigns here as well.

We will range widely in our topics in this chapter. After roughly characterizing the political and macroeconomic policy views of liberals and conservatives, we'll examine the font of liberal activist fiscal policy itself, the work of British economist John Maynard Keynes. We will look at the interaction between politics and economics in the context of a potential tradeoff between inflation and unemployment. Because deficits and the federal debt are now back on the front pages of American newspapers, we'll explain the relationship between these two concepts. We'll examine issues such as supply-side economics, the balanced budget amendment, and the effectiveness of macroeconomic policy, and place these in a political and macroeconomic context. First, we turn to a (perhaps overly) simplified characterization of political views.

Political Philosophy and Macroeconomic Policy Instruments

What follows is a gross oversimplification, but a useful one. We divide the world into two political groupings, liberals and conservatives. We choose not to use the names of specific political parties, such as Democrats and Republicans, even though Democrats are more likely to be liberals and Republicans are more likely to be conservatives. We begin this political bifurcation in Figure 12-1 with a listing of views on some key general political issues.

Issue	Liberal	Conservative
Equity vs. Efficiency	Equity	Efficiency
Socialist-Capitalist Continuum	Left of Center	Right of Center
Market Capitalism	Flawed	Divinely Inspired
Attitude Toward Market Capitalist Economy	Needs Government Help	Government Should Leave Alone
Most Important Government Action	Redistribution	Enforce Property Rights
What People Need	Government Help	Freedom & Initiative
Preferred Government Level	Federal	State & Local

Figure 12-1
Some Liberal and Conservative Political Views

Liberals are first and foremost concerned about equity or fairness, especially in the distribution of income and wealth. By contrast, conservatives worry more about the size of the total economic pie, rather than the relative size of each piece of it. A conservative would think that all is well if everyone's piece of economic pie is getting bigger, even though some pieces are growing faster than others. A liberal worries about the relative size of each piece of the economic pie, regardless of the absolute size of the smallest piece.

In Chapter 1, we discussed the economic systems command socialism and market capitalism and noted that economies usually lie somewhere on a continuum between polar cases of each. Following convention, if we place command socialism at the left end of a continuum and market capitalism at the right end, political liberals are left of center and political conservatives are to the right.

The third distinction in Figure 12-1 relates to the second. Liberals see market capitalism as flawed. While "divinely inspired" might be a little strong in describing a conservative's view of market capitalism, it's close. From this view comes the next characteristic. If market capitalism is flawed, it needs government help to improve it and make it better. Because conservatives applaud the functioning of free markets, they think the government should leave the economy alone. Conservatives believe that government intervention in the economy will make things worse, not better.

Given their philosophies, liberals and conservatives differ on the most important action or activity of government. Liberals favor redistribution in the form of progressive taxation and transfer policies, while conservatives, who like market outcomes, seek to ensure that they occur through government enforcement of private property rights.

Liberals see people who are victims of market capitalism and believe that government should help them. Conservatives believe everyone can succeed in market capitalism if they are given freedom to choose their actions and use their own initiative. Finally, for government action in the economy, liberals would prefer federal government action, while conservatives would prefer to rely on state and local government.

Now that we have laid out the broad general outline of liberal and conservative politics, we can proceed to describing political views on macroeconomic policy. These are summarized in Figure 12-2. Liberals view macroeconomic instability as one of the important flaws of market capitalism. In this view, the economy is prone to booms and busts that can last for long periods of time without active, discretionary macroeconomic policy. Liberals are activists in macroeconomic policy. Consistent with their admiration of market capitalism, conservatives think the economy operates with a self-correcting mechanism. They believe that more often than not macroeconomic policy creates more instability in the economy than it removes, and would argue for constraints on policy, not active discretion in using it.

Characteristic	Liberals	Conservatives
General Philosophy	Activist (encourage discretionary policy)	Nonactivist (constrain policy makers)
Preferred Fiscal Policy Instrument	Government Spending ↑ Taxes on Rich for Restraint	None
Preferred Change in Government Spending	↑ Domestic Social Spending ↓ Defense Spending	↑ Defense Spending ↓ Domestic Spending
Preferred Change in Taxes	↑ Taxes on Rich ↓ Taxes on Middle Class	↓ Marginal Tax Rates
Preferred Member of Fed Board of Governors	Reduce Unemployment	Maintain Price Stability

Figure 12-2
Politics and Macroeconomic Policy

Because they believe that fiscal policy does more harm than good, conservatives have no preferred fiscal policy instrument. Liberals, on the other hand, prefer government spending to help people when the economy needs stimulus, and increased taxes on the rich if the economy requires restraint. Furthermore, we see in the third line of Figure 12-2 that if government spending needs to rise, liberals would wish to increase domestic social spending. If we need to reduce government spending, cuts in defense expenditures are warranted. While conservatives are generally opposed to changing government spending for fiscal policy reasons, if pressed, they would want to increase defense spending and decrease domestic social spending.

On tax policy, liberals and conservatives differ greatly. For expansionary fiscal policy, liberals would favor reducing taxes on the middle class, and for contractionary policy they would advocate increasing taxes on the rich. On the other hand, conservatives want to lower taxes as a matter of principle, no mater where we are in the business cycle, as long as we do it by lowering marginal tax rates at all income levels.

Finally, in choosing members of the board of governors of the Fed, or in just constructing a general macroeconomic policy, liberals would favor policy makers who focused on reducing unemployment, even if it means an increase in inflation. By contrast, conservatives view price stability as the overriding macroeconomic policy goal, believing that if the Fed maintains the purchasing power of money, employment and unemployment in the economy will take care of themselves.

This is quite a difference of opinion on politics and macroeconomics. We will see later in the chapter that just this difference in political philosophy among elected officials in the U.S. government can make it difficult to undertake fiscal policy. But before we get to this and other issues, we'll explore the origins of macroeconomics and macroeconomic policy itself in the work of British economist John Maynard Keynes.

Keynesian Macroeconomics and Macroeconomic Policy

Many would argue that John Maynard Keynes (rhymes with gains) was the most influential economist of the 20th century. Before Keynes, macroeconomics as we know it did not exist. Keynes' most influential work was a book entitled *The General Theory of Employment, Interest and Money*. The book appeared in 1936, when people in economies around the world could vividly remember the hardships of the Great Contraction from 1929 to 1933, and the lingering high rate of unemployment throughout the 1930s.

Keynes had a radical view at the time, that the capitalist economy was prone to long periods of low or negative growth and high unemployment. According to Keynes, government had an obligation to stimulate the depressed economy through expansionary fiscal policy. Keynes was a liberal, a socialist, and how one views Keynes' economics and policy often stem from one's political philosophy. In southern Idaho, a bastion of conservative politics, if you stand on a street corner and shout "Keynes," rotten potatoes will fly in your direction.

Along with his policy prescriptions, Keynes offered a model of the economy that emphasized fluctuations in what we now call aggregate demand. We show a simplified version of a Keynesian macroeconomic model in Figure 12-3. The left-hand column includes definitions of many symbols we have seen previously. DI is the only new one, and this stands for disposable income, income after taxes. The right column of Figure 12-3 contains nine equations or manipulations of them. Equation (1) is our now familiar definition of aggregate demand, the sum of consumption, investment and government spending plus net exports. We need go no further than this equation to introduce the essence of Keynesian economics. Suppose C, I, and NX fall, as they did in the Great Contraction from 1929 to 1933. This would lower aggregate demand and cause a recession. But if G increased enough, it could counteract the reduction in aggregate demand. This is fiscal policy using government spending as the policy instrument.

<u>Definitions</u>

C	consumption
I	investment
G	government spending
NX	net exports
DI	disposable income
Y	national income or GDP
T	taxes
AD	aggregate demand
a	autonomous spending
b	marginal propensity to consume

<u>Algebraic Manipulations of Equations</u>

(1) $AD \equiv C+I+G+NX$

(2) $AD=Y$

(3) $Y=C+I+G+NX$

(4) $DI \equiv Y-T$

(5) $C=a+bDI$

(6) $C=a+b(Y-T)$

(7) $Y=a=b(Y-T)+I+G+NX$

(8) $Y-bY=a-bT+I+G+NX$

(9) $(1-b)Y=a-bT+I+G+NX$

$$(10)\ Y=\underbrace{\left(\frac{1}{1-b}\right)}_{\text{Multiplier}}[a-bT+I+G+NX]$$

Figure 12-3
A Simple Keynesian Macroeconomic Model

Equation (2) is simply one that forces macroeconomic equilibrium in the model, where aggregate demand equals national income or output. Or in terms of our more elaborate model from the last chapter, aggregate demand equals aggregate supply. We can now substitute the Y from equation (2) into equation (1) to obtain equation (3). We define disposable income formally in equation (4). Equation (5) is a simple linear consumption function. Consumption is a function of (depends on) an autonomous component, a, plus a constant, b, multiplied by disposable income. We call b the marginal propensity to consume, the additional consumption generated by additional disposable income. We assume b is positive, but less than 1.

Equation (6) comes from substituting Y-T from equation (4) for DI in equation (5). Now we perform our last substitution by inserting the consumption function of equation (6) into equation (3). We now have equation (7), which has Y terms and I, G, T and NX. Note that we have not made these latter terms a function of disposable income. In modeling, we call these variables exogenous. They depend on factors other than income in the economy, factors we hold constant until we wish to change them.

In equation (8), we collect all the Y terms on the left side of the equation, and, in equation (9), we factor out the Y term. By dividing both sides of equation (9) by (1-b), we get equation (10) at the bottom of the figure. The fraction 1/(1-b) is greater than 1, because 1-b is less than one. This means that any change in the exogenous or autonomous variables inside the brackets on the right hand side of equation (10) causes a larger change in equilibrium income and output in the economy. The dynamic implications of the model are summarized in Figure 12-4. A change in the any of the exogenous or endogenous components of the model will be magnified by the multiplier to create changes in income and output.

$$\Delta Y = \left(\frac{1}{1 - b}\right) \quad \text{times} \quad \begin{pmatrix} \Delta a \\ -b\,\Delta T \\ \Delta I \\ \Delta G \\ \Delta NX \end{pmatrix}$$

Figure 12-4
Multipliers from a Simple Keynesian Model

Note that the sign of the tax effect is negative. Of course, this is because an increase in taxes withdraws spending from the economy. The tax multiplier is dampened by the b term, however, as some of the increase in taxes would not have been spent anyway. Don’t worry about understanding this last tricky concept now. It’s interesting because it influences something called the balanced budget multiplier, but this is beyond out scope here. The important thing to recognize is that G and T, our fiscal policy variables, can have an effect opposite of the others, and changing G and T in this regard is discretionary fiscal policy.

Ironically, in the real world, you likely will be exposed to the Keynesian multiplier effect not in discussions of macroeconomic policy but in regard to local economic development policy. Imagine the economy of a town or region. We could construct a similar model for this economy. Here, net exports become much more important in the economy than they are in the national economy. Local economic boosters usually support enticing to the region new businesses that export from the region. This is a boost in exogenous spending in the economy, and it will have a multiplier effect on income in the area. Economists in the economics subdiscipline regional economics develop and employ elaborate regional economic models that are used to calculate these multiplier effects. You are much more likely to encounter this aspect of Keynesian economics in discussions with the local Chamber of Commerce than you are in discussions with the Congress of the United States.

The Phillips Curve

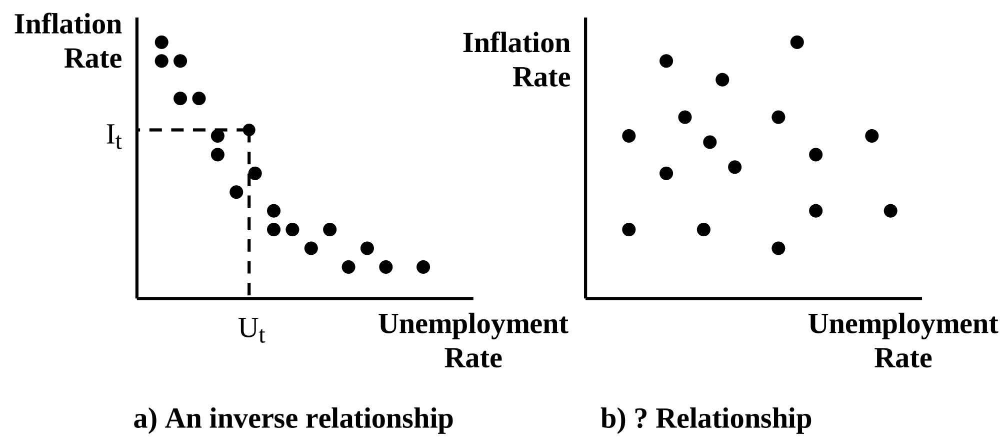


Figure 12-5
Contrasting Scatter Diagrams

In 1960, following the pioneering work of British economist A.W. Phillips, U.S. economists and later Nobel Laureates Paul Samuelson and Robert Solow discovered a relationship between an economy's rate of inflation and its unemployment rate. This relationship came to be known as the Phillips curve. Samuelson and Solow did this by analyzing scatter diagrams of points such as those in the two panels of Figure 12-5. Let the dots in these graphs represent the annual values of both inflation and unemployment, one on each axis of the graph, e.g., I_t and U_t in panel a) of the figure, where t identifies some year.

If we had to infer a relationship between the annual rate of inflation and the unemployment rate in panel a), we would conclude that an inverse relationship exists. Historically, it looks like high rates of inflation are associated with low rates of unemployment, and vice versa. If we look at panel b) of Figure 12-5, we would infer that no clear relationship exists. In fact, panel a) is representative of the relationship in the economy from 1950-1970, while panel b) looks more like the data for 1970 – 2000. And here lies the story of the Phillips curve.

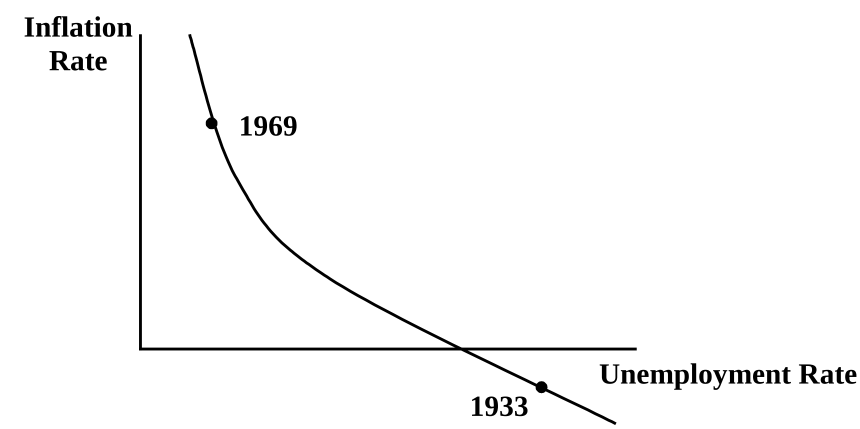


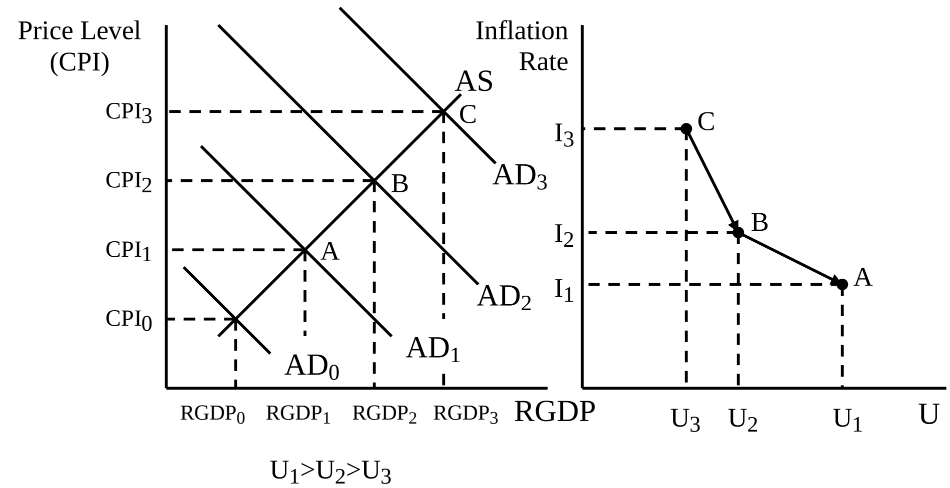
Figure 12-6
The Traditional Phillips Curve

We show a stylized Phillips curve in Figure 12-6. As the unemployment rate increases, we see lower rates of inflation in the economy. The Phillips curve eventually crosses the unemployment axis and inflation turns to deflation. We've labeled one point 1969 and the other 1933 to suggest just two combinations of inflation and unemployment on a Phillips curve. The Phillips curve is merely an empirical regularity discovered with statistical analysis of data. But why does this relationship exist? We address this question in the next section.

The Phillips Curve and Changes in Aggregate Demand

The macroeconomic theory of the Phillips curve is easily explained with our model of the last chapter, the ADAS model. Consider the ADAS model in panel a) of Figure 12-7. Note the four levels of real output shown on the horizontal axis and the three levels of the unemployment rate corresponding to three of them. $RGDP_1$ results in an unemployment rate of U_1. $RGDP_2$ results in an unemployment rate of U_2, and so on. Of course, as RGDP rises the unemployment rate falls, so $U_1 > U_2 > U_3$, as shown in the figure. Note also the four levels of the CPI appearing on the price level axis. We can infer rates of inflation on the axis only by comparing changes in the CPI from a given period. For example, start the economy with the CPI equal to CPI_0, with aggregate demand represented by AD_0. We can conclude that more inflation would occur if the price level rose to CPI_2 instead of CPI_1, and even more inflation still from CPI_0 to CPI_3.

If we match the amount of inflation to get to points A, B and C in the ADAS model in panel a) with the corresponding unemployment rates occurring at the different equilibria, we obtain something like panel b) of the figure. If the price level moves from CPI_0 to CPI_3, we have an inflation rate in panel b) we call I_3. The large increase in aggregate demand, from AD_0 to AD_3 also generated enough real output and related employment to reduce the unemployment rate to U_3. But if the increase in aggregate demand is only from AD_0 to AD_2, we have lower inflation, I_2 in panel b), and a higher unemployment rate generated by the real output level $RGDP_2$. This is very much a Phillips curve relationship. As long as fluctuations in the economy come from fluctuations in aggregate demand, we can expect to observe a normal looking Phillips curve in the inflation and unemployment data.



a) ADAS model with AD changes

b) A Phillips Curve Relationship

Figure 12-7
The ADAS Model and the Traditional Phillips Curve

Supply Shocks, Inflation Expectations and the Vanishing Phillips Curve

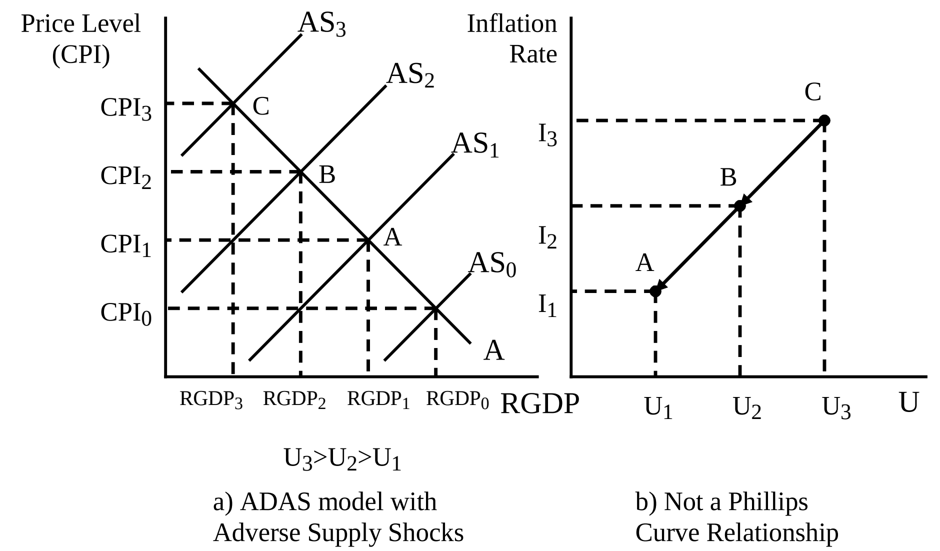


Figure 12-8
Supply Shocks and the Vanishing Phillips Curve

Fluctuations in aggregate demand can generate a normal Phillips curve. But what if fluctuations in the economy come from the supply side, through the effects of supply shocks. This event is pictured in panel a) of Figure 12-8. By arguments similar to those we just made in the case of changes in aggregate demand, we now discover that higher inflation rates and higher unemployment rates can occur together. For example, an inflation rate I_3, associated with an increase in the price level from CPI_0 to CPI_3, would result in an unemployment rate U_3. A lower inflation rate I_2, associated with an increase in the CPI only to CPI_2, would result in an unemployment rate of U_2. This is not a normal Phillips curve relationship, as higher inflation is associated with higher unemployment. The supply shocks of the 1970s and constant revision upward of price level expectations caused the well-behaved Phillips curve of the earlier part of the century to vanish in the 1970s.

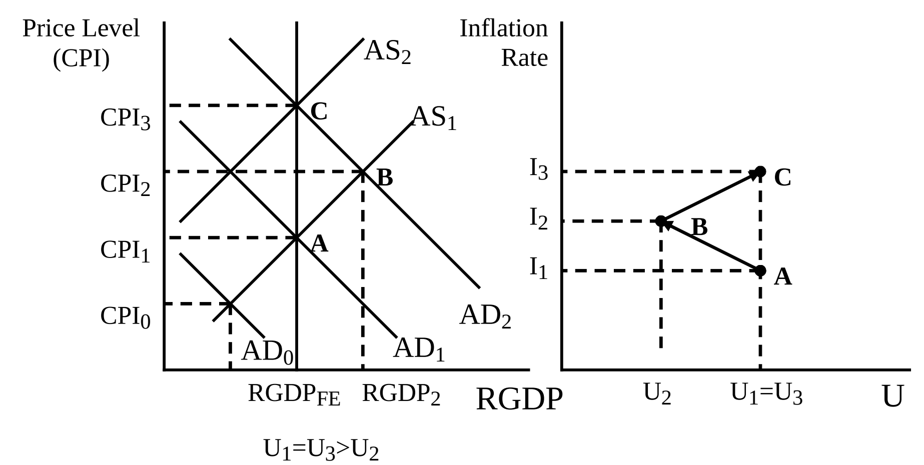


Figure 12-9
Booms, the Self-Correcting Mechanism and the Breakdown of the Traditional Phillips Curve

We can understand another factor in the vanishing Phillips curve by recalling our theory of the self-correcting mechanism in a boom. We show this in Figure 12-9. As aggregate demand increases and the economy expands beyond full employment to $RGDP_2$, the rate of inflation increases to I_2 from I_1. This is a normal Phillips curve relationship, as always is the case from changes in aggregate demand. But the price level and real output combination where the AD curve intersects the AS curve at point B is unsustainable. Workers and businesses will revise upward their price level expectations and the AS curve will shift up to AS_2. This is more inflation and a higher unemployment rate, another abnormal Phillips curve relationship.

The self-correcting mechanism in a boom also destroys the Phillips curve as a menu of policy choice. Before the Phillips curve vanished, policy makers considered it a set of choices. Liberals would choose lower unemployment, even if it meant higher inflation. Conservatives would strive for lower inflation, even if it meant higher unemployment. But now, economic theory suggests that supply shocks and changes in price level expectations can make the choice more difficult, and the self-correcting mechanism in a boom can make the tradeoff between inflation and very low unemployment impossible.

The Federal Budget Deficit (Surplus) and the Federal Debt

When you spend more than you earn, when your outgo is more than your income, you have to borrow to finance your deficit. When you spend less than you earn, when your outgo is less than your income, you have a surplus, and you can either save or pay off previous debt. The federal government operates the same way. We derive the relationship between the federal budget deficit and the debt from the federal budget constraint in Figure 12-10.

$$\underbrace{\text{Taxes}+\text{Borrowing}}_{\text{Sources of Funds}}=\underbrace{\text{Government Spending}+\text{Transfers}+\text{Interest on Debt}}_{\text{Uses of Funds}}$$

$$\underbrace{\text{Borrowing}}_{\text{New Debt}}=\underbrace{\text{Government Spending}+\text{Transfers}+\text{Interest on Debt}-\text{Taxes}}_{\text{Deficit (+) or Surplus (–)}}$$

Figure 12-10
The Federal Government Budget Constraint and the Deficit (Surplus) and the Debt

The top line of Figure 12-10 shows the government's budget constraint. The sources of funds must equal the uses of funds. We find two sources of funds on the left-hand side of the equation, taxes and borrowing. Taxes involve coercive public finance, while borrowing is voluntary. No one is required to lend to the government, but everyone must pay his or her taxes. For simplicity, we ignore user fees, asset sales and other minor sources of government funds. On the right hand side of the budget constraint, we find the government's uses of funds. The government spends on goods and services, it transfers funds from one American to another, and it pays interest on its debt.

In the second line of Figure 12-10, we simply subtract taxes from both sides of the equation. When the uses of funds exceed tax revenues, the right hand side shows a deficit. If taxes exceed spending, transfers and interest, the deficit is negative. A negative deficit is called a surplus. Borrowing remains on the left-hand side of the equation. If the deficit is positive, the government must issue new U.S. Treasury debt to cover the shortfall. Government deficits must be financed, and when they are, debt results.

History and Politics of the Federal Budget Deficit (Surplus) and the Debt

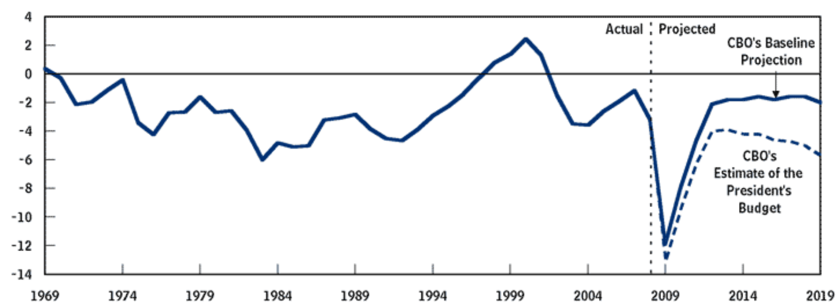


Figure 12-11
Deficit (surplus) as a Percent of GDP, 1969-2019

We see a 40 year history of the U.S. federal budget deficit in Figure 12-11. To give a better picture of the relative magnitude of the deficit (surplus), we show it as a percent of GDP. The budget had a slight surplus in 1969, the first one since the late 1950s. The red ink trended upward (more negative budget balance in Figure 12-11) through the recessions of the 1970s and then rose dramatically in the early 1980s during the Reagan administration. The Reagan tax cuts and defense buildup led to a deficit of nearly 6% of GDP in 1983. The long economic expansion in the 1980s reduced the deficit to just less than 3% of GDP by the end of the decade. Slowing economic growth and eventually the 1992 recession increased the deficit as a percent of GDP in the early 1990s. The long expansion of the middle and late 1990s, however, caused a steady decline in the deficit as a percent of GDP, so much so that we had rare surpluses from 1998 through 2001. In the next section, we will explain this phenomenon through the relationship of tax revenue to tax rates and tax bases, but here we will focus on the political debate engendered by these years of surplus.

The surpluses of the second Clinton administration led to interesting political debates in the 2000 presidential campaign about what to do about them. True to his conservative label at the time, Governor Bush emphasized an across the board cut in marginal tax rates. Vice President Gore, a liberal, favored a smaller tax cut targeted at middle income taxpayers. Governor Bush advocated some increases in military spending, while Vice President Gore suggested an increase in transfer payments, especially a prescription drug plan for people on social security. Both candidates mentioned reduction of the federal debt, as well. The views in this debate came right out of the components of the federal budget constraint in Figure 12-10 and our characterization of liberals and conservatives in Figures 12-1 and 12-2.

After the 2000 election, a version of Mr. Bush's tax cuts was implemented and combined with the 2001 recession and spending pursuant to the September 11, 2001, terrorist attacks, and eventually the Iraq war, the deficit rose again as a percent of GDP. Before the onslaught of the most recent recession and financial crisis, starting in December 2007, the deficit as a percent of GDP was trending downward again, registering a (modern day) modest 1.2% of GDP in fiscal year 2007.

The recession was already wreaking havoc on the federal budget balance when the financial crisis hit, Barak Obama became president, and the Democrats achieved large majorities in both houses of congress. A combination of general expansionary fiscal policy, a deeper recession, financial system rescue efforts, and a very ambitious program of government activism have led to the prospect of historic deficits as a percent of GDP. Indeed, the Congressional Budget Office (CBO) projected in March 2009 that the 2009 budget deficit would be 13.1% of GDP. These are uncharted waters for deficit levels since the end of World War II. As we see in the rightmost portion of Figure 12-11, CBO projects deficit reduction in future years, but that is the end of the good news. CBO sees the deficit averaging in the mid 5% range from 2010 to 2019. Never in the last 60 years have deficits of this magnitude persisted for so long. The dotted line in Figure 12-11 shows CBO's more dire projection of the deficit as a percent of GDP if President Obama's budget were to be implemented. Few economists or normal persons of any political persuasion would find these projections encouraging.

It's not surprising that a representative democracy might tend toward chronic budget deficits. Politicians gain support by doing things for their constituents. Usually, doing things requires spending. No electoral candidate wants to run on a platform of lower spending and higher taxes. Of course, we should not be too quick to blame the politicians in this regard, for we the people elect them, and they try to give us what we want, government programs that benefit us and taxes on someone else to pay for them.

The Relationship Between Tax Rates and Tax Bases

In the last section, we noted that economic growth in the 1990s helped generate surpluses that first occurred in 1998. But economic growth is defined as growth in real GDP, and real GDP growth is almost the same thing as real income growth. Because much of the federal government's revenue comes from income and payroll taxes, growth in the economy means growth in tax revenue, even if income tax rates remain unchanged. This is a specific example of a relationship between tax revenue and tax bases. We generalize this relationship in Figure 12-12.

Tax Revenue = Tax Base X Tax Rate

Income Tax Revenue = Taxable Income X Income Tax Rate

$+\Delta$ Income Tax Rate $\rightarrow$ $-\Delta$ Taxable Income
$-\Delta$ Income Tax Rate $\rightarrow$ $+\Delta$ Taxable Income

Assuming that the tax base remains the same, if the tax rate changes, will lead to bad revenue forecasts.

Figure 12-12
Tax Revenue, Tax Rates, and Tax Bases

As seen in the top line of the figure, tax revenue comes from a tax rate multiplied by a tax base. For example, sales tax revenue is the product of the sales tax rate and the sales tax base, taxable sales. Property tax revenue is the product of the property tax rate and the value of taxable property. We focus on the income tax in the second line of the Figure 12-12. Income tax revenue is the product of taxable income, the tax base in this case, and an income tax rate.

Suppose we wished to estimate the reduction in income tax revenue that would result from a reduction in the income tax rate. We might be tempted to multiply the existing tax base times the new lower tax rate. Economists would argue that this would surely overestimate the reduction in tax revenue. When we change a tax rate, its tax base does not remain the same, because a tax rate is an economic incentive that affects behavior. The third line of the figure summarizes this point. A reduction in a tax rate, especially a marginal tax rate, will lead to an increase in taxable income. An increase in the tax rate will have the opposite effect, a decrease in taxable income.

Lets look at an example. A marginal income tax rate indicates how much of an additional dollar of income will go to the government, and how much the generator of that additional income can keep. If marginal tax rates fall, the incentive to earn income increases, and people and businesses respond by earning more. A lower tax rate leads to less tax evasion, as well. When tax rates rise, we see less income generated and more tax evasion.

While an inverse relationship exists between tax rates and tax bases, the effect of changes in tax rates on the direction of the change in tax revenue remains unclear. Does the effect of a lower rate outweigh the effect of a larger tax base it creates? If so, tax revenue will fall. But what if the tax base responds greatly to a reduction in the tax rate? It is theoretically possible that the expansion of the tax base can outweigh the reduction of the tax rate, such that tax revenue actually increases. This is the issue of supply side economics we address in the next section.

Economics, Politics, and Supply-Side Economics

In the early 1980s, an economist by the name of Arthur Laffer popularized the relationship between tax rates and tax bases with a model that came to be called the Laffer curve. We present a version of this model in Figure 12-13. In the Laffer curve model, we have the marginal tax rate on the horizontal axis and tax revenue on the vertical axis. Let's focus first on the extreme ends of the Laffer curve. Obviously, with a 0% marginal tax rate no tax revenue will be generated, because the tax rate is zero. Also, when the tax rate is 100%, no tax revenue will be generated, because at a 100% tax rate

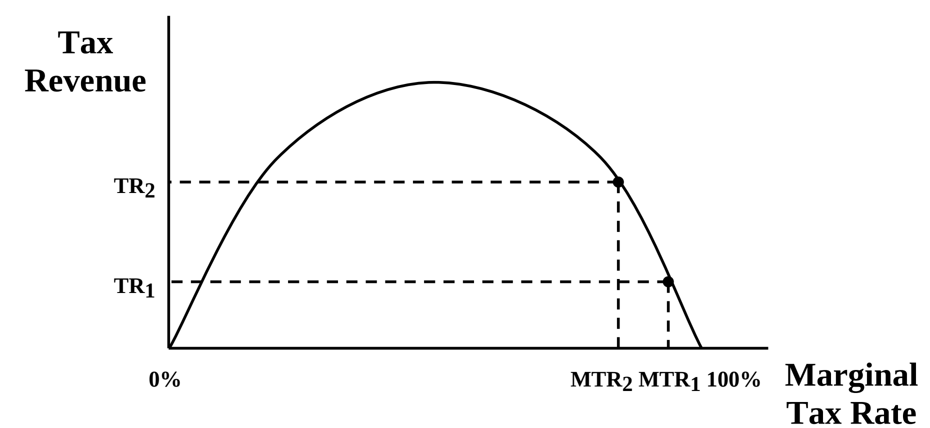


Figure 12-13
Supply-Side Economics and the Laffer Curve

the tax base will be zero. To see this point, just ask yourself if you would work if the government took 100% of your paycheck. Not likely. Because tax revenue is positive between tax rate of 0% and 100%, the Laffer curve must first rise and then fall as the marginal tax rate increases.

Now consider a cut in marginal tax rates. If initial marginal tax rates are high, such as MTR_1 in the model, tax revenue will rise from TR_1 to TR_2 when the tax rate is lowered to MTR_2. If marginal tax rates are low enough, however, so that we are on the rising portion of the Laffer curve, a cut in marginal tax rates will lower tax revenue.

Much political debate about tax policy in the United States occurs between "Supply-Siders," those who believe we are on the declining portion of the Laffer curve, and "Deficit Hawks," who believe we are on the rising portion. Some politicians, such as 1996 presidential candidate Robert Dole, even change their views as new political and economic information becomes available. Senator Dole, a longtime adamant Deficit Hawk became a Supply-Sider in the course of the 1996 campaign. We would stray too far afield here to attempt to conclude whether this conversion was politically or economically motivated.

Should We Put Constraints on Fiscal Policy?

Deficits year in and year out in the federal budget lead many conservatives to conclude that fiscal policy works only on the stimulus side of the equation. Even Keynes himself suggested that restraint was necessary in an overheated economy. This view suggests that American fiscal policy is like a car with an accelerator but no brakes, not a good car to drive. Conservative fiscal policy critics also suggest that, in addition to the bias toward stimulus, fiscal policy will likely result in the wrong action at the wrong time. The difficulty of correctly timing fiscal policy derives from the lags involved in applying it. We show these lags in Figure 12-14.

Lag	Reason for Lag
Recognition Lag	Takes time to recognize where the economy is in the business cycle
Implementation Lag	Takes time for Congress and the President to develop spending and taxation legislation
Effect Lag	Takes time for fiscal policy to have an effect in the economy

Figure 12-14
Lags in Fiscal Policy

Both fiscal and monetary policy suffer from the recognition lag. We never really know at any point in time just exactly how the economy is doing. It takes a few months, at best, to gather all the economic information. In fact, because of this lag, economic forecasters have to forecast the present level of economic variable, not just the future. Forecasts are inaccurate by nature.

The implementation lag is especially burdensome in fiscal policy. Political wrangling between liberals and conservatives, and between proponents of one program or another, can delay fiscal policy implementation. Monetary policy has a much shorter implementation lag. Finally, once a policy is put in place, it takes time for that policy to have an effect on the economy. The sum of these lags can lead to a worst-case fiscal policy. If fiscal stimulus occurs in an expansion and fiscal restraint occurs in a contraction, we accentuate the business cycle rather than dampening it.

A best and a worst-case scenario for fiscal policy timing appear in Figure 12-15. In panel a), we see the result of a well-timed fiscal policy that succeeds in damping the business cycle. In panel b), we see badly timed fiscal policy actually exacerbating the business cycle. In this latter case, fiscal policy is actually destabilizing, rather than stabilizing.

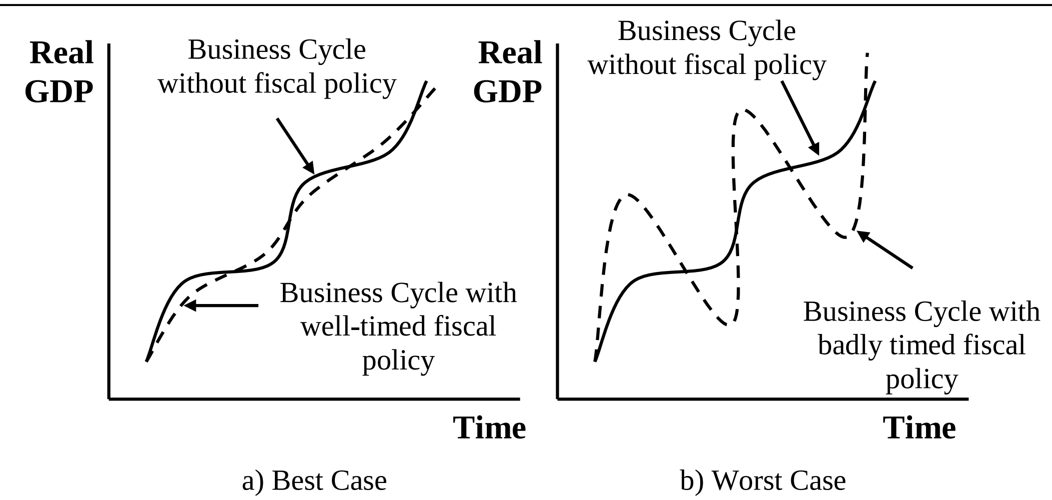


Figure 12-15
Best and Worst-Case Scenarios for Fiscal Policy

A bias toward fiscal stimulus in representative democracy and the difficulty in timing it have lead many conservatives to advocate a balanced budget amendment (BBA) to the U.S. Constitution, requiring a balanced budget at all times unless the Congress votes a deficit with a super majority vote or declares war. From a macroeconomic standpoint, a BBA has one major problem. It inactivates automatic stabilizers that work in the economy.

We have already seen that the tax base falls when taxable income falls. In a recession, taxable income falls, and even without changing tax rates tax revenue falls automatically. Likewise, as the economy heads into a downturn government programs to help the unemployed kick in to increase transfers and increase the deficit. We show this effect in Figure 12-16. As the economy expands toward a peak, the deficit falls. As the economy contracts toward a trough, the deficit increases. These effects happen automatically, without any action on the part of the President of the Congress. With a BBA, the President and the Congress would have to raise taxes and decrease government spending and transfers when an economy starts into a downturn. This is destabilizing fiscal policy.

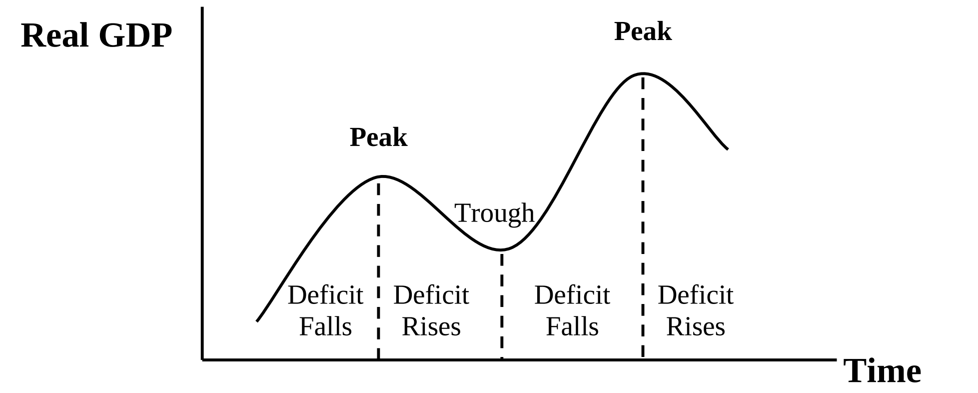


Figure 12-16
Automatic Stabilizers, the Deficit and the Business Cycle

Index

E

F

G

H

I

K

L

M

N

T

U

V

W

Z