EIILM UNIVERSITY
SIKKIM

BUSINESS ECONOMICS

Subject: BUSINESS ECONOMICS **Units: 4**

SYLLABUS

Introduction to Business Economics, Scarcity, Choice and Efficiency, Demand Analysis, Supply Analysis, Equilibrium of Supply and Demand, Elasticity of Supply and Demand

Utility Approach, Preference Approach in the context of consumer choice, Indifference Curve, Price Effect and Consumer Surplus

Theory of Firm, Firm and its Production, Law of Variable Proportions, Cost Analysis, Economies of Scale

Perfect Competition, Monopoly, Imperfect Competition, Monopolistic Competition, The Role of Government in the Market Economy, Pricing Strategy, Acquiring and Using Market Powers, Natural Monopoly, Profit Maximization and the Perfectly Competitive Firm

Demand and Supply in Factor Market, Labour, Land, Capital and Natural Resources, Income and Wealth

Suggested Readings:

1. Karmel, P.H. & Polasek, M.: Applied statistics for economists

2. Spiegel, M.R.: Theory & Problems of Statistics, Schaum's outline series, McGraw Hill ub.Co.

3. Spiegel, M.R.: Probability and Statistics

4. Freund: Mathematical Statistics

COURSE OVERVIEW

Economics has proved itself as a basic discipline; its application has a wide range. New areas of application are continuously being discovered where the principles, tools and techniques of economics are being used as the logic of reasoning in business. Without the knowledge and understanding of economics, no business, government, nation, any international body or for that matter any organization, including the NGOs, can function in today's world.

In other I can say that there's a need for basic training in economics followed by application in evaluating the rationality and optimality of business decisions taken by any agent. Hence, you being the student of BBA need this training. Therefore, this subject "Business Economics" is included in your curriculum.

The students on completion of the course shall develop the following skills and competencies:

1. Examine concepts and analysis of **basic business economics**
2. Investigate the behaviour of economic agents in **the consumer markets** in the context of market relationships.
3. Understand the various aspects of the **business organization** in the light of the managerial decision-making.
4. Develop a critical awareness of the various **market structures** aimed at tackling price and output determination.

BUSINESS ECONOMICS

CONTENT

BUSINESS ECONOMICS-I

CONTENT

BASIC BUSINESS ECONOMICS

Economics has been recognized as a special area of study for over a century. Virtually all four-year colleges offer courses in economics and most allow students to major in the subject. Economists maintain high profiles in governments, and they have been well-represented among the highest appointees in the federal government of the United States. The press reports on their doings and sayings, sometimes with praise and admiration, sometimes with ridicule and scorn. Economics and economists are words that almost everyone has heard of and uses. But what exactly is economics? Very few people can give a good definition or description of what this field of study is all about.

If ordinary citizens cannot give a good definition or description of economics, they can be excused because economists long struggled to define their field. In addition, in recent years, the subject matter that economists have studied has expanded, making its boundaries less defined. In recent years, for example, economic journals have published papers on topics such as sex, crime, slavery, childbearing, and rats. It is not surprising, then, that one economist, in a lighter moment, suggested that economics can be defined as "what economists do."

Defining economics as "what economists do" does not tell us anything we did not already know. A good definition must explain what it is that makes economics a distinct subject, different from physics or psychology. One should not expect to find a short definition that conveys with absolute clarity all there is to know about economics (or else there would be no reason to spend hours learning about it). Neither should one believe that there is only one correct definition possible. Many good definitions are possible, and each will focus on some important aspect of the subject. To use an analogy, there is not one spot from which one can best view Niagara Falls. Each viewpoint obscures some features and emphasizes others. There are, of course, some spots that are clearly superior to others, but people can disagree about which is the very best spot.

This group of readings examines definitions of economics and explains what those definitions mean.

After you finish this chapter, you should be able to:

- Explain what the term "invisible hand" means and who first used it.
- Give at least one common definition of economics.
- State what Malthus thought about population growth.
- Explain how Popper defines scientific statements.
- Distinguish between positive and normative statements.
- Explain what scarcity, choice, and self-interest have to do with economics.

LESSON 1:
INTRODUCTION TO BUSINESS ECONOMICS

As the title suggests, the purpose of this lesson is to help you understand what economics is about and what you can hope to learn by undertaking a study of economics. Since you will learn what economics is about as you progress through the course, you might wonder if this lesson is really important. The degree of success that you will experience in your study of economics will be determined, to a large extent, by your motivation and your intent to learn. People generally learn more when their study is being carried out with a particular intent. In the first half of this lesson you will read about various economic problems and ideas. You will probably find at least some of the problems and ideas to be interesting. You may then study economics with an intent to acquire knowledge and reasoning abilities that will help you to understand these ideas and solve these problems.

1.1 What is economics ?

One of the earliest and most famous definitions of economics was that of Thomas Carlyle, who in the early 19th century termed it the "dismal science." What Carlyle had noticed was the anti-utopian implications of economics. Many utopians, people who believe that a society of abundance without conflict is possible, believe that good results come from good motives and good motives lead to good results. Economists have always disputed this, and it was the forceful statement of this disagreement by early economists such as Thomas Malthus and David Ricardo that Carlyle reacted to.

Another early definition, one which is perhaps more useful, is that of English economist W. Stanley Jevons who, in the late 19th century, wrote that economics was "the mechanics of utility and self interest." One can think of economics as the social science that explores the results of people acting on the basis of self-interest. There is more to man than self-interest, and the other social sciences—such as psychology, sociology, anthropology, and political science—attempt to tell us about those other dimensions of man. As you read further into these pages, you will see that the assumption of self-interest, that a person tries to do the best for himself with what he has, underlies virtually all of economic theory.

At the turn of the century, Alfred Marshall's Principles of Economics was the most influential textbook in economics. Marshall defined economics as **"a study of mankind in the ordinary business of life; it examines that part of individual and social action which is most closely connected with the attainment and with the use of the material requisites of wellbeing. Thus it is on one side a study of wealth; and on the other, and more important side, a part of the study of man."**

Many other books of the period included in their definitions something about the **"study of exchange and production."** Definitions of this sort emphasize that the topics with which economics is most closely identified concern those processes involved in meeting man's material needs. Economists today do not use these definitions because the boundaries of economics have expanded since Marshall. Economists do more than study exchange and production, though exchange remains at the heart of economics.

Most contemporary definitions of economics involve the notions of choice and scarcity. Perhaps the earliest of these is by Lionell Robbins in 1935: **"Economics is a science which studies human behavior as a relationship between ends and scarce means which have alternative uses."** Virtually all textbooks have definitions that are derived from this definition. Though the exact wording differs from author to author, the standard definition is something like this: **"Economics is the social science which examines how people choose to use limited or scarce resources in attempting to satisfy their unlimited wants."**

In other words **"Economics** is the *science of choice* — the science that explains the choices that we make and how those choices change as we cope with scarcity."

By now you must have got an idea that **scarcity** is central in these definitions. Now let's examine *scarcity*.

1.2 Scarcity and Choice

Scarcity means that people want more than is available. Scarcity limits us both as individuals and as a society. As individuals, limited income (and time and ability) keep us from doing and having all that we might like. As a society, limited resources (such as manpower, machinery, and natural resources) fix a maximum on the amount of goods and services that can be produced.

Scarcity requires **choice.** People must choose which of their desires they will satisfy and which they will leave unsatisfied. When we, either as individuals or as a society, choose more of something, scarcity forces us to take less of something else. Economics is sometimes called the study of scarcity because economic activity would not exist if scarcity did not force people to make choices.

When there is scarcity and choice, there are **costs**. The cost of any choice is the option or options that a person gives up. For example, if you gave up the option of playing a computer game to read this text, the cost of reading this text is the enjoyment you would have received playing the game. Most of economics is based on the simple idea that people make choices by comparing the benefits of option **A** with the benefits of option **B** (and all other options that are available) and choosing the one with the highest benefit. Alternatively, one can view the cost of choosing option **A** as the sacrifice involved in rejecting option **B**, and then say that one chooses option **A** when the benefits of **A** outweigh the costs of choosing **A** (which are the benefits one loses when one rejects option **B**).

The widespread use of definitions emphasizing choice and scarcity shows that economists believe that these definitions focus on a central and basic part of the subject. This emphasis on choice represents a relatively recent insight into what economics is all about; the notion of choice is not stressed in older definitions of economics. Sometimes, this insight yields rather clever definitions, as in James Buchanan's observation that an economist is one who disagrees with the statement that whatever is worth doing is worth doing well. What Buchanan is noting is that time is scarce because it is limited and there are many things one can do with one's time. If one wants to do all things well, one must devote considerable time to each, and thus must sacrifice other things one could do. Sometimes, it is wise to choose to do some things poorly so that one has more time for other things.

1.3 What Is A Science?

Should we accept claims of economists who say they are scientists? To decide, we must first know what **science** is.

Philosopher Karl Popper's widely accepted definition of science says that a statement is scientific only if it is open to the logical possibility of being found false. This definition means that we evaluate scientific statements by **testing** them, by comparing them to the world about us. A statement is nonscientific if it takes no risk of being found false; that is, if there can be no way to test the statement against observable facts or events. Popper called this distinction the "line of demarcation."

An implication of Popper's definition is that one can never be completely sure that any scientific theory is true. Accepted scientific theory is only theory that has not yet been contradicted by evidence, though the future may bring a contradiction. For example, we cannot be absolutely sure that the statement, "The sun will rise in the east tomorrow" is true because it is a scientific statement. We can easily think of a logical possibility that would refute it—a sunrise in the west. We have great confidence that such an event will not happen because the sun has always risen in the east. However, the fact that all previous experience has been consistent with the statement does not prove that the statement will never be refuted.

Popper saw the growth of scientific knowledge as a process of **conjecture and refutation**. Someone originally comes up with a way of explaining a set of facts; a conjecture or guess or theory about how the facts are related. If further observation is inconsistent with the theory, the theory is considered refuted and a new theory or conjecture must be found. In contrast, if the original explanation is nonscientific, it will never be refuted and there will never be any need to change beliefs.

Most economists see their discipline as scientific in Popper's sense of the word. Economic theory makes statements about how facts fit together, and there are constantly new sets of facts arising that allow one to test the theory to see whether the facts are as theory predicts. However, this process is more difficult for economists than it is for physical scientists.

Unlike physical scientists, economists can almost never use controlled experiments to gather facts with which to test theories. Rather they must use whatever facts the world gives them and rely on statistical procedures to draw conclusions. Though statistical procedures let economists hold some variables constant to see the effect of other variables, just as a controlled experiment does, they are subject to serious limitations. If there are variables that the theory says are important, but they cannot be measured or they can be measured only imperfectly, statistical procedures may give misleading results. Or the procedures may fail if the theory is uncertain exactly which of the many possible variables that may be involved must be controlled. One strength of a properly done controlled experiment is that there is no need to list all the factors that are controlled. The procedure is such that only one factor, or a small and known group of factors, is different between the control and experimental groups. Given these difficulties, it is not surprising that controversy about whether a theory is supported or rejected by the facts can last for many years in economics.[1]

There is a minority of economists, however, who do not see economics as scientific in Popper's sense. A group of economists called the Austrian school, for example, has argued that economics starts with assumptions and that economic theory is the logically deduced results of those assumptions. If the theory does not fit the facts, one cannot conclude that the theory is wrong, but only that it is inappropriate to apply the theory in that particular situation because the initial conditions do not agree with the assumptions of the theory.

Besides distinguishing between scientific and nonscientific statements, one can make a positive/normative distinction.

1.4 Positive and Normative

Economists make a distinction between positive and normative that closely parallels Popper's line of demarcation, but which is far older. David Hume explained it well in 1739, and Machiavelli used it two centuries earlier, in 1515. A **positive** statement is a statement about **what is** and that contains no indication of approval or disapproval. Notice that a positive statement can be wrong. "The moon is made of green cheese" is incorrect, but it is a positive statement because it is a statement about what exists.

A **normative** statement expresses a judgment about whether a situation is desirable or undesirable. "The world would be a better place if the moon were made of green cheese" is a normative statement because it expresses a judgment about **what ought to be**. Notice that there is no way of disproving this statement. If you disagree with it, you have no sure way of convincing someone who believes the statement that he is wrong.

Economists have found the positive-normative distinction useful because it helps people with very different views about what is desirable to communicate with each other. Libertarians and socialists, Christians and atheists may have very different ideas about what is desirable. When they disagree, they can try to learn whether their disagreement stems from different normative views or from different positive views. If their disagreement is on normative grounds, they know that their disagreement lies outside the realm of economics, so economic theory and evidence will not bring them together. However, if their disagreement is on positive grounds, then further discussion, study, and testing may bring them closer together.

Economists can confine themselves to positive statements, but few are willing to do so because such confinement limits what they can say about issues of government policy. Both positive and normative statements must be combined to make a policy statement. One must make a judgment about what goals are desirable (the normative part), and decide on a way of attaining those goals (the positive part). Economists often see cases in which people propose courses of action that will never get them to their intended results. If economists limit themselves to evaluating whether or not proposed actions will achieve intended results, they confine themselves to positive analysis. (You should realize that although economists can speak with special authority on positive issues, even the best can be wrong.) However, most economists prefer a wider role in policy analysis, and include normative judgments as well. On normative issues economists cannot speak with special expertise. Put somewhat differently, addressing most normative issues ultimately depends on how one answers the following question: "What is the meaning of life?" One does not study economics to answer this question.

Most statements are not easily categorized as purely positive or purely normative. Rather ,they are like tips of an iceberg, with many invisible assumptions hiding below the surface. Suppose, for example, someone says, "The minimum wage is a bad law." Behind that simple statement are assumptions about how to judge whether a law is good or bad (or normative statements) and also beliefs about what the actual effects of the minimum wage law are (or positive statements).

Next we discuss unintended consequences.

1.5 Unintended Consequences

The conventional definitions of economics ignore an important aspect of the field. Economists are not interested in examining every case of actions based on costs and benefits, but only on those that have some sort of unexpected or **unintended consequences**. Because we live in systems so complex that we cannot fully understand them, our choices can have system-wide implications that we neither intended nor expected. Economics starts with individuals making choices based on self-interest, but it is primarily interested in how these actions affect society as a whole. Do these choices lead to chaotic results or to harmonious ones?

Their concern with unintended consequences of human choice and action leads economists to argue that good results do not necessarily come from good intentions, and that good intentions do not necessarily lead to good results. In contrast, parts of our popular culture believe that intentions determine results. For example, people who try to find a conspiracy behind all the world's problems, whether that conspiracy be of Communists, Jews, bankers, the CIA, or multinational oil companies, start with a belief that bad results must come from bad people with bad intentions.

As with any other field of study, economics has had a history and there are books that attempt to trace the trail of economic thought back to its origins. Though the trail can be traced back to the ancient Greeks, it is a difficult trail to follow prior to 1776. In 1776, Adam Smith published *The Wealth of Nations,* a book that was clearly about economics and that inspired a large number of books, pamphlets, and articles in the next 50 years. Before this book most ideas about economics were scattered in writings that were mostly about politics or ethics or philosophy, not in books that were clearly about economics. Yet if one looks at the topics and theories that modern economics textbooks contain and compares them to those things Smith discussed, one is struck by how little of contemporary economics comes directly from Smith. What does come from Smith is a concern about and an interest in unintended consequences.

The most famous term in the *Wealth of Nations* is **"invisible hand."** Smith used this term only once, in the following quotation:

"...[B]y directing that industry in such a manner as its produce may be of the greatest value, he intends only his own gain, and he is in this, as in many other cases, led by an invisible hand to promote an end which was no part of his intention. Nor is it always the worse for the society that it was not part of it."

Another quotation makes clearer what unintended consequences the invisible hand leads to:

"Every individual is continually exerting himself to find out the most advantageous employment of whatever capital he can command. It is his own advantage, indeed, and not that of the society, which he has in view. But the study of his own advantage naturally, or rather necessarily leads him to prefer that employment which is most advantageous to society."

Smith was not sophisticated in the level of economic theory that he used. He did not understand concepts that are considered basic today, such as the model of supply and demand. (Alfred Marshall developed the modern treatment of supply and demand a century after Smith.) In his comprehensive survey of economic theory, Joseph Schumpeter dismisses Smith as a theorist, saying,

"The fact is that the *Wealth of Nations* does not contain a single analytic idea, principle, or method that was entirely new in 1776."

Though the idea that there could be systematic unintended consequences was "in the air" at the end of the eighteenth century, no one articulated it as well as Smith did. Because he expressed so well the idea that these unintended consequences are of vital importance for understanding how a society works, Smith has often been called "the father of economics." His book is concerned with a question that has interested economists for two hundred years: Under what conditions are actions based on self-interest beneficial to society? Much of economic theory has been developed and improved in an effort to get better answers to this question.

The two most influential economists in the generation after Adam Smith were David Ricardo and Thomas Malthus. Though they disagreed about many things, they were in general agreement about the topic of population growth, and it was for his writings on this topic that Malthus is best known. Malthus believed that there was a tendency for human populations to grow more rapidly than the food supply could be increased. Land was fixed in amount, and more food could be produced either by tilling it more intensively or by adding less-productive land to tillage. In either case an extra hour of labor brought less

than an average return of food. The implication of these two different growth rates is clear—eventually a segment of the population would face starvation, and this would cut the growth rate of population.

Malthus' argument on population is pregnant with possible unintended consequences. Suppose that society aids its most needy members by giving them food. As a result, they can survive and reproduce. Helping the poor would, according to Malthus' argument, increase population and in the future lead to even larger numbers on the verge of starvation. Hence, charity would be self-defeating. All attempts to improve society seem doomed to failure, according to a strict reading of the Malthusian argument, a truly "dismal" conclusion, as Carlyle noted.

However, Malthus and Ricardo were wrong when they applied this argument to human populations. The predictions they made did not occur. They underestimated both the capability of technology to improve crop yields and the future of birth rates. As it became apparent that they were wrong, economists lost their interest in the study of population. It seemed not to have the unintended consequences for them to explore. Since that time, economists have occasionally developed other clever theories with intriguing possibilities of unintended consequences, only to find that they too were in conflict with real-world experience. The reader should bear this in mind as he learns what contemporary economists believe. There is a possibility that today's secure truth may be tomorrow's embarrassing mistake.

Now some advice for you how to study of economics.

1.6 How to Study Economics

Agatha Christie wrote a series of mystery novels in which she challenged the reader to outwit her fictional heroine, Miss Jane Marple. By the end of the book, the reader has the same facts that Miss Marple has. But the facts do not speak for themselves. Rather it is Miss Marple's ability to look at those facts in a special way, to see something significant where most readers see nothing, which lets her solve the mystery.

Facts in economics, as in an Agatha Christie novel, need to be organized in some way before they can tell one anything. By themselves they are meaningless. Thus, the study of economics involves more than a memorization of facts. Economics tries to organize facts with its theory. Good theory tells us which facts are important and which are not, and what is cause and what is effect. **The study of economics involves learning how to organize facts the way economists do.**

People who do not understand economics still try to make sense of the world around them by trying to see pattern in the facts they observe. Sometimes they use a simplistic "good-versus-bad" model. In a good-versus-bad model there are two conflicting groups who are classified as good people and bad people. These groups are usually involved in a **zero-sum** game: one person's gain is another's loss. Further, evil motives, possessed by the bad people, lead to bad results unless these people are in some way controlled. Good motives lead to good results.

An example of a good-versus-bad viewpoint was expressed at a town meeting of a small Indiana community during the winter of 1977. The meeting focused on the natural-gas shortages that the community was facing. One citizen declared that the town faced not an energy problem, but a pricing problem. He noted that several years previously there had been shortages of gasoline at 40 cents per gallon but no shortages at 60 cents. Therefore, he declared, there must have been a conspiracy at that time by oil companies to increase prices as there was now by gas producers. The events he observed do fit into a good-versus-bad framework. He saw a bad result. He saw a bad motive—the desire for profit seems to many people the same as greed or avarice. To connect motive and result, he inferred the existence of a conspiracy.

Though a good-versus-bad model is sometimes appropriate (especially in small-group situations), economists are very reluctant to use it. The economic model of supply and demand gives a more sophisticated interpretation of the gasoline shortage; one that is depersonalized and unemotional with no bad groups involved. This model suggests that, in cases of shortages, one should search for government regulation of prices. The good-versus-bad model does not suggest that such regulation is something one should look for. In fact there were price restrictions in place at the time, and such restrictions can lead to shortages. The good-versus-bad view of the world is attractive because we are able to understand the model at a very young age and because we see the model used so often: in fairy tales, in comic books, in movies, and in television shows, among other places. Because we know how to use this model, and because our culture discourages use of alternative explanations such as fate or mystery, it is easy to fall back to this model if we do not have a more sophisticated model to explain our world.

Economic issues affect us all and most people have opinions about them. These opinions may be based on a good-versus-bad view, some other non-economic framework, or simply slogans that are often repeated. Often the hardest problem students have in learning about how economists interpret the world is to unlearn their old, non-economic views.

Unlearning old ways of thinking can be difficult, as a well-known example illustrates. In the late 15th century Christopher Columbus believed that by sailing a relatively short distance to the West, he could reach Asia. Contrary to popular myth, it was Columbus, not his critics, who had an outdated view of the world. He believed that the world was much smaller and that Asia was much larger than they actually are; his critics in their guesses were much closer to the truth.

Columbus made four trips to the Caribbean, but he never realized the significance of what he had found. He died believing that he had found a short cut to the Far East. Rather than use the facts he had before him to alter and improve his ideas of world geography, he insisted on keeping his old views and trying to make the facts fit in.

Economic education involves learning to see reality from new perspectives. Sometimes these new perspectives may surprise you. They may even shock you. But if you take the time to look at the reasoning behind these economic ways of looking at

things, you will find that they consist of carefully-thought-out-and-applied common sense.

To summarize the lesson I can say that the Economics is essentially a subject that looks at choices - how individuals, governments and businesses make them and what the consequences of making those decisions are. It is likely to be a strong likelihood that every issue raised in the class involves some form of decision or choice - for example, if fines were the answer to poor attendance - the choice being to remain absent and get fined or to attend the class and avoid the fine - the question may then be how much of a fine is necessary before those who choose to remain absent feel the cost of doing so is too great? All this can be precisely summarized through the following figure :

Figure 1.1 The Economic Problem

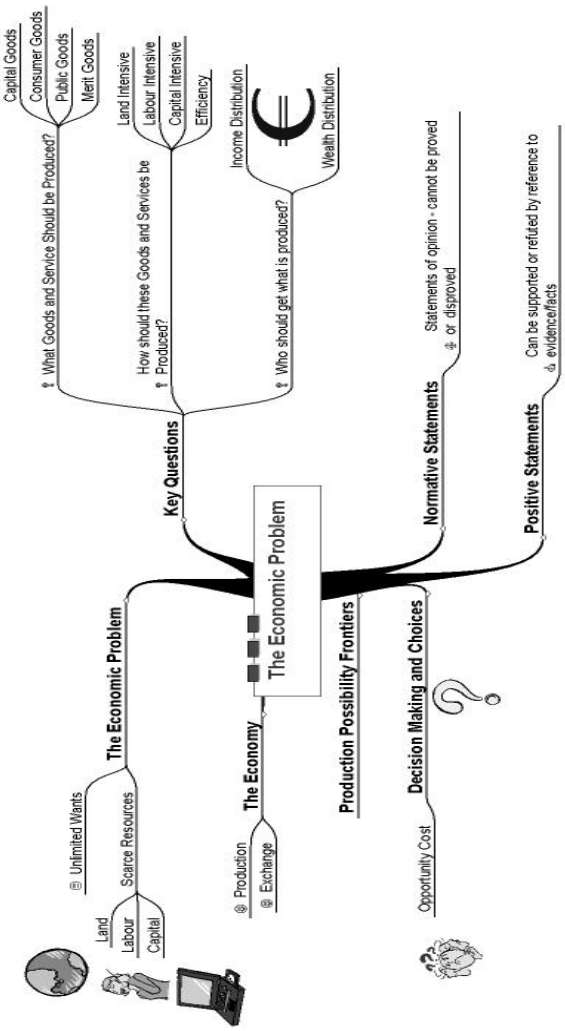


Figure 1.1 The Economic Problem

Notes

LESSON 2:
SCARCITY, CHOICE AND EFFICIENCY

Lesson 2 attempts to further your understanding of what economics is by delving more deeply into the concept of scarcity and introducing the most important notions of economics: opportunity cost and economic efficiency. As you will see, these concepts are fundamentally involved with the problem of how best to allocate scarce resources

2.1 Scarcity

Economics is the study of how economic agents or societies *choose* to use *scarce* productive resources that have *alternative uses* to satisfy wants which are *unlimited* and of *varying degrees of importance.* The main concern of economics is economic problem: its identification, description, explanation and solution, if possible. The source of any economic problem is scarcity. Scarcity of resources forces economic agents to choose among alternatives. **Therefore economic problem can be said to be a problem of choice and valuation of alternatives.** The problem of choice arises because limited resources with alternative uses are to be utilized to satisfy unlimited wants, which are of varying degrees of importance. Had the resources like human, natural, capital, etc. not been scarce, there would have been no problem of choice and hence no economic problem at all. Therefore, the root cause of all economic problems is scarcity.

In another words you can say that **Scarcity** means that you want more than is available. Scarcity limits us both as individuals and as a society. As individuals, your limited income (and time and ability) keep you off from doing and having all that you might like. As a society, limited resources (such as manpower, machinery, and natural resources) fix a maximum on the amount of goods and services that can be produced for you.

Scarcity is a central concept in economics. Resources scarcity is defined as there being a difference between the desire and the demand for a good. This means that the collective desire of individuals for **goods and services** exceeds the **productive resources (natural, human, and capital)** available to satisfy those desires. What this means is that a good is scarce if people would consume more of it if it were free. In other words, the things of **value** that people **want** are virtually unlimited, while the productive resources necessary to produce these things are limited. Every society, rich or poor, must determine how to best use its scarce productive resources to produce goods and services. This is the **basic economic problem.**

Table 2.1 Following could be an suggestive list of Desired Goods & Limited Resources :

Economic Goods (wants)	Limited Resources (scarcity)
Food	Land
Clothing National Defense	Cotton & other resources Human Resource
Education	No. of institutes, fee you can afford…..

Let us take an example, you may want to own gold jewelery. However, the amount of gold available is limited, so it is necessary to make choices as to how it is allocated. In a market economy, this is achieved by trade. Individuals trade resources between themselves to reallocate resources to where they are most wanted. In a smoothly operating market system, the rate of exchange between different resources, or price will adjust so that demand is equal to supply. One of the roles of the economist is to discover the relationship between demand and supply and develop mechanisms (such as pricing, incentives, or penalties) to achieve an optimal outcome (in terms of consumer welfare) between supply and demand.

Scarcity is a relative concept. It can be defined as excess demand i.e. demand more than the supply. For example. unemployment is essentially the scarcity of jobs. Inflation IS essentially scarcity of goods.

Another concept which you need to understand is of **CHOICE.**

2.2 Choice

Because goods and services are scarce, choices must be made. **Scarcity** - the available resources are insufficient to satisfy people's wants - is universal. All individuals, households, business firms, communities, nations - rich and poor alike - confront scarcity. The fundamental economic problem is the appropriate use of limited resources to produce the goods and services that we value most. Economics, therefore, can be defined as the study of the choices people make in order to cope with scarcity. Economists study (among other things) how societies perform the optimal allocation of these resources.

Scarcity requires **choice.** People must choose which of their desires they will satisfy and which they will leave unsatisfied. When we, either as individuals or as a society, choose more of something, scarcity forces us to take less of something else. Economics is sometimes called the study of scarcity because economic activity would not exist if scarcity did not force people to make choices.

The resources (also called **inputs** or **factors of production**) that can be used to produce goods and services are divided into four main categories:

- **Land**, the gifts of nature such as air, water, land surface and minerals lying beneath the earth's surface.
- **Labor**, the time and physical or mental effort devoted to producing goods and services.
- **Capital, goods made by people that** are used to produce other goods and services (factories, tractors, buildings, power plants, hand or power tools, machinery, equipment, transportation networks, etc). **Human capital** is the knowledge and skill people possess from education and vocational training. You are building human capital right now as you work towards your degree.

- **Enterpreneurial ability**, the resource that organizes land, labor and capital. The enterpreneur is the person who sets up a firm by combining all factors of production in order to produce a good or service. While labor receives wages or salaries for the work, the enterpreneur expects to receive profits for his efforts.

With the scarcity and choice in comes the **COST FACTOR.** Next what you need to understand is the concept of Cost.

2.3 Concept of cost

Till now you must be aware of that the root cause of economic problems is *scarcity*. Therefore, you need to be careful about utilization of each and every unit of scarce resources. To decide whether to use an additional unit of resource you need to know the additional output expected there from. Economists use the term marginal for such additional magnitude of output. Therefore marginal output of labor is the output produced by the last unit of labor, **marginal revenue is the additional revenue generated by an additional unit sold and marginal cost of production is the cost incurred for producing an additional unit of output.** While using the marginal concept we should be careful of the nature of relationship between the variables. In the above situations labor, sales and output produced are independent variables and output, revenue and cost are dependent variables respectively. In the same way if sales depend on advertisement, you talk of 'marginal sales of advertisement'; but if advertisement depends on the sales revenue, then you talk of 'marginal advertisement of sales' .

The concept of marginalism assumes that the independent variable changes by a single unit. In practice the independent variable may be subjected to "chunk changes" rather than unit changes. A contractor working on a turnkey project may change the labor employed not by one, but by tens and hundreds. Similarly, the costs and benefits of computerization are not subject to marginal analysis. In such situations the concept of incrementalism is more useful. In the above situations we talk about incremental output of labor and incremental costs and benefits of computerization respectively. In fact, incrementalism is more general whereas marginalism is more specific. All marginal concepts are incremental concepts; but all incremental concepts need not be confined to marginal concepts alone

The cost of any choice is the option or options that a person gives up. For example, if you gave up the option of playing a computer game to read this text, the cost of reading this text is the enjoyment you would have received playing the game. Most of economics is based on the simple idea that people make choices by comparing the benefits of option **A** with the benefits of option **B** (and all other options that are available) and choosing the one with the highest benefit. Alternatively, one can view the cost of choosing option **A** as the sacrifice involved in rejecting option **B**, and then say that one chooses option **A** when the benefits of **A** outweigh the costs of choosing **A** (which are the benefits one loses when one rejects option **B**).

The true cost of anything that is scarce is its **opportunity cost**, what is given up to get it. In other words, the opportunity cost of an action is the highest valued alternative forgone.

Now when you know the concept of Opportunity Cost with respect to an individual. I will go ahead with the concept with reference to a firm and then with the entire society.

2.3.1 Scarcity and Choice for a Single Firm

The **production possibilities frontier** (PPF) shows the different combinations of various goods that can be produced given the available resources and existing technology. The PPF marks the boundary between combinations of goods and services that can be produced and combinations that cannot.

Different resources are not equally effective in producing different goods. Thus, along the PPF, producing more of one good has increasing opportunity costs. Most activities in the real world are subject to increasing opportunity costs.

The **opportunity cost** of an action is the highest valued alternative forgone. On the PPF, the the opportunity cost of producing more of one good (e.g., soyabeans) is the output of the other good that must be forgone (e.g., wheat). The opportunity cost of a bushel of soyabeans is the number of bushels of wheat that must be forgone per bushel of soyabeans; therefore, opportunity cost is a **ratio**. The opportunity cost of a bushel of wheat is the inverse of the opportunity cost of a bushel of soyabeans. The following table shows the opportunity cost of producing wheat in the place of soybean and vice versa.

Table 2.2 - Calculation of the opportunity cost

			OPPORTUNITY COST OF:	
POINT	SOYABEANS	WHEAT	SOYABEANS	WHEAT
1	40	0	(38 - 0 / 40 - 30) = 38/10	-
2	30	38	(52 - 38 / 30 -20) = 14/10	(40 - 30 / 38 - 0) = 10/38
3	20	52	(60 - 52 / 20 -10) = 12/10	(30 - 20 / 52 - 38) = 10/14
4	10	60	(65 - 60 / 10 - 0) = 5/10	(20 - 10 / 60 - 52) = 10/12
5	0	65	-	(10 - 0 / 65 - 60) = 10/5

2.3.2 Scarcity and Choice for the Entire Society

Economic growth is the expansion in production. Two factors cause economic growth:

- Technological progress is the development of new goods and services and better ways to produce goods and services
- Capital accumulation refers to the growth in a society's capital resources.

The greater the rate of capital accumulation and/or technological process, the more rapidly the PPF expands, that is, the more rapid is economic growth. Economic growth is **costly**. The opportunity cost is incurred because resources are devoted to manufacturing capital goods and developing new technologies rather than to producing goods for current consumption. Nations that incur the cost of devoting more of their resources to capital accumulation or technological change grow more rapidly than nations that choose not to pay the cost and thus devote fewer resources to such purposes.

Form this discussion you must have got an idea that success of an individual, or a firm, or the entire nation depends on the

Effective utilization of the resources. Now I will take up the concept of Efficiency.

2.4 The Concept of Efficiency

Efficiency is a relative term. It is vital that this point be understood. Efficiency is never absolute; it is always relative to some criterion. This can be seen when one asks if farms are more efficient in the United States or China. The farming techniques in China are more efficient than those in the United States when measured in terms of output per unit of land, output per unit of fossil fuel, or output per unit of machinery. The farms in the United States are far more efficient in terms of output per man-hour. The statement that farms in one country are more efficient than farms in another makes no sense unless the criterion on which efficiency is measured is given. In other words I can say that "Efficiency is the relationship between what an organization or an economy produces & what it could feasibly produce".

The criterion for **economic efficiency** is **value**. A change that increases value is an efficient change and any change that decreases value is an inefficient change. A situation that is economically efficient may be inefficient when judged on different criteria. An example may make this concept clear.

Value is subjective. A thing has value only if someone wants it. How then can we know if value is maximized? If there is some change that makes someone feel better off, but making this change does not make anyone feel worse off, then the original situation was not one of highest value. Improvement was possible. When the highest value is reached, then any possible change that helps anyone must harm someone else. This way of defining economic efficiency, **Pareto optimality**, is named after Vilfredo Pareto, an early mathematical economist.

Economists are interested in economic efficiency for two reasons, one positive and the other normative. The positive reason is based on the observation that people search for value. We see this search for value vividly illustrated in the occupations of pimp, drug pusher, and hit man; given enough money, any occupation, no matter how immoral or risky, will attract people. On the theoretical level, we have seen this search for value in discussing utility maximization and profit maximization. The search for value is the driving force of market (and perhaps most nonmarket) economies. If there are situations in which there is unexploited value, that is, value that is possible but which no one obtains, the economist needs to explain why someone does not find a way to capture this value.

The normative reason stems from a desire to make policy recommendations. It is possible to discuss some aspects of policy without normative assumptions. An economist can predict, for example, whether a policy will or will not achieve the goals set for it. But economists often want to do more. They often want to compare two policies or two situations and decide which is better. To decide which is better requires some sort of basis for ranking situations. Thus, if they want to ask whether government regulation of utility prices, a tariff on steel, or a program to train unskilled workers helps society, economists need a criterion on which to base their answer. Economists generally use the criterion of economic efficiency to evaluate situations, though they often supplement it with other considerations because economic efficiency is not the only way to judge the relative merits of two situations.

The value maximized in the notion of economic efficiency reflects the goals people have. The concept of economic efficiency treats all goals as equally valid; no goals are considered better than other goals (with one exception—envy—discussed in the following paragraph). Not everyone agrees. Judging goals has been a central feature of the Judeo-Christian tradition. Generally, this tradition has condemned as immoral goal-seeking that emphasizes the most narrow individualism such as hedonism. To be moral, people must take into consideration the well-being of some others as a goal, including family or clan members and others who are members of a community grouping.

Production efficiency means that more of one good cannot be produced without decreasing the production of another good. Production efficiency occurs only when production takes place on the frontier line. Because another good must be given up, there is a **tradeoff**. If you are at a point 1 on table 1.2, production is inefficient because there are unused or misallocated resources.

Resources are **unused** when they lie idle but could be working. For example, you can leave some of the land used for the cultivation of soyabeans idle or some workers might be unemployed. Resources are **misallocated** when they are assigned to tasks for which they are not suitable. For example, you can assign land best suited to soyabean cultivation to wheat cultivation, or assign skilled soyabean workers to work in wheat cultivation. But yau can get more soyabeans **and** more wheat from the same inputs (i.e., land and/or labor) if we reassigned them to tasks that closely match their skills.

If you produce at a point 2, 3 or 4, you can use your resources more efficiently to produce more soyabeans and more wheat or more of **both** soyabeans and wheat.

From this I can very easily say that any individual, or organization, or an economy should know the answer to the five big questions, discussed below:

2.5 The Five Big Questions

Every society must figure out what is referred in economics as the "how", "what", "when", "where" and "for whom" to produce:

1. **How to produce or How to utilize its resources efficiently** - it is the choice among different resource combinations and techniques used in the production of a good or service. A good or service can be produced with different resource combinations and techniques; the problem is which of these to use. Since resources are limited, when a greater quantity is used to produce a particular good or service, less quantity is available for the production of another good or service. The problem facing society is choosing the right resource combination and production technique so that the cost in terms of the resources used for each unit of the good or service it decides to produce will be minimal. "How to produce": Because the price of a resource reflects its relative scarcity, the best way to produce a good or service is to ensure the least money cost of production. If

the price of a resource rises relative to the price of others used in the production of the particular good or service, producers will switch to another production technique: the one that uses less of the more expensive resource. The opposite holds true when the price of resource falls relative to the price of others.

2. **What to produce or What combination of goods and services to produce -** Since resources are scarce, no economy can produce as much of every good or service as desired by everyone. More of a good or service means less of others. So, society must choose which goods and services to produce and in what quantities. "What to produce": the price mechanism ensures that only those goods and services for which consumers are willing to pay a price sufficiently high to cover at least the full cost of production will be supplied by producers. A higher price induces producers to increase the quantity supplied of a good. Alternatively, a fall in price will induce producers to decrease the quantity supplied of a good.
3. **For whom to produce :** The economy will produce those goods and services that satisfy the wants of those consumers who can afford them. The higher the income of consumers, the more the economy will be geared to produce those goods and services they want and are willing to pay for them. "How much of each good to distribute to each person" - The problem of how to divide up what has been produced among the consumers, that is, how many of the consumers' wants can be satisfied. Scarcity ensures that society cannot satisfy the wants of all its members.
4. **When to produce :** The economy will produce the goods and services when they are needed most. So that they earn maximum profit.
5. **Where to produce :** This relates to the decision regarding the place of production to yield maximum profit. Eg, if you produce nearer to the raw material then the cast of inputs will be less, but you produce nearer to the market the cast of transportation of output will be less.

All individuals, organizations and nations can produce all the goods and services required by them, but the point is who can produce it with minimum inputs and maximum outputs. This is where the specialization starts. Now let me tell you about specialization and the cooperative advantage achieved through specialization.

2.5.1 Specialization & Comparative Advantage

People, businesses and nations can produce for themselves all the goods and services they consume, or they can concentrate on producing one good or service (or, possibly, a few goods or services) and then trade with others, that is, exchange some of their own goods or services for those of others. **Specialization** is the concentration on the production of only one good or service, or a few goods or services.

The principle of **comparative advantage** states that each nation (or individual) should specialize in the production of the good or service in which he is more efficient (or less inefficient). Stated differently, an individual or a nation has a comparative advantage in producing something if he can produce it at a lower opportunity cost than anyone else. This stems from the fact that people's abilities differ and, as a result, different people have different opportunity costs of producing a particular good or service.

It should be noted that it is not possible for anyone to have a comparative advantage in everything. Thus, gains from specialization and trade are always available when opportunity costs are different. Specialization requires a system of exchange to enjoy the fruits of comparative advantage. A voluntary exchange must yield mutual gains, that is, to make both parties better off.

This concept of exchange is the mother of ***markets.*** Now you will have a look on how market and prices work and how they are coordinated to get maximum comparative advantage.

2.5.2 Markets, Prices and the Coordination Tasks

Markets bring together buyers and sellers of goods and services. A **market** is any arrangement that enables buyers and sellers to get information and to do business with each other. **Prices** of goods and of resources, such as labor, machinery and land, adjust to ensure that scarce resources are used to produce those goods and services that society demands.

Much of economics is devoted to the study of how markets and prices enable society to solve the problems of how, what, when, where and for whom to produce, and this is the coordinate task to find the optimum mix of the following:

1. What ?
2. When ?
3. Where ?
4. How ?
5. For Whom ?

The widespread use of definitions emphasizing choice and scarcity shows that economists believe that these definitions focus on a central and basic part of the subject. This emphasis on choice represents a relatively recent insight into what economics is all about; the notion of choice is not stressed in older definitions of economics. Sometimes, this insight yields rather clever definitions, as in James Buchanan's observation that an economist is one who disagrees with the statement that whatever is worth doing is worth doing well. What Buchanan is noting is that time is scarce because it is limited and there are many things one can do with one's time. If one wants to do all things well, one must devote considerable time to each, and thus must sacrifice other things one could do. Sometimes, it is wise to choose to do some things poorly so that one has more time for other things.

2.6 Introduction to Production Possibility Frontiers

I am sure that by now you know that ***scarcity necessitates choice*.** **More of one** thing means less of something else. **The opportunity cost of using scarce resources for one thing instead of something else is often represented in graphical form as a production possibilities frontier.** The opportunity cost of producing (or Consuming) one good is how much of the alternative good must be sacrificed. Similarly, the **per-unit opportunity cost** tells us how much of a good is sacrificed in order to gain *one additional unit* of an Alternative good.

In other words, If a firm can produce two or more outputs or can produce output in two or more periods, a production possibility frontier can describe the possible combinations of output that can be attained for a given set of inputs.

Now you can very comfortably say that, the **Production Possibilities Frontier (PPF)** is a graphical representation which shows the maximal combinations of two goods that can be produced during a specific time period given fixed resources and technology and making full and efficiency use of available factor resources. A PPF is normally drawn as concave to the origin because the extra output resulting from allocating more resources to one particular good may fall. This is known as the law of diminishing returns and can occur because factor resources are not perfectly mobile between different uses, for example, re-allocating capital and labour resources from one industry to another may require re-training, added to a cost in terms of time and also the financial cost of moving resources to their new use. This cost, you know is called as, opportunity cost. The formula for calculating ***Per Unit Opportunity Cost (PUOC)*** is as :

$$\text{Per-Unit Opportunity Cost} = \frac{\text{Number of Goods Given Up}}{\text{Number of Goods Gained}} \text{ Units of Good Given Up}$$

Consider the following hypothetical PPF:

Production Point	Good X	Good Y
A	0	15
B	2	12
C	4	7
D	6	0

Using the above formula, the PUOC as we move from point A to point B is as follows

$$\text{PUOC} = \frac{\text{Number of Units of Good Y Given Up}}{\text{Number of Units of Good X Gained}} \text{ Units of Good Y Given Up} = 3/2\ Y = 1.5\ Y$$

Scarcity is the basis of many economic concepts because it constrains or limits our behavior. Let us explore the notion of constrained behavior by starting with the simplest sort of economic structure, suppose that YOU are alone on an island.

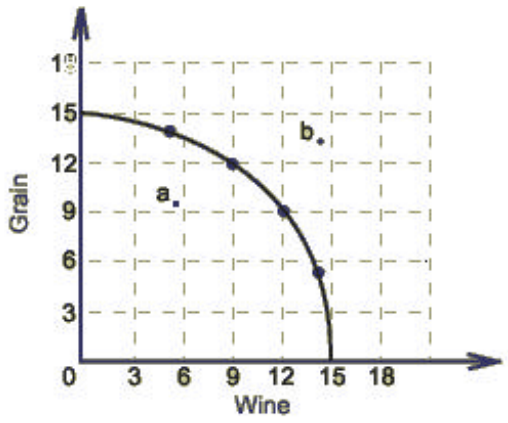
Now each day YOU have enough time to produce 15 thousand bottles of wine or 15 thousand bushes of grain. Notice that YOU cannot have both i.e., wine and grain. If YOU use your time to produce wine, you do not have that time to produce grain. If you want both wine and grain, you can devote some time to both. If, for example, you spend half of the day producing wine and the other half for producing grain, you can have 7500 bottles of wine and 7500 bushes of grain.

Table 2.3 Production Possibility Table

Wine (thousands of bottles)	Grain (thousands of bushels)
0	15
5	14
9	12
12	9
14	5
15	0

A list of all the possible combinations of wine and grain open to YOU makes up your **production possibilities**. The production-possibilities **frontier** separates outcomes that are possible for an individual (or a group) to produce from those which cannot be produced. Because YOU cannot exchange, your production-possibilities frontier is also your **consumption-possibilities frontier**. The consumption-possibilities frontier (sometimes called the budget constraint) is the line indicating which outcomes are affordable and which are not affordable. The graph below illustrates your production-possibilities frontier (and his consumption-possibilities frontier). Be sure you understand that the information in the table above is exactly the same as the information in the graph below—these are two different ways of presenting that information.

Figure 2.1 Production Possibility Frontier



The slope of the frontier in the graph above measures the costs you are facing. In order to get extra wine, YOU must sacrifice some grain, and vice versa. Notice that there is no money involved; cost does not depend on money, but rather exists whenever there is scarcity and choice. In economics, the cost of anything refers to whatever is given up in order to get that thing. The cost of going to college, for example, includes not only the money a person spends on tuition (which could be spent on something else), but also includes the time spent studying and going to classes. The value of this time can be estimated by computing the amount of income a person could earn if he did not go to college.

An example of a conventional PPF is shown in the diagram above which shows potential output of Wine and Grains from a given stock of labour and capital. Combinations of the two goods that lie within the PPF are feasible but point 'a' show an output that under-utilizes existing resources or where resources are being used inefficiently. Combinations of the two goods that lie on the PPF are feasible and can be produced using all available factor inputs efficiently. In the PPF diagram above, the combination of output shown by point 'b' is unattainable given current resources and the productivity of the available factor inputs**.** The PPF shows all efficient combinations of output for this island economy when the factors of production are used to their full potential. The economy could choose to operate at less than capacity somewhere inside the curve, for example at point a, but such a combination of goods would be less than what the economy is capable of producing. A combination outside the curve such as point b is not possible since the output level would exceed the capacity of the economy. The shape of this production possibility frontier illustrates the principle of increasing cost. As more of one product is

produced, increasingly larger amounts of the other product must be given up. In this example, some factors of production are suited to producing both wine and grain, but as the production of one of these commodities increases, resources better suited to production of the other must be diverted. Experienced wine producers are not necessarily efficient grain producers, and grain producers are not necessarily efficient wine producers, so the opportunity cost increases as one moves toward either extreme on the curve of production possibilities.

If you reflect on this table, you will see the importance of scarcity. You can think of the production-possibilities frontier as the way economists visualize scarcity. Which of the options will YOU choose? I cannot tell because I can only compute costs from this information, not benefits. The favorite assumption of economists is that individuals base their actions on the costs and benefits that they see. Benefits depend on the goals YOU have, and the production-possibilities frontier has no information about them

2.6.1 Shift in the PPF

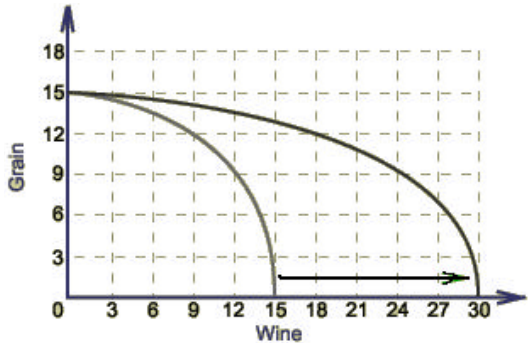
The production possibility frontier will shift when:

a. There are improvements in productivity and efficiency (perhaps because of the introduction of new technology or advances in the techniques of production).

b. More factor resources are exploited (perhaps due to an increase in the available workforce or a rise in the amount of capital equipment available for businesses to use).

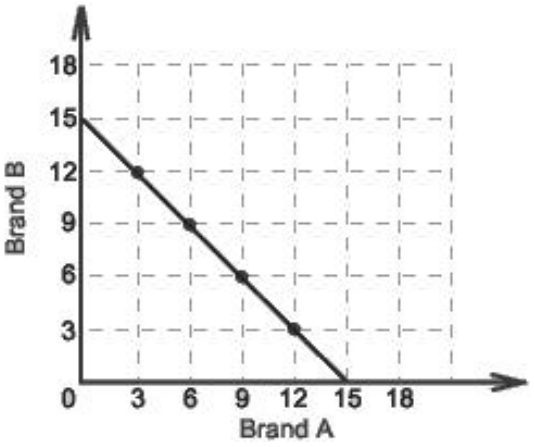
In our example illustrated in the second diagram 2.2 below we see the effects of a change in the state of technology that allowed the wine producers to double their output for a given level of resources. Further suppose that this technique could not be applied to grain production, i.e., resources allocated to grains are same as above. The real cost of WINE will fall – there has been a change in the opportunity cost The impact on the production possibilities is shown in the following diagram:

Figure 2.2 Shifted Production Possibility Frontier



In the above diagram, the new technique results in wine production that is double its previous level for any level of grain production. Finally, if the two products are very similar to one another, the production possibility frontier may be shaped more like a straight line. Consider the situation in which only wine is produced. Let's assume that two brands of wine are produced, Brand A and Brand B, and that these two brands use the same grapes and production process, differing only in the name on the label. The same factors of production can produce either product (brand) equally efficiently. The production possibility frontier then would appear as follows:

Figure 2.3 Production Possibility Frontier for Very Similar Product



Note that to increase production of Brand A from 0 to 3000 bottles, the production of Brand B must be decreased by 3000 bottles. This opportunity cost remains the same even at the other extreme, where increasing the production of Brand A from 12,000 to 15,000 bottles still requires that of Brand B to be decreased by 3000 bottles. Because the two products are almost identical in this case and can be produced equally efficiently using the same resources, the opportunity cost of producing one over the other remains constant between the two extremes of production possibilities.

2.7 The PPF and Economic Efficiency

An efficient production point represents the maximum combination of outputs given resources and technology – clearly the PPF is a useful way of illustrating this idea. The economy efficiency can be classified into following :

2.7.1 Allocative Efficiency

An economy achieves allocative efficiency if it manages to produce the combination of goods and services that people actually want. For allocative efficiency to be achieved we need to be on the PPF - because at points which lie within the frontier, it is possible to raise output of both goods and improve total economic welfare. The definition of Pareto Efficiency is an allocation of output where it is impossible to make one group of consumers better off without making another group at least as worse off.

2.7.2 Productive Efficiency

Productive efficiency is defined as the absence of waste in the production process. When the production of the two goods lies on the frontier, anywhere on the frontier is deemed to be production efficient and production inside frontier is inefficient. Productive efficiency requires minimizing the opportunity cost for a given value of output. When there is an outward shift of the PPF perhaps due to improvements in productivity or advances in the state of technology, then the opportunity cost of production falls and society can now produce more from given resources.

2.7.3 Distributive Efficiency

We achieve distributive efficiency if we get the goods and services produced to those who actually want or need them. Where we are on the production possibility frontier has little real bearing on distributive efficiency, we tend to use the concept to make comment on allocative and productive efficiency. But when an economy achieves economic growth leading to an outward shift in the PPF, economists have concerns over the distribution of gains in output and whether or not an improvement in average living standards has benefited the majority of consumers or whether there has been an increase in inequality and relative poverty.

2.8 Comparative Advantage

Comparative Advantage addresses a situation where two individuals or (in this case) countries are able to benefit from specialization and trade. Below, we work through an example involving two countries, Country A and Country X, where each country (first) attempts to meet domestic demand by producing only what is needed, and then (second) follows the Law of Comparative Advantage.

Country A produces compact cars and luxury cars and is able to achieve the following production possibilities (below). The table is written to reflect 10 different production/consumption choices (written as column/choice A through column/choice J).

	A	B	C	D	E	F	G	H	I	J
Compact cars	0	2	4	6	8	10	12	14	16	18
Luxury cars	9	8	7	6	5	4	3	2	1	0

To meet domestic demand, Country A must produce at pt. E (i.e. column E). Moving from pt. E to pt. D, Country A would have to give up producing 2 compact cars in order to produce 1 more luxury car. Because this country is fully employed, the only way to get more luxury cars is by taking workers out of compact car production and putting them into luxury car production. Doing this between pts. D and E causes 2 less compacts to be built. Therefore, the opportunity cost of that additional luxury car is 2 compact cars.

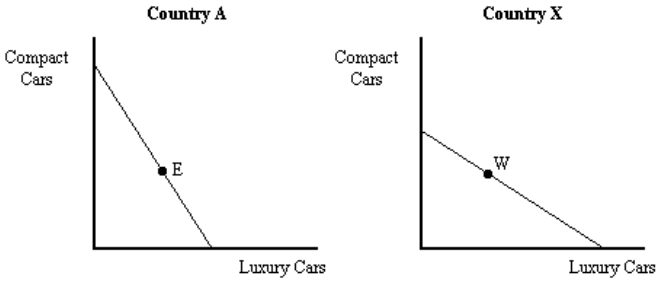
Moving from pt. E to pt. F, Country A must give up producing 1 luxury car in order to produce 2 more compact cars. Therefore, the opportunity cost of each (1) additional compact car is ½ of a luxury car.

If we consider any other pair of points, we find that the opportunity is always the same (for each good) no matter where we start. This implies constant opportunity costs, and tells us that the PPC here is a straight line.

Country X produces compacts and luxury cars as well. Their PPC relationship is:

	Q	R	S	T	U	V	W	X	Y	Z
Compact cars	0	1	2	3	4	5	6	7	8	9
Luxury cars	18	16	14	12	10	8	6	4	2	0

To meet domestic demand in Country X, it is necessary to produce at pt. W. If we were to illustrate Country A and X's production possibilities on a graph, then we would get the following.



Calculating the opportunity cost here, as we did above, we get:

Opportunity cost of each (1) additional compact car = 2 luxury cars

Opportunity cost of each (1) additional luxury car = 1/2 of a compact car

Now, let's compare the opportunity costs between countries for luxury cars:

(A) Opp cost of each 1 luxury car = 2 compact cars

(X) Opp cost of each 1 luxury car = 1/2 of a compact car

for compact cars:

(A) Opp cost of each 1 compact car = 1/2 of a luxury car

(X) Opp cost of each 1 compact car = 2 luxury cars

Country X gives up fewer compact cars when producing an additional luxury car, while Country A gives up fewer luxury cars when producing an additional compact car. Therefore, the opportunity cost of producing compacts is lowest in Country X, and the opportunity cost of producing luxury cars is lowest in Country A.

When Country A has a lower opportunity cost associated with producing something, then A is said to have a comparative advantage in producing that item. Therefore, A has a comparative advantage in producing compacts, while X has a comparative advantage in producing luxury cars.

The Law of Comparative Advantage says the following, "By specializing in the production of a good where a first country has a comparative advantage, the first country can trade with another country (who specializes in something that the first country doesn't have a comparative advantage in) and become "better off".

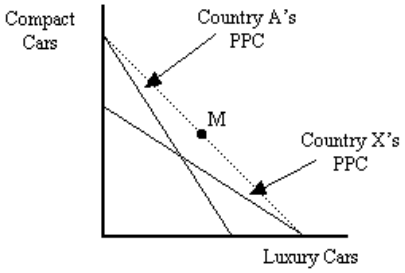
Suppose Countries A and X specialize where they have comparative advantage. Country A switches from pt. E to pt. J, while X switches from pt. W to pt. Q. We see this on the following table:

Country A	Domestic Demand	Specialize	Country X	Domestic Demand	Specialize
Compacts	8	18	Compacts	6	0
Luxury cars	5	0	Luxury cars	6	18

Country A now has 10 more compacts than needed domestically, whereas X has 12 more luxury cars than needed domestically. Assume that these countries are willing to trade on a 1-for-1 basis, and that A sends 9 compacts to X, in exchange for 9 luxury cars. That gives us the following result:

Country A	Before Trade	After Trade	Country X	Before Trade	After Trade
Compacts	8	9	Compacts	6	9
Luxury cars	5	9	Luxury cars	6	9

Do these countries benefit from specialization and trade? Yes, both have more of each good after trade than before trade. By specializing and trading, these countries can "consume" compact cars and luxury cars in amounts that would not be possible if these countries tried to meet domestic demand alone. That is, these countries are able to consume outside their PPC, even though they can't produce outside of it. This is illustrated in the graph below.



Country A's PPC and Country X's PPC are combined on the same graph. Consumption occurs at pt. M, a point that lies along the (green) dotted consumption possibilities line. Pt. M exists outside of each country's ability to produce, but by using the Law of Comparative Advantage, doesn't exist outside of each country's ability to consume.

Following are some important concepts of economics which you came across in this chapter :

1. **Opportunity Cost** The idea is that anything you must give up in order to carry out a particular decision is a cost of that decision. This concept is applied again and again throughout modern economics. If (God forbid) you were to learn only one of the Principles of Economics thoroughly, this should be the one.
2. **Scarcity** According to modern economics, scarcity exists whenever there is an opportunity cost, that is, where-ever a meaningful choice has to be made.
3. **Production Possibility Frontier** The production possibility frontier is the diagrammatic representation of scarcity in production.
4. **Comparative Advantage** A very important principle in itself, and a key to understanding of international trade the principle of comparative advantage is at the same time an application of the opportunity cost principle to trade.
5. **Discounting of Investment Returns** Another application of the opportunity cost principle that is very important in itself, this one tells us how to handle opportunities that come at different times.

In this lesson you have gone through:

- ***Scarcity and Efficiency : the Twin themes of economics***
  - ***Opportunity Cost***
  - ***Science of Choice***
- ***The five big questions that economists seek to answer.***
  - ***What?***
  - ***Why?***
  - ***When?***
  - ***Where?***
  - ***For Whom?***
- The Production-Possibility Frontiers

Notes

TUTORIAL 1

1. Multiple Choice Questions

1. From the viewpoint of economics, your college education can be thought of as an investment in a factor of production. Which factor is most appropriate?

- **i.** natural resources
- **ii.** labor
- **iii.** physical capital
- **iv.** human capital
- **v.** entrepreneurship

2. A Production Possibilities Curve (PPC) illustrates the concept of scarcity. Which item will be most likely to result in a shift of the PPC outward, indicating the ability to produce more goods?

- **i.** an increase in population
- **ii.** a decrease in the price of steel
- **iii.** reducing the federal debt
- **iv.** signing a trade agreement with China
- **v.** making more consumption goods

3. Which of the following questions is outside the scope of economics?

- **i.** What is the likely impact of the transfer of Hong Kong back to the Chinese?
- **ii.** When a paper mill closes in a small southern town, who is most likely to be out of work for a long time?
- **iii.** When you graduate from college, how might you best go about choosing a job?
- **iv.** How should society balance the needs of the environment against the needs of industry?
- **v.** None of the above; all are within the scope of economic study.

4. The study of economics is generally divided into two major sub-divisions: macroeconomics and microeconomics. Which statement is correct about the division?

- **i.** Macroeconomics deals with unemployment, inflation, the budget deficit, and the trade deficit.
- **ii.** Macroeconomics deals only with individual markets.
- **iii.** All the topics in macroeconomics are bigger than those in microeconomics.
- **iv.** Microeconomics is limited to the study of individual choices while macroeconomics deals with group decisions.
- **v.** Only macroeconomics deals with prices.

5. Which of the following is **not** a way we can use the study of macroeconomics?

- **i.** to understand how a national economy works
- **ii.** to understand the grand debates over economic policy
- **iii.** to decide between two types of automobiles when we are buying a new car
- **iv.** to make informed business decisions
- **v.** to help decide which candidate for office is most likely to have a successful economic policy

6. Which of the following is **not** a way we can use the study of microeconomics?

- **i.** to understand how markets work
- **ii.** to understand the full impact of our trade deficit with Japan
- **iii.** to make personal or managerial decisions
- **iv.** to evaluate the merits of specific public policies
- **v.** to help decide between two automobiles when we are buying a new car

7. The process of “thinking like an economist” involves three basic items. Which of the following five does **not** belong?

- Economists use assumptions to simplify matters.
- Economists deal only in items which have prices.
- Economists explore the relaltionship between two variables, holding other variables fixed.
- Economists think in marginal terms.
- Economists consider opportunity costs.

8. If a society is operating on its Production Possibilities Curve (PPC)with respect to thousands of computers and numbers of space missions, and is producing 300,000 computers and 5 space missions, in order to increase the number of space missions it must

- give up some computers.
- produce more computers as well.
- pay scientists more.
- shift federal spending from military to science.
- develop a new type of rocket.

9. The problem of scarcity,

- exists only in market economies.
- could be eliminated if we could force prices to fall.
- means that there are shortages of some goods.
- exists because human wants exceed available resources.
- can be eliminated by government intervention

10. If Josiah is producing inside his Production Possibilities Curve (PPC) then he

- can increase production of goods with no increase in resources.
- is fully using his resources.
- is optimizing.

- is unaffected by costs and technology.
- can do no better than he is currently doing.

11. The following table gives production possibilities for an economy that can produce two goods: lobsters and boats. Graph the production possibilties curve (PPC), given the information in Table 1-1, and putting lobsters on the horizontal axis. Use your graph to answer the first four questions. As this economy produces more and more lobsters, the slope of the PPC:

Table 1-1

	Lobsters	Boats
A	0	10
B	100	9
C	200	7
D	300	4
E	400	0

- increases.
- is constant.
- decreases.
- cannot be determined.

12. If the economy is producing at point C:

- we can produce more lobsters and more boats.
- we cannot produce any more lobsters.
- we cannot produce any more boats.
- we can produce more lobsters only by giving up some boats.
- the economy is not efficient.

13. At point B, to get one more boat, this economy must:

- give up one lobster.
- give up nine lobsters.
- give up 100 lobsters.
- increase the economy's resources.
- discover a new technology.

14. If this economy is producing 50 lobsters and 8 boats, then:

- the economy could produce more lobsters without giving up any boats.
- the economy could produce more boats without giving up any lobsters.
- the economy could produce more boats and more lobsters.
- the economy is not operating efficiently.
- all of the above

15. Economics is the study of:

- stock markets.
- money.
- self-interest.
- scarcity.
- all of the above

16. Which of the following is NOT an example of a marginal decision?

- Is it worth $2 to buy this extra slice of pizza?
- If I study for one more hour, how much will it raise my grade?
- If I hire ten workers to produce tables, what will be the average cost per table?
- If I drive slightly faster, what will be the change in my gasoline comsumption?
- All of the above are marginal questions.

17. If Y = 200 + 2X, what is the slope of this line?

- 100
- 200
- 2
- 1/2
- none of the above

18. If Y = 600 - 3X, what is the slope of this line?

- -1/3
- 1/3
- 3
- -3
- 600

19. If Y = 600 - 3X, what is the vertical intercept?

- 3
- 200
- 300
- 600
- none of the above

20. If Y = 600 - 3X, what is the horizontal intercept?

- 100
- 200
- 300
- 600
- none of the above

2. Long Answer Questions

i. Looking at the world in which you find yourself in college, imagine a typical class day. What examples of the ***factors of production*** are used to produce the class? What are examples of natural resources, labor, physical capital, human capital, and entrepreneurship?

ii. "And now for something really interesting...." Near and dear to the heart of every student is a market — the market for textbooks. Write a short essay explaining how the market for textbooks works on your campus — explaining what exactly takes place between you and the bookstore, both at the beginning and at the end of the term.

iii. Economists are often criticized for making assumptions. Why are assumptions necessary? To think about this, you might consider an assumption that is often made: people are rational. Do you think that people are rational, and how could you construct a model of irrational behavior? Would that be a better assumption to make?

THE CONSUMER MARKET SUPPLY AND DEMAND

There's a famous saying that "learned economist," is the one who always answers "supply and demand" in response to every question.

Not fair, but ...

It's true that the "theory of supply and demand" is a central part of economics. It is widely applicable, and also is a model of the way economists try to think most problems through, even when the theory of supply and demand is not applicable.

When people's actions are based on self-interest, people respond to incentives, that is, to costs and benefits. When the costs of an activity are raised or the benefits reduced, people do less of the activity. Economists have found that they can use this simple idea of action based on costs and benefits to construct a model (or theory) that explains how many markets work. This model, the model of supply and demand, is perhaps the most basic of the models economists use to explain the world around us.

Given the model's importance in the way modern economists think, it is surprising that one does not find the model in the writings of Adam Smith, David Ricardo, Thomas Malthus, or John Stuart Mill, though all of these pioneers in economics used the words "supply" and "demand" frequently. The modern supply-and-demand model does not appear until 1890, when Alfred Marshall published his *Principles of Economics.*

This group of readings explores economic terms and concepts that follow directly from supply and demand curves and that are important building blocks for other groups of readings. It begins with the concept of elasticity, which measures how people respond to changes. An elasticity computation can be used whenever a measurable change in something causes a measurable change in behavior. We meet the most commonly used elasticity measures: price elasticities of supply and demand, income elasticity, and cross-price elasticity. We then see how value can be represented on a demand-curve graph and meet the very important concept of marginal: examining how marginal, total, and average revenue are related. Finally we learn that elasticity and marginal revenue are related by means of a simple equation.

After you complete this chapter, you should be able to:

- Define demand, supply, inferior good, normal good, substitute, complement, law of demand, price taker, price searcher, market-clearing price.
- Distinguish between changes in demand and changes in quantity demanded.
- Distinguish between changes in supply and changes in quantity supplied.
- Predict how changes in factors such as income, prices of substitutes, prices of inputs, etc. affect the supply and demand curves and equilibrium quantity and price.
- Explain why we can treat the demand curve as positions of buyer equilibrium and the supply curve as positions of seller equilibrium.
- Compute price elasticities of supply and demand when given the curves in the form of a table.
- Explain what is meant when one says demand is elastic or when one says demand is inelastic.
- Define income and cross-price elasticity, and explain what they measure.
- Compute marginal revenue when given total revenue, and vise versa.
- Compute average revenue when given total revenue, and vise versa.
- Explain why marginal revenue is the slope of the total revenue curve.

Recognize the area that represents total revenue on a demand or supply graph.

LESSON 3:
DEMAND ANALYSIS

Before starting the lesson let me ask some questions to you :

Why does someone like Michael Jordan make more money per season than the rest of his team combined? Why are diamonds expensive? Why do heart surgeons make more money than sanitation workers? You probably guessed it, supply and demand. This unit will look at supply and demand and how they interact in the marketplace to determine the prices we pay for the goods and services we purchase.

Prices influence both buyers and sellers into making economic decisions. If the price for computers goes down, it will stimulate more demand to purchase computers. If the price of corn goes up, it will stimulate farmers into producing more corn. This is how the marketplace works. This section will look at the market processes that influence the demand side of the equation.

3.1 Introduction to Demand

A market exists when buyers and sellers interact to exchange products. Supply and demand analysis describes what happens in only some of these interactions. To use supply and demand analysis, we need markets in which there are many buyers and sellers, each small relative to the overall market. Also, both buyers and sellers must be well informed, and buyers and sellers must form distinct and separate groups. If buyers are a group distinct from the sellers, we can analyse how the act separately from how sellers act. Only when we have looked at these two groups separately will we combine them and see how they interact.

What determines the amount of a product that people are willing and ready to buy during some period of time? For example, what determines the amount of hamburger purchased in Chicago during a week? Economists answer such questions by examining the costs and benefits of buying the product. When any of the costs or benefits change, the amount of the product that people will buy should also change.

The benefits a person gets from a product depend on his **goals.** These goals are referred to in many ways in discussions of demand. The words **"tastes," "wants," "needs," "preferences,"** and **"usefulness"** all refer to goals. When people's goals change, the amount of benefit they get from the good changes, and this will cause them to change the amount of the good they want to buy.

Goals (or preferences or tastes) depend on many factors, such as the age of people and the amount of education they have. Social custom is an important determinant of preferences and can account for many differences in demand among groups. One can explain the large differences in squid sales in Japan and the United States, or the large differences in consumption of horse meat in Europe and the United States, almost entirely in terms of differences in preferences caused by differences in social custom.

The most obvious cost a person bears in buying a product is the price of the product. Price reflects cost because people have a limited amount of funds that they can spend, and if they spend their money on one thing, they cannot spend it on another. When the price of a product goes up, the amount of other things that a person must give up in order to buy the product rises. As a result, we expect people to buy more hamburger if the price is $1.00 per pound than if it is $2.00 per pound.

The amount of income a person receives affects the cost of buying an item because it determines which options a person must give up when buying a product. If a person with a low income spends $5000 for a trip around the world, he will have to cut back on food, clothing, or shelter. The same trip will cause a person with a high income to cut back on a very different set of options.

Increases in people's incomes raise consumption of most products. These products are called **normal** goods. There are some products, however, that people use less of as their income increases; these products are called **inferior** goods. Public transportation is an example—as people's incomes rise, they stop riding the bus and drive their own cars. Blue jeans were once another example—people with higher incomes bought them less frequently than people with lower incomes. It was because they were a symbol of "working-class" clothes that they were adopted by the radical left in the 1960s, and from there they moved into high fashion.

Prices of related goods also influence how much of a product people buy. Goods that are **substitutes** satisfy the same set of goals or preferences. An example of a substitute for hamburger is pork. If pork prices are high, people are tempted to shift away from pork to hamburger, and if pork prices are low, people are tempted to shift from hamburger to pork. The opposite of a substitute is a **complement,** a good that helps complete another in some way. Catsup and hamburger buns are complements to hamburger, and if they are priced low enough, consumption of hamburger may rise. Sometimes goods are such good complements that they are sold together and we think of them as a single item. Left shoes and right shoes are an example.

There are other factors that influence the amount of a particular product that people are willing to buy, such as the number of consumers in the market and their expectations about future prices, incomes, and quality changes. To get a complete list for any product might be time consuming and difficult, but it is not necessary because we want to focus on the relationship between price and the quantity of a product that people are willing to buy during some interval of time. To do this, we will assume that all other factors are held constant.

3.2 Demand Schedule & Demand Curve

The relationship between price and the amount of a product people want to buy is what economists call the **demand curve.** This relationship is inverse or indirect because as price gets higher, people want less of a particular product. This inverse relationship is almost always found in studies of particular products, and its very widespread occurrence has given it a special name: the **law of demand.** The word "law" in this case does not refer to a bill that the government has passed but to an observed regularity.

There are various ways to express the relationship between price and the quantity that people will buy. Mathematically, one can say that quantity demanded is a function of price, with other factors held constant, or:

Qd = f(Price, other factors held constant)

A more elementary way to capture the relationship is in the form of a table. The numbers in the table below are what one expects in a demand curve: as price goes up, the amount people are willing to buy decreases. This tabular representation is known as **Demand Schedule** (A widget is an imaginary product that some economist invented when he could not think of a real product to use in illustrating an idea.)

A Demand Schedule	
Price of Widgets	Number of Widgets People Want to Buy
$1.00	100
$2.00	90
$3.00	70
$4.00	40

The same information can also be plotted on a graph, where it will look like the graph below.[1] This graphical representation is known as **Demand Curve**.

3.3 Law of Demand

The graph above also demonstrates the law of demand. **The law of demand states that as price decreases, quantity demanded increases.** An inverse relationship exists. The law of demand is dependent on ceteris paribus- all other factors remaining unchanged.

These other factors are the assumptions of the law as well.

1. Price of related goods should remain unchanged.
2. Income of the consumer should not change.
3. Taste, preferences & fashion should not change.
4. All the units of product in question are homogeneous.

In economics, the term utility refers to the measure of satisfaction received from consuming a good or service. The law of demand does not go on for infinity. There are limits. The **law of diminishing marginal utility** describes how the last item consumed will be less satisfying than the one before. This means at some point, no matter how low the price is, consumers will purchase less.

3.4 Change in Demand & Shift in Demand

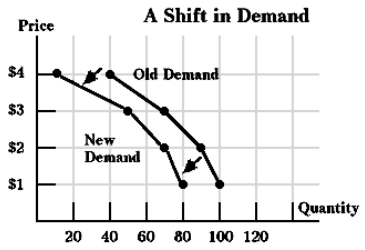
A **change in quantity demanded** can be illustrated by a movement between points along a stationary demand curve. Once again, demand is influenced by price. On the demand curve above, this is seen in the movement from point A to point B.

A shift in demand can also occur. A **shift in demand** refers to an increase (rightward change) or decrease (leftward change) in the quantity demanded at each possible price. This shift is influenced by non-price determinants. An example of an increase and a decrease in demand are pictured below.

If one of the factors being held constant becomes unstuck, changes, and then is held constant again, the relationship between price and quantity will change. For example, suppose the price of getwids, a substitute for widgets, falls. Then, people who previously were buying widgets will reconsider their choices, and some may decide to switch to getwids. This would be true at all possible prices for widgets. These changes in the way people will behave at each price will change the demand curve to look like in the table below.

A Demand Curve Can Shift	
Price of Widgets	Number of Widgets People Want to Buy
$1.00	[100] becomes 80
$2.00	[90] becomes 70
$3.00	[70] becomes 50
$4.00	[40] becomes 10

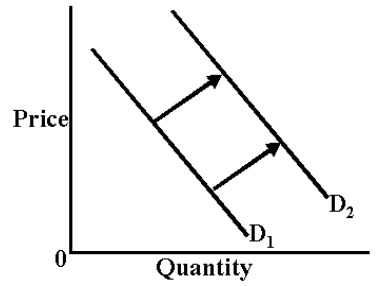
These are the same changes shown in a graph.



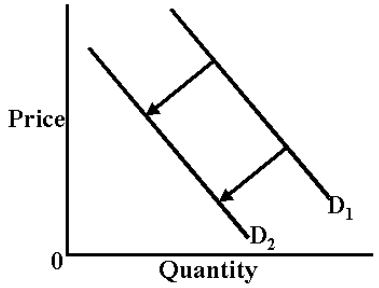
For all theoretical purpose we will assume the demand curve to be a straight line. Shift in demand can either be increase or decrease, shown in graph below

- **Increase in demand**, resulting from increase is the price of substitute, or increase in income, or change in taste & preferences etc. the graph will be one as below.
- **Decrease in demand**, resulting from decrease is the price of substitute, or decrease in income, or change in taste & preferences etc. the graph will be one as below.

Shift in Demand: Increase



Shift in Demand: Decrease



The most important distinction to keep in mind is that a change in quantity demanded is a movement along a single curve, while a shift in demand involves the creation of a second curve.

3.3 Non-Price Determinants of Demand

There are other factors besides price that influence consumers to purchase products. A brief description of each is provided below.

1. **A Change in income.** If you receive a raise you are likely to increase your demand for goods. If you get laid off, your demand for goods will likely decrease. When income increases, consumers buy more. When income decreases, consumers buy less.
2. **A Change in taste**. Fads, fashions, and the advertising of new products influence consumer decisions. Think of hula hoops and Pokeman cards.
3. **A Change in the price of a substitute good.** A substitute good competes with another good for consumer purchases. Examples of substitute goods include juice and soda, margarine and butter, and audio cassette tapes and compact discs. If the price of soda increases too much, consumer may decide to drink juice instead.
4. **A Change in the price of a complementary good.** A complementary good is jointly consumed with another good. Examples include cars and gasoline, tuition and textbooks, and milk and cereal. If the price of milk increases dramatically, consumers will decide to purchase less milk, and consequently, less cereal.
5. **A Change in buyer expectations.** If consumers think the price of a good will increase in the future, they may decide to buy more of it now so that they pay less. Suppose that a storm damages a large part of the orange crop. Consumers may run out and buy all the oranges they can find in anticipating the price of oranges increasing.
6. **A Change in the Number of Buyers**. Population growth will increase the demand for products because the pool of consumers has grown. Population decline will have the opposite effect. Look at the Baby Boom generation and how they have affected demand for goods over the course of their lifetimes.

A good place to end is to summarize the lesson. **Demand** refers to the quantities of a product that people are willing and able to purchase at a given price during some period of time. The term ***quantity demanded*** refers to a point on the demand curve- the quantity demanded at a particular price. A demand curve can be used to illustrate the relationship between quantity demanded and price.

3.4 Summary

To sum up I can say always remember when we speak of "demand" we usually mean the *entire demand relationship,* that is, the entire demand curve or table. By contrast, the "quantity demanded" is the *particular point* on the demand curve, as in Figure 2 below, or the quantity in a particular line of the table

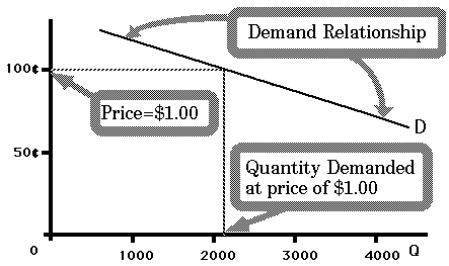


Figure 2: Demand Terminology

Notes

LESSON 4:
SUPPLY ANALYSIS

Why is it that farmers are more willing to grow certain crops one year and different crops the next? The price they receive for the crop they grow determines what seeds the farmers will sow. The information below will help you to understand the supply side of the equation.

4.1 Supply Defined

What determines the amount of a good or service that people are willing and ready to sell during some period of time? A discussion of exchange suggested that people sell things because it is a way, indirect but effective, of obtaining other things that they prefer. Sellers intend to make a profit from their sales, and economists assume that they want their profits to be as large as possible. Because profit is the difference between benefits in the form of revenues and costs, anything that influences revenues or costs can influence the amounts sellers want to sell.

Supply focuses on the producer of goods and services. **Supply refers to the quantities of a product that producers are willing and able to offer at a given price during some period of time.** Like demand, there are price and non-price determinants for supply. Producers make decisions on how much to supply based on profitability.

Revenue is found by multiplying the price of the product by the amount sold. A change in price changes revenues, and hence profits, so it is a major determinant of the amount sellers will want to sell. Because a higher price leads to higher profit, and a higher profit leads to a larger amount that sellers will want to sell, one expects that a greater quantity should be supplied when the price is higher. Thus, the relationship between quantity that sellers will sell and price should be direct or positive.

Though the positive relationship is almost always the case, there are a few exceptions. An example is labor; as wages go up, people may decide to enjoy their higher wages and work less. As a result, there is no law of supply that matches the law of demand.

The cost of something is what must be given up in order to get it. When costs are only monetary, they are easy to see. If the **price of an input** increases, the cost of the output will increase, and, other things held constant, profits will decrease. The seller will then have to decide if shifting part of his resources and effort to other products will improve his well-being.

Production costs are determined not only by the prices of inputs, but also by **technology.** Technology represents the knowledge of how inputs (such as labor, raw materials, energy, and machinery) can be combined to produce the product. If this knowledge increases so that people find cheaper ways to make the same output, then, other things held constant, profit increases and we expect sellers to respond by producing more.

Costs may be nonmonetary as well as monetary. For example, a farmer takes the expected price of soybeans into account in deciding how much corn to plant. If soybeans are expected to sell for a high price, then the farmer may find that shifting some of his land from corn production to soybean production will increase profit. The decision to plant corn means that the farmer gives up the opportunity to plant soybeans (as well as giving up the money for seed, fuel, equipment, and labor). Because we have defined cost as what must be given up to get something, the **prices of other goods** that sellers could otherwise produce and sell must be part of the calculation of the cost of production.

There are other factors that can influence the amount of a product that sellers will sell, such as the number of sellers, expectations about the future, and whether or not there are by-products in production that are valuable. (An example of a valuable by-product is cottonseed in the production of cotton. A farmer who produces cotton also gets cottonseed, which yields cottonseed oil, a widely used vegetable oil.) But as in the discussion of demand, the emphasis in the discussion of supply is on the relationship between quantity and price. To focus on this relationship, all other factors must be assumed to be constant.

The supply side of the equation also has a law**. The law of supply states that sellers will offer more of a good at a higher price and less at a lower price.** This law can also be graphically displayed.

4.2 The Supply Schedule & Supply Curve

The relationship between the quantity sellers want to sell during some time period (quantity supplied) and price is what economists call the **supply curve**. Though usually the relationship is positive, so that when price increases so does quantity supplied, there are exceptions. Hence there is no law of supply that parallels the law of demand.

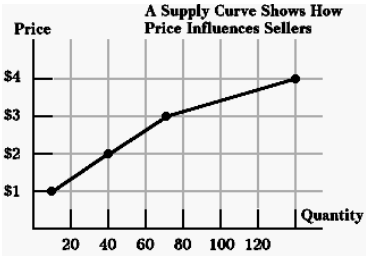
The supply curve can be expressed mathematically in functional form as

Qs = f(price, other factors held constant).

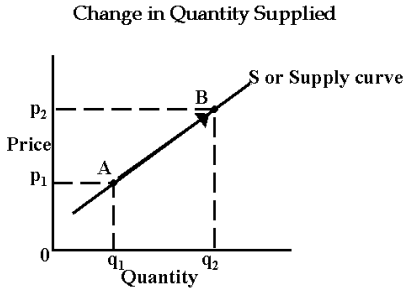
It can also be illustrated in the form of a table or a graph. The tabular representation is known as **Supply Schedule**, whereas graphical representation is called as **Supply Curve**

A Supply Schedule	
Price of Widgets	Number of Widgets Sellers Want to Sell
$1.00	10
$2.00	40
$3.00	70
$4.00	140

The graph shown below has a positive slope, which is the slope one normally expects from a supply curve.



For all practical purpose we assume this curve as a straight line. This will be clear when you examine the following curve :



A **change in Quantity supplied** occurs when there is a movement between points along a stationary supply curve. Once again, this movement is influenced by price. This change can be seen in the graph above with the movement from point A to point B.

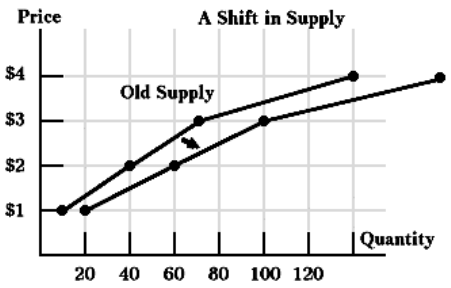
There can also be a **shift in supply**. A shift in supply refers to an increase (rightward change) or a decrease (leftward change) in the quantity supplied at each possible price. These shifts are influenced by non-price determinants.

4.3 Shift in Supply

If one of the factors that is held constant changes, the relationship between price and quantity, (supply) will change. If the price of an input falls, for example, the supply relationship may change, as in the following table.

A Supply Curve Can Shift	
Price of Widgets	Number of Widgets Sellers Want to Sell
$1.00	[10] becomes 20
$2.00	[40] becomes 60
$3.00	[70] becomes 100
$4.00	[140] becomes 180

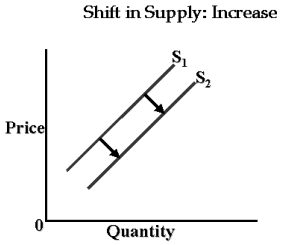
The same changes can be shown with a graph.



As I have already said that theoretical purpose we will assume the supply curve to be a straight line. This shift in demand can either be increase or decrease, shown in graph below

- **Increase in demand**, resulting from increase is the price of substitute, or increase in income, or change in taste & preferences etc. the graph will be one as below.

Decrease in demand, resulting from decrease is the price of substitute, or decrease in income, or change in taste & preferences etc. the graph will be one as below.



The most important distinction to keep in mind is that a change in quantity supplied is a movement along a single curve, while a shift in supply involves the creation of a second curve.

4.4 Non-Price Determinants of Supply

There are other factors besides price that influence producers to sell products. A brief description of each is provided below.

1. **Change in technology.** New, efficient technology makes it possible to offer more products at any possible selling price. Technology such as computers and robots have made it possible to reduce production costs and increase the supply of goods and services.
2. **Change in production costs**. A change in the cost of labor, or taxes, or a resource needed to produce a good, impacts the decisions of sellers on how much to produce.
3. **Change in the number of sellers**. An increase or decrease in the number of sellers can influence the production of goods and services. If the United States removes a restriction of foreign imports, then there are more sellers in the market.
4. **Change in supplier expectations**. Expectations of the future can influence the production of goods and services. If prices of a good or service is expected to rise in the future, sellers may hold back production in the present in the hopes of making more profit by selling more in the future. For example, if farmers think the future of the price of corn to decline, they will increase the present supply of corn, in the hopes of making more money now.

4.5 Supply Terminology

As with demand, economists separate changes in the amount that sellers will sell into two categories. A **change in supply** refers to a change in behavior of sellers caused because a factor held constant has changed. As a result of a change in supply, there is a new relationship between price and quantity. At each price there will be a new quantity and at each quantity there will be a new price. A **change in quantity supplied** refers to a change in behavior of sellers caused because price has changed. In this case, the relationship between price and quantity remains

unchanged, but a new pair in the list of all possible pairs of price and quantity has been realized.

Supply curves as well as demand curves appear much more concrete on an economist's graph than they appear in real markets. A supply curve is mostly potential—what will happen if certain prices are charged, most of which will never be charged. From the buyer's perspective, the supply curve has more meaning as a boundary than as a relationship. The supply curve says that only certain price-quantity pairs will be available to buyers—those lying to the left of the supply curve.

4.6 The Long and Short Run

Here's a complication: The supply relationship will depend on how long the suppliers have to adjust to a change in the price.

With respect to supply, time plays a role that it does not (in most cases) play in the case of demand. If there is plenty of time for the suppliers to adjust to a change in the price, we have a **long run** analysis. This means the sellers can invest and expand productive capacity, in response to a high price, or can gradually reduce the productive capacity by under-replacing worn-out equipment in the case of a low price. However, if the sellers are not sure the high or low price will continue for a long time, a **short run** analysis may be more appropriate. In a short run analysis, we treat the plant and equipment of the industry as inflexibly given. In that case, output can be increased only by using that fixed plant and equipment more intensively. Thus, we would expect the adjustment of supply to a change in price to be more complete in the long run than in the short run.

We do not ordinarily apply the long run versus short run distinction to demand, but there are some special cases where it might be important. For example, for durable goods such as cars, buyers might adjust less completely in the short run than in the long, since they can postpone replacement of their durable goods until the price comes down. In the long run, the goods wear out and so the consumers cannot postpone replacement long enough.

In summary,

- **In the *short run* the plant and equipment (productive capacity) of the industry are fixed**
- **In the *long run* sellers can change the productive capacity, in response to the price**

Next I will take up the Equilibrium of demand & supply.

Notes

LESSON 5:
EQUILIBRIUM OF SUPPLY AND DEMAND

5.1 Introduction to Equilibrium

Now that we have introduced the concepts of supply and demand separately — with illustrative examples of the demand for, and supply of, beer — it is time to move on from analysis to synthesis, and put the two concepts together. In economic theory, the interaction of supply and demand is understood as *equilibrium.*

We may think of demand as a force tending to increase the price of a good, and of supply as a force tending to reduce the price. When the two forces balance one another, the price would niether rise nor fall, but would be stable. This analogy leads us to think of the stable or natural price in a particular market as the "equilibrium" price.

This sort of "equilibrium" exists when the price is just high enough so that the quantity supplied just equals the quantity demanded. If we superimpose the demand curve and the supply curve in the same diagram, we can easily visualize this "equilibrium" price. It is the price at which the two curves cross. The corresponding quantity is the quantity that would be traded in a market equilibrium.

We have already developed two behavioral statements, or assertions, about how people will act. The first says that the amount buyers are willing and ready to buy depends on price and other factors that are assumed constant. The second says that the amount sellers are willing and ready to sell depends on price and other factors that are assumed constant. In mathematical terms our model is

Qd = f(price, constants)

Qs = g(price, constants)

This is not a complete model. Mathematically, the problem is that we have three variables (Qd, Qs, price) and only two equations, and this system will not have a solution. To complete the system, we add a simple equation containing the equilibrium condition:

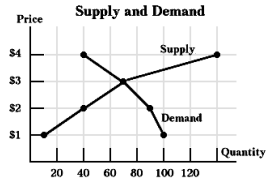
Qd = Qs.

In words, equilibrium exists if the amount sellers are willing to sell is equal to the amount buyers are willing to buy.

If we combine the supply and demand tables in earlier sections, we get the table below. It should be obvious that the price of $3.00 is the equilibrium price and the quantity of 70 is the equilibrium quantity. At any other price, sellers would want to sell a different amount than buyers want to buy.

Supply and Demand Together at Last		
Price of Widgets	Number of Widgets People Want to Buy	Number of Widgets Sellers Want to Sell
$1.00	100	10
$2.00	90	40
$3.00	70	70
$4.00	40	140

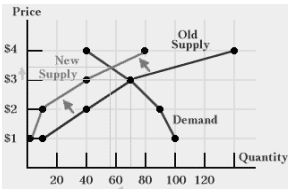
The same information can be shown with a graph. On the graph, the equilibrium price and quantity are indicated by the intersection of the supply and demand curves.



If one of the many factors that is being held constant changes, then equilibrium price and quantity will change. Further, if we know which factor changes, we can often predict the direction of changes, though rarely the exact magnitude. For example, the market for wheat fits the requirements of the supply and demand model quite well. Suppose there is a drought in the main wheat-producing areas of the United States. How will we show this on a supply and demand graph? Should we move the demand curve, the supply curve, or both? What will happen to equilibrium price and quantity?

A dangerous way to answer these questions is to first try to decide what will happen to price and quantity and then decide what will happen to the supply and demand curves. This is a route to disaster. Rather, one must first decide how the curves will shift, and from the shifts in the curves decide how price and quantity would change.

What should happen as the result of the drought? One begins by asking whether buyers would change the amount they purchased if price did not change and whether sellers would change the amount sold if price did not change. On reflection, one realizes that this event will change seller behavior at the given price, but is highly unlikely to change buyer behavior (unless one assumes that more than the drought occurs, such as a change in expectations caused by the drought). Further, at any price, the drought will reduce the amount sellers will sell. Thus, the supply curve will shift to the left and the demand curve will not change. There will be a change in supply and a change in quantity demanded. The new equilibrium will have a higher price and a lower quantity. These changes are shown below.



Shifting the Supply Curve Moves Us along the Demand Curve

What should one predict if a new diet calling for the consumption of two loaves of whole wheat bread sweeps through the U.S.? Again one must ask whether the behavior of buyers or

sellers will change if price does not change. Reflection should tell you that it will be the behavior of buyers that will change. Buyers would want more wheat at each possible price. The demand curve shifts to the right, which results in higher equilibrium price and quantity. Sellers would also change their behavior, but only because price changed.

Next we reflect on some assumptions we have made.

5.2 Assumptions

The supply and demand model **does not describe all markets**—there is too much diversity in the ways buyers and sellers interact for one simple model to explain everything. When we use the supply and demand model to explain a market, we are implicitly making a number of assumptions about that market.

For a supply curve to exist, there must be a large number of sellers in the market; and for a demand curve to exist, there must be many buyers. In both cases there must be enough so that no one believes that what he does will influence price. In terms that were first introduced into economics in the 1950s and that have become quite popular, everyone must be a **price taker** and no one can be a **price searcher.** If there is only one seller, that seller can search along the demand curve to find the most profitable price.[1] A price taker cannot influence the price, but must take or leave it. The ordinary consumer knows the role of price taker well. When he goes to the store, he can buy one or twenty gallons of milk with no effect on price. **The assumption that both buyers and sellers are price takers** is a crucial assumption, and often it is not true with regard to sellers. If it is not true with regard to sellers, a supply curve will not exist because the amount a seller will want to sell will depend not on price but on marginal revenue.

The model of supply and demand also requires that **buyers and sellers be clearly defined groups.** Notice that in the list of factors that affected buyers and sellers, the only common factor was price. Few people who buy hamburger know or care about the price of cattle feed or the details of cattle breeding. Cattle raisers do not care what the income of the buyers is or what the prices of related goods are unless they affect the price of cattle. Thus, when one factor changes, it affects only one curve, not both. When buyers and sellers cannot be clearly distinguished, as in the New York Stock Exchange, where the people who are buyers one minute may be sellers the next, one cannot talk about distinct and separate supply and demand curves.

The model of supply and demand also assumes that **both buyers and sellers have good information about the product's qualities and availability.** If information is not good, the same product may sell for a variety of prices. Often, however, what seems to be the same product at different prices can be considered a variety of products. A pound of hamburger for which one has to wait 15 minutes in a check-out line can be considered a different product from identical meat that one can buy without waiting.

In addition, the supply and demand model needs **well-defined private-property rights**. Elsewhere, we discussed how private-property rights and markets provide one way of coordinating decisions. When property rights are not clearly defined, the seller may be able to ignore some of the costs of production, which will then be imposed on others. Alternatively, buyers may not get all the benefits from purchasing a product; others may get some of the benefits without payment.

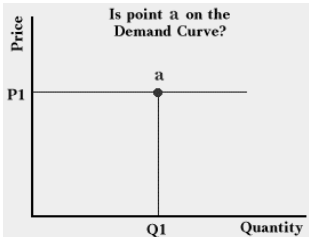
Finally, the supply and demand model requires **many buyers and sellers**. If there is only one seller, the seller can search along the demand curve of the buyers for the position that is most profitable. In this case, it is not just price that matters, but the slope of the demand curve as well. The seller in this case is not a price taker, but a price searcher.

Even if the assumptions underlying supply and demand are not met exactly, and they rarely are, the model often provides a fairly good approximation of a situation, good enough so that predictions based on the model are in the right direction. This ability of the model to predict even when some assumptions are not quite satisfied is one reason economists like the model so much.

Next we discuss the process of adjustment.

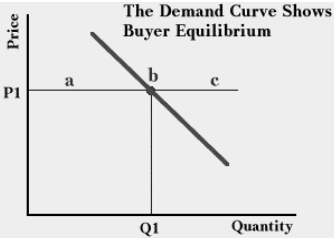
5.3 Buyers Equilibrium

We have developed the model of supply and demand as an equilibrium model. We have said nothing about how **adjustments** from disequilibrium to equilibrium take place. To develop this idea, it is useful to take still another view of supply and demand curves, to view demand as points of **buyer equilibrium** and supply as points of **seller equilibrium**.



Suppose that price is at **P1** in the graph above. Will point **a** be a point on the demand curve? If people would like to buy more than **Q1** at price **P1**, point **a** must lie to the left of the demand curve. In this case, some consumers are unhappy with the amount they have purchased, and will try to purchase more. If there is no more to purchase, some will attempt to offer more money for the product, or they will increase the time they devote to getting the product. The important idea is that if point **a** lies to the left of the demand curve, people will be unhappy with their situation and will change their behavior. If point **a** lies to the right of the demand curve, people will decide that they are buying too much of the product and will cut purchases. In conclusion, if a position is not on the demand curve, people will change their behavior, which indicates that only positions on the demand curve are positions of buyer equilibrium.

Similar reasoning explains why the demand curve can be considered a **boundary**. In the graph below, buyers are not in equilibrium at point **a**, but they can be held there and made to adjust in ways that do not change the money price. They cannot be held at point **c** unless there is some way to force people to buy a product when they do not want it.

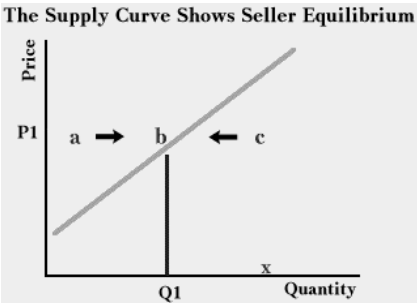


Point **b** in the graph is a position of buyer equilibrium because given price **P1**, people will be satisfied with **Q1** and will do nothing to change their behavior. Buyers would, of course, prefer a lower price than **P1**—they are always willing to move down the demand curve. However, this is not the issue here. Given **P1**, **Q1** is the preferred quantity.

A similar analysis holds for the supply curve.

5.4 Sellers Equilibrium

Just as the demand curve shows positions of buyer equilibrium the supply curve shows positions of **seller equilibrium**. At point **a** in the picture below suppliers find that they could increase profits (or reduce losses) by moving to the right to a larger quantity. If they could not increase profits by moving toward the right, they would stay at point **a**. Because they do not, they are not in equilibrium and on the supply curve but to the left of it. If they find that they could increase profits by cutting production, they are to the right of the supply curve and out of equilibrium. There is a quantity at the price **P1** that maximizes profits and toward which sellers will adjust. This point, shown as **b** in the picture, is on the supply curve.

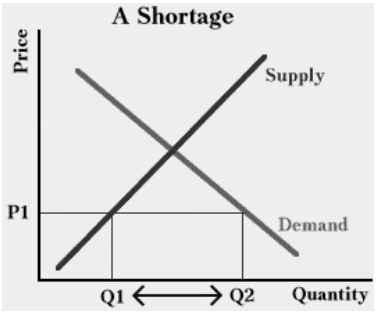


It is possible to force a seller to a position left of the supply curve. This is the case in which the seller would like to sell more at the given price, but for some reason can not. One reason might be that the buyers will not buy as much as the sellers would like to sell. It is virtually impossible—short of slavery—to force sellers to the right of the supply curve. This is the situation in which sellers are selling more than they want to at the given price. Thus, the supply curve represents a boundary facing the buyers. If buyers could force sellers to the right of the supply curve, they would find it advantageous to force sellers to a position such as x in graph above, which represents getting something for nothing.

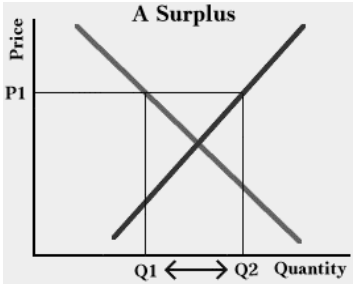
5.4.1 Shortage and Surplus

Sellers prefer higher prices to lower prices. Although all points on the supply curve represent points of equilibrium, not all are equally preferred by sellers.

The above analysis helps explain how an adjustment process takes place in the supply and demand model. If price is originally **P1** in the graph below, only **Q1** will be sold even though buyers would like to buy **Q2**. The difference **Q2 - Q1** represents a **shortage.** The sellers are in equilibrium in this situation because they can sell everything they want to sell at this price, but buyers are not. Some buyers who cannot obtain the product are willing to offer more, and sellers are always willing to accept a higher price. Therefore, the actions of the buyers, as they compete with each other to obtain the amount that is available, drive the price upward in this model toward market equilibrium.



If price is originally at **P2** in the picture below, only **Q1** will be sold because this is all that buyers will purchase, even though sellers are willing to sell more, **Q2**. The difference **Q2 - Q1** is called a **surplus.** In this situation the buyers are in equilibrium because they can buy all they want to buy at the going price. However, the sellers are not in equilibrium and will compete among themselves to get rid of the surplus. Some sellers will be willing to offer their product at a lower price. Buyers are always willing to move down the demand curve, so there is a tendency to move downward toward market equilibrium in the picture below.



If left to itself, a supply-and-demand market tends to adjust to the point where the supply and demand curves cross. The price at this intersection is called the **market-clearing price.** There is, however, the possibility that the existence of lags in the adjustment process may make the adjustment more complex than the previous discussion indicates.

Suppose that the price of cattle feed rises sharply. This event should affect the supply curve of cattle by shifting it to the left. The profitability of cattle production is reduced at each possible price, and some producers will drop out of the industry while others will curtail production. Looking at the curves, we see that prices should rise and quantity should drop.However, initially

prices might drop and quantity might rise, which is the exact opposite of the prediction from the supply and demand graph. The higher costs of feed will encourage farmers to raise fewer cattle, but as part of that cutback, they will temporarily send more cattle to the slaughterhouses. The prediction that supply-demand analysis gives will ultimately be correct, but it will not be correct in the process of adjustment.

More complicated adjustment patterns are possible. Suppose, for example, that higher beef prices shift the demand for pork to the right. Supply and demand analysis says that this should increase pork prices, and at the higher prices, farmers should produce more hogs.

However, hog production takes time, and will only happen if farmers expect the higher prices to continue for a long time. If pork producers do expect the higher prices to last, they may decrease the number of pigs sent to slaughter, further increasing price. A sow can either produce pork or baby pigs, but not both. If farmers expect high prices to last, they will keep their sows for piglet production.

In six months to a year, the baby pigs will have grown enough to go to market. If enough farmers had expected the high prices to last, they may have produced so many pigs that pork prices will now plunge to a level below that which is considered normal. The new, abnormally low price can then influence decisions that will not affect the price for many months. You should see that, once disturbed, a market with long time lags in production may bounce around for years before it finally finds its way back to equilibrium. If such a market is disturbed often enough, its prices and quantities will never come to rest at equilibrium levels.

Microeconomic discussion generally ignores adjustment problems, at least at the introductory level. Microeconomics assumes that markets clear, that is, they are always in equilibrium. Its analysis begins with the assumption that equilibrium has been reached and then asks questions about that equilibrium. However, adjustment problems are very important in macroeconomics . Macroeconomics cannot assume there are no adjustment problems, or else it assumes away one of the problems it wants to explain, unemployment. In fact, much of macroeconomics is about the forces that bump an economy away from equilibrium, and why, once it is away, it has problems reaching a new equilibrium.

5.4 Competition and Equilibrium

What we have seen is that the price will be in constant motion, up or down, except when quantity demanded is equal to quantity supplied. That is the position of rest.

Put another way, it is the price toward which competition pushes the price. At equilibrium, there is no competition either to buy or to sell, because everyone can buy or sell however much they may wish, *at the going price.* But whenever the market is away from equilibrium, competition will arise and tend to force it back.

Competition eliminates itself, by forcing the market into an equilibrium in which there is no need to compete. (This is a very different concept of competition than the biological “struggle for survival!)

Notes

LESSON 6:
ELASTICITY OF SUPPLY AND DEMAND

6.1 Price Elasticity

Businesses know that they face demand curves, but rarely do they know what these curves look like. Yet sometimes a business needs to have a good idea of what part of a demand curve looks like if it is to make good decisions. If Rick's Pizza raises its prices by ten percent, what will happen to its revenues? The answer depends on how consumers will respond. Will they cut back purchases a little or a lot? This question of how responsive consumers are to price changes involves the economic concept of **elasticity.**

Elasticity is a measure of responsiveness. Two words are important here. The word "measure" means that elasticity results are reported as numbers, or elasticity coefficients. The word "responsiveness" means that there is a stimulus-reaction involved. Some change or stimulus causes people to react by changing their behavior, and elasticity measures the extent to which people react.[1]

The most common elasticity measurement is that of **price elasticity of demand.** It measures how much consumers respond in their buying decisions to a change in price. The basic formula used to determine price elasticity is

e = (percentage change in quantity) / (percentage change in price).

(Read that as elasticity is the percentage change in quantity divided by the percentage change in price.)

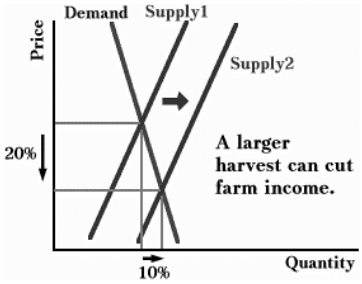
If price increases by 10% and consumers respond by decreasing purchases by 20%, the equation computes the elasticity coefficient as -2. The result is negative because an increase in price (a positive number) leads to a decrease in purchases (a negative number). Because the law of demand says it will always be negative, many economists ignore the negative sign, as we will in the following discussion.

An elasticity coefficient of 2 shows that consumers respond a great deal to a change in price. If, on the other hand, a 10% change in price causes only a 5% change in sales, the elasticity coefficient will be only 1/2. Economists would say in this case that demand is **inelastic.** Demand is inelastic whenever the elasticity coefficient is less than one. When it is greater than one, economists say that demand is **elastic.**

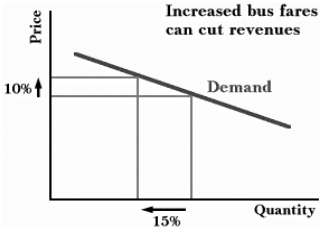
Products that have few good substitutes generally have a lower elasticity of demand than products with many substitutes. As a result, more broadly defined products have a lower elasticity than narrowly defined products. The price elasticity of demand for meat will be lower than the price elasticity of pork, and the price elasticity for soft drinks will be less elastic than the price elasticity for colas, which in turn will be less elastic than the price elasticity for Pepsi.

Time plays an important role in determining both consumer and producer responsiveness for many items. The longer people have to make adjustments, the more adjustments they will make. When the price of gasoline rose rapidly in the late 1970s as a result of the OPEC cartel, the only adjustment consumers could initially make was to drive less. With time, they could also move closer to work or find jobs closer to home, and switch to more fuel-efficient cars.

The concept of elasticity can help explain some situations that at first glance may seem puzzling. If American farmers all have excellent harvests, they may have a very poor year financially. They may be better off if they all have mediocre harvests. If a bus company decides it needs more revenue and tries to get it by raising fares, its revenues may decrease rather than increase.



In the case of the farmers, the key to their problem is that the demand curve for their products is quite inelastic. This means that if the harvest is unusually good, a large drop in price is necessary to encourage consumers to use the additional grain. If the elasticity coefficient is .5, for example, and the harvest is 10% larger than the previous year, then a 20% drop in prices will occur (assuming that the many things that we keep constant in drawing the demand curve have remained constant). Because this price reduction more than offsets the effect of the larger harvest, the average farmer's income drops.



For the bus company, the key is that demand is elastic. For example, suppose that the elasticity is 1.5. Then, if price is raised by 10%, quantity (ridership) must drop by 15%. But the drop in ridership more than offsets the increase in price, and so revenue will drop.

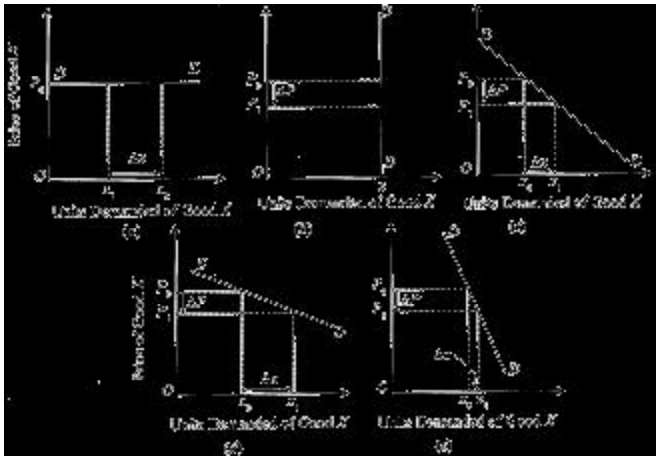
Just as we can measure how responsive buyers are to a change in price, we can measure how responsive sellers are. This measurement, the **price elasticity of supply**, has the same formula as price elasticity of demand, only the quantity in the formula will refer to the quantity that sellers will sell.

As with demand elasticity, supply elasticity depends on the amount of time available for adjustment. In the very short run, there may be no adjustments they can make, which would mean a perfectly vertical supply curve. For example, if on December 1 the price of apples doubles, there will be minimal effect on the number of apples available to the consumer. Producers cannot make adjustments until a new growing season begins. In the short run, producers can use their facilities more or less intensively. In the apple example, they can vary the amounts of pesticides, and the amount of labor they use to pick the apples. Finally, in the long run, not only can producers change their facilities, but they can also leave the industry or new producers may enter it. In our apple example, new orchards can be planted or old ones destroyed.

6.2 Types of Price Elasticity of Demand

Different commodities respond differently to changes in their price. A price change has relatively much less impact on quantity demanded *of* a necessity than it has on the quantity demanded *of* a luxury. In fact, it is the nature *of* a commodity which is responsible for differing elasticities *of* demand in case *of* different commodities. Conceptually price elasticities *of* demand is generally classified into the following categories:

1. ***Perfectly elastic demand*** *(e* = (0). Where no reduction in price is needed to cause an increase in quantity demanded [figure (a)];
2. ***Absolutely inelastic demand*** (e = 0). Where a change in price, however large, causes no change in quantity demanded [Fig. *(b)];*
3. ***Unit elasticity of demand*** (e = 1). Where a given proportionate change in price causes an equally proportionate change in quantity demanded (in this case the demand curve takes the form *of* a rectangular hyperbola. [Fig. *(c)];*
4. ***Relatively elastic demand*** (e > 1). Where a change in price causes a more than proportionate change in quantity demanded [Fig. *(d)]*
5. ***Relatively inelastic demand*** (e < 1). Where a change in price causes a less than proportionate change in quantity demanded [Fig. (e)].



Now comes the technical stuff, a discussion on how to compute price elasticity.

6.3 Computing Price Elasticity

(Warning: This section involves some simple algebra. If you are math-challenged, take it slowly and it will probably be OK.)

You can calculate elasticity in two different ways:

1. Point Elasticity
2. Arc Elasticity

When we try to use the equation in the first section to calculate elasticity coefficients, we run into a problem. If we look at an increase in price from $3.00 to $4.00, we have an increase of 33%. However, if we have a price reduction from $4.00 to $3.00, we have a reduction of 25%. Thus, there are two ways of viewing what happens between $3.00 and $4.00.

Suppose that when price of a product is $3.00, people will buy 60, but when price is $4.00, they will buy only 50. What is the elasticity over this segment of the demand curve, between prices $3.00 and $4.00? Should one start at the price of $3.00 and compute a price rise of 33 1/3% and a quantity decline of 16 2/3%, or

e =(16.67%)/(33.33%) = .5.

Or, should one start at the $4.00 price, and compute a price decline of 25% and a quantity increase of 20%, or

e = (20%)/(25%) = .8.

The formula mentioned in the previous section does not tell us which way to proceed, and it matters. To get around the problem of deciding which starting point to use, economists compute elasticity based on the midpoint, or in the example above, at a price of $3.50. The formula that does this is

e = (Change in quantity divided by average quantity) / (Change in price divided by

average price)

or

e = ((Q1 - Q2) / (Q1 + Q2)/2)) / ((P1 - P2)/((P1 + P2)/2)).

Putting the numbers from the previous example into this equation yields:

e = (60-50)/ (60+50)/2 divided by ($4-$3) / ($4+$3)/2

or

e = (10/55)/(1.00 /3.50) = (10/55)x(35/10) = 7/11 = .6363...

This formula is the formula for **arc elasticity,** or the elasticity between two points on the demand curve. As the two points get closer together, arc elasticity approaches **point elasticity**, the measure of elasticity preferred by professional economists.

With a bit of algebra, one can show that the equation for elasticity above can be rewritten as:

e = (1 / (Slope of Demand Curve)) multiplied by ((Average Price)/(Average Quantity))

Using this last equation, consider what happens when the slope gets steeper, which means that the slope becomes a bigger number.[1] Elasticity becomes smaller, which means that consumers are less responsive. As the demand curve approaches a vertical line, the slope approaches infinity and elasticity approaches zero. As the demand curve approaches a horizontal line, the slope approaches zero and elasticity approaches infinity.

One can also see what happens when the slope is constant, which means that the demand curve is a straight line. As one moves along the line, elasticity changes because average price and average quantity change. At the top of the demand curve, price is high and quantity is low, so elasticity is high. At the bottom price is low and quantity is high, so elasticity is low.

Wow. You've made it through all that technical stuff. Now, you can relax a bit and look at some other things that can be measured with the elasticity concept.

6.4 Other Elasticities

In addition to price elasticities of supply and demand, economists frequently refer to other elasticity measurements. **Income elasticity of demand measures the responsiveness of people's purchases to changes in income.** It is defined as

Income Elasticity = (percentage change in amount bought) divided by (percentage change in income)

Income elasticity measures whether a good is a normal or an inferior good. A product is a **normal good** when its income elasticity is positive, meaning that higher income causes people to purchase more of the product. For an **inferior good,** income elasticity is negative because an increase in income causes people to buy less of the product.

Cross-price elasticity, often simply called just cross-elasticity, **measures whether goods are substitutes or complements.** It looks at the response of people in buying one product when the price of another product changes. The formula for cross-price elasticity is

Cross-Price Elasticity = (percentage change in amount of A bought) divided by (percentage change in price of B).

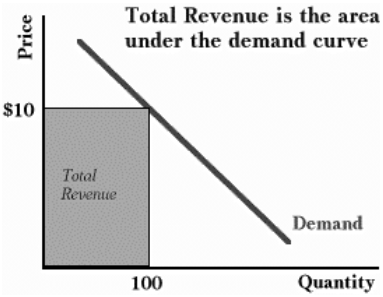
If goods are complements, cross-price elasticity will be negative. For example, if the price of gasoline rises, the sales of large cars will decline. The positive change in the denominator (bottom) is matched with a negative change in the numerator (top) of the equation. The result is therefore negative. If cross-price elasticity is positive, B is a substitute for A. For example, sales of Coke will fall if the price of Pepsi falls because some Coke drinkers will switch from Coke to Pepsi.

Next lets shift gears and see how revenue and the demand curve are related.

6.5 Revenue And Demand

The demand curve is a tremendously useful illustration for those who can read it. We have seen that the downward slope tells us that there is an indirect relationship between price and quantity. One can also view the demand curve as separating a region in which sellers can operate from a region forbidden to them. But there is more, especially when one considers what an area on the graph represents.

If people will buy 100 units of a product when its price is $10.00, as the picture below illustrates, total revenue for sellers will be $1000. Simple geometry tells us that the area of the rectangle formed under the demand curve in the picture is found by multiplying the height of the rectangle by its width. Because the height is price and the width is quantity, and since price multiplied by quantity is total revenue, the area is total revenue. The fact that area on supply and demand graphs measures total revenue (or total expenditure by buyers, which is the same thing from another viewpoint) is a key idea used repeatedly in microeconomics.



From the demand curve, we can obtain total revenue. From total revenue, we can obtain another key concept: **marginal revenue.** Marginal revenue is the additional revenue added by an additional unit of output, or in terms of a formula:

Marginal Revenue = (Change in total revenue) divided by (Change in sales)

According to the picture, people will not buy more than 100 units at a price of $10.00. To sell more, price must drop. Suppose that to sell the 101st unit, the price must drop to $9.95. What will the marginal revenue of the 101st unit be? Or, in other words, by how much will total revenue increase when the 101st unit is sold?

There is a temptation to answer this question by replying, "$9.95." A little arithmetic shows that this answer is incorrect. Total revenue when 100 are sold is $1000. When 101 are sold, total revenue is (101) x ($9.95) = $1004.95. The marginal revenue of the 101st unit is only $4.95.

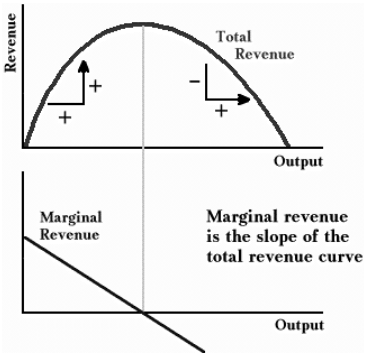
To see why the marginal revenue is less than price, one must understand the importance of the downward-sloping demand curve. To sell another unit, sellers must lower price on all units. They received an extra $9.95 for the 101st unit, but they lost $.05 on the 100 that they were previously selling. So the net increase in revenue was the $9.95 minus the $5.00, or $4.95.

There is a another way to see why marginal revenue will be less than price when a demand curve slopes downward. Price is average revenue. If the firm sells 100 for $10.00, the average revenue for each unit is $10.00. But as sellers sell more, the average revenue (or price) drops, and this can only happen if the marginal revenue is below price, pulling the average down.

The reasoning of why marginal will be below average if average is dropping can perhaps be better seen in another example. Suppose that the average age of 20 people in a room is 25 years, and that another person enters the room. If the average age of the people rises as a result, the extra person must be older than 25. If the average age drops, the extra person must be younger than 25. If the added person is exactly 25, then the average age will not change. Whenever an average is rising, its marginal must be above the average, and whenever an average is falling, its marginal must be below the average.

If one knows marginal revenue, one can tell what happens to total revenue if sales change. If selling another unit increases total revenue, the marginal revenue must be greater than zero.

If marginal revenue is less than zero, then selling another unit takes away from total revenue. If marginal revenue is zero, than selling another does not change total revenue. This relationship exists because marginal revenue measures the slope of the total revenue curve.



The picture above illustrates the relationship between total revenue and marginal revenue. The total revenue curve will be zero when nothing is sold and zero again when a great deal is sold at a zero price. Thus, it has the shape of an inverted U. The slope of any curve is defined as the rise over the run. The rise for the total revenue curve is the change in total revenue, and the run is the change in output. Therefore,

Slope of Total Revenue Curve = (Change in total revenue) / (Change in amount sold)

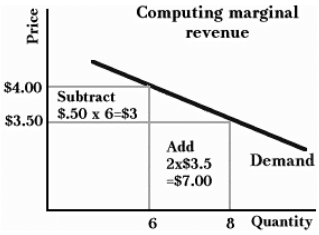
But this definition of slope is identical to the definition of marginal revenue, which demonstrates that marginal revenue is the slope of the total revenue curve.

Next we tie marginal revenue to elasticity.

6.6 From Elasticity to Marginal Revenue

(This is a moderately technical section that may trouble those who fear math, but it logically completes the chapter.)

Marginal revenue is the extra revenue from adding another unit of output. If a firm finds that when it sells six units, its revenue is 24, and when it sells eight, its revenue is 28, its extra revenue for adding two more units is four. Its marginal revenue, or the extra revenue for adding one more unit of production, will be two.



The graph above illustrates an alternative way to compute this extra revenue. When the firm sells six, it can charge a price of $4, but when it sells eight, it can charge only $3.50. (Thus, six units at $4 each gives a total revenue of $24 and eight units at $3.50 each gives a total revenue of $28.) When the firm sells the extra two units, it adds two units at $3.50 each, or $7 to its revenue. However, it also loses something because it had to lower the price on the six units it was previously selling. The loss is these six units times $.5 each, or $3. The net change in revenue is $7 less $3, or $4. Equation (6) says that to get marginal revenue, the change in total revenue ($4) must be divided by the change in output (2), which in this example gives us $2.

We have shown that marginal revenue can be computed as

(Change in Q)P + (Change in P)Q) divided by (Change in Q).

(This formula holds only approximately when changes are big, but becomes exact as the changes get very very small. Because the change in price will be negative, the second term in the numerator will be subtracted from the first.)

When changes in price and quantity are very, very small, the formula for price elasticity can be written as

e = ((Change in Q)/Q) divided by ((Change in P)/P)

If your algebra is fairly good, you should be able to use these two formulas to show that the following equation is true:

Marginal Revenue = Price (1 - 1/|elasticity|)

(Verbally, this says divide one by the absolute value of elasticity. Subtract this number from one. Then, take this second number and multiply it by price. The result is marginal revenue.)

This last formula says that if demand is inelastic (less than one), trying to sell more will reduce total revenue, whereas if demand is elastic (greater than one), trying to sell more will increase total revenue. This should make intuitive sense. If people are not sensitive to price, then one must reduce price a great deal to sell more, which means that total revenue declines.

Notes

TUTORIAL 2

1. Case study

Explore & Apply: Demanding Better Schools, Supplying Better Schools

According to an article from the *Wall Street Journal* (July 1, 2003) entitled "President Bush Renews Push for School-Voucher Program", the President "has rarely spoken out for vouchers since Congress rejected his proposal two years ago to strip federal funds from the worst-performing schools and make them available to parents for private education vouchers." However, in an address on July 1, 2003, he renewed his push by supporting legislation for a national "school choice" incentive plan.

According to the article, Bush is supporting a bill to allocate $75 million for a national "school choice" incentive plan, of which $15 million would go to the District of Columbia, largely because its students score lower on some standardized tests than those anywhere else in the nation. Part of the money allocated to the District would go to lower-income children enrolled in targeted public schools. The President said he hoped the D.C. program would become a national model for how private school choice can cause improvements in public schools.

Critics of vouchers, including teachers unions, contend that research does not show that students in voucher schools do better than those children who are not receiving vouchers. They argue that vouchers "drain money from public schools and too often end up supporting religious schools." Pointing to examples from Cleveland and Milwaukee, union officials assert "vouchers subsidize a choice that parents have already made—the choice to send their child to a private school."

Thinking Critically

Apply what you have read and answer the following questions:

1. Why might private-school choice make a difference in quality education in public schools?
2. What do critics say about vouchers?

2. Multiple choice questions

1 . Refer to the figure below. Which move illustrates the impact of a decrease in market price on market demand, all else the same?

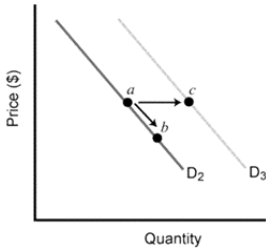
i. The move from *a* to *b.*

ii. The move from *a* to *c*.

iii. Both moves show the same result on demand.

iv. None of the above.

2 . Refer to the figure below. Assume that TVs and VCRs are two complement goods and that the diagram below represents the demand for VCRs. Which move would best describe the impact of a decrease in the price of TVs on this diagram?



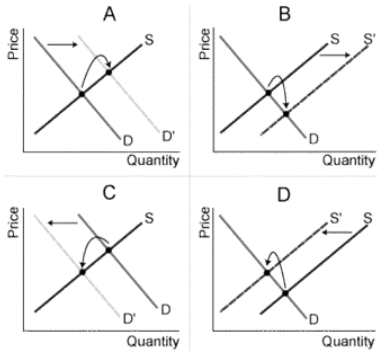
i. The move from *a* to *b.*

ii. The move from *a* to *c*.

iii. Both moves. Demand first moves from *a* to *b,* then from *b* to *c*.

iv. None of the above. Since this is the demand for VCRs, changes in the price of other goods would have no impact on it.

3 . Refer to the figures below. Which figure shows the impact of a decrease in income, assuming that the good in question is a normal good?



i. A.

ii. B.

iii. C.

iv. D.

4 . In a figure of supply and demand, can you describe more optimistic expectations on the part of business firms in the market?

i. Yes, by shifting the supply curve to the left.

ii. Yes, by shifting the supply curve to the right.

iii. Yes, by a move from one point to another along the supply curve.

iv. No, optimism is not a determinant of supply or demand.

5 . Factors that quickly and directly affect the market demand for a good or service do not include:

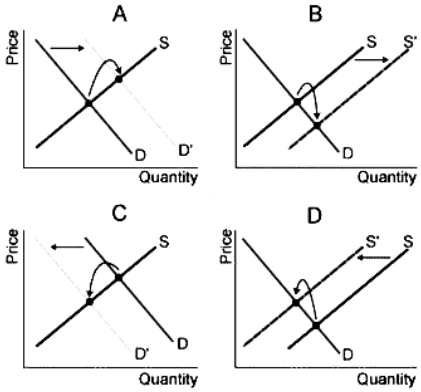
i. The price of related goods—substitutes and complements.

ii. The number of sellers, or business firms in the market.

iii. The number of buyers.

iv. The tastes and preferences of consumers.

6 . Which of the figures below would best describe the impact of an increase in the wages and input prices firms must pay in order to produce output?



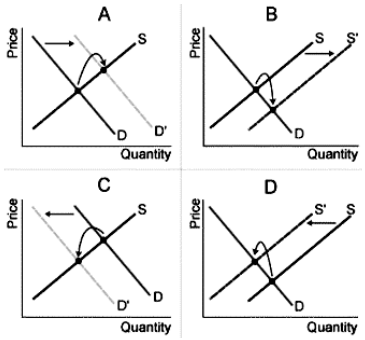
i. A.

ii. B.

iii. C.

iv. D.

7 . Refer to the figures below. In 1973, the Arab oil embargo resulted in a severe shortage of oil in the United States, and long lines at the gas pump. Which of the figures below would best describe the impact of the oil embargo on gas prices?



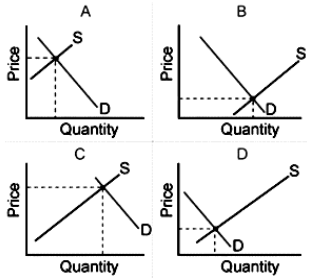
i. A.

ii. B.

iii. C.

iv. D.

8 . Refer to the figures below. Which of these markets depicts a situation of high market supply relative to market demand?



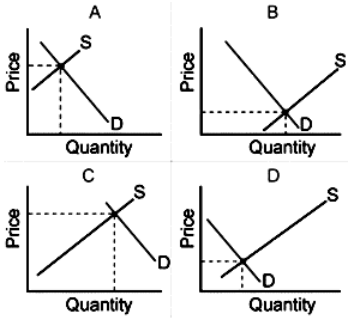
i. A.

ii. B.

iii. C.

iv. D.

9 . Refer to the figures below. In which of the markets below does the good in question appear to be relatively scarcer?



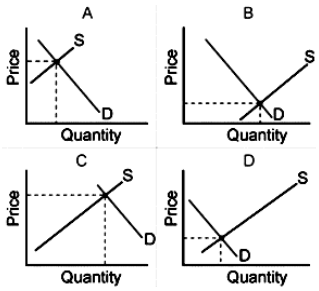
i. A.

ii. B.

iii. C.

iv. D.

10 . Refer to the figures below. When demand is high, equilibrium price is high. Choose the most applicable.



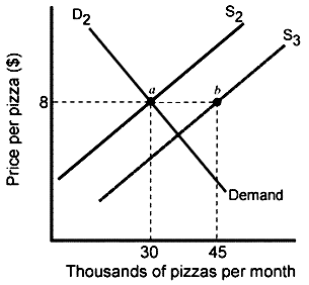
i. A.

ii. B.

iii. C.

iv. D.

11 . Refer to the figure below. Which of the following statements describing this figure is entirely correct?



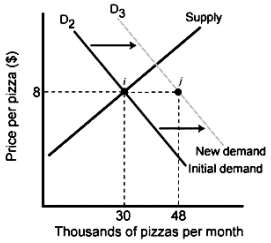
i. An increase in price has shifted the supply curve, and the price is likely to fall.

ii. After the increase in supply from S_2 to S_3, at a price of $8, there is an excess of quantity demanded over quantity supplied, which will cause the market price to increase.

iii. After the shift in supply, from S_2 to S_3, quantity demanded is likely to increase, arriving at a lower equilibrium price.

iv. After the shift in supply, quantity supplied will likely increase along the new supply curve, and the market will likely settle at equilibrium at a higher price and quantity.

12 . Refer to the figure below. Which of the following statements is correct?



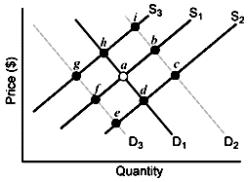
i. At $8, there is excess demand after the demand curve shifts from D_2 to D_3.

ii. After the market settles in equilibrium, quantity supplied will have increased.

iii. The shift in demand will create upward pressure on price.

iv. All of the above.

13 . Refer to the figure below. Starting at point a, an increase in demand with no change in supply moves equilibrium to:



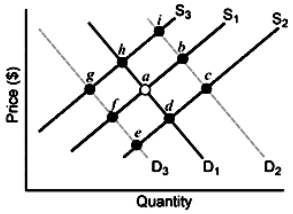
i. Point *e*.

ii. Point *c*.

iii. Point *h*.

iv. Point *b*.

14 . Refer to the figure below. Start at point a. A decrease in quantity demanded can be best exemplified by a move to:



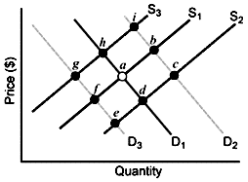
i. Point *g*.

ii. Point *d*.

iii. Point *h*.

iv. Point *f*.

15 . Refer to the figure below. An increase in demand accompanied by a decrease in supply moves equilibrium to:



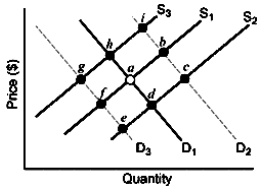
i. Point *c*.

ii. Point *i*.

iii. Point *d*.

iv. Point *b*.

16 . Refer to the figure below. To get to point g from point a, you need:



i. A decrease in demand and an increase in supply.

ii. Only a decrease in demand.

iii. A decrease in demand and a decrease in supply.

iv. Only a decrease in supply.

17 . If the supply and the demand for a good both rise, one thing is for sure:

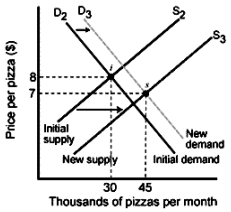
i. Market price will increase.

ii. Equilibrium quantity will increase.

iii. A surplus will result.

iv. Nothing. Nothing is for sure.

18 . Refer to the figure below. This figure shows that:



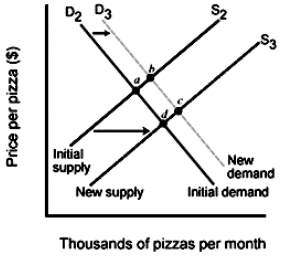
i. When the magnitude of a decrease in supply is greater than the magnitude of an increase in demand, equilibrium price will fall, and quantity will rise.

ii. When the magnitude of an increase in supply is greater than the magnitude of an increase in demand, equilibrium price will fall, and quantity will rise.

iii. When supply and demand both increase, price always decreases.

iv. In equilibrium, quantity demanded is not always equal to quantity supplied.

19 . Refer to the figure below. Start at point a. This figure demonstrates that:



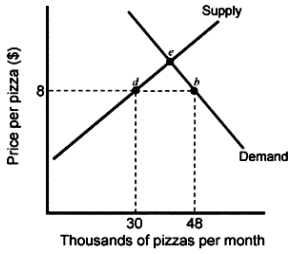
i. An increase in demand alone may move equilibrium to either point *b* or *c*.

ii. Both supply and demand must change in order to get to point *d.*

iii. Higher supply and higher demand result in higher prices.

iv. To get from point *a* to point *d*, there must be an increase in supply and an increase in quantity demanded.

20 . Refer to the figure below. Assume that the government has imposed the $8 price in this market. Which of the following statements is entirely correct concerning this figure?



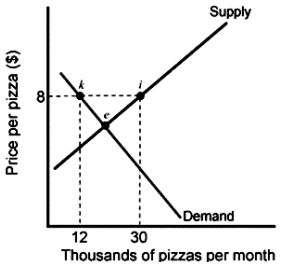
i. $8 is a minimum imposed price.

ii. $8 is a maximum imposed price.

iii. $8 is too low for equilibrium, therefore, both consumers and producers would benefit if the price rose to equilibrium.

iv. $8 creates an excess of quantity supplied over quantity demanded.

21 . Refer to the figure below. Assume that the government has imposed the $8 price in this market. Which of the following statements is entirely correct concerning this figure?



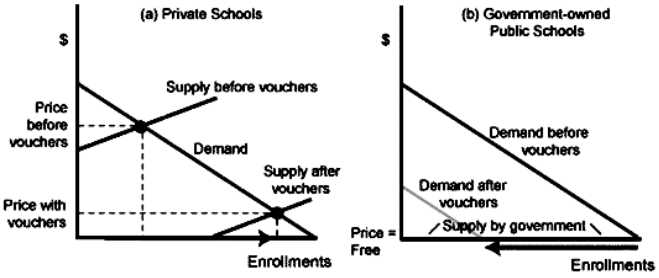
i. $8 is a price that results in a shortage, or a price above equilibrium price.

ii. $8 represents an imposed minimum price that creates a surplus.

iii. $8 is a price that generates an excess of quantity demanded over quantity supplied.

iv. All of the above.

22 . Refer to the figure below. According to the text, which of the following statements is correct about the effects of school vouchers on the figure below?



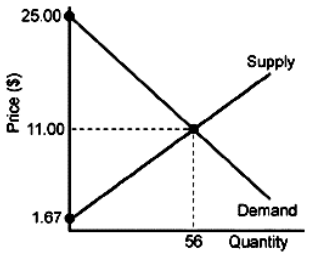
i. Vouchers increase both the price and the quantity demanded of private schooling.

ii. Vouchers cause the supply of private schooling to shift to the right.

iii. Vouchers increase the demand for government-owned public schools.

iv. Vouchers increase the demand for both private schooling and public schooling.

23. Use the information on the figure below to determine the slopes of the supply and demand lines.

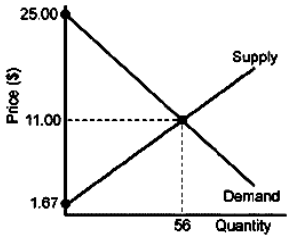


i. The slopes of the supply and demand lines are 1.67 and 25.00 respectively.

ii. The slopes of the supply and demand lines are +6 and – 4, respectively.

iii. The slopes of the supply and demand lines are +0.166 and – 0.25, respectively.

24 . Refer to the figure below. According to the information on the figure, the algebraic expression of the demand and supply lines are, respectively.



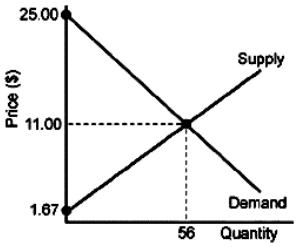
i. P = 25 – 56Q and Qs = 1.67 + 0.166Q

ii. Q = 100 – 4P and Q = –10 + 6P

iii. Q = 25 – 0.25P and Q = 1.67 + 0.166P

iv. P = 25 – 0.25Q, and P = 1.67 + 0.166Q.

25.Refer to the figure below. If the government imposes a maximum price of ten dollars in this market, there will be:



i. A shortage of 10 units.

ii. A surplus of 10 units.

iii. A shortage of roughly 19 units.

iv. A shortage of output, but there is insufficient information to estimate how much the shortage is.

Notes

THE CONSUMER MARKET
DEMAND AND CONSUMER BEHAVIOR

In this chapter we go a bit deeper in the economic interpretation of consumer demand, exploring two approaches to modeling the consumer's choices. the older approach, in terms of "utility," is taken first, and the more recent approach, in terms of "preferences," follows. The "preference" approach is the one used in more advanced economic theory and should be mastered by students who will pursue further study in economics.

A central idea of economics is that people make decisions by weighing costs and benefits. This idea can be stretched to explain the amounts of goods which people buy or sell. The key is to see a decision about the amount to buy or sell as a series of small decisions. To decide how much to buy, for example, a person should consider the costs and benefits of buying the first item and then the costs and benefits of buying the second, etc. Much of economic theory is based on this simple but vital idea.

This group of readings begins by looking at goals as the source of benefits and constraints of scarcity as the source of costs. It then shows how one can arrange information obtained from goals and constraints so that one can make decisions based on costs and benefits. It does this by introducing you to two rules, the maximization principle and the equimarginal principle. These rules can be derived mathematically and are in fact nothing but applied calculus. They also make common sense, and these readings stress the common-sense approach to these rules.

After you complete this unit, you should be able to:

- Define utility, tradeoff, cost, and law of diminishing marginal utility.
- Given income and prices, be able to form a budget constraint.
- Explain what a production-possibilities frontier shows, and what factors can move it.
- If given a table of total costs or benefits, be able to compute marginal costs and benefits, and vice versa.
- When given a table or graph showing costs and benefits for various levels of activity, use the maximization principle to find optimum positions.
- When given an income and a table showing marginal utilities and prices of two items, use the equimarginal principle to find optimum positions.
- Explain why the marginal cost curve measures the slope of the total cost curve.
- Interpret information on a graph showing a budget constraint and indifference curves.
- Define consumers' surplus and explain how it is measured on a demand curve.
- Define producers' surplus and explain how it is measured on a supply curve.
- Explain the paradox of value.

LESSON 7:
CONSUMER BEHAVIOUR – I (UTILITY APPROACH)

In the previous chapter, we saw that the consumer's demand curve could be traced to the "marginal benefit" the person gets from one more unit of consumption. The "marginal benefit" is the money value of the goods the consumer would give up to get one more unit — an application of the opportunity cost idea. That's correct and adequate for the purposes of Essential Principles of Economics, but not quite complete in the context of the Rational Dialog that is the history of economic thinking. The idea of "marginal benefit" as the basis of demand has a long history, and the related ideas that have been developed in that history are important in themselves. In the earliest development of the theory of demand, economists tried to tie demand to the one common factor that all goods and services have: utility. This may seem pretty vague, but, surprisingly, we can do a good deal with it. However, there are some criticisms of the utility approach, and many economists prefer to base demand theory on a concept of "preference." The "preference" theory is the one used in more advanced microeconomics courses. This chapter I will start with the benefit & utility approach in turn, and in next lesson we will deal with preference approach.

The theory of demand is also useful in some practical applications of economics known as cost-benefit analysis. In the theory of demand, we are asking the question, "What determines how much people are willing to pay for a good or service?" In answering that question, we also learn how to measure the benefits consumers get from what they consume. Benefits to consumers are important for cost-benefit analysis — after all, the purpose of production is to provide benefits to consumers, and everything else in cost-benefit analysis (and in economics) starts with that!

The terminology and approach we will use are based on practical cost-benefit analysis. In the long "rational dialog" of demand theory, economists have developed two other approaches, "utility theory" and "preference theory," each with its own terminology, logic, and diagrams. All three approaches come to the same practical conclusions. The later theories play important roles in more advanced economics and the student who expects to go on in the study of economics should understand them. They also provide insights that are important in themselves, on issues such as inflation, equity, and managing risk.

7.1 Benefit

Economists assume that people have goals and work to obtain those goals. Economists are vague about the goals people have, though they expect that material advancement (for self or family) is important for most, and they say nothing at all about the source of goals or their desirability. Nonetheless, the assumption that people strive to obtain their goals as best they can, given the limitations that the world imposes on them, forms the heart of virtually all economic theory.

Goals are complex and often a bit fuzzy. Often, it is hard to see exactly what goal a person has. For example, there are many people who smoke but say they want to quit. Such people are at war with themselves. One part pulls them one way and another part pulls them in an opposite direction. In addition, people often do not know what it is that they want, or if they do know what they want, they may not know how to obtain it. Most of us want to be happy, but many of us are unsure what that means in terms of how we should act. People spend time searching for goals, and there are institutions, most notably organized religion, which try to convince people that certain goals are more desirable than others.

In the early part of this century, Frank Knight emphasized the instability of goals. He wrote:

"Wants...not only are unstable, changeable in response to all sorts of influences, but it is their essential nature to change and grow; it is an inherent inner necessity in them. The chief thing which the common-sense individual actually wants is not satisfaction for the wants which he has, but more, and better wants."

Knight argued that "[e]conomic activity is at the same time a means of want-satisfaction, an agency for want- and character-formation, a field of creative self-expression, and a competitive sport."

These problems of defining goals have played little role in the way economists have gone about their business, and critics of economics say it is poorer as a result. Milton Friedman, one of Knight's students, states the way most economists proceed:

"Despite these qualifications, economic theory proceeds largely to take wants as fixed. This is primarily a case of division of labor. The economist has little to say about the formation of wants; this is the province of the psychologist. The economist's task is to trace the consequences of any given set of wants. The legitimacy of and justification for this abstraction must rest ultimately, as with any other abstraction, on the light that is shed and the power to predict that is yielded by this abstraction."

People will consider as a **benefit** anything that moves them closer to the goal they are seeking. A businessman running a business will consider revenue as a benefit if his goal is to make a profit. A person striving for material advancement will consider more belongings a benefit. A father who wants his family to be happy will consider the joy of his child from a gift a benefit.

But economists are not content to be this general. They want to discuss it in mathematical terms.

7.1.1 Utility Functions

Economists like to discuss goal-seeking in a mathematical terminology. When they talk about maximizing utility func-

tions, they are using an abstract, mathematical way of saying that people are trying to attain goals. A utility function, written as follows:

U = f(x1, x2,...xn)

means that items x1, x2, etc. to some nth x all contribute to a person's utility. Utility, as the word is used here, is an abstract variable, indicating goal-attainment or want-satisfaction. If a person has simple goals or objectives, such as accumulating material possessions, then x1 may represent cars, x2 fine furniture, x3 antiques, x4 land, etc. Anything that helps achieve goals gives utility, in the jargon of economists.

Though it is of no practical importance here, it should be noted that some economists disagree with the interpretation above. They see utility as a real psychic entity, just as happiness, joy, and satisfaction can be considered real psychic states. In this interpretation, the possibility of measuring utility exists, though the techniques have not been developed. If, however, utility is a fiction invented to allow us to talk about goal attainment in an abstract way, no general technique of measurement is possible.

Economists have been reluctant to examine goals that involve competition for **status**. Perhaps their reluctance is due to their emphasis on the mutual advantages of exchange. In an honest transaction both the buyer and sellers must benefit or else the transaction will not take place. However, quest for status is zero-sum. If one person rises in status, he does so at the expense of others who he passes up. Their reluctance to examine goals involving status has meant that economists have surrendered some interesting questions to other disciplines. For example, to what extent does economic development depend on the goals that people have? Do some goals stop economic development? Limited evidence seems to suggest that groups in which it is socially unacceptable for a person to rise relative to his neighbors have a harder time developing than those groups in which social mobility is acceptable. Other social sciences have emphasized pursuit of status, and it gives them a very different way of seeing the world than the way economists do.

Benefits are only part of the picture. The other part is constraints and costs.

7.1.2 Constraints And Costs

People have many goals that cannot be fully accomplished. Because people face **constraints** or limitations on their behavior, they must maneuver within those constraints.

Constraints come in many forms. Sometimes, they are mathematically imposed. Not everyone can be above average in intelligence, or athletic ability, or any other desirable trait. Not everyone can be a leader—there must be followers for leaders to exist. Not everyone can win, because the concept of winner implies that losers must also exist.

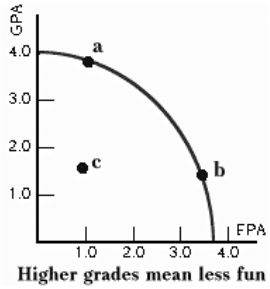
Time and biology impose the ultimate constraint on humans. A person's life is finite, almost always lasting less than a mere century. If one has many goals or has ambitious goals, the limited amount of time one has to live may make many impossible. This means, however, that the development of time-saving technologies, be it jet travel or packaged cake mixes, is liberating in the sense that it makes the constraint of time less pressing.

Income and wealth are other important constraints of which people are aware. Often our goals require us to make other people act in certain ways. Lacking the ability to force them, one may try to persuade them, or one may pay them. If one wants bread from the baker or beer from the brewer, one pays them. Exchange of this sort takes place because both parties to the exchange find the exchange beneficial, i.e., it helps each side move toward its goals. Most people claim that their money and income are not sufficient and that they need more, which means that they would like to influence the actions of many more people than they in fact can. Economists call this limitation the **budget constraint**.

Budget constraints reflect more basic constraints. People have limited abilities and limited time in which to earn income. In order to earn income, people sacrifice leisure time. In general, the existence of constraints means that people faced choices, and thus costs. It means that people cannot accomplish all their goals (satisfy all their wants), but must choose to forgo some goals in order to accomplish others.

This point is important enough to justify spending some time with examples. Consider the college student who wants to do well academically, yet also wants to have an interesting and exciting social life. The basic limitation that this student faces is time. Each day has but 24 hours, and each week has but seven days. If enough time is spent to achieve a really excellent grade-point average (GPA), say straight As, the student may have a poor or miserable social life, or a low fun-point average (FPA). If the student enjoys as high a FPA as time permits, there may be little or no time for study, and grades may be very poor. This notion of a **tradeoff,** that to get more of one thing, a person must sacrifice something else, is central to the way **economists view the world.**

The graph below illustrates the tradeoff that the student faces. Points to the right of the line are not attainable, whereas those to the left of the line and points on the line are. Point **a** represents a use of time with a great deal of studying and very little social life. Point **b** represents the opposite. Point **c** represents a poor use of time—the student in this case may be trying to study when everyone else is socializing and trying to socialize when most others are studying. In this case, a different use of time could improve both GPA and FPA.



Higher grades mean less fun

The tradeoff line in this example is curved. This curve indicates that, starting from a position of no study and all fun, devoting only a few hours to study has a big effect on GPA but reduces

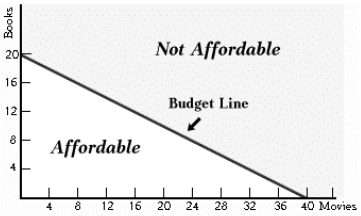
FPA only a little. Then, for each additional hour spent studying the increase in GPA becomes less and less and the drop in FPA becomes greater and greater. In other words, if the student is not studying much, an extra hour with the books will help his GPA a lot and not cost him much in lost fun. If the student is studying a lot already, an extra hour with the books will not help his GPA much, but will cost a lot in lost fun. Economists call this pattern "decreasing returns," and find its presence in a great many situations.

In our case of GPA versus FPA, there is neither money nor prices, but there are costs. The cost of a higher GPA is the loss of FPA that must be given up. The cost of a higher FPA is the loss of GPA that must be given up. The notion of cost in economics refers not just to money costs, but to all options, whether measured in money or not, that must be sacrificed to get something.

We next look at a budget constraint that reflects prices and money.

7.1.3 The Budget Line

Suppose, you have only Rs.100 to spend on two passions in your life: buying books and attending movies. If all books cost Rs.5.00 and all movies cost Rs.2.50 (these are simply assumptions to make the problem easier—as is the assumption that only two items are involved in the problem), the graph below shows the options open to you. The budget line is a frontier showing what you can attain. The budget line limits choices; it is due to scarcity. The cost of a book is Rs.5.00 or two movies. Spending money on a product means that money cannot be used to purchase another product. In the case of books versus movies, the tradeoff is a straight line because one more book always costs two movies, regardless of how many books you have already.

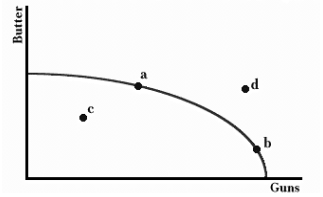


A budget line separates what is affordable from what is not affordable

You should be able to see that the **slope of the budget line depends only on the price of books relative to the price of movies.** If either books get cheaper or movies get more expensive, the budget line in the graph above will get steeper. If this is not immediately obvious, compute the possibilities open to a you with Rs.100 to spend if books and movies both cost Rs.5.00 (a case of more expensive movies), and the possibilities open to you with Rs.100 to spend if books and movies both cost Rs.2.50 (a case of cheaper books). Graphing the possibilities open to you with only Rs.50 to spend but with books costing Rs.5.00 and movies costing Rs.2.50 gives you a line that is to the left of the line in the graph above, but parallel to it, which means that it has the same slope. The amount of money available to spend does not determine the slope of the budget line; only the ratio of prices does that.

A famous example of a budget constraint is the case of guns versus butter. During the Second World War, the United States decided it needed to produce large amounts of armaments (guns). It shifted factories that previously produced goods for civilian use (butter) to the production of guns. This tradeoff could be represented as a move from a point such as **a** to a point such as **b** in the graph below, except that at the start of the war there was still a high level of unemployment left over from the recessions of 1929-33 and 1937-8 (a period better known as the Great Depression). Hence, the United States was not at the limit of what it could produce, but rather at a point such as **c,** which indicates that more of all goods could have been produced given the amount of resources and technology.

More for defense means fewer consumer goods



Though point **d** was a more desirable position than points **a** or **b**, it was unattainable given technology and resources. The limit to what is possible to produce is called the production-possibilities frontier. Its existence, which is a result of scarcity, indicates that there are costs to producing all goods and services. During World War II, the cost of producing thousands of tanks and jeeps was the virtual elimination of production of autos for civilian use. The cost of feeding millions of troops in the field was a less attractive diet for the civilian population.

The major idea in this section has been that all economic activity takes place within limitations or constraints. Because of these constraints, choosing results in sacrificed options. The options that are not taken, which sometimes can be measured in monetary terms, are costs.

Next, we combine the utility function and budget line to get ***utility theory***.

7.2 Utility

A choice involves deciding in favor of one option and discarding others. A budget constraint limits the options from which people can choose. To make the best decision, a person must choose the option that is both possible and that contributes most to the achievement of that person's goals. This section analyzes how people can make such choices.

Though it is easy to show the budget constraint with a table or graph, showing goals is a bit more difficult. For the purpose of illustrating some important ideas, this section will assume that ***goal-attainment can be measured in some unit of satisfaction or utility***. The table below gives an example by using an imaginary measurement called the **util.**

Benefits Measured in Utils		
Amount	Utils from Shirts	Utils from Hamburgers
1	11	8
2	20	15
3	27	21
4	31	26
5	32	30

Our second table expands the first to show utility for various combinations of shirts and hamburgers. Thus, one shirt and three hamburgers give 32 utils of satisfaction (because 11 utils from shirts + 21 utils from hamburgers equals 32 utils). The person gets the same level of satisfaction from five shirts and no hamburgers. The person whose wants are described in this table should find these two combinations of equal value, or, to anticipate a term, he will be **indifferent** between them.

A Utility Function						
Number of Shirts						
5	32	40	47	53	58	62
4	31	39	46	52	57	61
3	27	35	42	48	53	57
2	20	28	35	41	46	50
1	11	19	26	32	37	41
0	0	8	15	21	26	30
.	0	1	2	3	4	5
.			Number of Hamburgers			

The consumer wants to get as much utility as possible, but a budget constraint limits him. The table above the budget constraint is drawn so that the person can have only five items. Looking at all combinations possible, that is, to the left of the budget constraint (the numbers in red), one can see that the combination three shirts and two hamburgers maximize utility. This combination yields 42 utils, and no other combination that is allowed by the budget constraint gives more.

This simple problem can be solved in another way, with the maximization principle. The advantage of the second solution is that it gives insight into a whole range of problems.

The consumer wants to get as much utility as possible, but a budget constraint limits him. The table above the budget constraint is drawn so that the person can have only five items. Looking at all combinations possible, that is, to the left of the budget constraint (the numbers in red), one can see that the combination three shirts and two hamburgers maximize utility. This combination yields 42 utils, and no other combination that is allowed by the budget constraint gives more.

This simple problem can be solved in another way, with the maximization principle. The advantage of the second solution is that it gives insight into a whole range of problems.

The Maximization Principle

Often, it is impossible or difficult to list all the options and the budget constraint as the last section does. There can be simpler ways to approach this problem. The overview suggested that we could break the question into a series of questions. We begin by assuming that all money is used to buy shirts, which, in this example, means that the person buys five shirts. Then, we ask whether the person is better off buying one hamburger than buying none. To answer this question, we need to compute the costs and benefits of making this change. Using the example from the previous section, the added benefits of the first hamburger is eight utils. To compute the cost of making this change, we must remember the budget constraint. To get a hamburger, the person must sacrifice a shirt. Because he began with five, he will be left with four. As a result, the utility he gets from shirts will decline by one util, and this is the cost of adding a hamburger. Since the change adds benefits of eight utils at a cost of one util, this is a smart change to make.

Benefits Measured in Utils (from last section)		
Amount	Utils from Shirts	Utils from Hamburgers
1	11	8
2	20	15
3	27	21
4	31	26
5	32	30

Now, we ask if another change is worthwhile. What will the costs and benefits of adding a second hamburger be? The table says that the utility of two hamburgers is 15. Because eight of those utils come from the first hamburger, the added utility of the second hamburger is seven. Because the budget constraint forces the person to give up his fourth shirt in order to obtain this hamburger, utility from shirts will drop from 31 utils to 27 utils, a loss of four utils. Thus, the benefits of adding the second hamburger are seven utils, and the cost is a loss of four utils. Adding the second hamburger is also a smart move because it increases total utility.

The third hamburger is not worth obtaining. The benefit of adding the third is six utils (moving from 15 to 21 utils in the table). But this move requires the person to move from three shirts to two, and in this move, seven utils from shirts are given up. Because the cost of adding the third hamburger (seven utils) is greater than the benefits of this hamburger (six utils), the person should not add it.

Economists call the approach taken in the preceding paragraphs the **marginal** approach. Thinking on the margin means that a person is asking what the effects of small changes will be. In this approach one considers **marginal costs** and **marginal benefits.** The marginal cost of a change is the change in costs caused by the change. The marginal benefit of the change is the change in benefits caused by the change. The marginal approach suggests that one should make all the changes that increase benefits more than they increase costs (or that reduce costs by more than they reduce benefits). When all these changes have been made, one will find oneself at a point for which marginal costs equal marginal benefits. This rule for finding the best level of an activity is called the **maximization principle.**

Costs and Benefits of Hamburgers					
Number of Hamburgers	Marginal Benefit of Hamburgers	Marginal Cost of Hamburgers	Total Benefit of Hamburgers	Total Cost of Hamburgers	Net Benefit
1	8	1	8	1	7
2	7	4	15	5	10
3	6	7	21	12	9
4	5	9	26	21	5
5	4	11	30	32	-2

To see that the maximization principle does generate the largest net benefits, the problem of how many hamburgers to buy can

be analyzed with total costs and total benefits. This analysis is illustrated in the table above. Columns two and three show marginal costs and benefits, and the way in which they were obtained has been described in the previous paragraphs. Total benefits of hamburgers are taken from the first table.

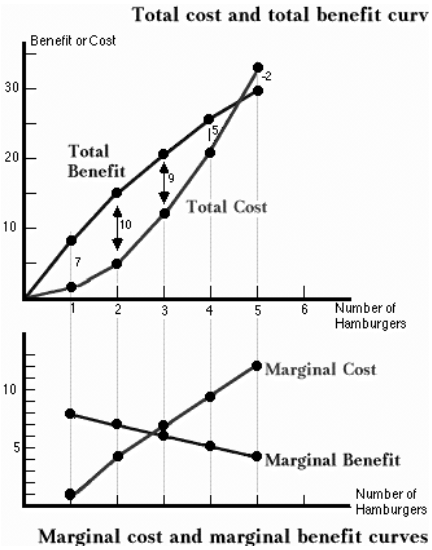
Total Costs are obtained from column two of the first table and depend on the budget constraint. The total cost of three hamburgers, for example, will be the lost utility of three shirts. Because five shirts give 32 utils, and losing three leaves only two giving 20 utils, the total cost of three hamburgers is 12 utils. The Net Benefit column in the second table is found by subtracting total cost from total benefit. At two hamburgers, the total utility will be ten utils higher than at the starting point of five shirts and no hamburgers.

You should see that if one has total cost one can obtain marginal cost, and if one has total benefit one can obtain marginal benefit, and vice versa. The formula for marginal cost is:

Marginal Cost = (Change in total cost)/(Change in activity)

Thus, if a business knows that the total cost of producing 98 shirts is Rs.398 and the total cost of producing 100 is Rs.400, the marginal cost of the 100th shirt is approximately Rs.2/2 = Rs.1.00. Notice that marginal cost is not the same as average cost, which is found by dividing total cost by output. Alternatively, if one knows the marginal cost or benefit, one can find the total cost or total benefit by adding up all the marginals. (Check the second table to see that this is so.)

These results can also be shown graphically. In the picture below the total costs and benefits from the second table have been graphed. The goal of the person, to maximize net benefits, requires that the person try to find the point where the total benefit curve is at its greatest vertical distance above the total cost curve. (Here is one case in which the person does not want to end up at the intersection. Can you see why?) At this point, the total cost and total benefit curves have the same slopes. Before this point, the total benefit curve is steeper, so they are moving apart. After this point, they are moving together, which means that the total cost curve has the steeper slope.



Marginal cost and marginal benefit curves

The slope of the total benefit (or cost) curve is the rise over the run, or the change in total benefit (or cost) divided by the change in the number of hamburgers. But the marginal benefit (cost) of hamburgers is also defined as the change in total benefit (cost), divided by the change in the number of hamburgers. Hence, the slope of the total benefit curve is marginal benefit and the slope of the total cost curve is marginal cost.

This idea is used to construct the marginal benefit and marginal cost curves in the bottom of the picture above. The marginal curves are obtained by graphing the slopes of the total curves. The point at which they cross corresponds to the level of activity for which the slopes of the total cost and total benefit curves are equal.

We have not exhausted the insights from this simple problem. We can also analyze the numbers with the equimarginal principle.

7.3 The Equimarginal Principle

At this point, you may think we have exhausted all the insights we can get from the hamburger-shirt problem. We have not. The table below contains columns showing the marginal utility of shirts and the marginal utility of hamburgers. These marginal utilities are obtained from our original example, which shows the total utility of one shirt, two shirts, etc. Marginal utility is the utility of the first shirt, the second shirt, etc. Thus, the utility of the fourth hamburger is found by subtracting the utility of four hamburgers from the utility of three hamburgers. Notice that the marginal utility of each good declines as more of it is used. This is a case of diminishing returns that has the special title of "**the law of diminishing marginal utility.**" It is based on everyday observation and introspection. After four beers, a fifth gives less pleasure than the fourth, a third hamburger gives less satisfaction than the second, etc.

The Equimarginal Principle, or How to Spend Your Last Rupee		
Number	Marginal Utility of Shirts	Marginal Utility of Hamburgers
1 (first)	11	8
2 (second)	9	7
3 (third)	7	6
4 (fourth)	4	5
5 (fifth)	1	4

Suppose that the person is not at the optimal solution of three shirts and two hamburgers. Suppose instead that he has two shirts and three hamburgers. Can we tell from the table that he has spent his money incorrectly?

We can. Shirts and hamburgers cost the same. Suppose that each costs Rs.1.00 and the person has Rs.5.00 to spend. Then the last Rs. spent on hamburgers gave the person only six utils, whereas the last Rs. spent on shirts gave him nine utils. The Rs. spent on shirts gave a much larger return, and if he could shift money from the area in which it is giving a low return to the area in which it has a high return, he will be better off. This is the basic idea of the **equimarginal principle**. Maximization occurs when the return on the last Rs. spent is the same in all areas. In terms of a formula, a person wants

(Marg. Benefit of A)/(Price of A) = (Marg. Benefit of B)/(Price of B)

The power of this idea can be shown if we change the original problem. Suppose that the person still has Rs.5.00 to spend, but the price of shirts doubles from Rs.1.00 to Rs.2.00. The old solution of three shirts and two hamburgers will no longer be affordable but will lie to the right of the budget line. To solve this new problem, two new columns must be added to our table: the marginal utility of shirts per Rs. and the marginal utility of hamburgers per Rs.. The table below adds them in columns MUs/(Price of Shirts) (the marginal utility of shirts divided by the price of shirts) and MUh/Price of Hamburgers).

The Equimarginal Principle, Continued

Number	Marginal Utility of Shirts	MUs / Price of Shirts	Marginal Utility of Hamburgers	MUh / Price of Hamburgers
1 (first)	11	5 1/2	8	8
2 (second)	9	4 1/2	7	7
3 (third)	7	3 1/2	6	6
4 (fourth)	4	2	5	5
5 (fifth)	1	1/2	4	4

The equimarginal principle tells us to maximize utility by selecting the highest values in the columns giving marginal utility per Rupee until our budget is used up. A person with only two Rupees should buy two hamburgers rather than one shirt because both eight and seven are larger than five and one half. A person with Rs.5.00, as in our example, should buy three hamburgers and one shirt. This decision does not quite equalize returns on the last Rupees spent on shirts and hamburgers, but it comes as close as possible. Any other combination would give less utility and would allow for further improvement. For example, if one bought two shirts and one hamburger, the extra satisfaction from a Rupee spent on shirts is only four and one half utils, whereas shifting money to hamburgers would allow one to get seven utils per Rupee.

7.4 Diminishing Marginal Utility

This illustrates a general principle that has much wider application in economics. In economics, we speak of a law or principle of diminishing marginal utility.

The "Law of Diminishing Marginal Utility" states that for ***any good or service, the marginal utility of that good or service decreases as the quantity of the good increases, ceteris paribus. In other words, total utility increases more and more slowly as the quantity consumed increases.***

This is "diminishing returns" from the viewpoint of the consumer, and is a general principle of economics. There might be a threshold before the principle applies. For example, the marginal utility of golf clubs might increase until you have a fairly full set. But *beyond some threshold*, marginal utility will diminish with increasing consumption of any good.

As we will see, there are other applications of "diminishing (marginal) returns" in other branches of microeconomics.

Here is another example of **Diamond & Water,** this is known as Paradox of Diamonds and Water: The marginal utility approach implies that when one commodity is very common, and the other is very scarce, a person would have good reason to pay more for the scarce good. The reason is that marginal utility for any good diminishes as the person consumes more of the good. Thus, if a good is scarce, the average person consumes only a little of it, and the marginal utility is (relatively) high. If the good is plentiful, the average person will have more of it, and so the marginal utility will be (relatively) low.

Of course, it was known that a person will pay more for a scarce good, and that's a matter of "common sense." But that "common sense" fact didn't sit well with Smith's idea of a "natural price" — it seemed to be an exception of some sort. When economists switched to the "marginal utility" approach, the commonsense fact that people will pay more for scarce goods no longer seemed exceptional. Instead, it is a central point of the theory of demand.

To make the idea clear, we need to say one more thing. In everyday life, marginal utility depends on the average consumption over a period of time. Take, for example, my marginal utility of pizza. After a couple of slices on a Thursday evening (Thursday is often pizza night) I won't have any more pizza, usually until next week. So my average rate of pizza consumption is roughly two slices a week. We could say that my marginal utility of pizza fluctuates over the week, but for practical purposes, it makes more sense to say that my marginal utility of pizza, per week, depends on the number of slices of pizza I have per week. Marginal utility in that sense determines what I'm willing to pay for pizza.

That is, if we use the marginal utility interpretation. But there are still some problems with marginal utility thinking, of course. The basic assumption was that the "**Consumer is Rational".** Next we see theory in practice with rational ignorance.

7.5 Rational Ignorance

Marginalism is an application of the basic idea of calculus, and though calculus was invented a century before Adam Smith, it was a century after Smith when economists realized its significance. This "marginalist revolution" greatly clarified economic theory. The better understanding of their theory has prompted economists to search for new areas in which to use it. We conclude this section with a visit to an area that economists have explored fairly recently.

A person purchasing a new car usually spends time learning about various makes of cars and shopping for prices. The more effort spent in these activities, the more one's knowledge about cars and their prices increases. Because time is limited, and spending time searching for information means that one cannot use that time for other purposes, there is a limit on how much knowledge is worthwhile to gain. After some amount of reading, talking to friends and acquaintances, and visiting automobile dealers, a person finds that the extra benefit of another hour spent on these activities is less than the value of that hour spent in other pursuits. When one judges that this point has been reached, one stops searching and makes a decision.

The amount of time people spend obtaining information differs from product to product. They will spend less time learning about the bicycle they give their child than they will learning about a new car, less time deciding which brand of soup to buy than in deciding which house to purchase, and less time deciding which brand of dog food is best for Rover than

in finding a college for their first-born. The larger the purchase, the larger the potential benefit of a few hours spent learning about the purchase.

The government has many policies that involve major sums of money. For example, a major weapons system in the defense department can cost Rs.50 billion. This amounts to about Rs.200 for every person in the United States, or Rs.1000 for a family of five. Yet few people spend much time studying these policies. A reason is that to understand them requires many hours of study, and the probability that an understanding of them will change them in any way is very small. Thus, for most citizens the benefit of learning about a program that does not directly affect them is small, the cost is large, and they end up not knowing much about the program. Economists say that these poorly informed citizens are **rationally ignorant.**

The phenomenon of rational ignorance is not confined to political affairs. There is vastly more to know than any one person can possibly know. To survive and prosper in the world, one must seek that knowledge which will benefit the seeker. Most people would consider someone a bit odd who was not planning to buy a car, but went from dealer to dealer trying to learn all he could about relative car prices in the name of intellectual curiosity. The behavior of most citizens suggests that they also consider odd the seeking of in-depth knowledge about the pros and cons of a specific government policy if that knowledge does not directly benefit the person who gets it. The hypothesis of the rationally ignorant voter suggests that people will be better informed about the choices they make in the marketplace than about those they make in the voting booth.

A look at costs and benefits not only explains why few citizens understand the subtleties of most government policies, but it also explains why about one half of the eligible voters in the United States do not vote. The probability that one's vote will be the crucial vote that decides an important election is small. Even if one's vote is the crucial vote that breaks the tie, one may not like the outcome—many people regret the way they voted when they compare actual performance with campaign promises. Given these small benefits compared to the costs of time and transportation that voting entails, it is not surprising that many people who are eligible to vote do not. What is surprising is that the percentage of people voting is not even smaller. It seems likely that there are other benefits to voting that have not yet been mentioned.

Politics is in many ways a spectator sport, with all the excitement and drama of football or baseball. Voting may be enjoyable in the same way as watching and cheering on a favorite ball team. Indeed, voting against a politician one does not like is enjoyable, even if it does not result in his defeat. Another explanation for voting is that people have a sense of public duty. They want to be good citizens, and voting may seem important regardless of its effect—the act of voting itself can be important as a symbolic act. One other possibility is that people may overestimate the importance of their vote and the probability that theirs will be the ballot that decides an election.

In contrast to elections in the United States, elections in the old Communist-bloc nations were predictable. There was no doubt about who would win. Yet, these countries reported impressive percentages—sometimes more than 99%, of their citizens voted. Anyone who understands how to reason in terms of costs and benefits should be able to explain the implications of very high participation rates in meaningless elections.

Keep the rationally ignorant voter in mind when interpreting polls that ask citizens their opinions about complex public issues. The idea that voters are rationally ignorant also has implications for how governments work.

7.6 Marginal Utility and Demand

The marginal utility approach resolves the "paradox of diamonds and water." There is no paradox: the scarcer good, diamonds, have the higher marginal utility, even though water gives the greater total utility. This opens the way to develop a theory of demand based on utility. But we aren't there yet.

Demand is a relation between **money** price and quantity purchased. So far, using our example, we have seen why a person might give up a large amount of water to get a diamond. That's not quite the same as giving up a large amount of money for a diamond. So one step we need to take is to translate from the barter of goods for goods to the exchange of goods for money. Let's check that out, in next lesson.

Notes

LESSON 8:
CONSUMER BEHAVIOUR – II PREFERENCE APPROACH

This lesson applies and extends the analysis of the rational consumer choice. It begins by abandoning the assumption that utility can be measured. Instead, a set of indifference curves represents a person's preferences, and that person strives to find the point on the highest indifference curve that his budget allows. The maximization and equimarginal principles are still there, but behind the scenes.

We then consider some implications of budget-line and indifference-curve analysis. The simple idea that constraints hurt leads to the important concept of present value. We also discover the concept of consumers' surplus and see how this concept dissolves the paradox of value. Finally, we add the concept of producers' surplus, and see how fights over the division of surpluses can reduce the total.

8.1 Preference Approach

Economists, like many other people, have been a bit skeptical about the idea that a person's satisfactions could be measured in a number, as the "utility" idea assumes. Twentieth-century economists have usually thought instead of "preference." Surprisingly, perhaps, the discussion of "consumers' preferences" can get quite technical and mathematical. We will instead take it at an intuitive level to get the flavor of the idea.

Think of a restaurant that sells barbecued chicken wings by the wing and french fries by the piece. The prices will be Rs.45 a wing and Rs3 a piece of fried potatoes. (I don't know of a restaurant that sells wings and fries this way, but it will be ok for the example. Let's consider some alternative menus that you could choose: no wings, one wing, two wings, and no fries, fifteen pieces of fries, or thirty pieces of fries. Taking all possible combinations, we have 3x3 = 9 alternative lunches. Utility thinking says that each combination will give you a definite amount of utility. The preference approach says that, while your satisfaction from consuming wings and fries may not be measurable as a number, you will be able to say whether he prefers two wings and fifteen pieces of fries to one wing and thirty pieces of fries. In general, you will be able to rank the alternatives as more or less preferable. Let's suppose your ranking of the nine alternatives looks like this: Table 8.1

		wings		
		0	1	2
	0	eighth	seventh	fifth
fries	15	sixth	fourth	third
	30	fourth	second	first

This ranking illustrates some ideas from the preference approach.

- First, preference is an order-ranking, not a number. This ranking from first preference to ninth applies specifically to these nine alternatives. If we were to consider more alternatives, the rankings might change, but in relative terms, they would be the same — no wings and 15 fries will always be ranked ahead of one wing and no fries.
- Second, your preferences are applied to combinations of the two goods. It is not that you prefer wings to fries. Rather, you prefer one wing with thirty fries to two wings with fifteen fries, and also prefers one wing with fifteen fries to two wings with no fries. These combinations are often called "market baskets" in economics, with the idea that the basket contains specific amounts of two or more goods.
- Third, given the amount of one good, more of the other good is preferred to less. For example, if you have one wing, 30 fries rank higher (second) than 15 (fourth). There could be limits to this, of course. If you are choosing the menu for a single meal, he might get full and prefer less to more. But, in more realistic examples, there will always be other goods beside wings and fries that you can spend your money on. So, even if you couldn't possibly eat another wing or another fry, there will be some goods that he does prefer more of, rather than less. That's good enough.
- Fourth, notice that 1 wing and 15 fries is in a tie with no wings and 30 fries for fourth place. When two alternatives come up with the same ranking, we say the consumer is "indifferent" between them, and that the two alternatives are "indifferent choices" or "indifferent alternatives." This "indifference" relationship is not something to be "indifferent" about! It proves to be a very useful idea in the preference approach.

8.1.1 Spending Decision

Now, let's see how your preferences influence your spending. We'll need to keep using Table 8.1:

Suppose that you have Rs.1.35 to spend for your lunch. You can afford 2 wings and 15 fries or 1 wing and 30 fries (or one wing and 5 fries or one wing and no fries or no wings and 15 fries). These alternatives rank third and second, so your "rational" choice is 1 wing and 30 fries. This is the choice that you most prefer, within your budget. By choosing it, we might say, you are "maximizing your preference." That's an awkward phrase, but it will have to do: in the preference approach, we say that a rational consumer maximizes her or his preferences within the limit of her or his budget.

It's almost as simple as that. But, of course, there are some important details to keep in mind — because they help to bridge the gap between the preference approach and the marginal benefits approach

8.1.2 Relative Preference

Remember, the preference rankings are really relative to the alternatives you are considering. If the only two alternatives are 1 wing and 30 fries or two wings and 15 fries, then 1 wing and 30 fries ranks first and 2 wings and 15 fries ranks second among

those limited alternatives. What we know (and what you know) is the relative preference ranking between the two alternatives: 1 wing and 30 fries is preferable 2 wings and 15 fries. Let's see what happens if you are allowed to consider still more alternatives — he can choose 0, 1, 2, or 3 wings. We'll show that in a new table, Table 8.2; but to make it easy for you to compare them we will put the two tables side by side:

Table 8.1

		wings		
		0	1	2
	0	eighth	seventh	fifth
fries	15	sixth	fourth	third
	30	fourth	second	First

Table 8.2

		wings			
		0	1	2	3
	0	tenth	Ninth	seventh	sixth
fries	15	eighth	Sixth	fifth	third
	30	sixth	Fourth	second	First

Since we have more alternatives to rank, some of the ranks are different, but the relative rankings are the same. Check that: for example, 2 wings and 15 fries ranks below 1 wing and 30 fries in each table.

But we have more information In Table 8.2. We now know that 2 wings and 30 fries ranks above 3 wings and 15 fries, which we didn't know before.

8.1.3 Indifferent Behaviour

Let's take a closer look at Table 8.2:

		wings			
		0	1	2	3
	0	tenth	ninth	seventh	sixth
fries	15	eighth	sixth	fifth	third
	30	sixth	fourth	second	first

Looking again, we see another tie. Now we have a three-way tie. No wings and 30 fries, 1 wing and 15 fries, and 3 wings and no fries are all tied for 6th place. These three menus, together, form what preference theory calls an ***"indifference curve" — a linking of all the combinations of goods and services that come up with the same ranking in a person's preference ranking.***

This "indifference" conception helps us to relate the preference approach to the marginal-benefit approach. What is the marginal benefit of the first wing? We can get that by traveling along the indifference curve corresponding to sixth place. By definition, the marginal benefit is the money value of the other goods you would give up to get that wing. Since you are indifferent between no wings and 30 fries (on the one hand) and 1 wings and 15 fries (on the other hand) we can conclude that You would give up 15 fries to get that first wing — no more and no less. So yours marginal benefit from one wing is the market value of 15 fries, that is, 45 cents. What is the marginal benefit of the second wing? We don't know, exactly, from this information. We would need to try more alternatives until we find one that ties in rank with 1 wing and 15 fries. But we can approximate, using the formula MB ≅ Δ benefits/Δ wings. Notice that you would give up all 30 fries to get all 3 wings. So the total benefit of 3 wings is the market value of 30 fries, 90 cents. Moving from 1 wing and 15 fries to 3 wings and no fries, we have Δ benefits of 90-45=45 cents and wings of 2, so MB ≅45/2 = 22.5 cents. We see that you experiences diminishing marginal benefits of consuming wings — just as we would think. Notice, we are still applying the good old "opportunity cost" concept. To say that you are indifferent between no wings and 30 fries (on the one hand) and 3 wings and no fries (on the other hand) is to say that 30 fries is an opportunity cost you are willing to give up to get 3 wings.

There is one thing we need to be cautious about. It turns out that the marginal benefit will be different if we start out from a different place. For example, suppose we started out with just 15 fries and no wings. We can see that you would NOT give 15 fries for that first wing — that would reduce him from eighth place in his preference ranking to ninth. So the marginal benefit of the first wing will be somewhat less than the 45 cents it was when you could start from 30 fries. This should not be a surprise, though. We can look at it from three points of view, and they all agree that the Marginal Benefits should depend on the starting point. First, starting from just 15 fries, you are going to be hungrier, so it might make sense if he wants to fill up a little more on fries. Second, you are richer in the first example than in the second — starting from 30 fries is starting from 90 cents, while starting from 15 fries is starting from 45 cents. The richer people are, the more they usually are willing to pay for the goods they buy — and we are measuring benefits in terms of the person's willingness to pay. Finally, we remember that the marginal benefit curve is the individual demand curve. When a person's income or wealth drops, their demand for most goods and services will decrease, and that's what has happened here — starting off from less income, the demand for wings is less.

8.1.4 Key Points of Preference Approach

Of course, this example is unrealistic. The numbers are too small — "real men" order wings by the dozen. Much more important, we have many more alternatives than we could list in a table, and we have to choose among combinations of many more than two goods, and time plays a role, so that I can average my dozen wings this week with my two dozen next week to consume an average of 1.5 dozen a week. All the same, the example has illustrated some key points about the preference approach:

- We don't need numerical measures of utility. It is enough if consumers can rank the alternatives they face in terms of better and worse, or first, second, third and so on.
- We know that people can do that, because they do it when they choose among those alternatives. By choosing they reveal their preferences.
- The alternatives are not the goods and services themselves, but different combinations of goods and services — different lunches or "market baskets."
- The rankings have to be consistent in several ways: more is preferred to less (perhaps up to a limit), and any two market baskets are ranked in the same way no matter what other alternative market baskets are included in the ranking. There are some other, more technical consistency requirements that we will not go into now.

- All the same, two or more market baskets can be ranked as a tie. Then we say the consumer is "indifferent" between them. That doesn't mean he doesn't care about the goods and services — it means he will give up one of the two market baskets willingly in exchange for the other. This is important because it helps us to apply the opportunity cost approach to estimate his marginal benefits and thus his demand for any good or service.

Perhaps this is enough to make it clear that we really can do it — we can express an individual's subjective benefits from consuming goods and services in terms of money. And that was a big step forward in the Reasonable Dialog of Economics.

This is, of course, no more than a taste of the preference approach. A good intermediate microeconomics course would go into a lot more detail, with some very careful diagrams. I'll go on with an application of preference thinking to equity in income distribution and job assignments — a topic economists don't often get into and, frankly, pretty far afield from most other texts — and then come back with the basics of the intermediate level, diagrammatic approach to preference, which is covered in a good many introductory texts as well. But, in Essential principles of Economics, these are all advanced topics, and you can skip out and go on to the next chapter any time without losing the thread.

8.2 Examples

Lets take some more examples to make the point clear:

1) We consider a very small economy consisting of two persons, Grasshopper and Ant, two jobs, a hard job and an easy job, and an income that can come in two sizes: large and small. The institutions of the society (the "rules of the game") link the large income to the hard job and the small income to the easy job. We suppose that Grasshopper is a bit of a lazybones. Grasshopper's preferences among jobs and incomes is shown by Table 8.3:

Table 8.3. Grasshopper's Preferences

		income	
		large	small
job	easy	1	2
	hard	3	4

That is, an easy job with a large income is Grasshopper's first preference (naturally enough) and the easy job with a small income comes next, the hard job with a large income next yet, and (again naturally enough) a hard job with a small income ranks lowest. Ant's preferences are different and are shown by Table 8.4:

Table 8.4. Ant's Preferences

		income	
		large	small
job	easy	1	3
	hard	2	4

At the extremes, Ant ranks the alternatives in the same natural way as Grasshopper does, an easy job with a large income first and a hard job with a low income last. In between, however, Ant ranks the other two choices in the opposite way, choosing a hard job and a large income over the reverse.

Now suppose that Ant is allocated the hard job, and the large income that comes with it, and Grasshopper is allocated the easy job with its low income. Then Ant has his second preference, while Grasshopper's allocation is Ant's third preference. Ant would not choose Grasshopper's portion over his own. Conversely, what Ant has is Grasshopper's third preference, and Grasshopper has his own second preference, so Grasshopper would not, either, choose Ant's portion if he could. Since each insect has a job-and-income package he *positively* prefers over the package the other insect has, the allocation between the two insects is said to be*superfair*. In general, if each insect were *indifferent* between her own package and the package the other insect enjoys, then, in superfairness theory, the allocation would be described as "*fair;*""when each positively prefers her own package, then the allocation is "superfair."

Suppose instead that by accident Ant had been assigned the easy job and Grasshopper the hard job. Now Grasshopper has his own third choice, but Ant has Grasshopper's second choice: Grasshopper "envies" Ant. Conversely, Ant has his own third choice, but Grasshopper has Ant's second choice: Ant "envies" Grasshopper. There is inequity all around. But the inequity is easily remedied by a market transaction: given the opportunity, Ant and Grasshopper will voluntarily exchange jobs. Then both superfairness and efficiency are established. In this simple economy, a free market equilibrium is superfair.

2) Now we consider the same small economy except for one change: the rules will be different, and a hard job will be associated with a small income, while an easy job has a large income. Perhaps that is because the easy job is highly assisted by technology, while the hard job is not. Anyway, in this small economy there is no fair or superfair assignment of jobs. No matter who gets the easy, high-pay job, he has the other insect's first choice, and the other insect has his own fourth choice. There is no free-market switch that will eliminate the inequity — switching jobs just changes the victim. We could say that it is the economic system itself — the rule for assigning jobs and incomes — that is inequitable, since, with a system like that, there can never be a fair assignment of jobs.

3) Now consider one more variation on the same small economy. This time we will again associate the large income with the hard job, but there are two Ants in the population and no Grasshopper. Let's call the two Ants Adam and Hillary. One of the Ants will have to be assigned the easy job/small income bundle. Let us suppose it is Adam. Adam finds that he is stuck with his own third preference, while Hillary has Adam's second preference. Adam "envies" Hillary and the allocation between them is inequitable. Switching the Ants will not help — one or the other of them will "envy" the other. Inequity is unavoidable in this example also.

4) In the second and third example inequity cannot be avoided in part because we have assumed that incomes come on only two indivisible sizes and are rigidly associated with effort supplies. To make the example a little less rigid, we might allow income to be divisible, while retaining the simple assumption that each job requires a fixed, larger or smaller, amount of

effort. Generalizing the first example above, we suppose that a Grasshopper will accept a hard job rather than an easy job on the condition that the hard job pays Rs.5 more, and an Ant will accept a hard job rather than an easy job on the condition that the hard job pays Rs.3 more. Let the "rules of the game" assign (by productivity?) an income of Rs.4 for an easy job and Rs.8 to a hard job. Otherwise, each of the two prefers more income to less. We have example 1 over again — the market assignment of jobs will be superfair.

5) Continuing with the preference valuations in paragraph 4, we generalize example 2, supposing that the "rules of the game" assign income of Rs.4 to a hard job and Rs.8 to an easy job. Once again, here is no fair or superfair assignment of jobs and incomes, since each insect must be assigned to his first or last preference. Suppose, however, that a benevolent economic planner assigns Grasshopper to the easy job and Ant to the hard job, then redistributes Rs.4 of the productivity-based income of Grasshopper to Ant. Ant now has a hard job and Rs.8 of income, while Grasshopper has an easy job and Rs.4. Ant does not prefer what Grasshopper has, net of the tax and transfer, since Grasshopper has Rs.4 less income now; and Grasshopper does not prefer what Ant has, since Ant has a hard job and only Rs.4 more income. Redistribution of income from the more productive to the less productive (but harder working) insect has restored equity in a case in which equity would otherwise be impossible.

6) Now generalizing example 3, we again suppose that the "rules of the game" assign income of Rs.8 to a hard job and Rs.4 to an easy job, but we have to allocate jobs between two Ants. Since only one can have the high-income job, one Ant will prefer the bundle the other has to her own bundle — inequity. Now, however, our planner assigns Hillary to the easy job and Adam to the hard job, the redistributes Rs.0.50 of Adam's income to Hillary. Now Hillary has the easy job and Rs.4.50 while Adam has the hard job and Rs.7.50. The difference in pay is just Rs.3.00, which makes both Ants indifferent between the hard job and the easy job with their associated incomes. The result is a fair (not superfair) allocation.

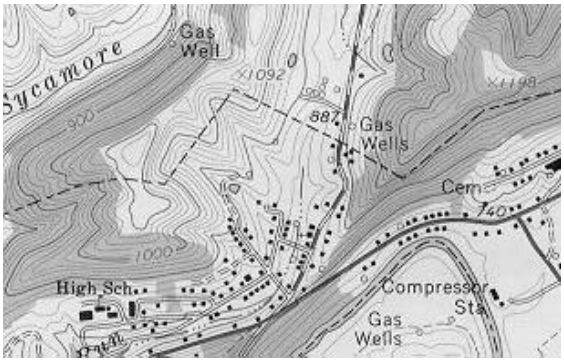
What examples 5) and 6) illustrate is that income redistribution may restore equity in a situation in which equity would be impossible without income redistribution. In some cases, market competition can lead to an equitable outcome, as example 1) shows. In some cases, market outcomes can be very far off from the neoclassical concept of equity, as example 5 illustrates — it was necessary in that case to redistribute half of Grasshopper's income to Ant to make the distribution of jobs and income equitable. In the real world, that much redistribution would distort the incentives to work and produce inefficiency. One economist who has studied these issues, Dr. William Baumol, believes that the loss of production involved would be very great indeed. In any case, it is clear that our real economy does not come very close to an equitable allocation of resources, and perhaps only limited progress in that direction is possible.

Once again, though, we can approach the study of economic equity without any numerical measures of "utility" — strictly through preference theory. That illustrates some of the wide applicability of preference theory in modern economics.

8.3 Indifference Curve

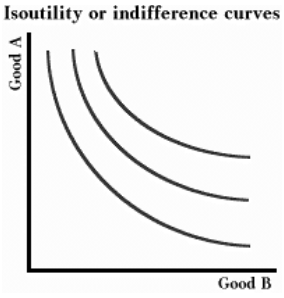
As usual, we can make these ideas more general and applicable if we visualize them with a diagram. One of the best ways to visualize a consumer's preferences is with an "indifference curve" diagram.

At first glance, using a graph instead of a table may not seem like a good way to proceed. Discussion of maximizing utility must involve at least three variables: the amount of good **A**, the amount of good **B**, and the level of utility. Graphs have only two axes, and three variables seem to require a three-dimensional construction rather than a two-dimensional one. However, there is a way around this problem, one that geographers use when they draw contour maps showing the three variables of longitude, latitude, and altitude. They show altitude with a series of lines or topographic contours such as those in the map below, which shows a hilly section of West Virginia.



The same method of construction can be used to show a utility map. A line will connect all possible combinations of good **A** and good **B** that show the same level of utility. This line is called an **isoutility** (**iso** is Greek and means "the same" or "equal") line or, more commonly, an **indifference curve.** In general, these isoutility lines will be curved, as in the graph below, if diminishing returns hold.

Figure 8.1



One does not need to measure utility in order to draw a graph such as that in the graph above. All one needs is the ability to **order** different levels of utility; that is, to say that bundle **A** is preferred to bundle **B**, or that bundle **B** is preferred to bundle **A**, or that the chooser is indifferent between the two bundles.

Indifference curves assume that individuals are consistent. If Jane prefers option **A** to option **B**, and if she also prefers

option **B** to option **C**, then she should prefer option **A** to option **C**. People with inconsistent behavior will not attain their goals as well as they could, and if behavior is too inconsistent, their behavior may show few regularities or predictable patterns.

Figure 8.2 is an Indifference Curve Diagram for wings and fries, partly based on the previous numerical example. The number of wings the person consumes is on the horizontal axis, and the number of fries is on the vertical axis. Thus, each point in the diagram corresponds to a particular number of wings and fries. For example, point A — with the asterisk — corresponds to 2 wings and 30 fries.

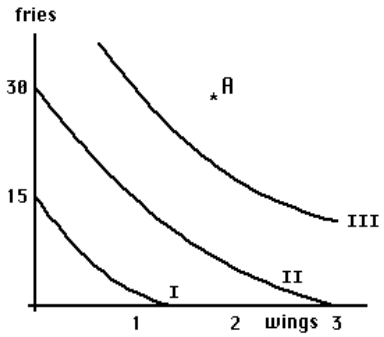


Figure 8.2: Indifference Curves for Wings and Fries

An indifference curve is a curve connecting points in the diagram in such a way that the consumer is indifferent between any two combinations shown by points on the curve. For example, consider the curve labeled II. It connects — among others — 30 fries and no wings, 15 fries and 1 wing, and 0 fries and 3 wings. We have already seen that these alternatives are tied for 6th place in the preferences, as shown in Table 8.2. Let's assume that all the other points on curve II also correspond with combinations of fries and wings (some in fractional amounts) that would all be tied with these three. Then Curve II is an indifference curve for this consumer. But it is not the only one, and in fact there are infinitely many indifference curve corresponding to any consumer's preferences.

We can always draw an indifference curve through the point in the diagram. Fractional quantities are OK. Remember, we may get fractional quantities when we average the person's behavior over time — so this is an advantage. We don't have to limit ourselves to whole numbers as we do in the numerical examples. If we choose two points, corresponding to two different combinations of wings and fries (or any other goods or services) either they will be on the same indifference curve or on two different indifference curves. If the same, then the person is indifferent between the two choices. If they are on different indifference curves, then one curve will be completely above the other, and that means the person prefers the combination on the higher curve.

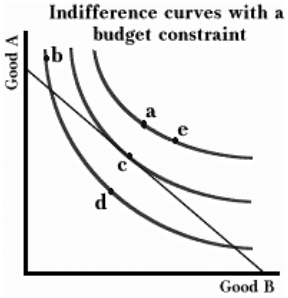
Next we see how demand curves come out of indifference curves and budget lines.

8.4 Indifference Curve & Demand

This section takes you a bit beyond where one needs to go in introductory economics, but it illustrates ***how indifference curves are used***.

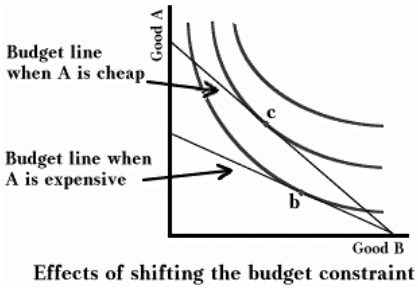
To show what the consumer should do to maximize utility, a budget line must be added to the preferences shown in the indifference curves. The picture below adds one. Point **a** is not attainable because it lies to the right of the budget line. The consumer is indifferent between points **b** and **d** because they lie on the same indifference curve, but point **d** is cheaper than **b** because **d** lies below the budget line. The consumer wants to get on the highest indifference curve affordable, and this will lead him to point **c.**

Figure 8.2



The effect of a rise in the price of good **A** is shown on the graph below. A higher price of **A** means that less of **A** can be purchased, and hence the budget line moves to the left, intersecting the vertical axis at a lower point. Point **c** is no longer possible and the consumer must move to a new position, which, assuming utility maximization, will be point **b**. Unless the indifference curves are peculiar, point **b** will represent less of good A than will point **c**, which is what the law of demand says will happen.

Figure 8.3



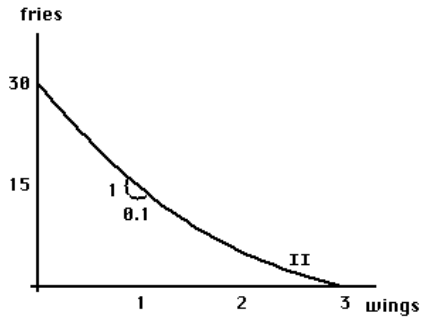
Effects of shifting the budget constraint

Looking at two different prices has produced two different points on an individual's demand curve. By varying the price of good **A**, other points could be found and an entire demand curve for one individual consumer constructed. The market demand curve is obtained by adding up the demand curves of all individuals.

8.5 Marginal Rate of Substitution

Let's step back and think about an indifference curve in a little more detail. Looking at Figure 4, below, we see that indifference curve II gets flatter and flatter as we move down it from the left to the right. The other indifference curves are similar — typically, the slope of an indifference curve changes, becoming flatter as we move from left to right.

Figure 8.4 : The Slope of an Indifference Curve



What is this telling us? At 1 wing and 15 fries, the slope of the indifference curve is 10 — that is, moving down the indifference curve, a reduction of one fry has to be compensated with 0.1 wing, in order to keep the consumer on the same indifference curve. (Remember, the scales on the two axes are different). As long as he is on the same indifference curve, the consumer is at the same level in his preference ranking — no better and no worse off. Putting it another way, the consumer is **willing** to give up one fry to get another tenth of a wing (since the exchange leaves him no better and no worse off). The slope of the indifference curve tells us something very important — how much of one good a person is willing to give up to get one more unit of the other good.

Economists have a term for the slope of an indifference curve. The term is the **"marginal rate of substitution."** What we have just seen is that, at 1 wing and 15 fries, the slope of the indifference curve is 10 — meaning the consumer would give up only one-tenth of a wing to get one more fry. To summarize:

Definition: The Marginal Rate of Substitution

The marginal rate of substitution between two goods is (minus) the slope of an indifference curve for the two goods.

Interpretation

The marginal rate of substitution tells us how much of one good a person will give up to get one more unit of the other good.

We have also seen that the marginal rate of substitution varies in a specific way. As the consumer has more of a good (moving from left to right) the marginal rate of substitution decreases (the slope gets flatter). This "law" of decreasing marginal rate of substitution brings to mind the "laws" of diminishing marginal benefit and marginal utility. Indeed, it is the same law in another form.

8.6 Income & Substitution Effect

Lets take an example of Potatoes as a Giffen Good, i.e. low priced – low quality product. Consumers are much better off with the lower price of potatoes that they buy more of other foods, and thus less of the inferior good, potatoes. Can we put this explanation to the test? Turning it around, the explanation says that, if people could buy potatoes at the lower price, but somehow they were no better off, then they would buy more potatoes, not less. (If it is being better off that reverses the law of demand, then we should see the law of demand back in full force if they were no better off).

With a lower price and the same money income, people are better off, in this example. But if the lower price were offset by a lower income, they would be no better off. There is a certain reduction of income that would leave the person neither better off nor worse of than before, even though the price is lower. That reduction in income is called the "compensating income variation" for the price drop. This is shown in Figure 5. To figure out the compensating income variation in the diagram, we just move the budget line N (for the new, low price) down to N', which just touches indifference curve I. Since the consumer is back on indifference curve I, she is no better off nor worse off than before — in terms of her own preferences.

With budget line N', we see that the consumer will buy quantity C of potatoes — more potatoes at the lower price. What we see is that the cut in price has increased the quantity demanded from A to C, but the increase in the purchasing power of income has cut the consumption of potatoes all the way from C back to B. So the testimony holds up — it seems that a Giffen good could be a reality in terms of preference theory.

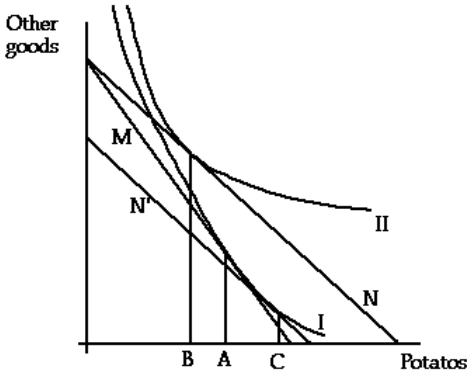


Figure 5: Income and Substitution Effect

8.7 Concluding Summary

The Reasonable Dialog on the theory of demand began with the idea that willingness to pay for a good or service might be related to the utility derived from the good or service — and quickly rejected it, since that idea seemed to be undercut by the "Paradox of Diamonds and Water." However, the alternative approach (the Labor Theory of Value) had its own problems, and nearly 100 years later, economists reconsidered the issue. Using the marginal approach — for the first time — economists of the 1870's realized that willingness to pay for **one unit** of a good or service would depend on the marginal utility, not the total utility. In the light of that understanding, the paradox was not a paradox at all. The marginal approach undercut the paradox, and restored the credibility of the utility approach. A "new economics" of consumer demand was created on the basis of marginal utility.

But there were other problems with the utility approach. Not everyone could accept the idea of direct numerical measurement of consumer satisfaction. What are the units? Are they the same for different people, or different, and if they are different for

different people, how can that be considered measurement? Some of the critics proposed an alterna

The theory of consumer choice that the indifference curves embody is an elegant construction with which economists frame problems. One of its weaknesses is that a great many outcomes are consistent with it—though a downward-sloping demand curve can be derived from it, so too can an upward sloping demand curve. Further, in recent years there has been a realization among economists that pictures such as those above may not be a good description of the decision-making process when people must make decisions with partial information, with fuzzy goals, under conditions of risk and uncertainty, and when options are difficult to compare. Finally, there do seem to be cases in which people systematically violate the rules that this theory says are rational.

Notes

LESSON 9:
CONSUMER BEHAVIOUR – III INDIFFERENCE CURVE

9.1 Indifference Curve Analysis

1. We can begin by examining the two good, single consumer case. Each consumer starts with a budget constraint, representing how one's income is spent on a set of goods and services. We'll assume that there are only two goods to consider in the typical consumer budget and that all of this consumer's income is spent on these goods.

The Budget Constraint is:

$I = P_1Q_1 + P_2Q_2$
(where I = income, P = price, Q = quantity for goods 1 and 2)

We can take this equation, rearrange it to get:

$Q_1 = -(P_2/P_1)(Q_2) + (I/P_1)$

What can we say about the rearranged budget constraint equation? First, we may notice that this rearranged budget constraint is an equation for a line (with a negative slope P_2/P_1 and vertical intercept I/P_1). Intuitively, we may recognize that the ratio of prices represents a comparison of the cost to consumers of one unit of each good. Therefore, in a sense, we can say that P_2/P_1 is the ratio of the marginal cost of goods 1 and 2 respectively. Recalling our macroeconomic discussion of price indexes, we see that I/P_1 is a measure of our good 1 purchasing power (i.e. how much of good 1 our income can buy). If P_1 falls, I/P_1 gets bigger - which means that we can buy more of good 1.

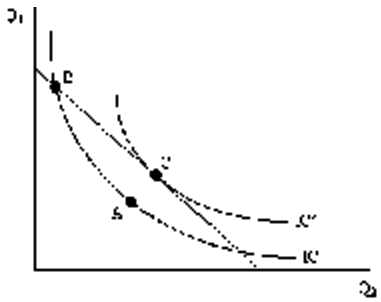
2. While the budget constraint represents how much a consumer is able to spend, we also need to know how much a consumer wants to spend on each good. That is, we need some information about this consumer's preferences regarding each good.

This information is found in an indifference curve. Indifference curves are drawn with two basic ideas in mind: (a) within certain limits, consumers always prefer more of everything to less (e.g. I'd prefer receiving 3 boxes of Cocoa Puffs and 2 boxes of Honeycomb to 2 and 1 box respectively); and (b) it is possible to derive the same satisfaction out of a variety of potential purchase combinations (e.g. when considering a potential cereal purchase, a consumer may be indifferent between buying 3 boxes of Cocoa Puffs and 2 boxes of Honeycomb versus 2 and 3 boxes respectively).

Therefore, by considering one's preferences, we see that consumers make purchasing decisions which depend upon the satisfaction (more formally, the utility) derived from a particular good. Each unit consumed (e.g. each box of cereal) in a given time period yields some sort of satisfaction. When we examine the amount of satisfaction derived from each unit consumed, we are considering something called marginal utility (MU). The slope of the indifference curve may be expressed as a ratio of the marginal utilities associated with each good (MU_2/MU_1). Rather than write this ratio, however, we can simplify things by calling it the marginal rate of substitution between goods 1 and 2 (MRS).

Where does equilibrium occur? Equilibrium occurs where the slopes of the indifference curve and budget constraint are equal. Mathematically, this occurs where $MRS = P_2/P_1$. This is an equilibrium point because at this point there is no reason to move away. The marginal rate of substitution can also be thought of as a ratio of marginal benefit that each good provides our consumer. Therefore, equilibrium in this setting involves equating the marginal benefit for two goods with their marginal cost. In simpler terms, we're saying that our consumer is getting out of each good exactly what they're worth.

We can demonstrate equilibrium graphically as well (see the graph below). Consider two different indifference curves: IC (the red curve) and IC' (the blue curve). Every point on IC (and IC') represents a different potential purchase of goods 1 and 2. As mentioned above, on each indifference curve our consumer is indifferent about purchasing any of the potential combinations along that curve. Consequently, along a particular curve, the only difference between each point is the amount of goods 1 and 2 that are purchased. The consumer is just as satisfied with any of the points on a given curve. Two things will determine which point gets selected: the consumer's income and the price of each good.



To find out where the equilibrium is, if one exists, we want to see if there is one point that is always prefered to every other point. We can begin by starting at a specific point (the one we pick isn't important). To keep things simple, we'll continue to assume that our consumer spends their entire income on these two goods. Start at point B, at the top of the Budget Constraint. Based on our discussion above, we know that points A and B provide this consumer with equal levels of satisfaction. That is, the consumer is indifferent between points A and B.

Although this consumer is indifferent between points A and B, this is not the case with points A and C. Point C is clearly better than point A for one important reason. At point C, our consumer gets more of both goods. As mentioned above, the basic idea behind these indifference curves (where each good's MU is greater than zero) is that "more is better". When comparing two points, like A and C, this is always true. When

you get more of one good but less of the other, it may be true but not necessarily so (e.g. our consumer is not better off when moving from A to B).

Thus far we know that our consumer is indifferent between A and B, but prefers C to A. Therefore, logic dictates that our consumer must also prefer C to B. No matter which point we start with, our result would be the same. In the end we realize that, if "all roads lead to point C", point C must be the equilibrium.

9.2 Indifference Curves and the Consumer Equilibrium

Let's assume that a representative consumer named Homer Simpson consumes beer and pork rinds in varying amounts. Assume further that the overall utility he derives from consuming these goods can be described by the utility function below. Note that this is just one possible example of a utility function, that there are many other possible functions we could have used instead.

$$(1)\ U(B,R)=\sqrt{B\cdot R}$$

We can use this utility function to derive Homer's indifference curve. By setting (1) equal to a specific number, we are saying that there are various combinations of B and R that yield a level of utility equal to that specific number. For example, suppose we set Homer's utility function equal to 100. We derive the indifference curve allowing 100 units of utility (i.e. utils) by rearranging the equation as

$$(1a)\ \sqrt{B\cdot R}=100$$

Now, solve (1a) for B by squaring both sides to get:

$$1(b)\quad B.R = 10{,}000$$

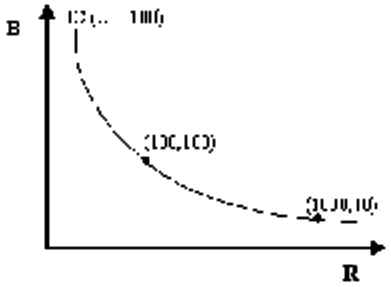
Second, we divide both sides of (1b) by the variable R.

$$1(c)\ \ B = 10{,}000/R$$

This is the equation for one indifference curve. As stated above, (1c) tells us the various combinations of beer and pork rinds that will provide Homer with 100 utils of satisfaction.

For example, if Homer consumes 10 units of beer, he needs to consume 1,000 units of pork rinds to get 100 utils of satisfaction. Of course, this equation also tells us that Homer would be indifferent between consuming that bundle of goods (10 units of beer and 1,000 units of pork rinds) and another one with 100 units of beer and 100 units of pork rinds. This is because both bundles provide 100 utils of satisfaction.

The graph that goes with (1c) is pictured below. The two different consumption points we just discussed are pictured too (with their coordinates reported as (R,B)). Both are on the indifference curve, both yield 100 utils of satisfaction.



Not knowing whether Homer will actually consume at either of these points, or whether he'll even consume on this indifference curve, we turn now to figuring out where Homer's consumption will actually occur. To do this we need a couple pieces of missing information: (a) the slope of the indifference curve, and (b) the budget constraint equation.

In a model where we examine two goods simultaneously, the slope of the indifference curve is going to be the marginal utility related to consuming more of one good divided by the marginal utility related to consuming less of the other good. While the utility along any indifference curve is constant, the marginal utility is not.

The marginal utility (MU) for each good above is given as:

$$MU_B = \frac{\sqrt{R}}{2\sqrt{B}}$$

$$MU_R = \frac{\sqrt{B}}{2\sqrt{R}}$$

The slope of the indifference curve, called the marginal rate of substitution, will be MU_R/MU_B. Note that the slope of this curve is negative (to see this mathematically, consider (1c)), which means we write the marginal rate of substitution for pork rinds and beer ($MRS_{R,B}$) as:

$$(2)\ \ MRS_{S,B} = -B/R$$

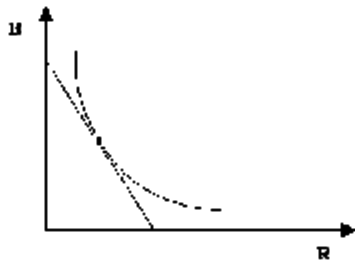
We'll assume that the price of beer is $4 and that the price of pork rinds is $2. Assume further that Homer's income is $200. The budget constraint is then given as:

$$(3)\ \ 4B+2R = 200$$

Rearranging (3), by solving for B, we get the following (rearranged budget constraint):

$$(3a)\ \ B = -0.5R+50$$

Noting that (3a) is the equation of a line (slope of –0.5, vertical intercept of 50), we can graph the indifference curve and budget constraint together. Equilibrium is attained where the (blue) indifference curve is tangent to the (red) budget constraint. This point is included in the graph.



The graph enables us to visually determine equilibrium, but also note the two conditions which must simultaneously occur when we are at this equilibrium point. Those conditions are:

- the slope of the budget constraint must equal the slope of the indifference curve (i.e. $MRS_{R,B} = -P_R/P_B$)
- our consumer must be on their budget constraint (i.e. 4B + 2R = 200)

With this in mind, we can now solve for equilibrium here. Substitute the values of the slopes into the first condition.

(4) $-B/R = -0.5$

Solve (4) for B.

(4a) $B = 0.5R$

Substitute (4a) into the budget constraint (for B).

(5) $4(0.5R) + 2R = 200$

Solve (5) for R. This is the equilibrium value for R (i.e. R*).

$R^* = 50$

Plug R* into the original budget constraint (or (4)), and solve for B. This is the equilibrium value for B (i.e. B*).

$4B + 2(50) = 200$

$B^* = 25$

Given Homer's budget constraint and utility function, Homer should consume 25 units of beer and 50 units of pork rinds. If he does this, then his overall utility will be:

$$\sqrt{25 \cdot 50} = 25\sqrt{2}$$

That is, Homer will experience about 35.4 utils of satisfaction from his 25 units of beer and 50 units of pork rinds.

9.3 Utility Max Application of the Implicit Function Theorem

Assume that a consumer named Homer Simpson consumes varying amounts of Duff beer and pork rinds

Let:

- units of beer consumed = B
- units of pork rinds consumed = R

Homer derives his utility from consuming these goods in accordance with the following utility function (where U = utility):

(1) $U = f(B,R)$

Homer's purchasing decision is limited by the following budget constraint (where p_i is the price of good i, and I is Homer's income):

(2) $_{PB}B + _{PR}R = I$

Note that (2) can be rearranged to become:

$$(2a)\ R = -\frac{P_B}{P_R} + \frac{I}{P_R}$$

Utility maximization leads us to the following equilibrium condition (which says that the slope of the indifference curve equals the slope of the budget constraint):

$$(3)\ \frac{MU_B}{MU_R} = \frac{P_B}{P_R}$$

(where MU_i = marginal utility of good i; which equals the derivative of the utility function with respect to good i)
Let us first take the total derivative of (1), the utility function. Upon doing so, we have:

$$(4)\ \frac{\partial U(B,R)}{\partial B}dB + \frac{\partial U(B,R)}{\partial R}dR = dk$$

(where k is a constant equal to some overall level of utility, such that $k \geq 0$)
Dividing both sides of (4) by dB yields:

$$(5)\ \frac{\partial U(B,\ R)}{\partial B}\cdot\frac{dB}{dB} + \frac{\partial U(B,\ R)}{\partial R}\cdot\frac{dR}{dB} = \frac{dk}{dB}$$

Because $dB/dB = 1$, and $dk/dB = 0$, we can simplify (5) to get:

$$(5a)\ \frac{\partial U(B,R)}{\partial B} + \frac{\partial U(B,R)}{\partial R}\cdot\frac{dR}{dB} = 0$$

Solving (5a) for dR/dB yields:

$$(5a)\ \frac{dR}{dB} = -\left(\frac{\dfrac{\partial U(B,\ R)}{\partial B}}{\dfrac{\partial U(B,\ R)}{\partial R}}\right)$$

At this point, we need to stop and ask what we've got thus far. In doing so, let's recall a couple of points made above. First, we note that the marginal utility of good i can be expressed as the first derivative of the utility function taken with respect to good i. Second, we note that an indifference curve's slope is equal (in the two-good case) to the ratio of the marginal utilities.

Because the righthand side of (6) involves the ratio of two derivatives of the utility function (each taken with respect to one of the goods consumed by Homer), the righthand side of (6) must be the slope of Homer's indifference curve. If the slope of Homer's indifference curve was set equal to the slope of his budget constraint, then we would have the consumer equilibrium expression given in (3).

To take the actual derivatives just mentioned, however, we need to assume a functional form for the utility function in (1). Let's assume a linear (additive) utility function for this example, the function given below (where q is a parameter that's greater than zero, a is a parameter that's between 0 and 1, and ln(i) = natural log of good i):

$$(1a)\ U = ? + a\ln(B) + (1-a)\ln(R)$$

If we take the derivatives described in (6) and substitute those derivatives into (3), then we have (recall that if $y = \ln(x)$, then $dy/dx = 1/x$):

$$(7)\ -\frac{\left(\dfrac{a}{B}\right)}{\left(\dfrac{(1-a)}{R}\right)} = -\frac{P_B}{P_R}$$

The two equations which describe the tangency point between Homer's indifference curve and his budget constraint are (7) and (2a). Using these equations together, we can solve for B* and R*. In their present form, those solutions are:

$$B^* = \frac{aI}{P_B}$$

$$R^* = \frac{(a-a)I}{P_R}$$

If we wish to go further and assume numerical values for the parameters in this model, then we could assume the following:

$\alpha = 0.5$ $\quad p_B = \$4$

$\theta = 100$ $\quad p_R = \$2$

$I = \$200$

Substituting into our solution above, the numerical values for B* and R* are:

B* = 25

R* = 50

These are the amounts of beer and pork rinds that will give Homer his maximum utility.

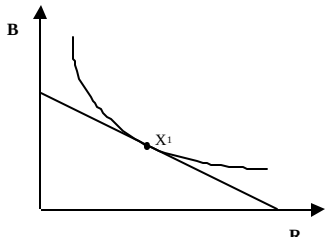
9.4 Substitution and Income Effects in the Indifference Curve model

Homer Simpson, our representative consumer, consumes varying amounts of beer and pork rinds. Assume that B = quantity of beer consumed, and that R = quantity of pork rinds consumed. Homer's utility function is given as: $U(B, R) = \sqrt{B \cdot R}$.

The marginal rate of substitution (which is the slope of Homer's indifference curve) between beer and pork rinds is given in absolute value as: $MRS_{R,B} = -B/R$. Recall that this can be derived from Homer's utility function. If we use a different utility function, then we get a different $MRS_{R,B}$.

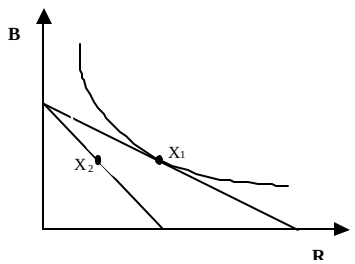
Assume further that the price of beer is \$4, the price of pork rinds is \$2, and that Homer's income is \$200. We can obtain Homer's budget constraint from this information, which we can rearrange as: B = -0.5R + 50.

A consumer equilibrium occurs in the graph below at pt. X_1, where the (blue) indifference curve is tangent to the (red) budget constraint.



It is possible to calculate the quantities of beer and pork rinds at this consumer equilibrium. After doing so, we would find that B* = 25 units and R* = 50 units.

How is the graph above affected when the price of pork rinds increases from \$2 to \$4? This change is shown on the graph below. The budget constraint becomes steeper and Homer moves to a new (pink) indifference curve and a lower level of utility at pt. X_2. If we calculate the new consumer equilibrium at pt. X_2, we would get B* = 25 and R* = 25.



Notice, however, that the price change included two actions. The movement from pt. X_1 to pt. X_2 involved a change in the marginal rate of substitution (i.e. a change in the slope of the indifference curve), *and* a change in utility (i.e. a change from the blue indifference curve to the pink indifference curve). This is different from a change in income, which only involves one change – a change in utility. These two actions form the analytical basis for what we call the substitution effect and the income effect.

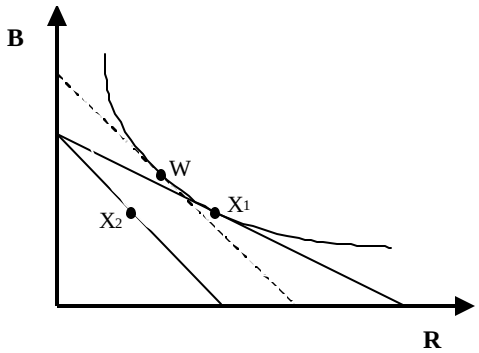
9.4.1 The Substitution and Income Effects

When prices rise, consumers lose purchasing power. What if the price of pork rinds goes up, but the government offers to compensate Homer for this loss of purchasing power. That is, Mayor Quimby offers to mail Homer a check, in an effort to keep Homer from feeling worse off. Homer still faces the higher pork rinds price, but doesn't experience a change in utility. That is, for Homer to be no worse off after the price increase, the government check must be large enough to keep Homer on his original indifference curve.

If the government check allows Homer to remain on his original indifference curve, will he just return to pt. X_1 and go back to buying 50 units of pork rinds? No. Even though Homer would return to his original indifference curve, he would also still face a different pair of prices. Therefore, we know that Homer must be located at a different point on that original indifference curve.

By taking the second graph above, and drawing a "hypothetical budget constraint", we can find this new point. This new constraint must satisfy two criteria. First, the constraint must be parallel to the new prices (where beer and pork rinds each cost \$4). Second, the constraint must be tangent to the original indifference curve.

The dotted line in the graph below satisfies these criteria, and so represents this new constraint. This line is tangent to Homer's original indifference curve at pt. W. This point reveals the quantities of beer and pork rinds that Homer would buy after receiving his government check (the check that keeps his utility constant). Of course, in real life, Homer would never get a check from the Mayor, but we will use pt. W to distinguish between the two actions (or effects) we noted as occuring with every price change.



How much would Homer consume at pt. W? The calculation is somewhat involved. First, note that the slope of Homer's new constraint is -1. Consequently, at pt. W, the slope of his original

indifference curve equals -1. If R/B = 1 at pt. W, then B = R at pt. W also. That is, we can ascertain that Homer will buy an equal amount of beer and pork rinds at pt. W.

Homer's original level of utility is $25\sqrt{2}$ (i.e. plug the original consumer equilibrium values

of B = 25 and R = 50 into Homer's utility function). To maintain Homer's original level of utility, then $\sqrt{B \cdot R} = 25\sqrt{2}$. That is, Homer will buy some combination of B and R that makes his utility function equal to . Recall that, at pt. W, Homer will buy an equal amount of beer and pork rinds. Therefore, we can rewrite as , which simplifies to . If , and B* = R*, then .

The new (hypothetical) budget constraint would be given as 4B + 4R = 200 + DI, where DI is the change in income necessary to keep Homer's utility constant. Plugging in B* and R* from the paragraph above, we find that DI = $82.84. That is, if Homer receives a check for $82.84, then Homer can continue to receive his original level of utility (i.e. utils) even though pork rinds are $2 more expensive now.

What are the substitution and income effects? The two effects are separated by pt. W. As the quantity of pork rinds changes between pt. W and pt. X_1 we observe the substitution effect. At pt. X_1, Homer consumes 50 units of pork rinds. At pt. W, Homer consumes $25\sqrt{2}$ units of pork rinds (i.e. about 35.36 units). The substitution effect associated with this price increase is represented by a decrease in quantity. That is, the substitution effect reveals a negative relationship between the price and quantity change. In fact, with every price change, we find this negative relationship within the substitution effect.

The income effect is measured as the quantity change attributed to moving from pt. W to pt. X_2. Between these two points, only utility changes, there is no change in the slope of the budget constraint. At pt. X_2, Homer consumes 25 units of pork rinds. The difference between pts. W and X_2 becomes $25\sqrt{2} - 25$, about 10.36 units.

Note that, like the substitution effect, there is a decrease in quantity within the income effect. Unlike the substitution effect, however, a negative relationship between price and quantity does not always arise within the income effect. For normal goods, the income effect reveals a negative relationship between price and quantity changes. That is, price increases lead to the income effect involving a decrease in quantity, and price decreases lead to the income effect involving an increase in quantity. Obviously, Homer considers pork rinds to be a normal good.

For inferior goods, we get the opposite result – the income effect involves a positive relationship between price and quantity changes. Any increase in price (decrease) would lead to the income effect yielding an increase in quantity (decrease).

Suppose the inferior good is highly inferior. For example, suppose we have a good where any small increase in price leads to a large, positive income effect. This would explain why a fairly large price change leads to an insignificant (overall) change in quantity. The inferior good's large income effect moves in the opposite direction of the substitution effect, causing the overall change (i.e. the sum of the two effects) to be very small.

In some cases, if a good is inferior enough, the positive income effect may be so large that it leads to price increases (decreases) being accompanied by overall quantity increases (decreases). When this occurs, we are dealing with a special (and rare) type of good known as a Giffen good. Giffen goods are so inferior that the income effect overwhelms the substitution effect, leading to the perverse result described above – where there is an overall positive relationship between price and quantity changes.

9.5 Application of Indifference Curves: Lump Sum vs. Per Unit Taxes

Assume we have a new representative consumer, Marge Simpson, who derives different levels of utility from buying varying quantities of beer and pork rinds (Q_B = quantity of beer, Q_R = quantity of pork rinds).

Her utility can be calculated from the following utility function:

$$U = \sqrt{Q_B \cdot Q_R}$$

This utility function implies that her indifference curves, IC, have a slope that is nonlinear. That slope, called the marginal rate of substitution between beer and pork rinds, can be calculated by plugging different quantities into the equation below:

$$MRS_{B,R} = -\frac{Q_R}{Q_B}$$

Let's assume further that Marge faces the following prices and that she has the income given below as well.

Price of beer = $4

Price of pork rinds = $2

Income = $200

We can start our analysis by asking this question:

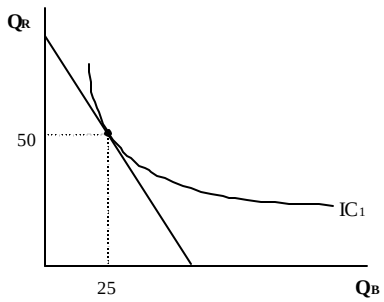
How many units of beer and pork rinds should Marge buy?

We know, from previous work/handouts, that there are two equations which will help us determine the answer to our question. Furthermore, we know that both must be true simultaneously. Those equations are:

(1) $4Q_B + 2Q_R = 200$ (Marge must be somewhere on her budget constraint, BC)

(2) $-4/2 = -Q_R/Q_B$ (the slope of Marge's BC must equal $MRS_{B,R}$ at equilibrium)

Solving (1) and (2) simultaneously (e.g. using algebraic substitution), we find that Marge will buy 25 units of beer and 50 units of pork rinds. Her indifference curve graph appears as:



Let's assume that the government is considering two different types of tax. The first tax is a per unit tax. That is, a tax that is

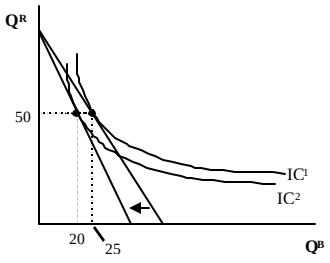
levied on the number of units of a specific good that are purchased by Marge. The second tax is a lump sum tax. That is, a tax of some fixed amount that does not correspond with the number of units Marge decides to buy of either good.

Suppose that a $1 per unit tax is levied on beer. This changes the actual price that Marge pays for each unit of beer. For each unit of beer that she purchases, she pays the sum of the equilibrium price and tax on that unit. For example, if the price is $4 per unit and the tax is $1 per unit sold, then Marge will pay $5 for each unit she buys. That changes the budget constraint equation (above) and, of course, the slope of the budget constraint as follows:

(1a) $5Q_B + 2Q_R = 200$

(2a) $-5/2 = -Q_R/Q_B$

Again, solving (1a) and (2a) simultaneously, we find that Marge will buy 20 units of beer and 50 units of pork rinds. The budget constraint shifts inward, and she moves to a new, lower indifference curve as follows:



We note that, with this per unit tax on beer in place, the government will raise $20 in tax revenue from Marge (i.e. Marge pays $1 tax for each of the 20 units she buys).

If the government decides to go with the lump sum tax, then Marge (and all other consumers) must pay (instead) a specific lump sum amount. It seems safe to assume that if the government raises the same amount of tax revenue from Marge with either tax, then the government will not have a preference as to which tax is used.

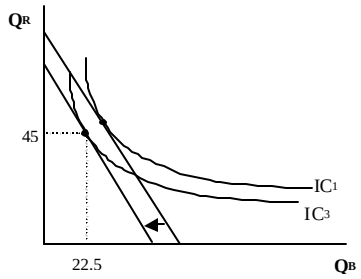
An important second question, however, is whether Marge has a preference. To answer this question, let's assume that the government levies a lump sum tax of $20 on Marge (equal to the amount she'd pay with the per unit tax).

The new lump sum tax would reduce her Marge's income by $20, but, unlike the per unit tax, will leave the price of beer unchanged. With her new, lower post-tax income ($180, instead of $200), Marge faces a new budget constraint. The budget constraint doesn't change in slope, but does shift inward because of a change in each intercept. This is reflected in the following equations:

(1b) $4Q_B + 2Q_R = 180$

(2b) $-4/2 = -Q_R/Q_B$

Again, solving (1b) and (2b) simultaneously, we find that Marge will buy 22.5 units of beer and 45 units of pork rinds. On a graph, we see the budget constraint shift in parallel as Marge moves to a new, lower indifference curve (IC_3):



The most straightforward method for determining whether Marge prefers one tax over the other is to calculate her utility in both situations. To do so, we utilize the utility function given at the beginning of this handout and then compare the utility associated with the per unit tax case and the lump sum tax case by plugging in the appropriate equilibrium values for Q_B and Q_R as follows:

Per unit tax: $U = \sqrt{20 \cdot 50} = 31.62$

Lump Sum tax: $U = \sqrt{22.5 \cdot 45} = 31.82$

Marge's utility is higher when the government raises $20 in tax revenue from a lump sum tax than when the government raises $20 from a per unit tax.

Why do we get this result? In other material, we've noted that price changes have a dual effect on a consumer's purchasing decision. There is both a change in relative prices, called the substitution effect, and a purchasing power change, called the income effect. When income changes, there is only one effect – the income effect. The per unit tax is comparable to a price change and so consumers react more with this tax than with the lump sum tax - when there is only a change in income. The difference in utility between these two cases represents a type of utility loss for Marge. Because deadweight loss arises when consumers substitute their consumption away from a particular good, the lump sum tax is thought to be more efficient, because the lump sum tax does not induce this kind of behavior (i.e. there is no substitution effect with a lump sum tax).

LESSON 10:
CONSUMER BEHAVIOUR – IV
PRICE EFFECT AND CONSUMER SURPLUS

In previous lesson you read the preference approach with the assumption of **'Rationality'**. Now in this lesson I will move a step ahead and take up some cases of violations of assumption of **'Rationality'**.

9.1 Quasi Rationality

Imagine that you purchased a ticket to a concert given by your favorite musical group. On the evening of the concert, a blizzard makes travel extremely hazardous. Would you go? Now imagine that you had been given this same ticket. Would you be more or less likely to travel to the concert in this case than in the previous case?

Or, suppose that you were given Rs.200 and these two options: A) a sure gain of Rs.50 or B) a 25% chance of winning Rs.200 and a 75% chance of winning nothing. Which would you choose? Now, suppose that you were given Rs.400 and these two options: C) a sure loss of Rs.150 or D) a 3/4 chance of losing Rs.200 and a 1/4 chance of losing nothing? Which would you choose?[1]

If you are like most people, you would be less likely to go to the concert if the ticket were given to you. However, this response is, according to the logic of economics, irrational. If you are like most people, you will choose A over B and D over C in the second example. Again, this is an irrational response. Examples such as these suggest that we should be cautious in assuming that people are rational calculating machines. There are cases in which people deviate from the behavior that the simple calculus of utility maximization says they should follow, and these deviations are predictable. The term "**quasi-rational**" has been given to them.

When given the concert example, many people find it difficult to believe that it should make no difference how one obtains the ticket.[2] The easiest way to see the logic of economics is to ask what determines the value of the ticket. Is it the cost of the ticket or is it the value of the concert? If it is the value of the concert (the economist's answer), it should be obvious that you should be equally likely to go in either case. If you want to say it is the cost, what is the value of a forged ticket that you bought? Does it have value because you spent money on it, even if it will not get you admitted to the concert?

If you analyze the second case above, you will see that situation A and C are identical, as are B and D. In A or C, one is offered a sure gain of Rs.250; and in B or D, one is offered a 25% chance of Rs.400 and a 75% chance of Rs.200. Yet the way of framing the question seems to trick the human mind into seeing these options in very different ways.

The economists and psychologists studying these anomalies have suggested that our mental abilities cannot process the economic information in our lives as the abstract logic of consumer choice says we should, and in order to deal with it, we develop **mental accounting systems**. Sometimes, these systems are more than mental, as when families have separate savings accounts for various items. They will often borrow money rather than dip into one of these special accounts, though a calculating-machine mind would never do that.

One feature of most mental accounting systems is that they start from a fixed point, usually the status quo. Changes coded as losses seem to have a greater emotional impact than changes coded as gains. As a result, if a situation is seen as an actual loss rather than as a gain not taken, it has a greater impact on people than if it is seen in the other way. However, economic theory says an actual loss and a gain not taken are equivalent.[3]

If we can be fooled by the way situations are framed, people selling things to us should be smart enough to take advantage of this computational defect. There are a number of situations in which this seems to happen.

We are more pleased with many small gains than one big gain of equal magnitude—we would rather get our Christmas presents in lots of boxes rather than one big one. There are innumerable sales pitches that promise something free if and only if we buy a product. If we think about this, we realize that nothing is free—we are paying for the complete package. Yet, the popularity of this type of sales pitch suggests that it works.

Alternatively, we are less affected by one big loss than a number of small ones of equal value. One of the appeals of credit cards is that they give us the bad news as one number. Also, sellers know that when we make a large purchase, they have an opportunity to sell us even more. If we are paying Rs.100,000 for a house, an extra Rs.1000 does not seem to be much to add on some conveniences. However, if we see the extra Rs.1000 as a completely separate transaction, we may react in a very different way.

Though the free trial with money-back guarantee is a way to signal quality. it also takes advantage of our mental accounting. Once we have an item at home and in use, it becomes part of the status quo. Giving it up is coded as an actual loss rather than a gain not taken, and affects us more.

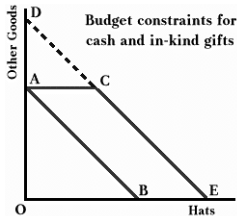
If you are still dubious about people relying on mental accounting rules, ask yourself why so many prices are in nines: Rs.9.95, Rs.19.95, Rs.999.99, etc. Why not simply round them up to an even number? The author has been immersed in economics for almost 30 years, and finds it amusing that he has mental accounting techniques that violate the logic of economic choice and that they are so deeply ingrained he cannot get rid of them. If you start examining how you view the world, you will probably find that you too often make decisions in ways that violate the logic of choice.

Having seen that one can harbor reservations about the traditional theory of choice, we next return to that theory to see why economists prefer cash to in-kind transfers.

In-Kind And Cash Transfers

Suppose that an eccentric millionaire decides to help a poor neighbor by giving him Rs.1000 worth of nontransferable hat certificates—certificates that could be used to buy hats but nothing else. Is this gift of the same value to the poor neighbor as a gift of Rs.1000 in cash? Economists answer that generally a gift **in kind** has less value than a cash gift because it has restrictions. A cash gift gives more options, and economists usually assume that more options never harm a person, but may help one.

The argument that cash gifts are superior to in-kind gifts can be shown with budget lines. In the graph below, the poor neighbor originally faces budget line **A-B.** A gift of Rs.1000 worth of hat certificates means the person could buy more hats than he could previously have bought, or if all income were spent on hats, he could buy **O-E** rather than the **O-B** that he could have had before. There is no increase in Other Goods that he can buy if he spends all his income on them. As a result, the new budget line is **A-C-E.** A gift of Rs.1000 in cash, on the other hand, would increase not only the number of hats that is possible, but also the amount of Other Goods. The new budget line with a cash gift is **D-C-E**. The dotted portion **D-C** represents the options that cash gives, but which the in-kind transfer does not allow. Because few people spend very much of their incomes on hats, most people prefer to be on the segment **D-C** rather than **C-E.** In this case, the gift in kind is less valuable to the recipient than a gift in cash.



This analysishas underlying assumptions, as all economic analyses do, and if these assumptions do not hold, the above conclusion may not either. First, it assumes that there is no cost in making decisions. Making decisions often involves gathering information, weighing it, and worrying about whether the decision is correct or not. Anyone who has had to choose between two good job offers knows the agony that making a decision can entail. Further, some people recognize that they do not make good decisions and hire others to do so for them. This shortcoming explains the career of financial advisor, in which people draw up budgets for others and in some cases make the actual purchases. Thus, if the giver knows of items that the recipient would like but does not know about, if the cost of making decisions is high, or if the individual is not capable of making decisions that get him to his goal, the recipient may be better off with the in-kind transfer than with a cash transfer.

Most gifts between friends are not for cash but are in-kind gifts. Most girls would be upset if their boyfriends gave them money for birthdays or Christmas. Gifts between friends are a way of cementing ties and give rise to obligations. They play an important role in helping small-group associations run smoothly. An in-kind gift indicates that the giver is interested enough in the recipient to learn about the recipient's likes and dislikes, and has spent time doing this and in obtaining the gift. The simple economic analysis of our graph does not capture these subtleties. Though anthropologists have studied gift-giving more than economists, some economists do consider it important. For example, Kenneth Boulding's *Economics of Love and Fear* argues that there are three mechanisms that help hold groups together: the integrative mechanism of gift-giving and love, voluntary and mutually advantageous exchange, and coercion and threat.

Third, the economic analysis of the picture above refers only to the recipients' preferences, not to those of the donors. A donor may disapprove of the goals that the recipient pursues, and thus may be unwilling to let him make choices. Economists are a bit unusual because they generally assume that people know what is best for themselves. As a result, most are reluctant to approve of desires of donors to deny choices to recipients.

A final reservation with the above analysis is that people often pursue goals that involve status, and pursuit of status is a zero-sum game. When one person rises in status, others must fall relative to him. If I am forced to help others, I will be less resentful if my help does not change their status, especially if those being helped are close to me in status. One way to make sure that my help does not change their status is to give them specific goods that have no status: things such as free health care, public housing, and free food. In contrast, giving others money allows them to purchase items that convey status, such as fancy cars and faddish brands of shoes and clothing. People who receive government aid and who have new cars are often intensely resented by those who almost qualify for the aid. Perhaps an important reason for the persistence of many forms of in-kind grants in the face of the opposition of so many economists is that the economists are ignoring the quest for status.

The question of whether in-kind transfers are better than cash transfers is important when governments devote hundreds of billions of dollars to transfers. At one time the government gave food to the poor; now it gives food stamps. The in-kind versus cash argument was involved in the switch (though it may not have been the decisive factor). The government attempts to help poor people by building and providing public housing; would the poor be better off if the government provided money instead of the housing? The government provides education by providing free schools, but students can rarely choose which free school they attend. Would they be better off if the government stopped producing education and instead provided them with a "tuition voucher," a sum of money which they could use at the school of their choice? The present system of transfers is a mixture of cash and in-kind transfers that is only partially explainable in terms of the recipients' welfare.

The topic of present value follows naturally from the discussion above.

9.1.1 Present Value

Suppose that someone will give you a gift of Rs.100, and will give it to you either now or in four years. Which is better, the money now or the money four years from now? The rule that

gifts with restrictions are of less value than gifts without restrictions suggests that money now is worth more than money in the future. Anything that one can do with the gift of Rs.100 four years from now one can do with Rs.100 now simply by saving it for four years. But there are many things that one can do with money now that one cannot do with money four years from now. Therefore, Rs.100 promised four years from now is not worth Rs.100 right now, but a smaller amount.

One of the things that can be done with money now is to invest it so that it will earn interest. Because this cannot be done with money four years from now, this option of foregone interest is a cost of waiting for the money. When this cost is measured, one sees the amount by which money in the future must be **discounted** to obtain its **present value.**

If the interest rate is 10%, Rs.100 now can be turned into Rs.110 one year from now. Thus, Rs.100 now and Rs.110 a year from now have the same value. (You may have to think about this for awhile.) This simple idea is vital in business and governmental decisions because a great many decisions have costs and benefits spread over time, and it is often necessary to compare sums in different time periods.

Computing the present value of future sums is nothing more than working compound interest problems backward. The formula for finding the future value of a present sum after one period is

(1) $P + Pr = F$

or

(2) $P(1 + r) = F$

where P is the present sum, r is the interest rate in decimal form, and F is the future sum. (Try the formula for P = Rs.100 and r = .10. You should get F = Rs.110.)

After two years the amount of money will be

(3) $F1(1 + r) = F2$

where F1 is the amount of money one year from now, and F2 is the amount of money two years from now. This may be rewritten as

(4) $(P(1 + r))(1 + r) = P(1 + r)^2 = F2.$

(Try this formula for P = Rs.100 and r = .10. F2 should be Rs.121.) Using the same logic, the future value three years from now will be

(5) $P(1 + r)^3 = F3$

and for any arbitrary period n, it will be

(6) $P(1 + r)^n = Fn.$

Simple algebra allows us to solve this equation if we have the time period and two of the three remaining variables. (Logarithms help a lot if we are solving for r.) In particular, if we have a future sum of money and want to find its present value, the last equation can be rewritten as

(7) $P = F / (1 + r)^n$

Using this formula in our case of Rs.100 four years from now and an interest rate of 10%, the present value is

$P = 100/(1.10)^4 = 100/1.464 = 68.30$

This means that Rs.68.30 invested at 10% will grow to Rs.100 four years from now. Therefore, Rs.68.30 now and Rs.100 four years in the future have equivalent value if the interest rate is 10%.

Present value analysis explains bond prices.

9.1.2 Bond Prices

Present value explains why the price of bonds on the bond exchange falls when interest rates rise and rises when interest rates fall. A bond is a contract. At the beginning of the contract, the lender pays a specified amount, usually Rs.1000 or a multiple of Rs.1000, to the borrower. The bond specifies when the Rs.1000 will be paid back, and how much will be paid as interest each year and all other terms of the agreement. A bond issued in 1983, due in 2003, and paying Rs.130 a year interest (a coupon rate of 13%) has the stream of payments illustrated in the table below. If the market rate of interest equals 13%, the 1983 value of the Get column will equal Rs.1000.

What a Bond Does

Year	Give	Get
1983	Rs.1000	Rs.0
1984	0	Rs.130
1985	0	Rs.130
1986	0	Rs.130
...	...	...
...	...	...
2001	0	Rs.130
2002	0	Rs.130
2003	0	Rs.1130

If the lender decides that he no longer wants to hold this bond, he cannot demand payment from the borrower because the contract does not give the owner of the bond the right to payment on demand. But he can sell it to someone else, and the price of this sale need not be Rs.1000. If market interest rates have risen since the original purchase of the bond, the present value of future payments in the Get column will drop, and so will the value of the bond. Another way of seeing this principle is to realize that if the market interest rate rises to 15%, there will be borrowers selling contracts for Rs.1000 that pay Rs.150 each year until the maturity of the bond (when the contract ends). It would be foolish to pay Rs.1000 to buy a contract that pays only Rs.130 a year.

The notion of present value may seem dry to someone who has never owned a bond or made business decisions. But for many corporate financiers, and for those who have money invested in bonds, it is a notion that provides a lot of excitement—both joy and woe—in their lives.

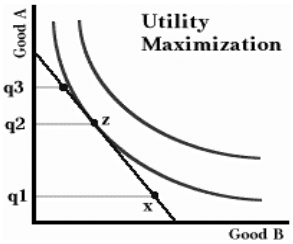
The notion of rational consumers leads to an important concept called consumers' surplus.

9.2 Consumers' Surplus

The assumption that consumers maximize utility leads to the downward-sloping demand curve. Actually, even non-rational or random behavior will lead to a downward-sloping demand

curve, as economist Gary Becker has demonstrated, but this demand curve does not have the same interpretation that a demand curve based on utility maximization (trying to attain goals) has.

Becker's argument is quite simple. Because the budget line is a constraint separating what is possible from what is not possible, even non-rational consumers face a budget constraint. Becker notes that if people randomly purchase goods, they will be randomly distributed, either along a budget constraint or within the area bordered by the budget constraint. (Becker considers both cases.) If the price of a good increases, the budget line will shift and a new random distribution of points will occur. The geometry of the situation implies that, on the average, people will buy less of a good as its price rises.



Though the demand curve of non-rational consumers will slope downward, it can no longer be interpreted as a locus of points of consumer equilibrium. With the assumption of utility maximization, the preferences and prices used to construct the graph above imply that **q2** is the amount of good A that is optimal for the consumer. If either more (**q3**) or less **(q1**) is being used, there is an incentive to change behavior because it would lead to better fulfillment of goals. However, if behavior is random and not concerned with fulfilling goals, point x is as good as point z. Thus, the argument that price controls have unintended results depends on the assumption that behavior is goal-directed.

Utility maximization suggests that the demand curve, because it measures buyer's willingness to pay, shows marginal benefits to buyers. The table below indicates that people will buy only one item if the price is Rs.5.00, or that people are willing to pay Rs.5.00 for the first item. They are not willing to pay Rs.5.00 for a second item, but only Rs.4.00. A second item has a smaller marginal benefit than the first because of the law of diminishing marginal utility. The equimarginal principle suggests that as price gets lower, consumers find that they must use more of an item to keep equality among marginal-utility-to-price ratios. Alternatively, as people use more of an item, its marginal utility drops, and so must its price if they are to stay in equilibrium.

A Demand Curve	
Price	Amount People Are Willing to Buy
Rs.5.00	1
Rs.4.00	2
Rs.3.00	3
Rs.2.00	4
Rs.1.00	5
Rs.0.50	6

This notion of the demand curve has an interesting implication known as the **consumers' surplus.** If in the table above consumers are buying three items, they must pay a total of Rs.9.00. But the total value to them is Rs.5.00 + Rs.4.00 + Rs.3.00 = Rs.12.00. There is a surplus value of Rs.3.00. In a more intuitive example, suppose that a person has been working in the hot sun all afternoon and is extremely thirsty. This person may be willing to pay as much as Rs.2.00 for a can of cold beer, but if he can buy it for only Rs..50, he thinks he has found a good deal and may buy two or three. The difference between the maximum a person would pay and the actual amount that he does pay is consumers' surplus. In other words, consumers' surplus is the difference between the **value in use** of an item and its **value in exchange.**

Notice that consumers' surplus is not related to the type of surplus that occurs in a market when price is above market-clearing price. Perhaps economists would have avoided this possible confusion if they had used a term other than consumers' surplus for this concept, but they did not and the term is now well-established.

In the early days of economics people puzzled over what was called the paradox of value. This paradox disappears once we understand consumer surplus.

9.2.1 The Paradox of Value

Why is it that some items that have relatively little use to society, such as diamonds, are extremely expensive, whereas others that are vital, such as water, are inexpensive. Adam Smith and other economists for a century after him struggled unsuccessfully to explain this **Paradox of Value**. Though Smith never unraveled the paradox of value, you can do it easily with a little help from the concept of consumers' surplus.

To see how this paradox is resolved, consider again the downward-sloping demand curve discussed earlier. As an item grows more abundant, its total use value to consumers, which is the entire area under the demand curve, rises; but its price, or its marginal value to consumers, declines. Thus, if two items in the table below are available, the total value to consumers is Rs.9.00 (or Rs.5.00 for the first and Rs.4.00 for the second), but the price or value in exchange is only Rs.4.00. If six are available, total use value rises to Rs.15.00, but exchange value (price) drops to Rs..50. Smith and his early followers missed this distinction between **marginal** and **total.** Thus, diamonds are scarce and have a high marginal value but a low total value. Another pound of diamonds has valuable uses that are not currently being met. Water is tremendously abundant and thus has a high total value and a low marginal value. Another gallon of water is not particularly important.

A Demand Curve	
Price	Amount People Are Willing to Buy
Rs.5.00	1
Rs.4.00	2
Rs.3.00	3
Rs.2.00	4
Rs.1.00	5
Rs.0.50	6

If there is a consumers' surplus, should there not also be a producers' surplus?

9.3 Producers' Surplus

Suppose that Charles considers a CD worth Rs.10 and Sam, who owns it, values it at only Rs.2.00. Sam agrees to sell it to Charles for Rs.5.00. We have seen that the value Charles gets but which he does pay for (Rs.5.00 in our example) is called consumers' surplus. But what of the Rs.3.00 of value Sam gets because he sold something worth only Rs.2.00 to him for Rs.5.00? There is a surplus here, and it is called either **producers' surplus** or **economic rent.**[1] Producers' surplus exists when actual price exceeds the minimum price sellers will accept.

Producers' surplus can appear as profit, but usually it takes a different form. Suppose, for example, that the price of corn has been Rs.2.00 per bushel for many years. Then it rises to Rs.3.00 per bushel and stays there. This higher price will draw more land into corn production, but this change is of no importance here. What is of interest is what happens to the farmers who were producing corn at Rs.2.00 per bushel and now find that they can sell corn at Rs.3.00. It certainly appears that these farmers are better off because a producers' surplus of Rs.1.00 per bushel has appeared that was not there before.

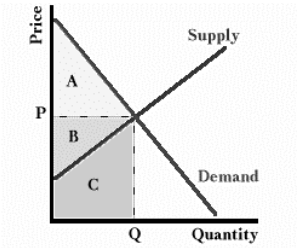
However, let us separate farming into two parts: working the land and owning the land. Suppose that a farmer does not own the land he works, but rents it. It then becomes unlikely that this farmer will benefit at all from the higher price of corn. If those working the land obtained the surplus, there would be competition for the right to work land that is especially suited to growing corn. This competition should raise the value of the land, and therefore it will be the landowners, not the cultivators, who benefit from the higher price of corn. Because the earliest case of producers' surplus analyzed was one in which land captured the surplus, the producers' surplus is often called economic rent.

Producers' surplus is usually captured by resource owners rather than by producers. Hence the producers' surplus is not the same as profit. The resources that capture the surplus are those that are especially good at producing the product in question or that have no other uses, and hence will be used for that product even when prices are low. Sometimes, the resource that captures economic rent is labor. The high pay that superstars in many fields earn is mostly producers' surplus. The basketball star paid Rs.1 million who would still play for Rs.25,000 earns Rs.975,000 in producers' surplus. There is an interesting conclusion to this observation: The reason basketball tickets are high-priced is not because star athletes have high salaries (as owners sometimes allege), but rather the salaries are high because fans are willing to pay so much for the tickets to see the stars play.

We conclude this group of readings by putting the producers' and consumers' surpluses together on a graph.

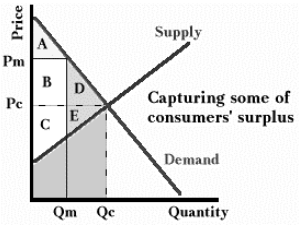
9.4 Producer And Consumer Surplus Together

The producers' and consumers' surpluses are illustrated with supply and demand curves in the figure below. The total value to consumers of quantity **Q** is represented by areas **A+B+C.** Because the consumers must pay **B+C**, only the area **A** is surplus for them. Producers get revenue of **B+C. B** is their surplus because only payments of **C** are needed to attract the resources necessary to produce quantity **Q**.



Producers' & consumers' surpluses

The concepts of consumers' and producers' surpluses are tools that can help analyze many situations. For example, is there any temptation for sellers to gang up on buyers? If sellers can raise the price, can they transfer some of the consumers' surplus to themselves? They can, and the graph below illustrates what happens. The consumers' surplus at price **Pc** is **A+B+D.** The producers' surplus at this price is **C+E.** By raising price to **Pm**, sellers cause the consumers' surplus to shrink to the area **A.** Area **B** is transferred from consumers to producers, but producers lose area **E**. If area **B** is greater than area **E**, this move benefits producers. The new producers' surplus is **C+B**. If sellers gang up on buyers, they are no longer price takers. Rather, the sellers leave the supply curve and search along the demand curve for the best deal. As a result, such behavior is called "**price searching**."



It is easiest for sellers to restrict output and raise price when there are very few sellers and many buyers. When there is **monopoly**, which means there is only one seller, economists expect the seller to act in this way. With many sellers, coordination of decisions becomes difficult (for the same reason that the problem of the commons can exist) and output restrictions become unlikely.

Alternatively, buyers can gang up on sellers and extract producers' surplus. They must restrict purchases to drive the price down. Again, this behavior is likely only when there are very few buyers and many sellers. When there is only one buyer, a **monopsonist**, economists expect it to restrict purchases.

What is good for the individual is not necessarily good for the group. Notice that the process of transferring the value of area **B** from consumers to producers in the second graph above causes consumers to lose area **D** and producers to lose area **E**, and no one gets this lost value. In the process of increasing their surplus by seizing area **B**, producers cause the value of total surplus to shrink. There is a conflict here between the interests of producers and society as a whole. This loss of value, which is not offset elsewhere in the system, is the essence of the economist's case against monopoly.

TUTORIAL 3

1. Case Study

Wealth and Health

Lead Story-dateline: The Economist, February 21, 2002.

Obesity is no longer just a western disease. It is becoming a problem in the developing world too. In 1991, 15% of Americans were obese; by 1999, that proportion had grown to 27%. Ten percent of Americans under 17 are obese. Physical activity has declined and diets have expanded. Meanwhile, poorer nations have enjoyed some success in their battles against malnutrition and famine. But, according to research presented at the annual meeting of the American Association for the Advancement of Science (AAAS), held on February 14th-19th in Boston, it is more a case of being out of the frying pan and into the fire. Obesity, anthropologists believe, could become an epidemic of the poor world as well as the rich one.

It now seems that even modest increases in wealth in countries undergoing industrialization are enough to tip their people from malnutrition to obesity. This increase in weight has been uneven as well as fast. In India, a survey of 83,000 women found that although 33% were malnourished, 12% were overweight or obese. The mix can even occur within a single household.

Another long-standing assumption about obesity in developing countries might also need to be overturned. In the developed world, obesity has become a disease of the poor, who eat chips and hamburgers while the wealthier sup on tofu and veggies, and hit the gym afterwards. The opposite was thought to hold true in poor countries, where as a matter of course only the upper classes could afford to eat enough to be overweight. A series of studies in Brazil show that in urban areas richer people eat more fruit and less sugar, and do more exercise, than their poorer compatriots. In Mexico and the Dominican Republic, a similar trend pertains.

Diabetes, heart disease and other so-called "diet-related non-communicable diseases" will join the list of ailments straining the public-health facilities of poor countries. The number of new cases of adult-onset diabetes in China and India already exceeds new cases in the whole of the rest of the world. An epidemic of cardiovascular disease lies heartbeats away. There is reluctance on the part of governments to spend resources on promoting diet and exercise while starvation is still a real threat.

Thinking about the future!

The availability of resources accessible to the rich versus the poor, whether across countries or among persons within a country, is astoundingly dissimilar, and hence, so too is consumer choice. Scarcity of resources underlies consumer choice, yet is insufficient in explaining idiosyncratic consumer diets. The article notes that in both advanced and less-developed countries, it is the poor whose diets are turning out obese adults. Other things equal, you might expect that lower income is linked to lower caloric intake. However, contributing to obesity among the impoverished are offsetting factors, including a predominance of high-calorie junk food in the diet, limited access to education and health care, exploitation, lack of political representation, and a multitude of stresses related to daily survival. Expectations of a bleak future can influence the poor to seek sources of immediate gratification. Sugar and fat-based diets can bring pleasures today. And when 'tomorrows' are perceived as bleak as 'yesterdays,' *having it your way* today at Burger King sounds increasingly like a rational consumer choice among the world's poor.

Talking it over and thinking it through!

1. With a cornucopia of food products available in U.S. and worldwide markets, why would rational consumers make dietary choices that tend to contribute adversely to their health?

2. With regard to consumer choice, is there any implication in the article that food is treated as a normal or inferior good?

Multiple Choice Questions

1. According to the marginal principle, if the marginal benefit is far greater than the marginal cost, than a rational individual will

- stop the activity and enjoy the current profits.
- stop the activity when marginal benefits are at their highest overall level.
- stop the activity when the marginal benefit is equal to the marginal cost.
- stop the activity only when there are no more remaining resources.
- none of the above are true.

2. Which of the following statements best describes total utility?

- Total utility decreases with each good a consumer purchases.
- Total utility increases at a decreasing rate with each good a consumer purchases.
- Total utility increases at an increasing rate with each good a consumer purchases.
- Total utility remains constant as the consumer purchases more goods.
- Total utility is the dollar value of all goods that a consumer purchases.

3. Roy recently bought a cat. He liked it so much that he bought two more. Eventually, when Roy had seven cats, he complained that it was not worth it to buy another. What is a possible explanation?

- The law of diminishing marginal utility is in effect.
- The supply curve for cats is downward sloping.
- A consumer will always buy positive amounts of all goods.
- His demand curve has a positive slope.
- Cats are an inferior good.

4. Kathy spends all her income purchasing pizza and soda. The price of pizza is $5, and the price of a soda is $1. What is the marginal cost to Kathy of purchasing one more pizza, in terms of utility?

- $1
- $5
- the marginal utility of one more soda
- (marginal utility per soda)(5 sodas)
- It depends on Kathy's income.

5. Marginal utility is the

- extra consumption divided by the amount of pleasure gained from the consumption.
- additional happiness gained by consuming one more unit of a good.
- enjoyment obtained by consuming all of a good.
- total satisfaction of the last unit consumed.
- average happiness from consuming some number of goods.

6. Claire's total utility from 3 scoops of ice cream is 100. Her total utility from 4 scoops of ice cream is 120. Claire's marginal utility for the fourth scoop is

- 115.
- 100.
- 30.
- 20.
- 10.

7. Tim likes rice. He begins to eat his first course of rice, but soon stops eating that and begins eating his baked beans instead. One could say

- the rice was undercooked.
- the beans have a higher marginal utility than the rice.
- Tim is not really trying to maximize his utility.
- for Tim, the rice now has a marginal utility of zero.
- the rice has a higher marginal utility than beans.

8. For Betty, the marginal utility of the first candy bar she eats is $2.50. The marginal utility of the eighth candy bar is $0. This means that if candy bars were free

- Betty would eat an infinite amount.
- Betty would probably not have any.
- Betty would have as many as she had time to consume.
- Betty would have at least eight candy bars.
- Betty would eat exactly 8 candy bars.

9. According to the utility-maximization rule, consumers will choose combinations of ice-cream and pizzas such that

- on the last unit of each good consumed, the marginal utility per dollar spent on ice-cream is equal to the marginal utility per dollar spent on pizza.
- on the last unit of each good consumed, the marginal utility per dollar spent on ice-cream is greater than the marginal utility per dollar spent on pizza.
- on the last unit of each good consumed, the marginal utility per dollar spent on ice-cream is less than the marginal utility per dollar spent on pizza.
- on the last unit of each good consumed, the marginal utility of ice-cream is equal to the marginal utility of pizza.
- none of the above are true.

10. At Disneyland, there is an admission price to get into the park, but after that, there is no charge per ride. You have to wait in line longer for some rides than others. Do all rides at Disneyland have the same price?

- Yes, because the price per ride is zero.
- Yes, because the price per ride is the admission fee that you paid to get in.
- No, the rides with longer lines cost more in terms of opportunity cost of time.
- No, because marginal utility per ride diminishes with every ride that you take, so additional rides are worth more to you.
- No, because marginal utility per ride diminishes with every ride that you take, so additional rides are worth less to you.

3. Long Answer Questions

1. Carefully explain the relationship between the individual demand curve, the marginal utility curve, and the marginal benefit curve.

2. Consider your monthly food budget. Suppose you only purchase two food items each month, apples and cereal. Apples cost $2.00 a pound, while cereal costs $3.00 a box. At your current level of consumption, you are purchasing 10 pounds of apples and 21 boxes of cereal. The marginal utility received from the last pound of apples and last box of cereal is 40 and 30, respectively. Are you maximizing your utility? If not, what actions could you take to maximize your utility? Explain how you arrived at your answer.

THE BUSINESS ORGANISATION
THEORY OF PRODUCTION

In this chapter we begin to investigate the theory of supply. Supply depends on the conditions of production, so we begin with a study of production, and specifically ask how outputs vary when the quantity of one input used varies.

Though economists are interested in many cases of unintended consequences, those unintended consequences that involve businessmen seeking their own gain have been at the heart of economic analysis since Adam Smith. Smith noted that

"It is not from the benevolence of the butcher, the brewer, or the baker, that we expect our dinner, but from their regard to their own interests. We address ourselves, not to their humanity but to their self love, and never talk to them of our necessities but of their advantages."

Since Smith, a great deal of intellectual effort has gone into exploring the question of under what conditions the interests of society will be served by businessmen seeking to make a profit—in fact, this is the core of microeconomics. The reading selections present background material to this exploration by explaining a large number of technical terms that economists use, and also by looking at a few of the simplifying assumptions they generally invoke.

After you complete this chapter, you should be able to:

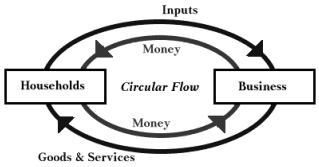
- Define production function, isoquants, marginal product, price discrimination, monopsonist, and the all-or-nothing demand curve.
- Define increasing, decreasing, and constant returns to scale.
- Distinguish between income and substitution effects.
- Distinguish between an individual buyer's demand curve and the industry demand, and between industry demand and the demand curve facing an individual seller.
- Compute marginal revenue from the demand curve of the seller when that demand curve is given in the form of a table.
- Compute marginal resource cost from the supply curve of the buyer when that supply curve is given in the form of a table.
- Explain why marginal resource cost equals price for a buyer who is a price taker.
- Explain why marginal revenue equals price for a seller who is a price taker, and why marginal revenue is less than price for a seller who is a price maker.
- Explain what the law of diminishing returns is and under what conditions it holds.
- Explain why the demand curve, the supply curve for resources, and the production function can be treated as boundaries.

LESSON 11:
THE FIRM AND ITS PRODUCTION

An **exchange** is a voluntary agreement between two people, in which each gives something to the other and gets in return something that he considers of greater value. When John and Jim exchange baseball cards, John gets cards that he considers more valuable than those he gives to Jim, and Jim gets cards that he considers more valuable than those he gives to John. Unless both sides of the exchange feel that the exchange benefits them, the exchange will not take place. Because both sides benefit, exchange is, in the terms of game theory, a **positive-sum game.**

An alternative to interaction by exchange is interaction that involves **coercion**. With coercion, the actions of one side are not voluntary but forced. If Jim takes baseball cards away from John and threatens to beat him up if he complains, we have interaction based on coercion. Economics focuses almost exclusively on interactions based on exchange and ignores those based on coercion. As a result, it has much more to say about markets than about government, which is the primary agent of coercion in society.

People engage in exchange to attain goals. Exchange is not just **take**; in order to get, one must **give**. People must do things that they do not want to do in order to get things that they desire. The unpleasant part of this process is work and production, and the pleasant part is consumption. Work and production are not pursued for their own sakes, but only because without them we cannot consume. This division of economic life is illustrated below in what Frank Knight called the Wheel of Wealth, but which is now more commonly known as **circular flow.**



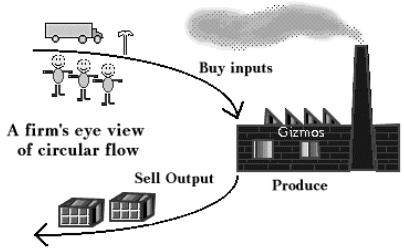
The circular flow diagram divides the economy into two sectors: one concerned with producing goods and services, and the other with consuming them. Resources are converted into goods and services by business, and in this transformed state travel back to consumers. Money flows in the opposite direction. These flows involve two markets in which exchange takes place: ***the resource or factor market in which business buys resources, and the goods and services market in which business sells goods.*** Some economists define a "factor of production" as the service of some resource**.** If resources are land, labor, and capital, the factors of production are the services of land, labor, and capital. I will ignore the distinction between resources and factors of production in the discussion that follows.

Both the model of supply and demand and consumer choice with utility analysis are key elements of an understanding of an exchange economy. However, this group of readings emphasizes the right side of the circular flow diagram, examining the business firm and the constraints or limitations that it must face in its fight for survival.

The economic theory of the firm is founded on the three fundamental tasks of a firm.

11.1 Three Fundamental Tasks

The right side of the circular flow diagram shows the **three fundamental tasks** of all firms in an exchange economy. First, a firm must obtain inputs. Inputs include raw materials, energy, machinery, office space, workers, and anything else needed to produce output. Second, the firm must combine or use inputs to produce output. Output may either be a tangible good such as a pair of shoes or an automobile, or a service such as a haircut or a medical checkup. Third, the firm must sell its output to someone else.



A firm that cannot do these three tasks well enough will not survive. When the automobile was developed in the early 20th century, firms that made carriages died because they could no longer do task three, selling their output, well enough. Almost all baskets sold in the United States are imported. Baskets are handmade, and no firm in the United States can hire workers at wages low enough to be able to compete with wages that are acceptable in some other countries. Here is a case in which American firms (or firms in any industrialized nation) have difficulty coping with the first task. The development of the electronics industry is a case in which the second task has changed. New technology allows firms to combine inputs to produce goods that were not possible just a few years ago.

The economic theory of the firm is an analysis of the way the firm must perform these three tasks to make a profit. Each task can be described in mathematical or graphical terms. Supply curves of resources describe the first task. They indicate how much the firm must pay for the amount of inputs it wants. The production function describes the second task. It tells how much output the firm can produce from a set of inputs. The demand curve for output describes the third task. It depends on people's wants or preferences and tells how much the firm can charge for output. Each of these mathematical ways of representing the three fundamental tasks can be seen as a constraint or limitation that the firm faces.

A supply-of-resources curve tells at what prices various amounts of a resource can be bought or hired. Though one can view it in a number of ways, it also can be viewed as a boundary. It tells the firm the minimum it can pay for any amount of a resource. Sellers of resources imposed this boundary on the firm, which must buy resources in order to produce. Points to the upper left of a supply curve are attainable, whereas those to the lower right are not.

The production function contains information about how much output can be obtained with various quantities of inputs. The production function is often discussed as a relationship between inputs and output, as its name implies. (Mathematically, a function is a special sort of relationship.) However, it too can be discussed as a boundary. It shows the maximum that can be produced with any combination of resources. Less than this maximum can be produced—one can always get nothing for something.

The demand curve can be viewed from a number of perspectives: as a relationship between price and quantity buyers will buy, as a locus of points of consumer equilibrium, as a measure of marginal benefit to the buyer, or as a boundary. This last view, that the demand curve represents a boundary that buyers impose on the seller, is one that is most useful when developing the theory of the firm. The demand curve limits the amount that sellers can sell at each price. Points to the left of the demand curve are attainable, while those to the right are not.

The remainder of this group of readings takes a closer look at these three constraints. The material is somewhat technical, but it contains much of the core of microeconomics.

Next we look more closely at the relationship between production and supply.

11.2 Production and Supply

In investigating the foundations of supply and demand, we will look at demand and supply as separate headings. It doesn't matter much, logically, which we take first. Historically, the first stages of the economists' Reasonable Dialog were focused more on supply. In investigating the foundations of supply, we are investigating the economics of production, and that was the central topic for the classical economists.

Adam Smith, we recall, had been very optimistic about the future economic development of the industrializing countries. With increased division of labor leading to higher wages and growing demand, he felt, production could continue to grow. However, Thomas Malthus criticized Smith's optimism. Malthus spoke for the pessimistic view, and, of course, Malthus is best known for his claim that increasing population would lead to poverty. In supporting this idea, Malthus began to study the limits on production. It was this study that has made his work important particularly for Neoclassical economics.

Limits on production stem from limited resources with a given technology. With a given technology, limited quantities of inputs will yield only limited quantities of outputs. The relationship between the quantities of inputs and the maximum quantities of outputs produced is called the "production function."

The "production function." is a relationship between quantities of input and quantities of output that tells us, for each quantity of input, the greatest output that can be produced with those inputs. Malthus didn't work out the details, but he clearly had this idea in mind as he originated the key concept of Diminishing Returns.

11.3 Diminishing Returns

As we recall, Malthus is best known for his pessimistic idea that population growth would force incomes down to the subsistence level. What we are interested in here is not his conclusion, but the reasoning that took him there.

Malthus argued that land is a fixed input, but the growth of population makes labor a variable input. Malthus proposed a general law of economics, **the Law of Diminishing Returns:** when a fixed input is combined in production with a variable input, using a given technology, increases in the quantity of the variable input will eventually depress the productivity of the variable input. (Malthus argued that decreasing productivity of labor would depress incomes).

Was Malthus right? The answer is, of course, yes and no.

There is plenty of evidence, both observational and statistical, that the Law of Diminishing Returns is valid. For example, agricultural economists have carried out experimental tests of the theory. They have selected plots of land of identical size and fertility and used different quantities of fertilizer on the different plots of land. In this example, land was the fixed input and fertilizer the variable input. They found that, as the quantity of fertilizer increased, the productivity of fertilizer declines. This is only one of many bits of evidence that the Law of Diminishing Returns is true in general.

On the other hand, in the two hundred years since Malthus wrote, on the whole, population has increased but labor productivity and incomes have not declined. On the whole, they have risen. What seems to have happened is that technology has improved. Malthus recognized that if technology improved (in agriculture, at least), that might postpone what he saw as the inevitable poverty as a consequence of rising population. Some economists, and other people, believe that the Malthusian prediction will eventually come true. Perhaps: what is clear is that in two hundred years it has not.

But that doesn't mean the Law of Diminishing Returns is wrong! A "law" such as this can be true in general but cannot be applied when its assumptions (such as an unchanging technology) aren't true. The "Law" isn't wrong — just inapplicable to that case.

There are many valid and useful applications of the Law of Diminishing Returns in economics. In this chapter we will look at two.

- We will use Diminishing Returns and related concepts to get a better idea of the meaning of the phrase "efficient allocation of resources" and some guidelines for efficiency in that sense.
- We will explore how a business firm should direct its production in order to get maximum profits, and that will give us a basis for a better understanding of the economics of supply.

First, though, we will need to look at production and diminishing returns in general in a little more detail.

10.4 The Production Function

Production is the transformation of inputs into outputs. Inputs are the factors of production — land, labor, and capital — plus raw materials and business services.

The transformation of inputs into outputs is determined by the technology in use. Limited quantities of inputs will yield only limited quantities of outputs. The relationship between the quantities of inputs and the maximum quantities of outputs produced is called the "production function."

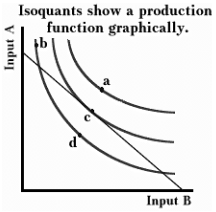
But how do these outputs change when the input quantities vary? Let's take a look at an example of a production function.

When most people think of fundamental tasks of a firm, they think first of production. Economists describe this task with the production function, an abstract way of discussing how the firm gets output from its inputs. It describes, in mathematical terms, the technology available to the firm.

A production function can be represented in a table such as the one below. In this table five units of labor and two of capital can produce 34 units of output. It is, of course, always possible to waste resources and to produce fewer than 34 units with five units of labor and two of capital, but the table indicates that no more than 34 can be produced with the technology available. The production function thus contains the limitations that technology places on the firm.

A Production Function			
Labor			
5	30	34	37
4	26	30	33
3	21	25	28
2	16	20	23
1	10	13	15
	1	2	3
Capital			

The production function can also be illustrated in a graph such as that below. This graph looks exactly like a graph of indifference curves because the mathematical forms of the production function and the utility function are identical. In one case, inputs of goods and services combine to produce utility; in the other, inputs of resources combine to produce goods or services. A curved line in the graph shows all the combinations of inputs that can produce a particular quantity of output. These lines are called *isoquants.* As one moves to the right, one reaches higher levels of production. If one can visualize this as a three-dimensional graph, one can see that the production surface rises increasingly high above the surface of the page; the isoquants indicate a hill. The firm must operate on or below this surface.



11.4.1 Marginal Productivity

Productivity, by definition, is a ratio of output to labor input. In most statistical discussions of productivity, we refer to the average productivity of labor:

$$\text{Average Productivity} = \frac{\text{Output}}{\text{Labor}}$$

Average labor productivity is an important concept, especially in macroeconomics. In microeconomics, however, we will focus more on the marginal productivity. We can think of the **marginal productivity** of labor as

the additional output as a result of adding one unit of labor, with all other inputs held steady and ceteris paribus.

In algebraic terms, an equally correct definition is:

$$MP = \frac{\triangle \text{Output}}{\triangle \text{Labor}}$$

Let's have a numerical example to illustrate the application of the theory. Suppose that:

- When 300 labor-days per week are employed the firm produces 2505 units of output per week.
- When 400 labor-days per week are employed the firm produces 3120 units of output per week.
- It follows that the change in labor input, • Labor, is 100.
- It also follows that the change in output, • Output, is 615.
- Applying the formula above, we approximate the marginal productivity of labor by the quotient 615/100 = 6.15.
- We can interpret this result as follows: over the range of 300 to 400 man-days of labor per week, each additional worker adds approximately 6.15 units to output.

Of course, if we had more information, we could get a closer approximation. For example, if we had the outputs for 310, 320, ... 390 man-days of labor, we could see how MP varies within the range 300-400. But we can be sure that the values will be in the neighborhood of 6.15.

Let's extend the numerical example in the page and see how marginal productivity varies over a wide range of labor inputs. Here is a hypothetical example of production with the inputs of land and labor held steady and varying quantities of labor, and the output and average and marginal productivities.

Labor	Output	Average Productivity	Marginal Productivity
0	0	0	
100	945	9.45	9.45
200	1780	8.90	8.35
300	2505	8.35	7.25
400	3120	7.80	6.15
500	3625	7.25	5.05
600	4020	6.70	3.95
700	4305	6.15	2.85
800	4480	5.60	1.75
900	4545	5.05	0.65
1000	4500	4.50	-0.45

11.4.2 Average and Marginal Productivity

We have put some stress on the difference between average and marginal productivity. Both are important, but for a model of

short-run profit-maximizing supply, marginal productivity is the more important. and the two are quite different.

Here are the average and marginal productivities for the same numerical example in the page before last. Notice how both average and marginal productivity decrease as the labor input increases. But the marginal productivity declines faster than the average productivity, pulling the average productivity down after it. The downward slope of the marginal productivity line expresses the Law of Diminishing Returns, and the downward slope of the average productivity is also a result of the law.

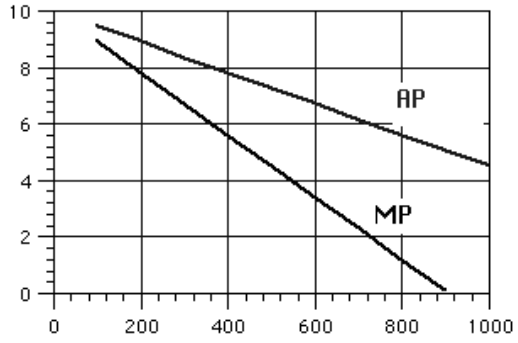


Figure 2: Average and Marginal Productivity

The relationship between average and marginal productivity in the diagram is important in itself, and we will see similar relationships in future chapters. So let's look at it a little more closely. Average and marginal productivity will not always have the same slope. In general,

- whenever average productivity is greater than marginal productivity, average productivity will slope downward.
- whenever average productivity is less than marginal productivity, average productivity will slope upward.

The diagram does not show any values where average productivity is less, but a more complicated example might, and then we would see the second part of the relationship visualized.

To understand the relationship, think of it this way: as we add labor input, one unit after another, we add a bit more to output at each step. When the addition is greater than the average, it pulls the average up toward it. When the addition is less than the average, it pulls the average down toward it.

Now let's analyse the above example to understand Law of Diminishing Returns .

11.5 The Law of Diminishing Marginal Productivity

In his discussions of the Law of Diminishing Returns, Malthus did not distinguish between average and marginal productivity. However, in modern economics, we think of diminishing returns primarily in terms of marginal, not average, productivity. Thus, we would state the law this way:

Law of Diminishing Returns (Modern Statement):

When the technology of production and some of the inputs are held constant and the quantity of a variable input increases continually, the marginal productivity of the variable input will eventually decline.

The inputs that are held steady are called the "fixed inputs." In these pages we are treating land and capital as fixed inputs. The inputs that are allowed to vary are called the "variable inputs." In these pages we are treating labor as the variable input.

Another way to express the law of diminishing returns, is that, as the variable input increases, the output also increases, but at a decreasing rate. The marginal productivity of labor is the rate of increase in output as the labor input increases. To say that output increases at a decreasing rate when the variable input increases is another way to say that the marginal productivity declines.

Here is a picture of the relationship between the variable input and the output in the numerical example in the previous table. Notice how the slope gets flatter: as the variable input increases, output increases at a decreasing rate. This is a visualization of **the Law of Diminishing Marginal Productivity.**

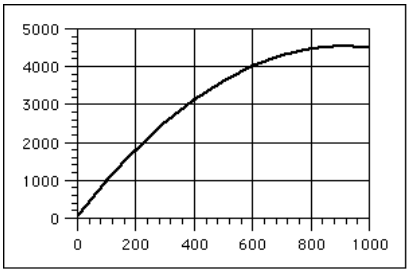


Figure 1. Production with Diminishing Returns

There is one rule that seems to hold for all production functions, and because it always seems to hold, it is called a law. The **law of diminishing returns** says that adding more of one input while holding other inputs constant results eventually in smaller and smaller increases in added output. To see the law in the table above, one must follow a column or row. If capital is held constant at two, the marginal output of labor (which economists usually call **marginal product** of labor) is shown in the table below. The first unit of labor increases production by 13, and as more labor is added, the increases in production gradually fall.

Marginal Product, as you already know, is the ***change in output with the increase of one additional unit of input.***

The Marginal Product of Labor	
Labor	Marginal Output
First	13
Second	7
Third	5
Fourth	5
Fifth	4

The law of diminishing returns does not take effect immediately in all production functions. It is possible for the first unit of labor to add only four units of output, the second to add six, and the third to add seven. If a production function had this pattern, it would have increasing returns between the first and third worker. What the law of diminishing returns says is

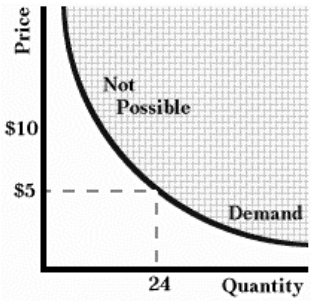
that as one continues to add workers, eventually one will reach a point where increasing returns stop and decreasing returns set in.

The law of diminishing returns is not caused because the first worker has more ability than the second worker, and the second is more able than the third. By assumption, all workers are the same. It is not ability that changes, but rather the environment into which workers (or any other variable input) are placed. As additional workers are added to a firm with a fixed amount of equipment, the equipment must be stretched over more and more workers. Eventually, the environment becomes less and less favorable to the additional worker. People's productivity depends not only on their skills and abilities, but also on the work environment they are in.

The law of diminishing returns was a central piece of economic theory in the 19th century and accounted for economists' gloomy expectations of the future. They saw the amount of land as fixed, and the number of people who could work the land as variable. If the number of people expanded, eventually adding one more person would result in very little additional food production. And if population had a tendency to expand rapidly, as economists thought it did, one would predict that (in equilibrium) there would always be some people almost starving. Though history has shown the gloomy expectations wrong, the idea had an influence on the work of Charles Darwin and traces of it still float around today among environmentalists. The second boundary that limits the firm is the demand curve for output.

11.6 Demand Curve for Output

Once a firm has produced a product, it must sell it. The **demand curve for output** describes the limitations the firm faces in doing this task. The demand curve for output is a constraint on the firm because it gives the maximum price that a firm can charge for each level of production. Thus, if the firm in the graph below wants to sell 24, it can do so by charging Rs.5.00 or any price that is lower. It cannot charge Rs.10.00 and still sell 24 because buyers will not allow it.



Demand as a boundry

The demand curve facing a firm depends both on the preferences of consumers and on how well other firms meet those preferences. One can derive a demand curve for an individual from a set of indifference curves showing the individual's preferences and a series of budget lines showing changes in price. To get a demand curve for the entire industry, one must add up all the demand curves of individuals. To get the demand curve for eggs, for example, one must add up the number of eggs that Smith and Jones and Nelson and all other consumers in the market want at each possible price.

When there is only one firm selling in a market, that firm is a **monopolist.** (The Greek root *mono-* means "one.") The demand curve for the monopolist is the demand curve for the industry. A monopolist is a **price searcher** or a price maker. It will search along the demand curve for the price-quantity pair that is most profitable. When there is more than one seller, the demand curve that a seller sees is not the same as the demand curve for the industry. The industry demand is split up among sellers. When there are only a few sellers, the sellers will still be price searchers or price makers. These sellers, or **oligopolists** (the Greek root *oli-* means "few"), are price makers because each recognizes that if it wants to sell more, it must lower its price.

However, the demand curve of each oligopolist will be more elastic than the demand curve for the industry as a whole. Suppose, for example, that there are two firms in an industry, each produces 50 units of output, and the elasticity of the industry demand curve is one. If one firm increases its output by 10% to 55, the industry output increases to 105, which is a 5% increase. Since the price elasticity of demand is one, price must decline by 5%. But for the original firm, a 10% increase in production and a 5% decline in price indicate a price elasticity of two, not one.

As firms get more and more numerous in an industry, the demand curve each sees gets more and more elastic. When there are a great many sellers in the market, a change of output by any one of them has an insignificant effect on price. To each firm, the demand curve will look perfectly flat—the firm will seem able to sell whatever amount it wants at a fixed price. In this case, each firm is a **price taker** and sells in a perfectly competitive market. An example of this type of market is the market for wheat. There are a great many wheat farmers in many countries, and none has any noticeable control over the price at which it can sell in the world wheat market.

However, even when there are a great many sellers, each firm may have a downward-sloping demand curve. If buyers must expend time and effort to discover prices or the characteristics of the product, they will pick a seller and stay with it as long as they find the exchange satisfactory. These downward-sloping demand curves of small sellers are a result of the ambiguous definition of industry. The products most firms produce differ in some way, such as in quality, service, or location, from the products of other firms in the industry.

From the viewpoint of the firm, it is not the demand curve, but the child of the demand curve, the marginal revenue curve, which is of vital importance. **Marginal revenue** is the extra revenue a seller gets when it produces and sells another unit. For the price taker, the marginal revenue curve is the demand curve. For the farmer who can sell corn at Rs.4.00 a bushel, the extra revenue from selling another bushel is Rs.4.00. The demand curve for this farmer is flat at Rs.4.00, and so is his marginal revenue curve.

The table below illustrates why marginal revenue will be less than price for a price searcher. If the firm charges Rs.3.00, it can sell one unit and total revenue will be Rs.3.00. If it sells one more unit, it will be forced to cut price to Rs.2.00 and total

revenue will rise to Rs.4.00. Selling the extra unit adds only Rs.1.00 to revenue. Although the second unit sold for Rs.2.00, the firm had to cut the price it was previously receiving for the first unit by Rs.1.00, so the net increase in revenue was only Rs.1.00. By similar logic, selling the third unit reduces total revenue by Rs.1.00, so marginal revenue is -Rs.1.00.

Demand and Marginal Revenue		
Price	Quantity	Marginal Revenue
Rs.3.00	1	.
.		Rs.1.00
Rs.2.00	2	.
.		-Rs.1.00
Rs.1.00	3	.

The previous analysis assumes that the firm can charge only one price. If it can charge more than one price, charging higher prices to those willing and able to pay them and lower prices to others, it can move the marginal revenue curve closer to the demand curve, increasing profits (or reducing losses). This pattern of pricing is called **price discrimination.**

Economists generally assume that the demand curve is fixed, but many businesses do not regard it that way. It can vary seasonally, with the general level of business activity, or with a trend. The demand for turkeys has a pronounced seasonal movement. The demand for automobiles changes when there is a recession. The demand for baby food follows the trends in birth rate.

Business also may be able to move its demand curve through advertising. Advertising may simply give people information, it may change their goals, or it may change their perception of the product. For the firm it does not matter which happens. The result is the same—good advertising moves the demand curve to the right.

The demand curve can move for other reasons. If a firm lowers its price and later raises it back to its previous level, it may find that sales at the old price have changed. The lower price may attract new customers who have not tried the product before, and who find they like the product enough to stick with it when the old price is restored. Alternatively, some customers may expect prices to be cut again sometime in the future, and may decide to postpone purchases until it happens again. The opposite can happen if the firm temporarily raises price. It may encourage some customers to try substitutes, which they may find suit them better than the original product. Or it may encourage customers to buy more when the price comes down to prepare for any future increase.

The firm may also be able to change its demand curve by changing the characteristics of its product.

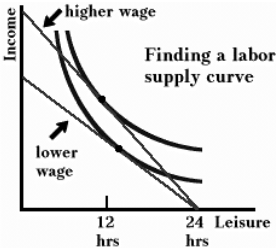
Finally, many firms sell several products that may be interrelated, and any pricing decision on one product will have effects not only on that product but on others. For example, the prices that General Motors charges for Chevrolets will affect the demand curve for Pontiacs.

The third and final boundary the firm faces is the supply curve for resources.

11.7 Supply of Resources

The third task of the firm is to obtain resources needed to produce a product. For each resource, a supply curve shows limitations that the firm faces. These supply curves are based on the preferences of sellers and on the actions of other firms which use the resource.

Because a market demand curve can be derived from utility curves and a budget line, it may seem surprising that a supply curve can also be derived from the same procedure. To see that it can, consider the graph below which shows indifference curves for income and leisure. Income is desirable because one can obtain other desirable things with it, and leisure is desirable because it lets one enjoy income.



In the world we live in, greater amounts of income must be purchased with more work—which means less leisure. The tradeoff between leisure and income, shown by the budget line, depends on the wage rate. If the wage rate is Rs.10, options open to an individual include 24 hours of leisure and no income, or 20 hours of leisure and Rs.40 of income, or 10 hours of leisure and Rs.140 of income.

To get a supply curve for labor, one must see what happens if the wage rate changes. Two wage rates are shown in the graph above. At the higher wage rate, the individual wants less leisure (which means he will work more as wages rise), but one could as easily draw indifference curves that show the amount of leisure rising as wages rise (which means he will work less as wages rise). Higher wages have two effects on the leisure-work decision, and these two effects pull in opposite directions. A higher wage rate increases the benefits of working, causing people to substitute work for leisure. This is called the **substitution effect** and is caused by changes in the slope of the budget line. Higher wages also increase income, and people want more leisure with a higher income. This is called the **income effect** of a price change, and is caused by changes in the distance of the budget line from the origin.

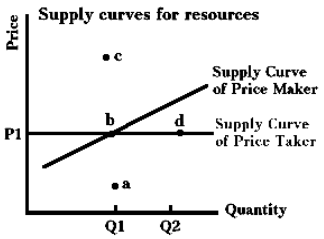
Usually the substitution and income effects reinforce each other, but they pull in opposite directions and almost cancel each other out in the case of labor. Most economists believe that the market supply curve for labor (found by adding up all the supply curves of individuals) is close to a horizontal line.

Supply curves for other resources can be obtained in similar ways. The supply curve for capital, for example, depends on decisions of people to consume now or to consume in the future. People who prefer to consume in the future will save and make funds available to finance capital. Their "time preference" determines the shape of indifference curves. The slope of the budget line depends on the interest rates. The

budget line tells how much one can get in the future if one sacrifices consumption now.

Although the market supply curve for labor may be almost vertical, few firms see this supply curve. If many other firms also are buying labor, what one firm does may have little effect on the overall market. If the firm is so small in the market that it can see no effects at all on the wage from its hiring decisions, it is a price taker. If it has some effect on wages, so that when it wants to hire more, it finds that wages rise, the firm is a price maker. The extreme case of a price maker is a **monopsonist**, the case of only **one buyer** in the market. The supply curve for a resource that a monopsonist sees is the same as the market supply curve.

The supply curve for a resource is a constraint or boundary on the firm because it shows the minimum that the firm can pay for a level of the resource. If the firm is a price taker in the resource market, it will face a horizontal supply curve such as that in the graph below. This curve indicates that any number can be bought at **P1** with no effect on price. There is no way the firm can attain point **a** even though it might prefer to pay less than **P1** because no one will sell at less than **P1**. Sellers will not sell because we assumed that there were a great many other buyers of the resource who will pay **P1**. Point **c** is possible, but a waste of money because the same amount of the resource could be bought for less.



If the firm is one of a few buyers or the only buyer of a resource, it may face a supply curve that slopes upward, making it a price searcher. It can obtain quantity **Q1** if it pays **P1**, but it must pay more than **P1** if it wants quantity **Q2**.

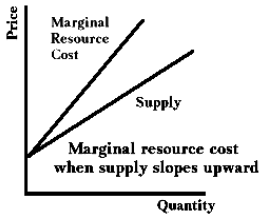
You should have noticed that there are similarities between a supply curve for a resource and a demand curve for output. Both are boundaries, and the curve the firm faces may differ from the market curve. The similarity goes further, because there is a counterpart for marginal revenue called "*marginal resource cost"* which measures the extra cost to the firm of hiring one more unit of the resource.

When a firm is a price taker, marginal resource cost is the same as the price of the resource. If the firm can hire as many workers as it wants at Rs.10 per hour, then hiring one more hour of labor adds Rs.10 to costs. Marginal resource cost and the supply of labor are both horizontal lines in this case.

When a firm is a price searcher facing an upward-sloping supply curve, the extra cost of hiring another unit of the resource is different from the price of the extra unit. The table below illustrates the reason for this difference. If the firm wants to buy two units, it cannot pay Rs.1.00 and get two. It must be willing to pay Rs.2.00 for each. However, the added cost of the second unit is not Rs.2.00, but Rs.3.00. This can be shown by comparing the total cost of two units and one unit, or Rs.4.00 less Rs.1.00. The added cost of the second unit is not only the two rupees that must be paid for it, but an added rupee for the first one. By the same logic, the added cost of a third unit is Rs.5.00.

Computing Marginal Resource Cost		
Quantity	Price	Marginal Resource Cost
1	Rs.1.00	Rs.1.00
2	Rs.2.00	2.00 + 1.00
3	Rs.3.00	3.00+2.00

The marginal resource cost curve lies above an upward-sloping supply curve because of the assumption that the firm can pay only one price. This is often a realistic assumption. If the firm hires two clerks who do exactly the same work, and pays one Rs.4.00 per hour and the other Rs.6.00 per hour, the lower-paid one will be unhappy and may refuse to work for the lower pay. On the other hand, many firms do not publicly disclose what they pay various people, and discourage employees from discussing salaries. To the extent that people are unaware of what others are earning, the firm may be able to pay different prices for the same resource, or in economists' jargon, to price discriminate. Price discrimination will pull down the MRC curve in the graph closer to, or perhaps even onto, the supply curve.



Next, we view a table that summarizes the many concepts in you have read in this lesson.

The major theme of the previous reading selections has been that in order to make a profit, a business must deal with three constraints or boundaries: the demand curve for output, the production function, and supply curves for inputs. The table below shows how these concepts are related to the functions of the business and to the marginal concepts that are central in microeconomics.

The Constraints Facing the Firm				
Task:	Economic Concept	Limitation Imposed by	Relationship between	Marginal Concept
Buying inputs	supply of resources	resource owners	price (money) and amount of resource	marginal resource cost
producing output	production function	technology	inputs and output	marginal product
selling output	demand curve	customers	output and price (money)	marginal revenue

The importance of these three constraints can be seen in a product that no firm can produce at a profit. Any one of three changes, if large enough, can change the situation and make a profit possible. The cost of the inputs may drop in price enough so that the product is profitable. Or technology may improve enough to make the product profitable. Or people

may increase the amount they are willing to pay for the product enough so that it is profitable. These three changes are changes in the supply-of-resources curves, the production function, and the demand curve, respectively.

11.8 Application

Diminishing returns plays an important part in the efficient allocation of resources. For efficiency, of course, we want to give more resources to the use in which they are more productive. But, as we give more resources to a particular use, we will observe diminishing returns — that use will become less productive. That may sound frustrating, but in fact it leads to a very important principle we can apply to the problem of efficient allocation of resources. We'll explore this in the next few pages and we will see the same principle again and again in the other chapters in this book.

We are not done with the theory of the firm, but only just begun. Next you will study the detailed analysis of economies of scale.

Notes

LESSON 12:
LAW OF VARIABLE PROPORTIONS

12.1 Law of Variable Proportions

Law of variable proportions, sometimes also referred to as the *law of diminishing returns*, this "law" is really a generalization economists make about the nature of technology when it is possible to combine the same factors of production in a number of different proportions to make the same product. The law states:

When increasing amounts of one factor of production are employed in production along with a fixed amount of some other production factor, after some point, the resulting increases in output of product become smaller and smaller.

(That is, first the marginal returns to successive small increases in the variable factor of production turn down, and then eventually the overall average returns per unit of the variable input start decreasing.) Since the law assumes that the available quantity of at least one factor of production is fixed at a given level and that technological knowledge does not change during the relevant period, the law of diminishing returns normally translates into a statement about the **short-run** choice of production possibilities facing a firm (since in the longer run it is virtually always possible for the firm to acquire more of the temporarily "fixed" factor — building an additional factory building, buying additional land, installing additional machines of the same kind, installing newer and more advanced machinery, and so on.)

A simple example of the workings of the law of diminishing returns comes from gardening. A particular twenty by twenty garden plot will produce a certain number of pounds of tomatoes if the gardener just puts in the recommended number of rows and plants per row, waters them appropriately and keeps the weeds pulled. If the gardener varies this approach by adding a pound of fertilizer to the topsoil, but otherwise does everything the same, he can increase the number of pounds of tomatoes the garden plot yields by quite a bit (notice the amount of land is being held fixed or constant). If he adds two pounds of fertilizer (rather than just one), probably he can get still more tomatoes per season, but the increase in tomatoes harvested by going from one pound to two pounds of fertilizer is probably smaller than the increase he gets by going from zero pounds to one (diminishing marginal returns). Applying three pounds of fertilizer may still increase the harvest, but perhaps by only a very little bit over the yields available using just two pounds. Applying four pounds of fertilizer turns out to be overdoing it — the garden yields fewer tomatoes than applying only three pounds because the plants begin to suffer damage from root-burn. And five pounds of fertilizer turns out to kill nearly all the plants before they even flower.

Another similar example of diminishing returns in an industrial setting might be a widget factory that features a certain number of square feet of work space and a certain number of machines inside it. Neither the space available nor the number of machines can be added to without a long delay for construction or installation, but it is possible to adjust the amount of labor on short notice by working more shifts and/or taking on some extra workers per shift. Adding extra man-hours of labor will increase the number of widgets produced, but only within limits. After a certain point, such things as worker fatigue, increasing difficulties in supervising the large work force, more frequent breakdowns by over-utilized machinery, or just plain inefficiency due to overcrowding of the work space begin to take their toll. The marginal returns to each successive increment of labor input get smaller and smaller and ultimately turn negative.

The law of diminishing returns is significant because it is part of the basis for economists' expectations that a firm's short-run marginal cost curves will slope upward as the number of units of output increases. And this in turn is an important part of the basis for the law of supply's prediction that the number of units of product that a profit-maximizing firm will wish to sell increases as the price obtainable for that product increases.

12.2 Production with Two Variable Inputs: Isoquants

In the first part of this lesson you viewed the following production function:

$Q(L, K) = 1 * L^{0.5} * K^{0.5}$

If you haven't done so already, go through the previous lesson first.

The concept of marginal productivity is central to economists' understanding of efficient allocation of resources. For an illustrative example, consider a farmer who has two fields to plant. He can grow a crop of corn (let's say) on each of them, but has a limited amount of labor to allocate between them. Let us say that the farmer can spend 1000 hours of labor, total, on the two fields. If he spends one more hour of labor on the north field, that means he has one hour less to spend on the south field.

Here are the production functions for the two fields.

Table 12.1

Labor Input and Output on Two Fields			
North Field		South Field	
labor	output	labor	output
0	0	0	0
100	9500	100	12107
200	18000	200	23429
300	25500	300	33964
400	32000	400	43714
500	37500	500	52679
600	42000	600	60857
700	45500	700	68250
800	48000	800	74858
900	49500	900	80679
1000	50000	1000	85715

Allocation of Labor and Total Output on Two Fields		
Labor on North Field	Labor on South Field	total output in bushels of com
0	1000	85000
100	900	89600
200	800	92400
300	700	93400
400	600	92600
500	500	90000
600	400	85600
700	300	79400
800	200	71400
900	100	61600
1000	0	50000

We can visualize the production functions for the two fields. Figure 1, below, shows the production function for the relatively fertile south field with a vertically dashed purple curve, and the less fertile north field with a solid green curve. As we see, the south field can always produce more, with the same amount of labor, as the north field can.

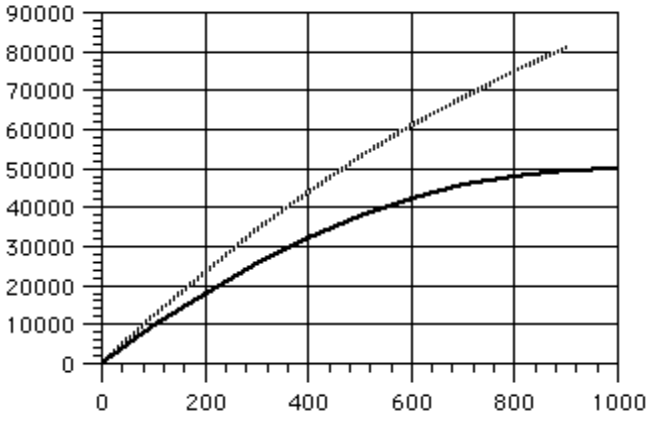


Figure 12.1. Production Functions for Two Fields

12.3 The Problem of Allocation

The farmer's "allocation problem" is: How much labor to commit to the north field, and how much to the south field? One "common sense" approach might be to abandon the infertile north field and allocate the whole 1000 hours of labor to the south field. But a little arithmetic shows that this won't work. Here is a table that shows the correlated quantities of labor on the two fields, and the total output of corn from both fields taken together.

Table 12.2

We see that the farmer gets his largest output by allocating most, but not all, of his labor to the south field. Because of the principle of diminishing returns, however, he shouldn't put all his resources into the one field, but divide the labor resource (unevenly!) between the two.

But how much should go to the north plot, and how much to the south plot?

We can visualize the efficient allocation of resources with a graph like this one. The labor used on the infertile north field is measured on the horizontal axis and the total output from both fields in shown on the vertical axis. (We are assuming, of course, that all labor not used on the North field is used on the South field). The dark green curve shows how total output changes as we shift labor from the north field to the south field. Thus, the top of the curve is the interesting spot — that's where we get the most output. In this example, that's the efficient allocation of resources between the two fields.

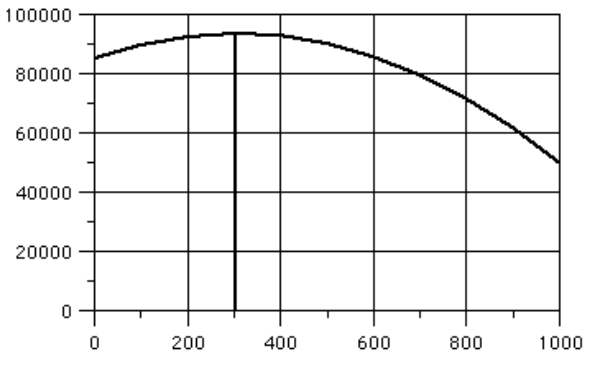


Figure 12.2: Maximum Production

It's easy to see that we should put some labor to work on the north field — but not too much. The vertical orange line shows that the maximum output — the top of the dark green curve — comes when about 300 labor days are allocated to the north

field and the rest, 700 labor days, to the south field. And that's exactly right.

It's pretty easy to see where the maximum is in this simple example. But in a more realistic example, in which there could be many more than just two dimensions, it's harder to visualize. We need a rule that we can apply in more complex, realistic examples, a rule that will tell us if we have or don't have an efficient allocation of resources.

That's where the economist's "marginal approach" comes in. The objective is to get to the top of the hill. You could call "the marginal approach" the "bug's-eye view." Think of yourself as a bug climbing up that production hill in the picture. How will you know when you are at the top?

12.4 Marginal Productivity and Allocation

If you were a bug, you couldn't see much. Perhaps you couldn't see to the top of the hill. But you would be able to tell if you were going up, or down, or neither. So you would just keep going as long as you were going up, and stop when you were neither going up nor down. That's the way a bug gets to the top of a hill.

If you were a farmer with two fields, it's a little more complicated, but the same principles apply: take it step by step. However much you may be producing, ask yourself "What would happen if I were to take one worker away from the North Field and put her to work on the South Field? How much less will the North Field produce? The answer to that question is the marginal productivity of labor on the North Field. How much more will the South Field produce? The answer to that question is the marginal productivity of labor on the South Field. So the move of labor from the North Field will increase production if the marginal productivity on the North Field is less than the marginal productivity on the South Field. Like the bug, you want to keep moving in that direction as long as production keeps getting greater, that is, as long as the marginal productivity on the North Field is less than the marginal productivity on the South Field. And you stop when further movement won't get you any higher on the hill, that is, when the marginal productivities are equal on the two fields.

So now, let's visualize the marginal productivities for these two fields. But this time we will do it a slightly different way. We will measure the labor used on the infertile north field from left to right on the horizontal axis. Then, what's left is what's available for the north field, so we will measure the labor used on the south field from left to right — from 1000 hours down to zero. The marginal product on the north field is shown with the green line, and the marginal product on the south field with the vertical-dashed purple line. (Remember, the marginal productivity on the south field decreases as the labor input on the south field gets bigger, so the marginal productivity on that field increases as labor used on the field gets smaller, as it does here). Here it is:

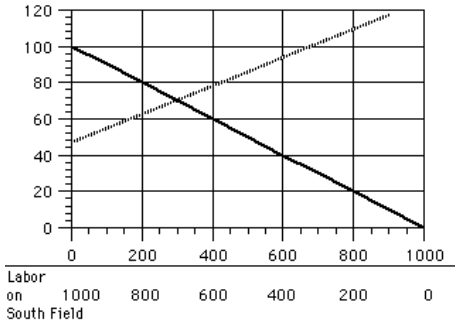


Figure 12.3: Marginal Productivity and Efficient Allocation

12.5 Marginal Productivity

Figure 6 shows the most efficient allocation of resources in this case. It is to allocate 300 hours of labor to the north field, and 700 hours to the south field, as shown by the vertical red-orange line. For maximum output, labor is allocated so that the marginal productivity of labor on the north field is equal to the marginal productivity of labor on the south field.

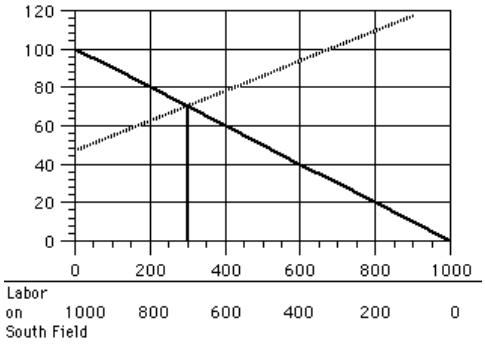


Figure 12.3: Efficient Allocation

To see why this works, think it through in reverse: what happens if the allocation of labor is not 300 to the north field and 700 to the south field? For example, suppose 200 hours are allocated to the south field and 800 to the north field. This puts us to the left of the orange line — and we read off the diagram that the marginal productivity of labor on the north field is 80 bushels of corn, while the marginal productivity on the south field is about 62. Remembering the definition of marginal productivity, that means: if the farmer spends one additional hour on the north field, he will gain 80 bushels, while spending one less on the south field will cost him 62 bushels, leaving a net gain of 18 bushels. What has happened is that spending 800 hours of labor on the south field has pushed the "diminishing returns" on that field so far that it is less productive **at the margin** than the north field. and that will be true anywhere to the left of the orange line, since, in that range, the marginal productivity on the north field is always greater than the marginal productivity on the south field.

Now let's see what happens if the allocation is to the right of the most efficient one — for example, suppose the farmer were to allocate 600 hours to the north field and 400 to the south field. Looking at the diagram, we see that the marginal produc-

tivity on the north field is 40 while the marginal productivity on the south field is 90. Thus, moving an hour of labor from the north field to the south field will yield a gain of 90-40=fifty bushels of corn. And the farmer will continue to gain as he moves toward the efficient allocation from the right, because, in that range, the marginal product on the south field is always bigger than the marginal product on the north field.

We have seen that the farmer can gain by reallocating his labor from either side toward the efficient output. Once he has 300 hours of labor on the north field and 700 on the south, the farmer cannot increase his output any further. That is why we think of it as the "efficient" allocation of resources.

12.6 Marginal Productivity and the Equimarginal Principle

This is a quite general principle, which we may state as follows.

Rule:

When the same product or service is being produced in two or more units of production, in order to get the maximum total output, resources should be allocated among the units of production in such a way that the marginal productivity of each resource is the same in each unit of production.

This example may also be a little clearer example of what we mean by "efficient allocation of resources." In the example, we have a tiny economy, consisting of one farmer and two plots of land. When the marginal productivities on the two plots are equal, this tiny economy has an "efficient allocation of resources." Of course, real economies are more complex, but the principles governing the efficient allocation of resources are the same.

This rule has a name: it is the Equimarginal Principle. The idea is to make two things equal "at the margin" — in this case, to make the marginal productivity of labor equal on the two fields. As we will see, it has many applications in economics. In more complicated cases, we will have to generalize the rule carefully. In this example, for instance, we are allocating resources between two fields that produce the same output. When the different areas of production are producing different kinds of goods and services, it will be more complicated. But a version of the Equimarginal Principle will still apply.

Notes

LESSON 13:
THEORY OF FIRM

13.1 Theory of the Firm

In developing the supply and demand approach to economics, economists first worked out the basis of the demand curve. By treating the demand for a product or service as a rational decision by a (primarily) self-interested individual or family, economists were able to understand the relation of the demand for one product or service to the demands for other products and services and to many other forms of economic activity. It was natural to apply the same approach to supply. As a first step, we need to think about the decision-makers in supplying goods and services, and what a "rational decision" to supply goods and services would mean. In economics, this is often called the "Theory of the Firm."

In the remainder of this lesson we will apply the concepts of marginal productivity and diminishing returns to the theory of the firm. First we will talk a bit about business firms and their role in a market economy, then we will return to the marginal productivity approach.

A firm is a unit that does business on it's own account. (Firm is from the Italian, "firma," a signature, and the idea is that a firm can commit itself to a contract). Thus, the firm is the decision-maker in supplying goods and services.

There are three main kinds of firms in modern market economies:

13.1.1 Proprietorships

A proprietorship (or proprietary business) is a business owned by an individual, the "proprietor." Many "Mom and Pop stores" — and other "Mom and Pop" businesses — are proprietorships. Some proprietorships are too small even to employ one person full time. Craftsmen, such as plumbers and painters, may have "day jobs" and work as self-employed proprietors part time after hours. Computer programmers and others may also do that. At the other extreme, some proprietary businesses employ many hundreds of workers in a wide range of specializations. In a proprietorship, the proprietor is almost always the decision-maker for the business.

13.1.2 Partnerships

A partnership is a business jointly owned by two or more persons. In most partnerships, each partner is legal liable for debts and agreements made by any partner. Of course, this requires a great deal of trust, and thus partners generally know one another well enough to have that sort of trust. Family partnerships are very common for that very reason. (There are now a few "limited partnerships" in which some partners are protected from legal liability for the agreements made by others, beyond some limits). In many cases, one partner is designated as the managing partner and is the main decision-maker for the business.

13.1.3 Corporations

A corporation has two characteristics that distinguish it from most proprietorships and partnerships:

- Limited liability
- Anonymous ownership

Limited liability means that the owner of shares in a corporation cannot lose more than a certain amount if the company fails. Usually the amount is the money paid to buy the shares. Anonymous ownership means that the owner of the shares can sell them without getting the permission of anyone other than the buyer. By contrast, in most partnerships, no one partner can sell out without getting the agreement of the other partners. In such a case the continuing partners will, of course, want to know about the new partner — he will not be an "anonymous owner." In a typical corporation, the shareholders formally elect a board of directors, who in turn select the officers of the company. One of these officers, often called the "president," will be the principle decision-maker for the firm, but he will be expected to make decisions in the interest of the shareholders.

While there are millions of proprietorships, typically very small, the biggest businesses are corporate and corporations are particularly important because of their size.

13.2 Objectives

As we recall, Malthus did not have firms in mind when he formulated the Law of Diminishing Returns. But this law has applications Malthus did not envision, and we will see how to apply the law to a business firm. In the Reasonable Dialog of economics in the nineteenth century, the development of these ideas was a bit indirect. In about the eighteen-seventies, economists were rethinking the theory of consumer demand. They applied a version of "diminishing returns" and the Equimarginal Principle to determine how a consumer would divide up her spending among different consumer goods. (We'll get into that in another lesson). That worked pretty well, and so some other economists, especially the American economist John Bates Clark, tried using the same approach in the theory of the firm. These innovations were the beginning of Neoclassical Economics.

Following the Neoclassical approach, we will interpret "rational decisions to supply goods and services" to mean decisions that maximize — something! What does a supplier maximize? The operations of the firm will, of course, depend on its objectives. One objective that all three kinds of firms share is profits, and it seems that profits are the primary objective in most cases. We will follow the neoclassical tradition by assuming that firms aim at maximizing their profits.

There are two reasons for this assumption. First, despite the growing importance of nonprofit organizations and the frequent calls for corporate social responsibility, profits still seem to be the most important single objective of producers in our

market economy. Thus it is the right place to start. Second, a good deal of the controversy in the reasonable dialog of economics has centered on the implications of profit motivation. Is it true, as Adam Smith held, that the "invisible hand" leads profit-seeking businessmen to promote the general good? To assess that question, we need to understand the implications of profit maximization.

13.3 Profit

Profit is defined as revenue minus cost, that is, as the price of output times the quantity sold (revenue) minus the cost of producing that quantity of output.

However, we need to be a little careful in interpreting that. Remember, economists understand cost as opportunity cost — the value of the opportunity given up. Thus, when we say that businesses maximize profit, it is important to include all costs — whether they are expressed in money terms or not.

For example, a cab-driver — the self-employed proprietor of an independent cab service — says: "I'm making a 'profit,' but I can't take home enough to support my family, so I'm going to have to close down and get a job." The proprietor is ignoring the opportunity cost of her own labor. When those opportunity costs are taken into account, we will find that he is not really making a profit after all.

Let's say that the cab-driver makes $500 a week driving his cab, after all expenses (gasoline, maintenance, etc.) have been taken out. Suppose he can get wages (including tips!) of $800 driving for someone else, with hours no longer and about the same conditions otherwise. Then $800 is the opportunity cost of his labor, and after we deduct the opportunity cost from his $500 net as an independent cabbie, he is actually losing $300 per week.

This is one of the most important reasons for using the opportunity cost concept: it helps us to understand the circumstances that will lead people to get into and out of business.

Because accountants traditionally considered only money costs, the net of money revenue minus money cost is called "accounting profit." (Actually, modern accountants are well aware of opportunity cost and use the concept for special purposes). The economist's concept is sometimes called "economic profit." If there will be some doubt as to which concept of profit we mean, we will sometimes use the terms "economic profit" or "accounting profit" to make it clear which is intended.

13.4 The John Bates Clark Model

Like any other unit, a firm is limited by the technology available. Thus, it can increase its outputs only by increasing its inputs. As usual, this will be expressed by a production function. The output the firm can produce will depend on the land, labor and capital the firm puts to work.

In formulating the Neoclassical theory of the firm, John Bates Clark took over the classical categories of land, labor and capital and simplified them in two ways. First, he assumed that all labor is homogenous — one labor hour is a perfect substitute for any other labor hour. Second, he ignored the distinction between land and capital, grouping together both kinds of nonhuman inputs under the general term "capital." And he assumed that this broadened "capital" is homogenous.

Of course, the simplifying assumptions aren't true — John Bates' Clark's conception of the firm is highly simplified, like a map at a very large scale. In more advanced economics, we can get rid of the simplifying assumptions and deal with a much more realistic "map" of the business firm. But for most of this book, we'll take that on faith, and stick to the simplified version Clark gave us. That will make it simpler, and the principles we will discover are sound and applicable to the real world in all its complexity.

In the John Bates Clark model, there are some important differences between labor and capital, and they relate to the long and short run.

13.4.1 Short and Long Run

A key distinction here is between the short and long run.

Some inputs can be varied flexibly in a relatively short period of time. We conventionally think of labor and raw materials as "variable inputs" in this sense. Other inputs require a commitment over a longer period of time. Capital goods are thought of as "fixed inputs" in this sense. A capital good represents a relatively large expenditure at a particular time, with the expectation that the investment will be repaid — and any profit paid — by producing goods and services for sale over the useful life of the capital good. In this sense, a capital investment is a long-term commitment. So capital is thought of as being variable only in the long run, but fixed in the short run.

Thus, we distinguish between the short run and the long run as follows:

In the perspective of the short run, the number and equipment of firms operating in each industry is fixed.

In the perspective of the long run, all inputs are variable and firms can come into existence or cease to exist, so the number of firms is also variable.

13.4.2 Assumptions

The John Bates Clark model of the firm is already pretty simple. We are thinking of a business that just uses two inputs, homogenous labor and homogenous capital, and produces a single homogenous kind of output. The output could be a product or service, but in any case it is measured in physical (not money) units such as bushels of wheat, tons of steel or minutes of local telephone calls. In the short run, in addition, the capital input is treated as a given "fixed input." Also, we can identify the price of labor with the wage in the John Bates Clark model. (In a modern business firm, we have to include benefits as well as take-home wages. The technical term for the total, wages and benefits, is "employee compensation.")

We will add two more simplifying assumptions. The new simplifying assumptions are:

- The price of output is a given constant.
- The wage (the price of labor per labor hour) is a given constant.

Putting them all together — just two kinds of input and one kind of output, one kind of output fixed in the short run, and given output price and wage — it seems to be a lot of simplify-

ing assumptions, and it is. But, as we will see in later lessons of the book, these are not arbitrary simplifying assumptions. They are the assumptions that fit best into many applications, and the starting point for still others.

Once we have simplified our conception of the firm to this extent, what is left for the director of the firm to decide.

13.5 The Firm's Decision

In the short run, then, there are only two things that are **not** given in the John Bates Clark model of the firm. They are the output produced and the labor (variable) input. And that is not actually two decisions, but just one, since labor input and output are linked by the "production function." Either

- the output is decided, and the labor input will have to be just enough to produce that output
  or
- the labor input is decided, and the output is whatever that quantity of labor can produce.

Thus, the firm's objective is to choose the labor input and corresponding output that will maximize profit.

Let's continue with the numerical example in previous lesson. Suppose a firm is producing with the production function shown there, in the short run. Suppose also that the price of the output is $100 and the wage per labor-week is $500. Then let's see how much labor the firm would use, and how much output it would produce, in order to maximize profits.

The relationship between labor input and profits will look something like this:

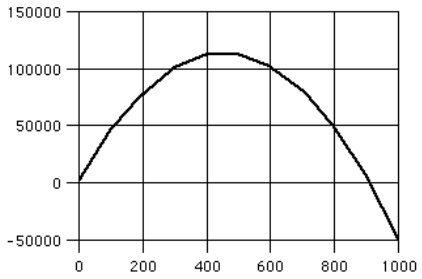


Figure 13.1: Labor Input and Profits

In the figure, the green curve shows the profits rising and then falling and the labor input increases. Of course, the eventual fall-off of profits is a result of "diminishing returns," and the problem the firm faces is to balance "diminishing returns" against the demand for the product. The objective is to get to the top of the profit hill. We can see that this means hiring something in the range of four to five hundred workers for the week. But just how many?

The way to approach this problem is to take a bug's-eye view. Think of yourself as a bug climbing up that profit hill. How will you know when you are at the top?

13.6 The Marginal Approach

The bug's-eye view is the marginal approach. However much labor is being employed at any given time, the really relevant question is, supposing one more unit of labor is hired, will profits be increased or decreased? If one unit of labor is eliminated, will profits increase or decrease? In other words, what does one additional labor unit add to profits? What would elimination of one labor unit subtract from profits?

We can break that question down. Profit is the difference of revenue minus cost. Ask, "What does one additional labor unit add to cost? What does one additional labor unit add to revenue?

The first question is relatively easy. What one additional labor unit will add to cost is the wage paid to recruit the one additional unit.

The second question is a little trickier. It's easier to answer a related question: "What does one additional labor unit add to production?" By definition, that's the marginal product — the marginal product of labor is defined as the additional output as a result of increasing the labor input by one unit. But we need a measurement that is comparable with revenues and profits, that is, a measurement in money terms. Since the price is given, the measurement we need is the Value of the Marginal Product:

Value of the Marginal Product

The Value of the Marginal Product is the product of the marginal product times the price of output. It is abbreviated VMP.

To review, we have made some progress toward answering the original question. Adding one more unit to the labor input, we have

increase in revenue = value of marginal product
increase in cost = wage

So the answer to "What will one additional labor unit add to profits?" is "the difference of the Value of the Marginal Product Minus the wage." Conversely, the answer to "What will the elimination of one labor unit add to profits?" is "the wage minus the Value of Marginal Product of Labor." And in either case the "addition to profits" may be a negative number: either building up the work force or cutting it down can drag down profits rather than increasing them.

So, again taking the bug's-eye view, we ask "Is the Value of the Marginal Product greater than the wage, or less?" If greater, we increase the labor input, knowing that by doing so we increase profits by the difference, VMP-wage. If less, we cut the labor input, knowing that by doing so we increase profits by the difference, wage-VMP. And we continue doing this until the answer is "Neither." Then we know there is no further scope to increase profits by changing the labor input — we have arrived at maximum profits.

Let's see how that works. Let's go back to the numerical example from earlier in the lesson, and assume that the price of output is $100 per unit and the wage is $500. In Figure 8, below, we have the value of the marginal product, $100*MP, and the wage for that example.

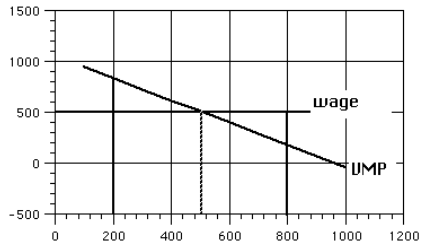


Figure 13.2

Now suppose that the firm begins by using just 200 units of

"If I were to increase the labor input to 201, that would increase both costs and revenues. By how much? Let's see: the VMP is 850, so the additional worker will add $850 to revenues. Since the wage is $500, the additional worker will add just $500 to cost, for a net gain of $350. It's a good idea to "upsize" and add one more worker.

On the other hand, suppose that the firm is using 800 units of labor, as shown by the other orange line. The manager asks herself, "If I were to cut the labor input to 799, that would cut both costs and revenues. By how much? Let's see: the VMP is 200, so the additional worker will add just $200 to revenues. Since the wage is $500, the additional worker will add just $500 to cost, for a net loss of $300. It's time to "downsize" and cut the labor force.

In each case, there is an unrealized potential, and the amount of unrealized potential is the difference between the VMP and the wage. The firm's profit potential will not be 100% realized until the VMP is equal to the wage. That's the "equimarginal principle" again.

13.7 The Equimarginal Principle

By taking the marginal approach — the bug's-eye view — we have discovered the diagnostic rule for maximum profits. The way to maximize profits then is to hire enough labor so that

VMP=wage

where p is the price of output and VMP = p*MP the marginal productivity of labor in money terms.

This is another instance of the Equimarginal Principle. The rule tells us that profits are not maximized until we have adjusted the labor input so that the marginal product in labor, in dollar terms, is equal to the wage. Since the wage is the amount that the additional (marginal) unit of labor adds to cost, we could think of the wage as the "marginal cost" of labor and express the rule as "value of marginal product of labor equal to marginal cost." But we will give a more compete and careful definition of marginal cost (of output) in the next lesson.

13.8 Profit Maximization

In our numerical example, suppose that the price of output is $100 per unit and the wage is $500 per worker per period. Then the p*MP, wage, and profits will be something like this:

Table 13.1

Labor	Marginal Productivity	p*MP	Wage	Accounting Profit
0	9.45	945	500	0
100	8.35	835	500	44500
200	7.25	725	500	78000
300	6.15	615	500	100500
400	5.05	505	500	112000
500	3.95	395	500	112500
600	2.85	285	500	102000
700	1.75	175	500	80500
800	0.65	65	500	48000
900	-0.45	-55	500	4500
1000			500	-50000

What we see in the table is that the transition from 400 to 500 units of labor gives p*MP=505, very nearly VMP=wage. And that is the highest profit. So the profit-maximizing labor force is about 500 units.

We can get a more exact answer by looking at a picture. Here is a picture of the profit-maximizing hiring in this example:

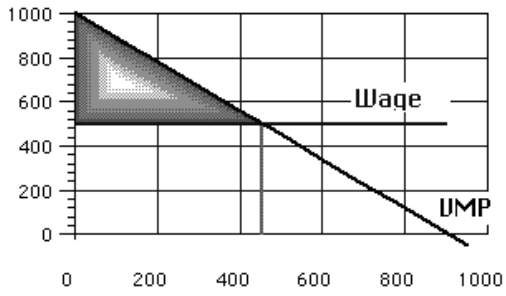


Figure 13.3: Maximizing Profit

The picture suggests that the exact amount is a bit less than 500 units of labor. The exact number is 454.54545454545 ... units of labor — a repeating decimal fraction.

Notice the shaded area between the VMP curve and the price (wage) line. n the picture, the area of the shaded triangle is the total amount of payments for profits, interest, and rent — in other words, everything the firm pays out for factors of production other than labor. The rectangular area below the wage line and left of the labor=454 line is shows the wage bill. Thus, the John Bates Clark model provides us with a visualization of the division of income between labor and property. We'll make use of this fact in exploring the economics of income distribution in the last Part of this lesson.

We can use the diagram also to understand why VMP=wage is the diagnostic that tells us the profit is at a maximum. Suppose the labor input is less than 500 — for example, suppose labor input is 200. Than an additional labor-day of labor will add about 7.8 units to output, and about $780 to the firm's sales revenue, but only $500 to the firm's costs, adding roughly $220 to profits. So it is profitable to increase the labor input from 200, or, by the same reasoning, from any labor input less than $500.

This difference between the VMP and the wage is the increase or decrease in profits from adding or subtracting one unit of labor. It is sometimes called the **marginal profit** and (as we observed in studying consumers' marginal benefits) the absolute value of the marginal profits is a measure of unrealized potential profits. That's why the businessman wants to adjust the labor input so that VMP-wage=0.

Let's try one more example. Suppose the labor input is 800 labor-days per week. If the firm "downsizes" to 799 labor-days, it reduces its output by just about 1.2 units and its sales revenue by about $120, but it reduces its labor cost by $500, increasing profits by about $380. Thus a movement toward the VMP=wage again increases profits by realizing some unrealized potential profit.

The formula VMP=wage is a diagnostic for maximum profits because it tells us that there is no further potential to increase the profits by adjusting the labor input — marginal profit is zero.

The marginal productivity rule is the key to maximization of profits in the short run. But now let's take a look at the long run perspective.

13.8 Increasing Returns to Scale and the Long Run

In microeconomics, we think of diminishing returns as a short run thing. In the long run, all inputs can be increased or decreased in proportion. Reductions in the marginal productivity of labor, due to increasing the labor input, can be offset by increasing the tools and equipment the workers have to work with. How will that come out, on net? The answer is — "it all depends!"

In the long run we define three possible cases:

Decreasing returns to scale

If an increase in all inputs in the same proportion k leads to an increase of output of a proportion less than k, we have decreasing returns to scale. Example: If we increase the inputs to a dairy farm (cows, land, barns, feed, labor, everything) by 50% and milk output increases by only 40%, we have decreasing returns to scale in dairy farming. This is also known as "diseconomies of scale," since production is less cheap when the scale is larger.

Constant returns to scale

If an increase in all inputs in the same proportion k leads to an increase of output in the same proportion k, we have constant returns to scale. Example: If we increase the number of machinists and machine tools each by 50%, and the number of standard pieces produced increases also by 50%, then we have constant returns in machinery production.

Increasing returns to scale

If an increase in all inputs in the same proportion k leads to an increase of output of a proportion greater than k, we have increasing returns to scale. Example: If we increase the inputs to a software engineering firm by 50% output and increases by 60%, we have increasing returns to scale in software engineering. (This might occur because in the larger work force, some programmers can concentrate more on particular kinds of programming, and get better at them). This is also known as "economies of scale," since production is cheaper when the scale is larger.

In introductory economics, we usually discuss these long run tendencies in the context of cost analysis, rather than marginal productivity analysis. However, increasing returns to scale, in particular, creates some complications for the application of marginal productivity thinking. Thus, I think there may be something to gain by exploring how increasing returns to scale goes together with marginal productivity. To keep it as simple as possible, we will look at a numerical example of a two-person labor market and a fictitious product that is produced with increasing returns to scale. Economists often like to talk about the production of "widgets," so our fictitious industry is the widget-tying industry.

13.8.1 Example of Production with Increasing Returns to Scale

Assumptions:

- Since this is a long run analysis, there is no fixed input. Indeed, for simplicity, there is only one input. Labor is the only input and is variable.
- Our small economy is populated by three people: Bob and John, workers, and Gordon, an entrepreneur (that is, a person who will organize a business if and only if it is profitable to do so).
  - Bob, working alone, can produce output worth 2000 per week.
  - Bob's opportunity cost is 2100 per week. (That means Bob can earn 2100 in producing some other good or service).
  - John, working alone, can produce 2000 per week.
  - John's opportunity cost is 2800 per week.
- If Bob and John work together, thanks to division of labor, they can produce 5500 per week. Suppose, for example, that Gordon sets up a Widget-Tying business and hires Bob and, later, John to do the work. This is an example of "increasing returns to scale" since input increases by 100% when the second worker is hired and output increases by 175% as a result.

Why would output increase more than in proportion to inputs? First, simply having four hands may increase productivity as the two men can simultaneously do different parts of the job. Second, each may concentrate on some part of the work, getting better at it with more practice, but leaving the other part to the other worker who also gains practice and skill in that part. (These were the kinds of advantages Adam Smith particularly stressed). Finally, each may concentrate on the tasks for which he has a greater inborn talent.

Notice that the two-person widget-tying operation uses resources with an opportunity cost of 2800+2100=4900 and produces output worth 5500, for a net increase in production of 600. Evidently, it is a good thing that such a team be organized.

13.8.2 The Dark Side of the Force

Increasing returns to scale are a powerful force for increasing productivity, but the problem of organizing them efficiently is "the dark side of the force." We have seen that an enterprise that yields a net gain of 600 to society cannot be organized, in this example, without producing a loss. The market system cannot take advantage of the potentiality for gain through division of labor and increasing returns to scale in this case. This possibility was discovered by an early 20th Century British economist named Arthur Charles Pigou, but despite 80 years of discussion, this analysis is not at all widely understood, even among professional economists. Pigou thought it might be a good idea for the government to subsidize enterprises with increasing returns to scale. In this case a subsidy of 150 would make the widget-tying enterprise profitable and produce a gain of 600 in national product.

There may be another solution. Since the widget-tying enterprise adds 600 to national output but loses at least 100, we might ask, what happens to the difference of 700? The answer is that Bob gets it. Bob is paid at least 2800 but his opportunity cost is only 2100, accounting for the difference of 700. Suppose that Bob and John were not paid the same wage, but, instead, each was paid his opportunity cost plus 100. The wage bill would then be 2200+3000=5200 and Gordon would finish with a profit of 300. Thus, wage discrimination may make it

possible for the widget-tying enterprise to exist when it cannot exist so long as each worker is paid the same wage for the same work.

The conclusions are surprising, and understandably, controversial — yet the numbers support them, both in this and more complicated and abstract examples.

1. Some people believe it is just that each person be paid according to her or his contribution, and interpret "marginal productivity" as the person's contribution. However, this may impossible when there are increasing returns to scale, as there may not be enough output to pay everyone on that basis.
2. Compromising, some would say that each person ought to be paid in proportion to her or his contribution, so that people are paid equally for the same work. That, too, may be impossible.
3. Discrimination or subsidy may be necessary to allow some socially useful activities to exist.
4. There may be no simple system of payment (such as supply and demand or equal pay for equal work) that will allow a socially useful enterprise with increasing returns to scale to exist.

I think this is the reason we have organizations. If there were no increasing returns to scale, there would be little reason for any business to employ more than one person. We would instead have an economy consisting of self-employed individuals, like a yeoman agricultural system. Instead we see an economic system consisting in part of large, complicated organizations with internal arrangements and payments systems that have little to do with contributions or marginal productivity, and may be discriminatory. From an abstract point of view, they may waste resources by not paying at the marginal productivity; but the benefits of increasing returns to scale are so great that, even falling far short of potential efficiency, they can still be very productive.

This is sometimes lost sight of by the organizations themselves. People naturally avoid complexity, and organizations sometimes try to set up simple, market-like internal payment and fund transfer systems, hoping that this will increase efficiency. But, as we have seen, this can fail badly in the context of increasing returns to scale (and that is the context of any large productive organization). We have recently been through such an experience at Drexel. A few years ago we went over to "revenue centered budgeting." The idea was to let the colleges retain a high proportion of the revenues they produce, through tuition, grants, contracts and so on. This would (it was felt) give the deans and college faculties more "incentive" to set up popular new programs and initiatives. However, it wasn't possible to let the colleges keep 100%, since some money is needed to run shared services like the computer center, student-life activities, and the library, not to mention the salaries of high administrators (and we wouldn't think of mentioning that). But it couldn't be made to work. If the proportion kept by the colleges was high enough to make it profitable for them to set up new programs and initiatives, there was not enough for the purposes of the central administration; while if the proportion taken by the central administration was enough to do its job, then the colleges were losing money on their new programs and initiatives — no incentive! So Drexel has moved away from "revenue centered budgeting" in practice, although there is still some work being done to try to work out a "revenue centered budgeting" system that will work. Here's a prediction based on the theory of increasing returns to scale: a revenue-centered budgeting system probably can be made to work, but it will be just as complex and frustrating than the centralized budgeting traditionally has been. That complexity and frustration (and large organizations) are the price we pay for the benefits of increasing returns to scale.

13.9 Summary

We have seen that the concept of marginal productivity and the law of diminishing marginal productivity play central parts in both the efficient allocation of resources in general and in profit maximization in the John Bates Clark model of the business firm.

The John Bates Clark model and the principle of diminishing marginal productivity provide a good start on a theory of the firm and of supply. In applying the marginal approach and the equimarginal principle to profit maximization, it extends our understanding of the principles of efficient resource allocation. Some key points in the discussion have been

- the distinction between marginal productivity and average productivity
- the "law of diminishing marginal productivity"
- the rule for division of a resource between two units producing the same product: equal marginal productivities
- the diagnostic formula VMP=wage, that tells us the input and output are adjusted to maximize profits in the business firm, in the short run
- In the long run, there may be increasing, decreasing, or constant returns to scale. Increasing returns to scale will complicate things somewhat for the marginal productivity approach.

This has given us a start on the theory of the business firm. But we will want to reinterpret the model of the firm in terms of cost — since the cost structure of the firm is important in itself, and important for an understanding of supply.

LESSON 14:
APPLICATIONS OF THEORY OF PRODUCTION

Production: the creation of any good or service that has economic value to either consumers or other producers. Production analysis focuses on the efficient use of inputs to create outputs. The process involves all of the activities associated with providing goods and services.

Managerial Questions:

1. Whether to produce or shut down
2. How much to produce
3. What input combination to use
4. What type of technology to use

Examples:

a. physical processing or manufacturing of material goods

b. production of transportation services

c. production of legal advice

d. production of education

e. production of invention (R & D)

f. production of bank loans

Production Function: a mathematical model relating the maximum quantity of output that can be produced from given amounts of various inputs.

or

a schedule (table, equation) showing the maximum amount of output that can be produced from any specified set of inputs, given the existing technology or "state of the art."

In short — the production function is a catalog of output possibilities.

$$Q = f(X,Y) \text{ or } (K,L)$$

Economic Efficiency:

The production function incorporates the technically efficient method of production — the latest technological processes are used. When economists use production functions they assume that the maximum level of output is obtained from any given combination of inputs; that is, they assume that production is technically efficient.

When producers are faced with input prices, the problem is not technical but economic efficiency: how to produce a given amount of output at the lowest possible cost. To be economically efficient, a producer must determine the combination of inputs that solves this problem.

What then is technical inefficiency? If, for example, an alternative process can produce the same amount of output using less of one or more inputs and the same amounts of all others, then the first process is technically inefficient.

If, however, the second process uses less of some inputs but more of others, the economically efficient method of producing a given level of output depends on the prices of the inputs. One might cost less but actually be less technically efficient.

Classifying Inputs:

Fixed: a fixed input is one that is required in the production process. The amount of the fixed input employed is constant over a given period of time regardless of the quantity of output produced.

Variable Input: one whose quantity employed in the production process is varied, depending on the desired quantity of output to be produced.

Time Frames:

Short-Run: a period of time in which one or more of the inputs is fixed.

Very Short-Run: all resources are fixed.

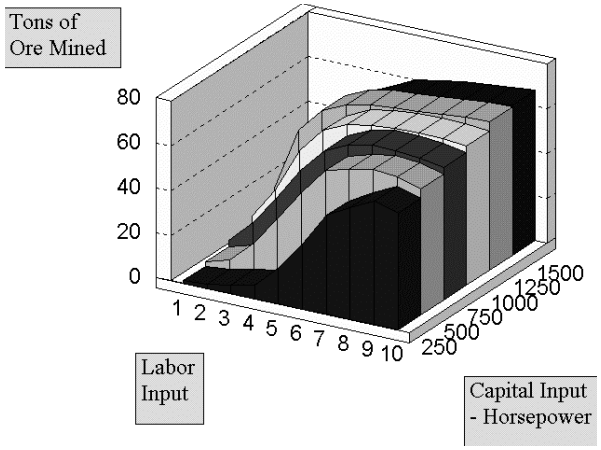
Long Run: time period long enough that all resources can be varied.

Ore mining example

Output = tons or ore mined

Capital (horsepower)

Labor								
	250	500	750	1,000	1,250	1,500	1,750	2,000
1	1	3	6	10	16	16	16	13
2	2	6	10	24	29	29	44	44
3	4	16	29	44	55	55	55	50
4	6	29	44	55	58	60	60	55
5	16	43	55	60	61	62	62	60
6	29	55	60	62	63	63	63	62
7	44	58	62	63	64	64	64	64
8	50	60	62	63	64	65	65	65
9	55	59	61	63	64	65	66	66
10	52	56	59	62	64	65	66	67



Labor	Output	MP $\Delta Q \div \Delta X$	AP $(Q \div X)$	Elasticity MP ÷ AP
0	0			
1	6	+6	6	1.0
2	16	+10	8	1.25
3	29	+13	9.67	1.34
4	44	+15	11	1.36
5	55	+11	11	1.0
6	60	+5	10	.50
7	62	+2	8.86	.23
8	62	0	7.75	0.0
9	61	-1	6.78	-.15
10	59	-2	5.90	-.34

Returns to Scale: The relation between output and variation in all inputs taken together.

Returns to factor: The relation between output and variation in only one of the inputs employed.

Total Product: The total output that results from employing a specific quantity of resources in a production system.

Marginal Product: the incremental change in total output that can be produced by the use of one more unit of the variable input in the production process.

-or-

The change in output associated with a unit change in one input factor, holding other inputs constant.

ΔQ change brought about by a change in

ΔX units of the variable input

Y remains fixed

$$MP_x = \frac{?\,Q}{?\,X}$$

-or, in continuous terms:

$$MP_x = \frac{\partial Q}{\partial X}$$

Average Product

$$AP_x = \frac{Q}{X}$$

the ratio to total output to the amount of the variable input used in producing the output.

Production Elasticity:

The percentage change in output resulting from a given percentage change in the amount of the variable input X employed in the production process with Y remaining constant.

-or-

The percentage change in output associated with a 1 (one) percent change in all inputs.

$$E_X = \frac{\%\Delta Q}{\%\Delta X}$$

$$= \frac{\Delta Q/Q}{\Delta X/X}$$

Rearranging Terms —

$$= \frac{\Delta Q/\Delta X}{Q/X}$$

$$= \boxed{\frac{MP_X}{AP_X}}$$

Law of Diminishing Marginal Returns:

The use of increasing amounts of a variable factor in a production process beyond some point, given the amount of all other production factors remains unchanged, will eventually result in diminishing marginal returns in total output.

This observation is easily verified by reviewing the slope of the marginal product curve.

As the number of units of the variable input increases, other inputs held constant, there exists a point beyond which the marginal product of the variable input declines.

Note:

a) not a mathematical theorem

b) empirical assertion

Summary: The concepts of total and marginal product and the law of diminishing returns to a factor are important in identifying efficient as opposed to inefficient input combinations.

Determining the optimal use of the variable input

With one of the inputs (Y) fixed in the short run, the producer must determine the optimal quantity of the variable input (X) to employ in the production process.

This is a study of Marginal Revenue Product (MRP) and Marginal Factor Cost (MFC). That is, this is a study about the role of revenue and cost in the production system.

MRP

The conversion from physical to economic relations is accomplished by multiplying the MP of input factors by the MR resulting from the sale of goods or services produced to obtain a quantity known as the *Marginal Revenue Product* of input:

Def: the amount that an additional unit of the variable input adds to total revenue, -or-

The economic value of a marginal unit of a particular input factor when used in the production of a specific product.

$$\boxed{MRP = \frac{\Delta TR}{\Delta X}}$$

where ΔTR is the change in total revenue associated with the given change (ΔX) in the variable input.

MRPx is equal to the marginal product of X (MPx) times the marginal revenue (MRQ) resulting form the increase in output obtained:

$$MRP = MP_X \cdot MR_Q$$

Example

Units	TP	MP	MR at $5
1	3	3	$15
2	7	4	20
3	10	3	15
4	12	2	10
5	13	1	5

If the addition of one more laborer to a work force would result in the production of two incremental units of a product than can be sold for $5, the MP of labor is 2, and its MRP is $10 (2 x $5).

Marginal Factor Cost

Def: MFC is the amount that an additional unit of the variable input adds to total cost

$$MFC = \frac{\Delta TC}{\Delta X}$$

Optimal Input Level

Given the marginal revenue product and marginal factor cost, we can compute the optimal amount of the variable input to use in the production process.

Recall that prior discussions on optimality and marginal analysis suggest that an economic activity should be expanded as long as the marginal benefits exceed the marginal costs. The optimal point occurs at the point where the marginal benefits are equal to the marginal costs.

$$MRP_X = MFC_X$$

Single Input System

Profit maximization requires production at a level such that marginal revenue equals marginal cost. Because the only variable in the system is input L, the marginal cost of production is:

$$MC = \frac{\Delta TC}{\Delta Output}$$

$$MC = \frac{P_L}{MP_L}$$

Since marginal revenue must equal marginal cost at the profit-maximizing level, MR_Q can be substituted for MC_Q

$$MR_Q = \frac{P_L}{MP_L}$$

solving for PL yields:

$$PL = MR_Q \times MP_L$$

-or-

$$PL = MRPL$$

The profit maximizing firm will always employ an input up to the point where its marginal revenue product equals its cost.

Note : if $MRP_L > P_L$ the expand labor usage.

if $MRP_L < P_L$ the cutback labor usage.

The Ore-Mining Example Revisited

1. Firm can employ as much labor as it needs by paying workers $50 per period (the labor market is considered perfectly competitive).
2. Firm can sell all the ore it can produce at a price of $10 per ton.

   MRP = MFC = $50
3. At less than 6 workers, MRP > MFC and the addition of more workers will increase revenues. Beyond 6, the opposite is true.

Labor Input	Total Prod. Tons of Ore	MP of Labor	TR or P • Q	MR $\frac{\Delta TR}{\Delta Q}$	MRP MP•MR	MFC
0	0		0			
1	6	6	60	10	60	50
2	16	10	160	10	100	50
3	29	13	290	10	130	50
4	44	15	440	10	150	50
5	55	11	550	10	110	50
6	60	5	600	10	50	50
7	62	2	620	10	20	50
8	62	0	620	10	0	50

Production functions with 2 variable inputs

Using the Ore Mining example, assumed that both capital and labor are now variable.

Production Isoquant

A production isoquant is either a geometric curve or an algebraic function representing all the various combinations of the two inputs that can be used in producing a given level of output, -or-

An isoquant is a curve (a locus of points) showing all possible combinations of inputs physically capable of producing a given fixed level of output.

Marginal rate of technical substitution

Def: the rate at which one input may be substituted for another input in the production process, -or-

the rate at which one input is substituted for another along an isoquant.

The rate of change of one variable with respect to another variable is given by the slope of the curve relating the two variables. Thus, the rate of change of input Y with respect to X — that is, the rate at which Y may be substituted for X in the production process — is given by the slope of the curve relating Y to X. This is the slope of the isoquant.

Since the slope is negative and one wishes to express the substitution rate as a positive quantity, a negative sign is attached to the slope.

$$MRTS = \frac{Y_1 - Y_2}{X_1 - X_2} = \frac{\Delta Y}{\Delta X}$$

For example, in the Ore Mining problem, moving from 3 to 4 workers yields an MRTS of 250 (horsepower).

$$MRTS = -\frac{750-500}{3-4} = 250$$

Stated differently, for every unit of labor added 250 horsepower may be discharged without changing total output.

Note that we can show the MRTS to equal the ratio of the marginal products of X and Y; remember:

$$\Delta Y = \frac{\Delta Q}{MP_Y}$$

and,

$$\Delta X = \frac{\Delta Q}{MP_X}$$

substituting these in above yields,

$$MRTS = \frac{MP_X}{MP_Y}$$

The optimal combination of inputs

The firm must make two input choice decisions:

1. Choose the input combination that yields the maximum level of output possible with a given level of expenditure.
2. Choose the input combination that leads to the lowest cost of producing a given level of output.

This occurs when, in any constrained optimization problem, we choose the level of each activity whereby the marginal benefits

from the last unit of each activity per dollar cost of the activity are equal. This is known as the *equimarginal* criterion.

$$\frac{MP_X}{C_X} = \frac{MP_Y}{C_Y}$$

-or-

$$\frac{MP_X}{P_X} = \frac{MP_Y}{P_Y}$$

The optimal combination of inputs in either the cost-minimization or output-maximization problem is a function of the relative prices of the inputs.

Changes in input prices

Assume that a firm is producing with the most cost minimizing combination of labor and capital. This is an efficient operation. From our development above, we know that:

$$\frac{MP_X}{C_X} = \frac{MP_Y}{C_Y}$$

Now suppose that the price of input X rises while the price of input Y is unchanged and the original combination of inputs MPX and MPY are unchanged. At the original combinations, the increase in CX makes:

$$\frac{MP_X}{C_X} < \frac{MP_Y}{C_Y}$$

Substitution Effect:

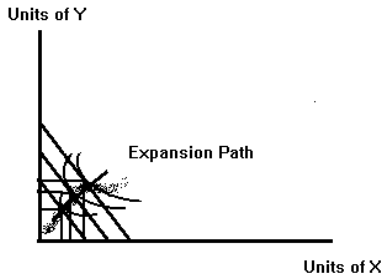
If the firm wishes to produce the same level of output, it will increase Y and decrease X as it moves along the isoquant.

If the input-price ratio changes, firms substitute toward the input that becomes relatively less expensive and away form the input that becomes relatively more expensive. In the case of labor and capital, if wages/interest increases (decreases), K/L increases (decreases) at each level of output. This change in the K/L ratio is called the *substitution* effect.

Isoquant and Isocost Combinations

Optimal input proportions can be found graphically for a two-input, single-output system by adding a budget line or isocost curve (a line of constant costs) to the diagram of production isoquants.

Each point on a budget line represents some combination of inputs (X and Y) whose cost equals constant expenditure.



The *expansion path* is the optimal input combinations for increasing output. Note that the proportions in which the inputs are combined need not be the same for all levels of outputs. Stated differently, the expansion path *shows how factor proportions change when output changes*, with the factor-price ratio held constant.

The Decision Making Principle:

To minimize the cost (expenditure) of producing a given level of output with fixed input prices, the producer must combine inputs in such quantities that the marginal rate of technical substitution of capital and labor is equal to the input ratio (the price of labor to the price of capital).

Returns to scale

Production theory also offers a means for analysis of the effects on output of changes in the scale of production.

An increase in the scale of production consists of a *simultaneous* proportionate increase in *all* the inputs used in the production process. The proportionate increase in the output of the production process that results from the given proportionate increase in all the inputs is defined as the physical returns to scale.

$$\varepsilon_Q = \frac{\%\Delta Q}{\%\Delta X} = \frac{\partial Q}{\partial X} \bullet \frac{X}{Q}$$

$\varepsilon Q > 1$ increasing

$\varepsilon Q = 1$ constant

$\varepsilon Q < 1$ diminishing

1. Increasing: Output goes up proportionately more than the increase in input usage.
2. Constant: Output goes up by the same proportion as the increase in input usage.
3. Decreasing: Output goes up proportionately less than the increase in input usage.

Estimation of production functions

One of the more common approaches utilizes the Cobb Douglas production function method.

$$Q = aL^{\rho 1}K^{\rho 2}$$

1. Both inputs are required to create output.
2. MRTS will diminish as required by production theory

Logarithmic Specification

$$Ln\ Q = Ln\ \alpha + \beta_1 Ln L + \beta_2 Ln K$$

Elasticity of Production

$$E_L = \frac{MP_L}{AP_L}$$

where;

$$MP_L = a^{\rho 1}L^{\rho 1^{-1}}K^{\rho 2} \quad \text{and,}$$

$$APL = \frac{aL^{\rho 1}K^{\rho 2}}{L} = aL^{ß_1 - 1}K^{ß_2}$$

thus

$$E_L = \frac{aß_1 L^{\rho 1\text{-}1}K^{\rho 2}}{aL^{ß_1 - 1}K^{ß2}} = ß1$$

The Exponents - Returns to Scale

1. Increasing: $\beta_1 + \beta_2 > 1$
2. Constant: $\beta_1 + \beta_2 = 1$
3. Decreasing: $\beta_1 + \beta_2 < 1$

Example

$Q = 1.01^{0.75} K^{0.25}$

Q was an index of physical volume of manufacturing; L was an index of the average number of employed wage earners only (that is, salaried employees, officials, and working proprietors were excluded); and K was an index of the value of plants, buildings, tools and machinery reduced to dollars of constant purchasing power.

The sum of the exponents were restricted to one (constant returns to scale).

Later Studies by Cobb and Douglas

$Q = .84\ L^{0.63}\ K^{0.30}$

A one (1) percent increase in labor input results in about a 2/3 percent increase in output, and a one (1) percent increase in capital input results in approximately a 1/3 percent increase in output.

The sum of the exponents is slightly less than one (1). Seems to indicate the presence of decreasing returns to scale, however, the sum is not significantly different from 1.0; hence, it really confirms constant returns to scale.

A Three Variable Model

$$Q = aL_{\rho}^{\rho_1} L_n^{\rho_2} K^{\rho_3}$$

Q is the value added by production plants over 18 industries

L_n is non-production work-years

L_p is production work-hours

K is gross book values of depreciable and depletable assets

Empirical Estimation of a Production Function for: Major League Baseball

by CE. Zech as published in the American Economist

In an attempt to quantify the factors that contribute to the team's success, a Cobb-Douglas production function was developed using data from the 26 major league baseball teams in 1977. Output (Q) was measured by team victories. Inputs from five different categories were included in the model:

- Hitting: batting average and power (home runs)
- Running: stolen base record
- Defense: fielding percentage and total chances accepted
- Pitching: earned run average (ERA) and strikeouts-to-walks ratio.
- Coaching: Lifetime won-lost record and number of years spent managing in the major leagues.
- Dummy: NL = 0, AL = 1

Variable	Eq. 1	Eq. 2	Eq. 3	Eq. 4
Constant	.017	.018	.010	.008
Dummy	-.002	-.003	.004	.003
B avg	2.017	1.986	1.969	1.927
HomeRuns	.229	.299	.208	.215
Stolen Bases	.119	.120	.110	.112
Strikeouts/Walks	.343	.355	.324	.334
TotField Chances	1.235	1.200		
Field %			5.62	5.96
Mngr.W/L		-.003		-.004
Mngr. Years	-.004		-.004	
R^2	.789	.790	.773	.774

Findings:

1. Hitting average contributes almost six times as much as pitching to a team's success. Contradict traditional wisdom?
2. Home runs contribute about twice as much as stolen bases to a team's success.
3. Coaching skills are not significant in any of the regression equations.
4. Defensive skills are not significant in any of the regression equations.

The sums of the statistically significant variables in each of the four equations range from 2.588 to 2.709. Because these are all greater than 1.0, the baseball production functions exhibit *increasing returns to scale*.

Notes

LESSON15:
COST ANALYSIS - I

15.1 Introduction to Cost

We can look at the business firm from at least two points of view: productivity, inputs, and outputs (as we have just done) or outputs and costs. In advanced microeconomics, these two points of view are called "duals." They are equally valid, but they point up different things. They are also opposites from a certain point of view — the higher the productivity, the lower the costs. By looking at the firm from the point of view of costs, we shift our perspective somewhat, and gain a much more direct understanding of supply.

We also look more directly at the difference between the long and short run. In the short run, we have two major categories of costs:

- fixed costs, and
- variable costs

In the long run, however, all costs are variable. Thus, we must study costs under two quite different headings. Costs will vary quite differently in the long run and in the short.

15.2 Fixed & Variable Cost

Variable costs are costs that can be varied flexibly as conditions change. In the John Bates Clark model of the firm that we are studying, labor costs are the variable costs. Fixed costs are the costs of the investment goods used by the firm, on the idea that these reflect a long-term commitment that can be recovered only by wearing them out in the production of goods and services for sale.

The idea here is that labor is a much more flexible resource than capital investment. People can change from one task to another flexibly (whether within the same firm or in a new job at another firm), while machinery tends to be designed for a very specific use. If it isn't used for that purpose, it can't produce anything at all. Thus, capital investment is much more of a commitment than hiring is. In the eighteen-hundreds, when John Bates Clark was writing, this was pretty clearly true. Over the past century, a) education and experience have become more important for labor, and have made labor more specialized, and b) increasing automatic control has made some machinery more flexible. So the differences between capital and labor are less than they once were, but all the same, it seems labor is still relatively more flexible than capital. It is this (relative) difference in flexibility that is expressed by the simplified distinction of long and short run.

Of course, productivity and costs are inversely related, so the variable costs will change as the productivity of labor changes.

Here is a picture of the fixed costs (FC), variable costs (VC) and the total of both kinds of costs (TC) for the productivity example in the last unit:

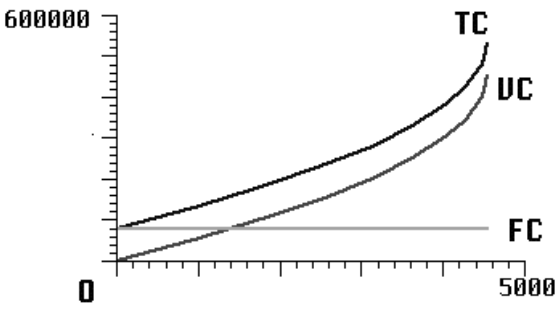


Figure 15.1

Output produced is measured toward the right on the horizontal axis. The cost numbers are on the vertical axis. Notice that the variable and total cost curves are parallel, since the distance between them is a constant number — the fixed cost.

15.3 Opportunity Cost

What is the connection between the distinction we have just made — fixed vs variable costs — and opportunity cost, the key concept in some earlier chapters?

In economics, all costs are included — whether or not they correspond to money payments. If we have opportunity costs with no corresponding money payments, they are called **implicit costs**. The implicit costs (as well as the money costs) are included in the cost analysis we have just given.

There is some correlation between implicit costs and fixed or variable costs, but this correlation will be different in such different kinds of firms as

a factory owned by an absentee investor

This is the easiest case to understand. All of the labor costs to the absentee investor are money costs, including the manager's salary. If the investor has borrowed some of the money he invested in the factory, then there are some money costs of the capital invested — interest on the loan. However, we must consider the opportunity cost of invested capital as well. The investor's own money that he has used to buy the factory is money that she could have invested in some other business. The return she could have gotten on another investment is the opportunity cost of her own funds invested in the business. This is an implicit cost, and in this case the implicit cost is part of the cost of capital and probably a fixed cost.

a "mom-and-pop" store

A "mom-and-pop" store (family proprietorship or partnership) is a store in which family members are self-employed and supply most of the labor. Typically, "Mom" and "Pop" don't pay themselves a salary — they just take

money from the till when they need it, since it is their property anyway. As a result, there are no money costs for their labor. But their labor has an opportunity cost — the salary or wages they could make working similar hours in some other business — and so, in this case, the implicit costs include a large component of variable labor costs.

a large modern corporation

The corporation has relatively few implicit costs, but generally will have some. All labor costs will be expressed in money terms (though benefits and bonuses have to be included), since the shareholders don't supply labor to the corporation as "Mom and Pop" do in a family proprietorship. It will pay interest to bondholders and dividends to shareholders. But the dividends aren't really a cost item — they include profits distributed to the shareholders. Moreover, the typical corporation will retain some profits and invest them within the business, a "plowback" investment. Conversely, shareholders may take a large part of their payout in appreciation of the stock value — and plowback investment is one reason for the appreciation. Thus we would say that the corporation has a net equity value, that is, that the corporation "owns" a certain amount of capital that it invests in its own business (very much like the absentee owner in the first example). This capital has an opportunity cost, and that opportunity cost is an implicit cost. The stockholders, who own the corporation, ultimately receive (as dividends or appreciation) both the opportunity cost of the equity capital and any profit left over after it is taken out.

15.4 Unit Cost

Costs may be more meaningful if they are expressed on a per-unit basis, as averages per unit of output. In this way, we again distinguish

Average Fixed Cost (AFC)

This is the quotient of fixed cost divided by output. In the numerical example we are using, when output is 4020 (in the table) fixed cost is 80000, so AFC is 80000/4020=19.9

Average Variable Cost (AVC)

This is the quotient of average cost divided by output. In the example, at an output of 4020 the variable cost is 350000, giving AVC of 350000/4020=87.06.

Average Total Cost (ATC or AC)

This is the quotient of total cost divided by output. In the example, with 4020 of output total cost is 430000, so AC is 430000/4020 = 106.96 = 87.06+19.90.

Once again, I will illustrate these concepts with the numerical example of the firm from the previous chapter

Here are the average, average variable, and average fixed costs for our example firm.

Table 15.1

Q	AC	AFC	AVC
945	138	85	53
1780	101	45	56
2505	92	32	60
3120	90	26	64
3625	91	22	69
4020	95	20	75
4305	100	19	81
4480	107	18	89
4545	117	18	99

Here are the average cost (AC), average variable cost (AVC) and average fixed cost (AFC) in a diagram. This is a good representative of the way that economists believe firm costs vary in the short run.

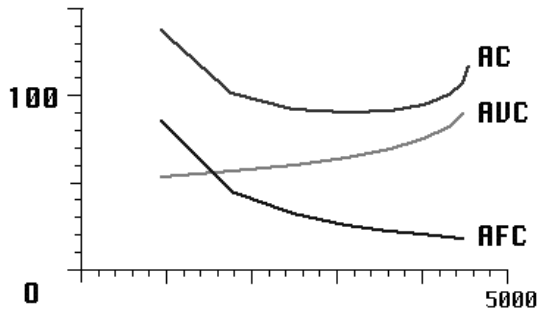


Figure 15.2

Notice how the average fixed costs decline as the fixed costs are "spread over more units of output." For large outputs, however, average variable costs rise pretty steeply. The idea is that with a limited capital plant and thus limited productive capacity — in the short run — costs would rise much more than proportionately to output as output goes beyond "capacity." The average total cost, dominated by fixed costs for small output, declines at first, but as output increases, fixed costs become less important for the total cost and variable costs become more important, and so, after reaching a minimum, average total cost begins to rise more and more steeply.

13.5 Marginal Cost

As before, we want to focus particularly on the marginal variation. In this case, of course, it is marginal cost. Marginal cost is defined as

$$MC=\frac{\Delta C}{\Delta Q}$$

As usual, Q stands for (quantity of) output and C for cost, so ΔQ stands for the change in output, while ΔC stands for the change in cost. As usual, marginal cost can be interpreted as the additional cost of producing just one more ("marginal") unit of output.

Let's have a numerical example of the Marginal Cost definition to help make it clear. Total cost is 280000 for an output of 3120, and it is 33000 for an output of 3625. So we have

$$\Delta C = 330000-280000=50000$$

and

$$\Delta Q = 3625 - 3120 = 505$$

so that

$$\frac{\Delta C}{\Delta Q} = \frac{50000}{505} = 99.01$$

for a marginal cost of $99.01 for the next unit produced. As usual, this is an approximation, and the smaller the change in output we use, the better the approximation is.

Here is the marginal cost for our example firm, along with output and average cost.

Table 15.2

Output	Average Cost	Marginal Cost
0	0	9.45
945	137.57	52.91
1780	101.12	59.88
2505	91.82	68.97
3120	89.74	81.30
3625	91.03	99.01
4020	94.53	126.58
4305	99.88	175.44
4480	107.14	285.71
4545	116.61	769.23

Here is a picture of marginal cost for our example firm, together with average cost as output varies.

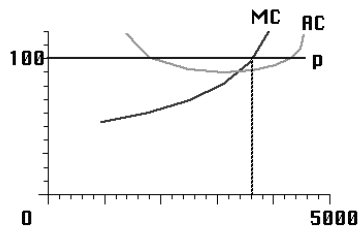


Figure 15.3

As before, the output produced is measured by the distance to the right on the horizontal axis. The average and marginal cost are on the vertical axis. Average cost is shown by the curve in yellow, and marginal cost in red. Notice how the marginal cost rises to cross average cost at its lowest point.

13.6 Maximization of Profit & Cost

We can now give another rule for the maximization of profits. The new rule is really just the same rule as we saw before, only now we state it in terms of price and costs. It is the equimarginal principle in yet another form.

The question is: "I want to maximize profits. How much output should I sell, at the given price?"

The answer is**: increase output until**

$$p = MC$$

The point is illustrated by the following table, which extends the marginal cost table in an earlier page to show the price and the profits for the example firm.

Table 15.3

Output	Average Cost	Marginal Cost	price	profit
0	0	9.45	100	0
945	137.57	52.91	100	-35503.65
1780	101.12	59.88	100	-1993.60
2505	91.82	68.97	100	20490.90
3120	89.74	81.30	100	32011.20
3625	91.03	99.01	100	32516.25
4020	94.53	126.58	100	21989.40
4305	99.88	175.44	100	516.60
4480	107.14	285.71	100	-31987.20
4545	116.61	769.23	100	-75492.45

Notice how profits are greatest (at 32516.25) when the marginal cost is almost exactly equal to the price of $100. This occurs at an output of 3625, with marginal cost at 99.01. The profit-maximizing output would be very slightly more than 3625.

13.7 Supply & Cost

We have discovered the principle of supply for the individual firm.

Remember: what is supply? It is the relation between the price and the quantity that people want to sell. For an individual firm, that is: the relation between the price and the quantity the firm wants to sell.

So we ask: at a given price, how much will a (profit- maximizing) firm want to sell? The answer: enough so that the price is equal to marginal cost. In other words, the marginal cost curve is the supply curve for the individual firm.

13.8 Shutdown Point

As long as the firm produces something, it will maximize its profits by producing "on the marginal cost curve." But it might produce nothing at all. When will the firm shut down?

The answer goes a bit against common sense. The firm will shut down if it cannot cover its variable costs. So long as it can cover the variable costs, it will continue to produce.

This is an application of the opportunity cost principle. Just because fixed costs are fixed, they are not opportunity costs in the short run — so they are not relevant to the decision to shut down. Even if the company shuts down, it must pay the fixed costs anyway. But the variable costs are avoidable — they are opportunity costs! So the firm will shut down it it cannot meet the variable (short run opportunity) costs. But as long as it can pay the variable costs and still have something to apply toward the fixed costs, it is better off continuing to produce.

It is important not to confuse shut-down with bankruptcy. They are two different things. If a company cannot pay its interest and debt payments (usually fixed costs), then it is bankrupt. But that doesn't mean it will shut down. Bankrupt firms are often reorganized under new ownership, and continue to produce — just because they can cover their variable costs, and so the new owners do better to continue producing than to shut down.

13.9 Long Run Cost

Thus far we have not considered the long run in cost theory. We will now think a bit about the long run, using the concept of average cost.

We have defined "the long run" as "a period long enough so that all inputs are variable." This includes, in particular, capital, plant, equipment, and other investments that represent long-term commitments. Thus, here is another way to think of "the long run:" it is the perspective of investment planning.

So let's approach it this way: Suppose you were planning to build a new plant — perhaps to set up a whole new company — and you know about how much output you will be producing. Then you want to build your plant so as to produce that amount at the lowest possible average cost.

To make it a little simpler we will suppose that you have to pick just one of three plant sizes: small, medium, and large. Here's the way they look in a picture:

Here are the average cost curves for the small (AC1), medium (AC2) and large(AC3) plant sizes:

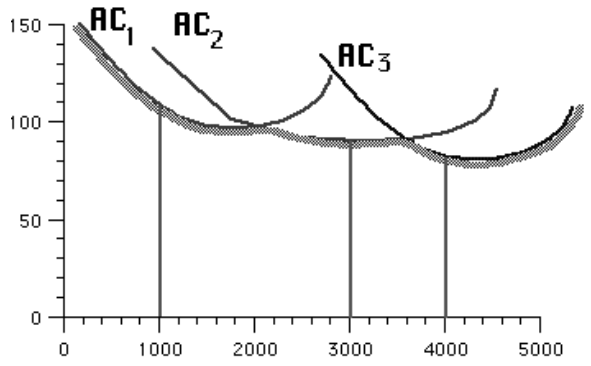


Figure 15.4

If you produce 1000 units, the small plant size gives the lowest cost.

If you produce 3000 units, the medium plant size gives the lowest cost.

If you produce 4000 units, the large plant size gives you the lowest cost.

Therefore, the long run average cost (LRAC) — the lowest average cost for each output range — is described by the "lower envelope curve," shown by the thick, shaded curve that follows the lowest of the three short run curves in each range.

More realistically, an investment planner will have to choose between many different plant sizes or firm scales of operation, and so the long run average cost curve will be smooth, something like this:

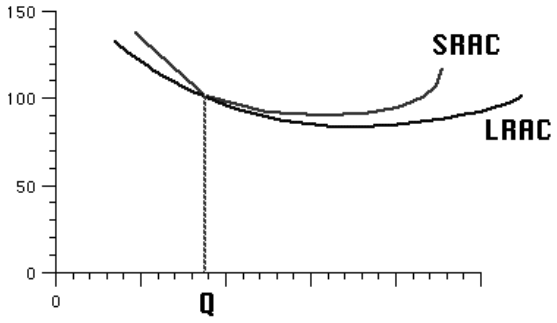


Figure 15.5

As shown, each point on the LRAC corresponds to a point on the SRAC for the plant size or scale of operation that gives the lowest average cost for that scale of operation.

13.10 Returns to Scale

In our pictures of long run average cost, we see that the cost per unit changes as the scale of operation or output size changes. Here is some terminology to describe the changes:

increasing returns to scale = decreasing cost

average cost decreases as output increases in the long run

constant returns to scale = constant costs

average cost is unchanged as output varies in the long run

decreasing returns to scale = increasing costs

average cost increases as output increases in the long run

Here are pictures of the average cost curves for the three cases:

increasing returns to scale = decreasing cost

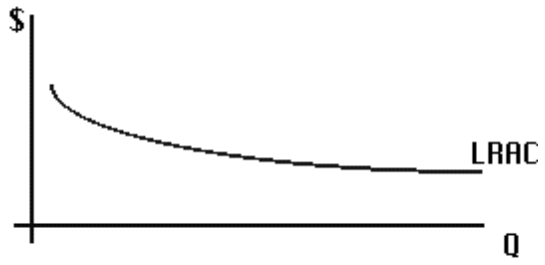


Figure 15.6

constant returns to scale = constant costs

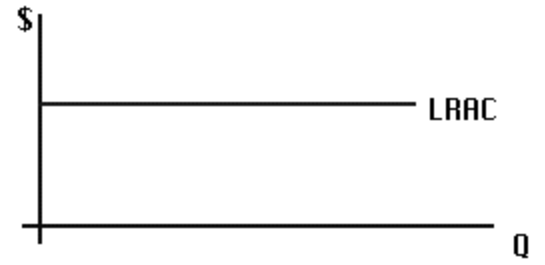


Figure 15.7

decreasing returns to scale = increasing costs

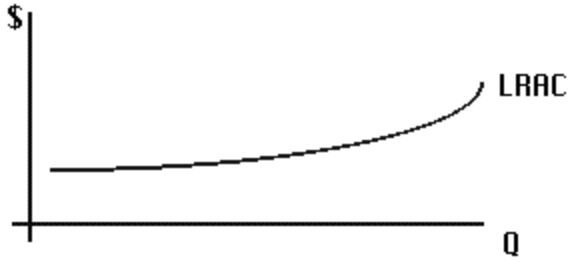


Figure 15.8

3.9.1.1 Increasing Returns to Scale

Economists usually explain "increasing returns to scale" by indivisibility. That is, some methods of production can only work on a large scale — either because they require large-scale machinery, or because (getting back to Adam Smith, here) they require a great deal of division of labor. Since these large-scale methods cannot be divided up to produce small amounts of output, it is necessary to use less productive methods to produce the smaller amounts. Thus, costs increase less than in proportion to output — and average costs decline as output increases.

Increasing Returns to Scale is also known as "economies of scale" and as "decreasing costs." All three phrases mean exactly the same.

13.9.1.2 Constant Returns to Scale

We would expect to observe constant returns where the typical firm (or industry) consists of a large number of units doing pretty much the same thing, so that output can be expanded or contracted by increasing or decreasing the number of units. In the days before computer controls, machinery was a good example. Essentially, one machinist used one machine tool to do a series of operations to produce one item of a specific kind — and to double the output you had to double the number of machinists and machine tools.

Constant Returns to Scale is also known as "constant costs." Both phrases mean exactly the same.

13.9.1.3 Decreasing Returns to Scale

Decreasing returns to scale are associated with problems of management of large, multi-unit firms. Again with think of a firm in which production takes place by a large number of units doing pretty much the same thing — but the different units need to be coordinated by a central management. The management faces a trade-off. If they don't spend much on management, the coordination will be poor, leading to waste of resources, and higher cost. If they do spend a lot on management, that will raise costs in itself. The idea is that the bigger the output is, the more units there are, and the worse this trade-off becomes — so the costs rise either way.

Decreasing Returns to Scale is also known as "diseconomies of scale" and as "increasing costs." All three phrases mean exactly the same.

In our examples, the LRAC is (more or less roughly) u-shaped, like this:

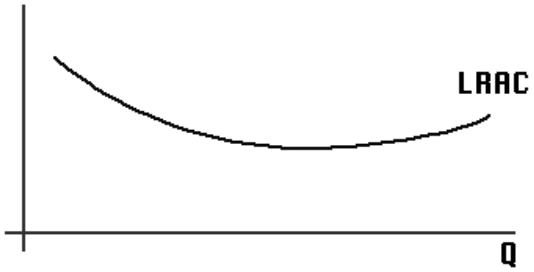


Figure 15.9

The idea is that:

- for small outputs, indivisibilities predominate, and so long run average cost declines with increasing output

- for intermediate outputs, operations can be expanded roughly proportionately, while tendencies to increasing and decreasing costs — if any — offset one another.
- for large outputs, the problems of management predominate, and so long run average cost increases with increasing output.
- That's reasonable — but we should recall that it is pretty much a guess, and may or may not apply in a particular case!

15.10 Summary

By thinking in terms of cost, rather than productivity, we gain several points of understanding of supply:

- While total and average cost concepts have their uses, the most important in the short run is marginal cost.
- To maximize profits, the firm will increase output to the point where marginal cost equals a given price.
- Therefore, the marginal cost curve is itself the supply curve, in the short run.
- The firm will shut down, however, if it cannot cover its variable cost.
- We may think of the long run as a perspective of investment planning.
- Long run average cost is the "lower envelope" of all the short run average cost curves for different plant or firm sizes.
- We may observe increasing costs, decreasing costs, or constant costs as output increases, in the long run — or all three, depending on output.

Notes

LESSON 16:
COST ANALYSIS - II

You have studied some common concepts of cost in previous lesson, now I will take up some other concepts of cost and utility in this lesson.

1. Explicit Costs and Implicit Costs

Economists classify these types of costs: explicit (=accounting) costs, and implicit costs. Explicit costs are out of pocket, obvious kinds of costs, e.g. expenses on books, tuition, as, etc.. Implicit costs are not really expenses you incur, but involve income or values you are giving up by not doing something that you could have chosen to do. For instance, if you decide to go to school full time instead of working a $20,000 job, you are giving up earning $20,000. This is your implicit cost. Sample problem: Suppose you are running a small business and incur the following expenses: labor = $80,000; raw materials = 30,000; finance charges on a loan = $3,000. You are not paying explicit rent, because you own the building you are operating in. If you would rent it out, however, you could be earning $12,000. You also estimate your own time to be worth $25,000. What are your expenses?

Answer: **Explicit costs =** $80,000 + $30,000 + $3,000 = $113,000.
Implicit costs = $12,000 + $25,000 = $37,000.
Total economic costs = explicit + implicit costs = $150,000

2. Accounting Vs Economic Costs

Accountants have been primarily concerned with measuring costs for financial reporting purposes. As a result, they define and measure cost by the historical outlay of funds that takes place in the exchange or transformation of a resource.

Economists have been mainly concerned with measuring costs for decision-making purposes. The objective is to determine the present and future costs (or resources) associated with various alternative courses of action.

In calculating the cost to the firm of producing a given quantity of output, economists include some additional costs that are typically not reflected in financial reports.

- explicit cost are considered by both groups
- implicit costs are considered by economists:
- opportunity cost of time
- opportunity cost of capital

Economic Profit = Tot Rev - Exp Cost - Imp Cost

3. Accounting versus Economic Profits

To calculate accounting and economic profits we need to know the company's total revenue. Let's suppose it is $140,000. Then, accounting profits are: total revenue minus explicit costs or: $140,000 - $113,000 = $27,000. Economic profits are: total revenue minus total economic costs or; $140,000 - $150,000 = -$10,000 (i.e. a loss of $10,000).

The above firm reaps a positive accounting profit; but the negative economic profit indicates that from an economic point of view, the owner should discontinue the operation.

4. Total and Per Unit Costs

Seven cost functions which we will discuss are:

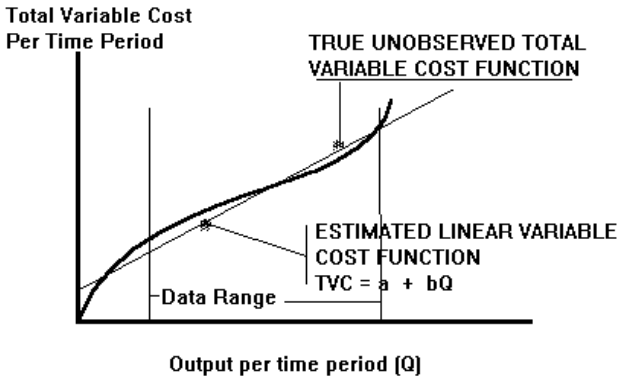
1. **Total Variable Cost (TVC)** – The cost of all variable resources

   Examples: Cost of labor, materials, office supplies

2. **Total Fixed Cost (TFC)** – The cost of all fixed inputs

   Examples: Cost of the building, large pieces of machinery, certain taxes

3. **Total Cost (TC)** – This the sum of TVC and TFC.
4. **Average Variable Cost (AVC)** – This is variable cost per product.
5. **Average Fixed Cost (AFC)** – This is fixed cost per product.
6. **Average Total Cost (ATC)** – This is total cost per product.
7. **Marginal Cost (MC)** – This is the cost of producing an additional unit of the product



5. Cost Calculations

Using the above abbreviations and Q for the quantity of output:
ATC = TC/Q
AFC = TFC/Q
AVC = TVC/Q
MC = change in TC/change in Q

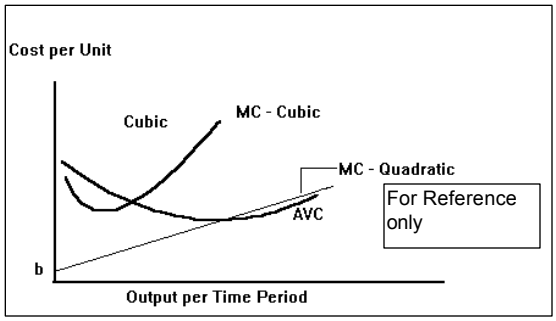
Example: Let's suppose you are making 50 bottles of wine each week. You know that your fixed costs add up to $300, and your variable costs amount to $900. You also know that if you were to make an extra 5 bottles, your total cost would rise by $60. What is your total cost; average total cost; average total cost; average variable cost; average fixed cost; and marginal cost?

Answer: Total cost = $300 + $900 = $1200
ATC = $1200/50 = $24
AVC = $900/50 = $18
AFC = $300/50 = $6

MC = $60/5 = $12

A table with cost data might show the following:

Q	TC	TFC	TVC	ATC	AFC	AVC	MC
45	1150	300	850	25.55	6.66	18.88	
50	1200	300	900	24.00	6.00	18.00	10
55	1260	300	960	22.91	5.45	17.45	12
60	1380	300	1080	23.00	5.00	18.00	24



- **Ore Mining Example: Cost Data**

Q	L	VC	FC	TC	AFC	AVC	ATC	MC
0	0	0	150	150	150	0	150	
6	1	50	150	200	25.00	8.33	33.33	8.33
16	2	100	150	250	9.38	6.25	15.63	5.00
29	3	150	150	300	5.17	5.17	10.34	3.85
44	4	200	150	350	3.41	4.55	7.95	3.33
55	5	250	150	400	2.73	4.55	7.27	4.55
60	6	300	150	450	2.50	5.00	7.50	10.00
62	7	350	150	500	2.42	5.65	8.06	25.00
62	8	400	150	550	2.42	6.45	8.87	ERR
61	9	450	150	600	2.46	7.38	9.84	-50.00
59	10	500	150	650	2.54	8.47	11.02	-25.00

a. rent of the ore-mining equipment = $0.20 per horsepower (750 x 0.20 = $150)

b. cost of each worker employed is $50 per period.

1. TC = FC + VC 2. AFC = eq FCQ

3. AVC = $\frac{VC}{Q}$ 4. ATC = eq TCQ or AFC + AVC

5. MC = eq TVCQ , note the following

TVC = a + bQ, where

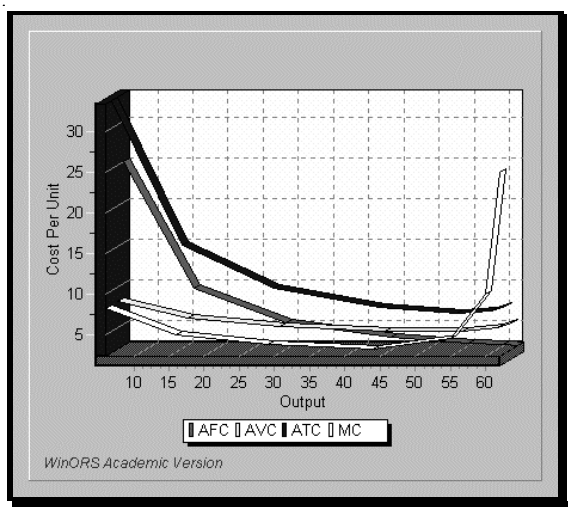
b = eq TVCQ = MC

6. Cost Elasticity e = eq %TC%Q = eq TCQ . eq QTC

e < 1 increasing returns to scale

ε = 1 constant

ε > 1 decreasing



6. Cubic Form (with Quadratic MC)

Cubic — Assumes that both marginal cost and average variable cost functions have U-shapes.

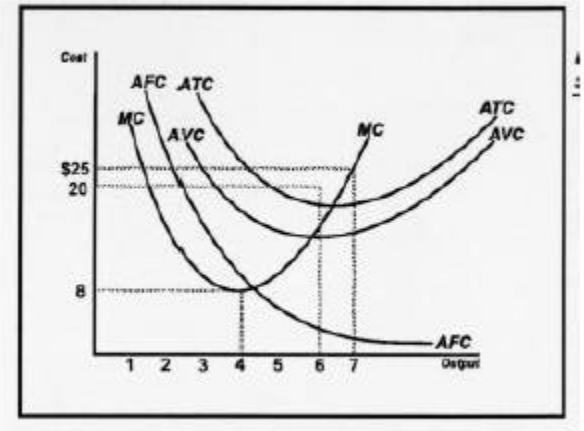


Fig. 2

The curves above show typical shapes of a firm's total cost, total variable cost and total fixed cost curves. Total fixed cost is constant at $50 for all levels of production. Total cost and total variable cost increase with higher levels of output. Note that total fixed cost and total variable cost always add to total cost. The data used in these graphs (figures 5.5a and 5.5b) is taken from the table in

Relationship of TFC, TVC, and TC

TFC is constant and unaffected by output level

TVC is always increasing; increasing at a decreasing rate first and then at an increasing rate

TC has same shape as TVC; but higher by the vertical distance equal to TFC

Note: These characteristics based on "typical" production function; shapes of cost curves depends on characteristics

of the underlying production function

Average & Marginal cost curves

AFC – always declining but at a decreasing rate

ATC and AVC – U-shaped; declining at first, reaching a minimum, and then increasing at higher output levels

Vertical diff. bet. ATC and AVC is the AFC (changes with output level)

ATC and AVC have minimum points at different output levels

MC – generally increasing; with a "typical" production function, MC decreases over a short range before starting to

increase

MC hits AVC and ATC at its minimum points

as long as marginal is below average, average is decreasing (and vice-versa)

8. Short-Run Cost Functions

In addition to measuring the costs of producing a given quantity of output, economists are also concerned with determining the behavior of costs as output is varied over a range of possible values.

The behavior of costs is expressed in terms of a cost function.

9. Total Cost Function

Total Cost = sum of all costs; FC + TVC

10. Long-Run Cost Functions

The long-run cost function is obtained directly from the production function by finding the expansion path. Remember, the expansion path for a production process consists of the combinations of inputs X and Y for each level of Q that satisfy the optimality criterion:

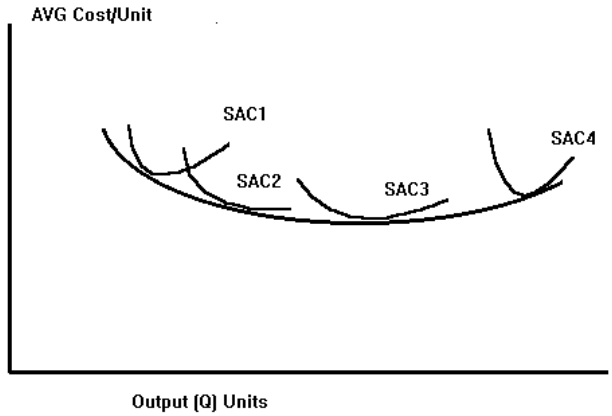
$$\frac{MP_x}{C_x} = \frac{MP_y}{C_y}$$

Over the long-run planning horizon, the firm can choose the combination of inputs that minimizes the cost of producing a desired level of output. Using the existing production methods and technology, the firm can choose:

- the plant size,
- types and sizes of equipment,
- labor skills, and
- raw materials

that, when combined, yield the lowest cost of producing the desired amount of output.

The long-run average cost function (LAC) consists of the lower boundary (envelope) of all the short-run cost curves.

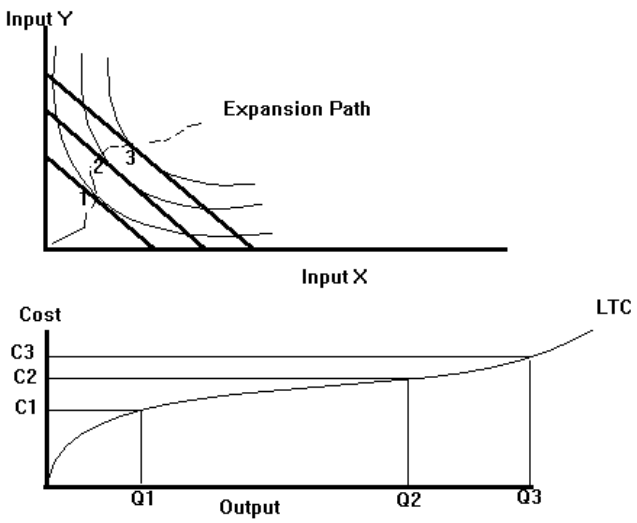


The relationship between LTC, LMC, and LAC is as follows — The long-run cost function is obtained directly from the production function by first finding the expansion path for the given production process.

The expansion path for a production process consists of the combinations of inputs X and Y for each level of output Q that satisfy the optimality criterion developed earlier.

Recall we derived the condition that the MRTS between two inputs must be equal to the ratio of the unit costs of the two inputs for a given input combination to be an optimal solution to either the output-maximization or cost-minimization problem.

$$MRTS = \frac{C_x}{C_y}$$



Graphically, the optimal input combination occurred at the point where the production isoquant was tangent to the isocost line.

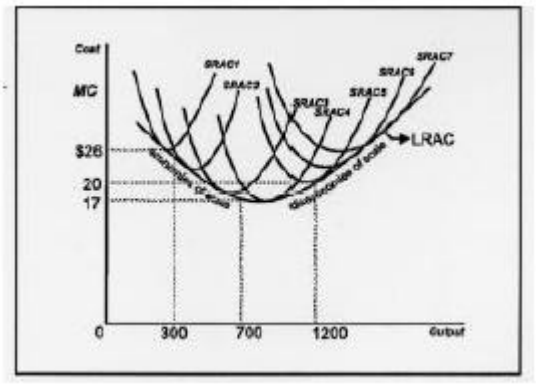
NOTE:

$$LAC = \frac{LTC}{Q} \text{ and } LMC = \frac{\Delta LTC}{\Delta Q}$$

11. The Long Run Average Cost Curve

The long-run average cost curve is derived from a number of short-run average cost curves. For each fixed plant size (short run), you look at the lowest costs for that size plant. These bottom portions of the different short run cost curves make up the long run average cost curve.

Fig. 5.6



A firm's long run average cost curve is the "envelope" of many short run average cost curves. All inputs are variable and the firm has the choice of building or changing to a variety of plant or facility sizes. A small operation (SRAC1) which wants to produce 300

units will have average costs of $26. A larger one, which produces 700 units, can produce each product for $17 (economies of scale). When the firm gets too large (SRAC6), average costs rise to $20 (diseconomies of scale).

12. Increasing, Decreasing, and Constant Returns to Scale

Note that increasing returns to scale is closely associated with the concept economies of scale (the downward sloping part of the long run average total cost curve.) Decreasing returns to scale relates to diseconomies of scale (the upward part of the curve).

Increasing returns to scale occurs when a firm increases its inputs, and a more than proportionate increase in production results. For example, one year a firm employs 200 workers and 50 machines and produces 1000 products. A year later it increases the number of workers to 40 0and the machines to 100 (inputs doubled) and the output rises to a level of 2500 (more than doubled).

Increasing returns to scale is often accompanied by decreasing long run average costs (economies of scale). A firm which gets bigger may experience this because of increased specialization, more efficient use of large pieces of machinery (for example, use of assembly lines), volume discounts, etc.

Decreasing returns to scale happens when the firm's output rises by less than the percentage increases in inputs. In the last example, had the firm's output risen to 1500, we would experience decreasing returns to scale.

Decreasing returns to scale can be associated with rising long run average costs (diseconomies of scale). An organization may become too big, thus creating too many layers of management, too many departments, and too much red tape. This lead to a lack of communications, inefficiency, and delays in decision making.

Constant returns to scale occurs when the firm's output rises proportionate to the increase in inputs. What would output have to be for this to take place?

13. Marginal Utility

Marginal utility is the additional satisfaction one gets from consuming one more item of a good or service. Satisfaction is measured in utils (use your imagination). Let's say that you are about to eat a pizza consisting of 6 slices. The first piece might give you 140 utils of satisfaction (you were starving!). The second slice, your hunger somewhat satisfied after the first, yields you only 60 utils. The third's down to 20 utils and a possible fourth, if forced upon you, could produce negative additional utils (your total would drop.)

14. The Law of Diminishing Marginal Utility

Referring back to the example in the previous objective, you can see that the marginal utility declines as the person consumes more slices. This is typical of almost all (beer and other substances known to be harmful to body and soul may be sole exceptions…) consumption: the more you have of something the less the additional unit is worth to you.

This phenomenon explains why if you are shopping in a supermarket you only buy a limited quantity of goods. As the marginal utility of, for example, the fifth orange declines, you may decide that this orange is not worth your additional expense.

15. Application of Cost Concepts

Profit-maximizing rules (Short-run):

If price or MR > min. ATC (or TR>TC), produce output where MR =MC

If min AVC < price or MR < min. ATC (or TVC < TR < TC), produce where MR = MC;

But will incur a loss between zero and TFC (loss minimized)

If price or MR < min ATC, do not produce and loss will equal TFC

Profit-maximizing rules (Long-run):

If price or MR > ATC, produce where MR=MC

If price or MR < ATC, stop production and sell fixed assets

16. Isoquants and the Producer's equilibrium

When producing a good or service, how do suppliers determine the quantity of factors to hire? Below, we work through an example where a representative producer answers this question.

Let's begin by making some assumptions. First, we shall assume that our producer chooses varying amounts of two factors, capital (K) and labor (L). Each factor was a price that does not vary with output. That is, the price of each unit of labor (w) and the price of each unit of capital (r) are assumed constant. We'll further assume that w = $10 and r = $50. We can use this information to determine the producer's total cost. We call the total cost equation an isocost line (it's similar to a budget constraint).

The producer's isocost line is:

$10L + 50K = TC$ (1)

The producer's production function is assumed to take the following form:

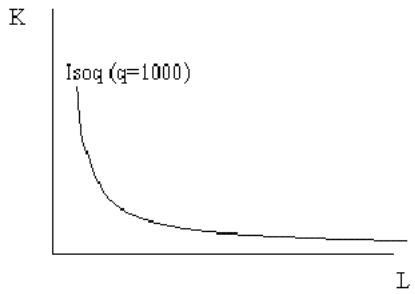
$q = (KL)^{0.5}$ (2)

Our producer's first step is to decide how much output to produce. Suppose that quantity is 1000 units of output. In order to produce those 1000 units of output, our producer must get a combination of L and K that makes (2) equal to 1000. Implicitly, this means that we must find a particular isoquant.

Set (2) equal to 1000 units of output, and solve for K. Doing so, we get the following equation for a specific isoquant (one of many possible isoquants):

$K = 1{,}000{,}000/L$ (2a)

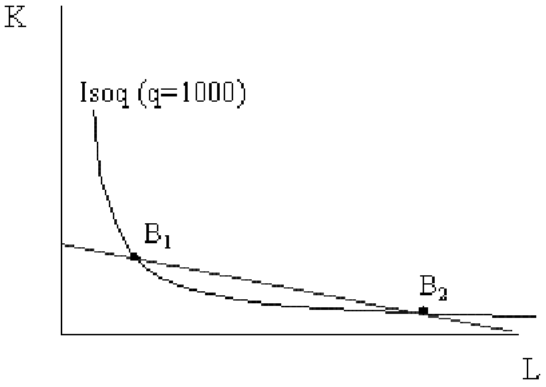
For any given value of L, (2a) gives us a corresponding value for K. Graphing these values, with K on the vertical axis and L on the horizontal axis, we obtain the blue line on the graph below. Each point on this curve is represented as a combination of K and L that yields an output level of 1000 units. Therefore, as we move along this isoquant output is constant (much like the fact that utility is constant as we move along an indifference curve).



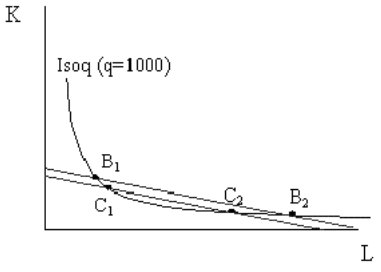
How much K and L should the producer hire? The answer is that our choice must exist somewhere on this isoquant. If each possible choice must lie on this isoquant, then our basis for choosing one "best" combination should be to choose the least cost combination. Let's experiment with some different possibilities, by plugging values for L and K into the isocost equation above. Each combination should yield output of 1000 units. Several combinations, and their specific total cost are given as follows:

L	K	TC
1,000	1,000	60,000
5,000	200	60,000
10,000	100	105,000
100	10,000	501,000

Now, the firm's goal is to produce 1000 units at the lowest possible TC. The lowest total cost on the table is $60,000, so we can start there. There are two choices which yield this total cost. They are represented below as B_1 and B_2.



Is there a lower cost combination of L and K available? Yes, we can produce at a total cost of $52,500 by employing either 250 units of K and 4000 units of L, or 800 units of K and 1250 units of L. This appears on the graph below.



The second isocost, where TC = $52,500, rests below the former isocost, where TC = $60,000. If we continue to find

lower and lower levels of total cost that provide us with 1000 units of output, then we will clearly reach lower and lower points on the isoquant. Eventually, we can find a level of total cost that involves a tangency between the isocost and isoquant. This is pictured below at point A.

In addition to being the lowest cost combination of L and K that produces 1000 units of output, pt. A involves a tangency point between the isoquant and isocost. That is, the slopes of these curves at pt. A are equal. The slope of the isocost is w/r, or –1/5. The slope of the isoquant is the ratio of the marginal products, MP_L/MP_K, which is given as the marginal rate of technical substitution (MRTS). Using calculus, it is possible to derive the MRTS as –K/L. Point A satisfies the condition that K/L = 1/5.

We can solve for K* and L* at pt. A, using (1), (2a) and the fact that, at pt. A, K/L = 1/5. First, substitute (2a) into (1) and the equation K/L = 1/5. We're left with:

10L + 50(1,000,000/L) = TC
and

(1,000,000/L)/L = 1/5

Solve the second equation for L, substitute that result into the first equation to get the lowest value for TC (TC*).

TC* = $44,721.36

Once you have TC*, you can substitute this value into the isocost equation above (10L + 50,000,000/L = TC) and then solve for L* (rounded to the nearest whole number).

L* = 2,236

Going back to (1), we can substitute in L* and TC*, to get K*.

K* = 447

18. Economies of Scale

Declining long-run average costs over the lower part of the possible outputs are usually attributed to economies of scale. Economies of scale occur over the range of the long-run cost function which corresponds to increasing returns to scale of the production function. Where do economies of scale come from?

1. Plant Economies

- specialization in the use of labor and capital
- indivisible nature of many types of capital equipment
- purchase price of different sizes of equipment

2. Firm Economies

- materials procurement / quantity discounts
- economies in raising funds (capital procurement)
- sales promotion
- technological innovation
- management

19. Glossary of terms

Cost: The sacrifice incurred whenever an exchange or transformation of resources takes place.

Sunk Cost: A cost incurred regardless of the alternative action chosen in a decision-making problem.

Cost Function: A mathematical model, schedule, or graph that shows the cost (such as total, average, or marginal cost) of producing various quantities of output.

Opportunity Cost: The value of a resource in its next best alternative use. Opportunity cost represents the return or compensation hat must be foregone as the result of the decision to employ the resource in a given economic activity.

Marginal Cost: The incremental increase in total cost that results from a one-unit increase in output.

Cost Elasticity: A measure that indicates the percentage change in total costs associated with a 1-percent change in output.

Economies of Scale : Declining long-run average costs as the level of output for the firm (or production plant) is increased. The decline in costs is generally attributed to production or marketing advantages.

Diseconomies of Scale : Rising long-run average costs as the level of output is increased.

Minimum Efficient Scale : The output level at which long-run average costs are first minimized.

Capacity. The output level at which short-run average costs are minimized.

Operating Leverage: the use of assets having fixed costs (e.g. depreciation) in an effort to increase expected returns.

Notes

LESSON 17:
ECONOMIES OF SCALE

Economies of scale can be of two kinds: internal economies and external economies. *Internal economies of scale* are those which arise from the firm increasing its plant size. On the other hand, *external economies* arise outside the firm-from improvement (or, deterioration) in the environment in which the firm operates. The economies external to the firm may be realised from actions of other firms in the same or in another industry. While the internal economies of scale relate only to the long-run and determine the *shape* of the long-run cost curve, the external economies affect the *position* of the long-run cost curves.

17.1 Internal Economies

Internal economies are given in a summary form in Fig. 17.1, where these are categorised into real and pecuniary economies. *Real economies* arise when the quantity of inputs used for a given level of output decreases. While *pecuniary economics* are those savings in expenses which accrue to the firm in the nature of relatively lower prices paid for inputs and lower costs of distribution. These savings arise due to bulk buying and selling by the growing firm.

17.1.1 Real Economies of Scale

Real economies are of 4 kinds:

a. Production economies

b. Marketing economies

c. Managerial economies, and

d. Transport and storage economies.

a. Production economies. Production economies arise from *(a)* labour, *(b)* fixed capital, and *(c)* inventory requirements of the firm. These are:

I. Labour economies. Labour economies arise because of the following factors :

i. ***Division of Labour Economies*** . Larger output allows division of labour which reduces cost by increasing specialisation, by saving time (otherwise lost in passing from one operation to another), and providing good conditions for inventions of a great number of machines.

ii. ***Cumulative volume economies***. The technical personnel engaged in production tend to acquire significant experience from large-scale production. This 'cumulative volume' experience helps in higher productivity and, therefore, reduced costs

II. Technical economies. These are associated with fixed capital, which includes machinery and equipment. Such economies arise because of the following:

i. ***Specialised equipment*** . The production methods become more mechanized as the output scale increases. This would imply more specialised capital equipment and lower variable costs.

ii. ***Indivisibility*** The machinery and equipment generally have the property of indivisibility, which means that equipment is available only in minimum sizes or in definite ranges of size. When output is increased from zero to the maximum capacity level of the machine, the same machine and equipment are used. As a result the cost of machine is shared between more and more units of output. In short, as the output is increased, the machinery and equipment comes to be utilised more intensively and consequently the cost of production per unit declines.

iii. ***Integration of processes*** . The large size firms enjoy economies of *large machines.* Integration of processes occurs where one large automatic transfer ornumerically controlled machine can carry out a series of consecutive processes, saving labour cost and time required to set up the work on each of a series of successive specialised machines.

iv. ***Economies of increased dimensions*** , for many types of equipment both initial and running costs increase less rapidly than capacity *(e.g.,* tanks, blast furnaces and other static and mobile containers). These result in economies of increased dimensions. Any container whose external dimensions are doubled has its volume increased eight times, but the area of its surface walls would have increased only four times. This reduces material costs and, where appropriate, heat loss and surface, air and water resistance per unit.

v. ***Economies in set-up costs*** . The larger the scale of output, the more a multipurpose machinery is left to one set-up and, therefore, set-up costs of general purpose machines reduce.

vi. ***Economies of overhead costs*** . Obviously, the larger the scale of output, the lower the unit costs of initial fixed expenses which are need for a new business or a new product.

III. Inventory economies. Role of inventories is to meet the random changes in the input and output sides ofthe operations of the firm. It has been found that the input as well as output inventories increase at a rate lower than that of increase in output. These economies arise due to the phenomenon of *massed resources.*

b Marketing Economies. These economies arise because

I. The advertising expenditure is generally found to have increased less than proportionately with scale. Consequently, larger the output, smaller the advertising cost per unit. Similar situation prevails in case of other types of selling activities.

II. The development and adoption of new models and designs involve considerable expenses in R&D. The larger the output, more thinly this R&D expenditure spreads over output.

c. Managerial Economies. Managerial economies arise because:

I. Larger the firm, greater are the opportunities for the division of managerial tasks. The division of managerial tasks helps managers to specialise in their own areas of responsibility, thus leading to greater efficiency.

II. Teamwork experience. By working in a team, the managers of large firms tend to acquire a more comprehensive outlook as well as a quicker and better decision-making ability.

III. In a large firm, with decentralization in decision-making, the delays in the flow of information is reduced, thereby increasing the efficiency of management.

IV. Modern managerial and organisational techniques. Large firms provide opportunities for the introduction of modern managerial techniques and organisational restructuring. These help the management to increase efficiency.

d. Transport and Storage Economies. Storage costs obviously fall with the increase in the size of output, as it provides the *economies of increased dimensions* (discussed already). The transportation costs, on the other hand, involve an L-shaped average cost curve-transport unit costs falling up to the point of the full capacity and remaining constant thereafter.

17.1.2 Pecuniary Economies of Scale

These economies include the discounts that a firm can obtain due to its large size. These discounts may be in the nature of:

i. Lower raw material price due to bulk buying.

ii. Lower cost of capital, as banks usually place greater faith in the large firms and, therefore, charge lower rate of interest.

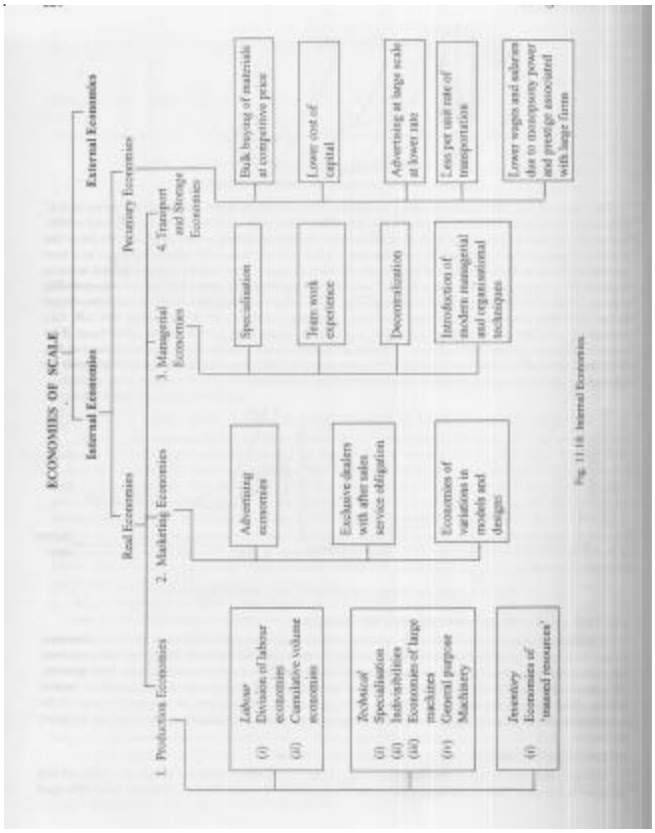
iii. Offers of lower rate's for advertising to large firms because of their large-scale advertising.

iv. Lower transportation rates due to bulk transportation.

v. In case the large firm is able to attain a size to gain monopsonistic power or is able to create an image of prestige to be associated with the firm it may be in a position to save on labour costs by paying lower wages and salaries.

17.2 External Economies

Like internal economies, external economies also help in cutting down production costs. With the expansion of an industry, certain specialised firms also come up for working up the by-products and waste materials. Similarly, with the expansion of the industry, certain specialised units may come up for supplying raw material, tools, etc., to the firms in the industry. Moreover, they can combine together to undertake research, etc., whose benefit will accrue to all the firms in the industry. Thus, a firm benefits from expansion of the industry as a whole. These benefits are external to the firm, in the sense that these arise not because of any effort on the part of the firm but accrue to it due to expansion of industry as a whole. In this sense these economies are external to the firm. All these external economies help in reducing production costs.

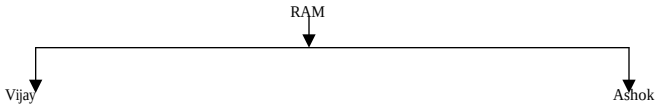


17.3 Diseconomies of Scale

When a firm continues to expand its size, a stage comes when diminishing returns to scale set in. As a firm expands beyond a level, it encounters growing diseconomies. These diseconomies more than cancel out the economies of large-scale production and cause average costs of production to start rising. Let us discuss in detail reasons for such a phenomenon.

Technical factors are unlikely to produce diseconomies of scale. If inefficiencies arise as a result of over large plant size then they can be avoided by replicating units of plant of a smaller size. In fact, *technical factors are more likely to 'limit' the sources of scale economies than to act as a source of diseconomies.*

When diseconomies of scale arise they are more likely to be associated with the *human and behavioral problems* of managing a large enterprise. Let us understand this with the he1p of a highly simplified organisation chart or a managerial hierarchy given in figure below. Both Vijay and Ashok (who may, for example, be the divisional managers) are responsible to Ram. If Vijay wants to communicate with Ashok he must follow the formal chain of command and pass his message through Ram, whose function is coordination. This will be a time-consuming process involving red-tapism but, given a large organisation with a large number of managers, such indirect coordination may be the only practical method of communicating while avoiding disorganization and chaos.

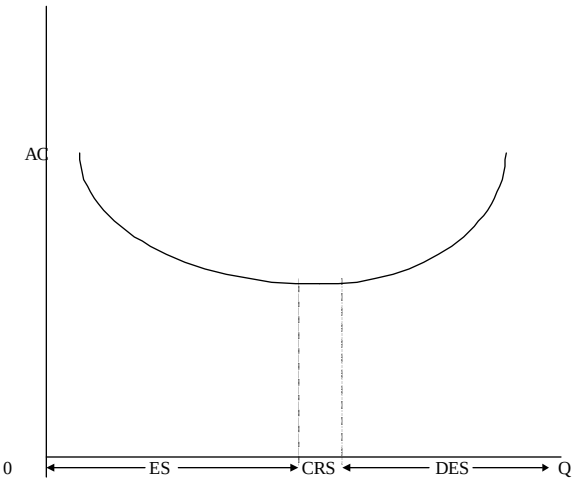


Now, if the firm grows further, then another layer of management must be inserted between Ram, Vijay and Ashok. This increases the chain of command and Vijay, in order to communicate with Ram, must pass his message through an intermediary. This increases the costs of communication and also introduces the problems of possible *message distortion* and *misinterpretation* with corresponding implications for organisational efficiency. These arguments can be well-explained with the help of Williamson's concept of *'control loss '*. The decisions taken by a top executive must be based on information passed across a series of hierarchical levels. In turn, the instructions based on this information must be transmitted down through these successive stages. This transmission results in a *serial reproduction loss* or distortion, of both the information and instructions. This may occur even when the individuals forming the hierarchy have identical objectives. Increases in the scale of the hierarchy result in reduction of the *quality* of the information reaching the top coordinator and of the instructions passed down by him to lower-level personnel. Moreover, since the capacity of the top administrator for assimilating information and issuing instruction is limited he can, after a point, only cope with an expansion of the hierarchy by sacrificing some of the details provided before the expansion. Thus, the *quantity of information received and transmitted per unit of output will be less after expansion than before it.* This is known as 'control loss'. As a result it can be argued that operating units will not adhere as closely to the top administrator's objectives of cost minimisation as they did before the expansion.

Secondly, there is the problem of morale and motivation of both management and labour force. It is often argued that due to lack of personal touch the *spirit* in a large firm is less than that in a small firm. The labour force is more closely identified with small firm and this results in improved productivity and greater overall loyalty to the organisation. Moreover, since management of a large firm may feel more secure they may become sluggish and develop lack of enterprise. This sluggishness is- absent in managers of small firms who see the generally present threat of being put out of business

In short, we may say that decreasing returns to scale will become operative when management becomes a problem. This problem is more serious in agriculture than in industry: as their operations expand the law of decreasing returns becomes operative earlier in agriculture than in industry

Economies/Diseconomies of Scale



ES: range of output encompassing economies of scale (decreasing unit costs; [includes increasing returns to scale])

CRS: range of output encompassing constant returns to scale (constant unit costs)

DES: range of output encompassing diseconomies of scale(increasing unit costs; [includes decreasing returns to scale])

17.4 The Concept Of Learning Curve

The learning curve analysis is based on the assumption that workers improve with practice, so the per unit cost of additional output declines. The reduction in cost due to this learning process is known as the *learning curve effect,* where learning curve graphically depicts the relationship- between labour cost and additional units of output.

Learning curve is measured in terms of percentage fall in marginal labour cost when output doubles. Table 17.6 presents data for a "80 per cent" learning curve. Each time output doubles, the cost of producing this additional output decreases to 80 per cent of the previous level. This implies 10 per cent reduction in addition to cost.

The pattern of reduction in factor cost is based on the following formula:

$$Lx=k.X"$$

where

x = production unit

Lx= units of labour hours for producing xth unit.

k = cost to produce first unit.

n = log slope/log 2, where slope equals the rate at which cost of producing additional unit declines,

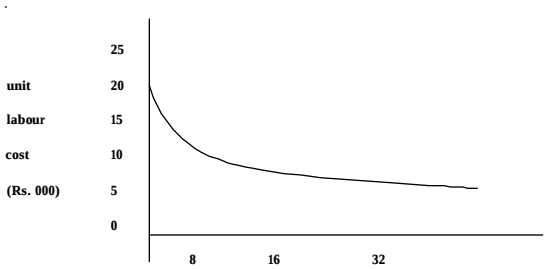
In a similar way we can find cumulative labour hours.

Table 17.1 : Measurement of Learning Curve

Output unit	*Labour hours*	*Cumulative*	*Cumulative*	*Cost of*	*Cumulative*
		labour hours	*average*	*labour hours*	*average labour*
			labour hours	*= col. 2 x Rs. 10*	*cost*
			= col. 3 -+- col. 1		*= co/. 4 x Rs. 10*
(1)	(2)	(3)	(4)	(5)	(6)
1	2,0000	2,000	2,000.0	20,000	20,0000
2	1,6000	3,600	1,800.0	16,000	18,0000
4	1,2800	6,284	1,558.5	12,800	15,5850
8	1,0240	10,692	1,336.5	10,240	13,3651
16	820	17,840	1,115.0	8,200	11,1500
32	650	29,357	917.4	6,500	9,1774
64	530	47,849	747.6	5,300	7,4776

With the help of cols. 1 and 5 in Table 17.1, we can plot the learning curve (Fig. 17.2).

The learning curve is expressed in terms of marginal labour cost. The average cost as expressed by cumulative average labour cost is also seen declining (col. 6, Table 17.6), showing the impact of improving efficiency of labour with practice.



17.5 Economies of Scope

While discussing the law of variable proportions and economies of scale, it was implicitly assumed that the firm produces only one product. In modem-day business we frequently encounter firms (like Hindustan Lever, P&G, Nestle, etc.) which produce more than one products. It has been observed that a multi-product firm often experiences *economies* or *diseconomies of scope.* If a single firm producing multiple products can together produce them cheaper compared to a situation where each product is produced by a separate firm, we say that the economies of scope exist in such a case. For example, if firm *A* produces 100 units of *X* and 500 units of Y per month at the total cost of Rs.1,00,000. While, on the other hand, suppose X and Y were produced by two separate firms: the cost of producing *100X* by firm *B* being Rs.25,000 and the cost of producing *500r* by firm C being Rs.90,000. Firm *A* then experiences economies of scope because its cost of producing both goods *X* and *r* together is Rs.1 ,00,000, which is less than the cost of producing them separately (Rs.25,OOO + Rs.90,000 = Rs. l, 15,000). This difference in the cost of producing goods jointly by a firm and producing them separately by separate firms can be used to measure the degree of *economies of scope.*

Degree of economies of scope =

$$\frac{TC(Q,) + TC(Q2) - Tc(QI + Q2)}{Tc(QI + Q2)}$$

where, *TC* (QI) = total cost of producing QI units of good 1 only;

TC (Q2) = total cost of producing *Q2* units of good 2 only;

TC (QI +*Q2)* = total cost of producing goods I and 2 jointly; producing QI units of goods 1 and *Q2* units of goods 2.

Thus, in the case of Firm *A* in example above, the degree of economies of scope equals

$$\frac{\text{Rs. 25, 000 + Rs. 90, 000 - Rs.l, 00, 000 - 0.15}}{\text{Rs.l, 00, 000}}$$

If the degree of economies of scope is positive it implies that economies of scope exist. When this measure becomes negative, it means that producing goods separately is cheaper than producing them together.

The main reasons for the existence of economies of scope are:

i. In case a firm produces several products, it is very likely that many of them use common production facilities and inputs. For example, a firm producing electrical goods may make use of the same testing or assembly line facilities for geysers, food-warmers, fans, irons, etc.

ii. It happens many a time that the production of one good results in by-products that can be sold by the producer, thereby gaining a cost advantage in the production of the main product. For example, a sugar mill gets molasses as a by-product, which it can sell directly in the market or can use in the production of liquor in its own distillery.

17.6 Summary

Economies of Scale

Are the factors that cause average cost to be lower in large-scale operations than in small scale ones

- **Specialisation** with a larger workforce it is possible to divide up the work and recruit and train individuals who exactly match the requirements. They can then become specialists.
- **Technical** firms benefit from being able to use machinery. Some items are only worthwhile being purchased and used when the fixed costs can be spread over a larger output.
- **Purchasing** As firms grow they can benefit from being able to buy in bulk.

Diseconomies Of Scale

Are the factors that cause costs per unit to increase as the scale of output increases.

- **Communication:** In larger organisations people have to be hired to pass on communication, extra bits of paper are used and the message has more opportunity to get distorted as it passes through more layers.
- **Co-ordination:** Problems occur as it is much more difficult to coordinate so many people. Empowerment might cause problems. The extra cost of meetings.
- **Motivation:** Being part of a larger organisation might cause a lack of motivation amongst some employees.

Are Bigger Firms Therefore Better?

The answer to that question is obviously 'it depends', as always! Clearly it depends on the opposing effects of the diseconomies and economies of scale. If the diseconomies outweigh the economies then the answer is 'no'. Vice-versa. In some cases the small firm has advantages over the larger firm, especially in terms of service, closeness to the market and their ability to exploit niches that will just not be profitable for larger firms to pursue.

Notes

TUTORIAL 4

1. Multiple Choice Questions

1. An upward-sloping short-run marginal cost curve shows that:

- with a fixed amount of capital, as more and more labor is added, output increases at a decreasing rate, and thus each unit is more expensive than the previous unit.
- with a fixed amount of labor, as more and more capital is added, output increases at a decreasing rate, and thus each unit is more expensive than the previous unit.
- with a fixed amount of capital, as more and more labor is added, output increases, and thus marginal cost increases.
- with a fixed amount of capital, as more and more labor is added, output decreases, and thus marginal cost increases.
- with a fixed amount of labor, as more and more capital is added, output increases at an increasing rate, and thus each unit is more expensive than the previous unit.

2. Consider a firm that has just built a small plant, which cost $4,000. Each unit of output requires $2.00 worth of materials. Each worker costs $7.00 per hour. Fill in the following table, and use it to answer the next four questions.

Table 1

Number of Worker Hours	Total Variable Cost	Total Cost	Output	Fixed Cost	Marginal Cost	Average Variable Cost	Average Total Cost
0	0			—	—	—	
50	400						
100	900						
150	1300						
200	1600						
250	1800						
300	1900						
350	1950						

After what level of worker hours does diminishing returns set in?

2

- 50
- 100
- 150
- 200
- 250

3. Based on Table 1, at what output is average variable cost minimized?

3

- 400
- 900
- 1300
- 1600
- 1800

4. Based on Table 1, at what output is average total cost (ATC) minimized?

4

- 900
- 1600
- 1800
- 1900
- 1950

5. Refer to Table 1. What is the relationship between marginal cost (MC), average variable cost (AVC), and average total cost (ATC) when the firm is producing 1300 units of output?

5

- MC is increasing, AVC is increasing, and ATC is increasing.
- MC is increasing, AVC is increasing, and ATC is decreasing.
- MC is increasing, AVC is decreasing, and ATC is decreasing.
- MC is decreasing, AVC is increasing, and ATC is decreasing.
- MC is decreasing, AVC is decreasing, and ATC is decreasing.

6. If marginal cost is increasing but less than average total cost, what do we know about average total cost?

6

- Average total cost is increasing.
- Average total cost is constant.
- Average total cost is decreasing.
- Any of the above could be true.
- not enough information to tell

7. If the firm has economies of scale in the long run, then the long-run average cost curve is:

7

- upward-sloping.
- horizontal.
- downward-sloping.
- It depends on whether there are diminishing returns.
- It depends on the minimum efficient scale.

8. Which of the following is the best explanation for diseconomies of scale?

8

- coordination problems
- indivisibilities
- gains from specialization
- diminishing returns

- Both Answers 1 and 4 are correct.

9. Which of the following statements is true?

9

- If marginal cost is decreasing, average variable cost must be decreasing.
- If marginal cost is decreasing, average total cost must be increasing.
- If average total cost is decreasing, marginal cost must be decreasing.
- If marginal cost is greater than average variable cost, average variable cost must be increasing.
- All of the above statements are true.

10. Which of the following statements is FALSE?

10

- If a firm has diminishing returns in the short run, it will have diseconomies of scale in the long run.
- If there are production indivibilities, the long-run average cost curve will be downward-sloping.
- The minimum efficient scale occurs when long-run average costs reach their lowest point.
- One reason for diseconomies of scale is increasing input costs.
- All of the above are false statements.

11. Refer to Figure 1. Diminishing returns set in at point:

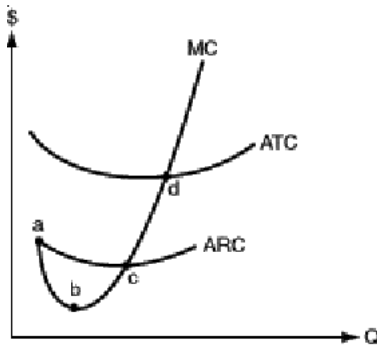


Figure 1

11

- a
- b
- c
- d
- Can't tell from the graph given.

12. Refer to Figure 1. At which point are there decreasing returns to scale?

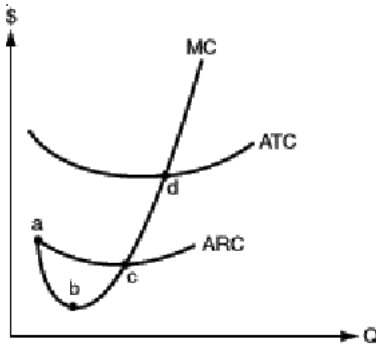


Figure 1

12

- a
- b
- c
- d
- none of the above

2 Production Consultant II: Carefully explain if the following statements are true or false: (Hint: you should draw a graph of the appropriate cost curves to help you answer.)

a. If average cost is decreasing, marginal cost must be decreasing.

b. If average cost is increasing, marginal cost must be increasing.

c. If there are diminishing returns, the short-run average cost curve must be positively sloped.

d. If there are diminishing returns, the marginal cost curve must be positively sloped.

Notes

TUTORIAL 5

1. Case Study

Technological Change and Government Subsidies

LEAD STORY-DATELINE: Business Week April 16, 2001

We tend to think of the microprocessor king Intel Corporation as an independent private company. Sometimes though, we wonder if anything is purely private or purely public, for that matter. After all, private airline companies land their planes at public airports. Private trucking companies use super-highways funded by the federal government. Why would it be different for Intel? In many ways, it is not-especially when it comes to technological change.

Since the first Sumerian clay tablets 5,500 years ago, the progress of humans is directly related to the ability to store, process, and retrieve information. In recent decades, that progress has been amazing. The power of chips has been doubling every 18 months, creating desktop super-computers, storehouses of digital data, and smarter cars and home appliances. Behind this progress in silicon technology are the steady advances in microlithography-the process used to "print" ever-smaller circuits on silicon wafers. Further expansion in the Information Age depends upon on a major leap in lithography because smaller is better.

Intel corporation leads a consortium of private companies, national laboratories, and academics in the development of extreme ultraviolet (EUV) radiation to reduce circuit lines to a minuscule 35 nanometers or even less. The microlithography machine is like a complex stenciling tool. It projects a circuit pattern through reducing lenses onto a high-sensitive coating on a silicon wafer. Wherever the light strikes, the coating hardens. Then the coating's soft regions can be etched away, leaving a maze of lines that further processing turns into millions of transistors and connections. Federally funded collaborations with the national labs helped solve many technical problems, and AT&T got $2 million from the Commerce Department, which it used to boost U.S. optics technology.

Then a crisis came. After gaining control of Congress in 1994, Republican lawmakers axed funding for joint research among national labs and private companies. Intel put together a consortium involving Sandia, Livermore, and Lawrence Berkeley national labs, along with chipmakers AMD, Infineon Technologies, Micron Technologies, Motorola, and equipment suppliers. Intel coughed up the lion's share of the $250 million budget, but won a guarantee that it would get the new lithography machines first, before other consortium members. Ironically, such feats may now be made even more difficult as the Bush Administration plans to withhold funding for one of the programs that jump-stared this project.

Thinking About The Future!

Japan has two giant lithography suppliers, Nikon Corporation and Canon Inc. To build a production-ready version of Sandia's experimental prototype, Intel in 1998 wanted to license the design to Japanese suppliers as well as to the (former) Silicon Valley Group Inc. (SVG), the main U.S. lithography company. Then, international trade concerns came to the fore. Washington politicians and Silicon Valley executives alike raised a fuss about handing the crown jewels to the Japanese. So the consortium brought in a Dutch company, ASM Lithography (ASML). As it turns out, SVG has now been acquired by ASML (though the deal had been delayed by national security concerns). There is a lot at stake. When these new chips are fabricated, they will make today's Pentiums seem like Model Ts. It could also deal a knockout blow to a rival electron-beam approach being developed by Nikon and IBM, although IBM has wisely now joined the EUV group as well.

Talking It Over and Thinking it Through!

1. In economics the subject of this article is the technology of production. What is a production function and what is happening to it in this new chip technology?
2. Does this idea of producing more with less conflict with the law of diminishing returns?
3. The prototype EUV machine, currently at Sandia National Laboratories is the result of 13 years of work and more than $250 million in research-and-development funding. Each new lithography machine will cost $40 million. These costs exceed what small corporations could fund. Does this kind of R&D expense have implications for corporate size and industrial concentration?
4. Are there long-term advantages to government funding of such expensive R&D through the national laboratories?

2. Multiple Choice Questions

1. Diseconomies of scale occur when

- a firm's average total cost increases with increased production.
- only one firm is producing the good that consumers wish to purchase.
- a firm pays a lower price for inputs as its level of production increases.
- a firm has a high level of initial sunk costs, such as manufacturing equipment.
- there are perfectly competitive markets.

2. Which of the following does NOT represent a firm's explicit costs?

- costs related to workers salaries
- the amount paid by the firm for employee health benefits
- the firm's total accounting costs

- the value of the CEO's time
- the electricity bill

3. The marginal cost curve intersects the short-run average cost curve at a level of output
- greater than the level of output at the minimum point of the short-run average cost curve.
- less than the level of output at the minimum point of the short-run average cost curve.
- equal to the level of output at the minimum point of the short-run average cost curve.
- greater or less than the level of output at the minimum point of the short-run average cost curve.
- where the marginal cost curve is unrelated to the short-run average cost curve.

4. In which of the following examples do firms NOT have to worry about "spreading the costs out?"
- an electric company with the high initial cost of installing electric lines
- a firm that incurs costs from indivisible inputs
- a seasonal fruit picking operation that often has to hire more workers during busy periods
- a small low-output potter who must purchase a new kiln before pots can be created
- an airline that just purchased a new plane

5. Which of the following is NOT an implicit cost?
- the impact on a firm's reputation of producing a poor product
- the opportunity cost of an entrepreneur's time
- the value of alternative inputs that could have been purchased
- the total amount a firm spends on advertising new products
- the opportunity cost of investment capital

6. Which of the following is NOT a reason why economies of scale may exist?
- it is cheaper for a firm to make only a small quantity of one item and charge a high price.
- firms are more productive due to increased input specialization as output increases.
- the cost of indivisable inputs can be spread over more units as output is increased.
- companies often face downward sloping long-run average cost curves.
- large start-up costs make higher output levels desirable.

7. At levels of output where the firm's short-run average cost curve is increasing,
- the marginal cost curve is above the short-run average cost curve.
- the marginal cost curve is below the short-run average cost curve.
- the marginal cost curve is equal to the short-run average cost curve.
- the marginal cost curve may be above or below the short-run average cost curve.
- none of the above are true.

8. Minimum Efficient Scale (MES) represents
- the least a firm can charge consumers and still make a profit.
- the greatest number of employees a firm can hire and still operate efficiently.
- the price at which all units supplied to the market will be purchased.
- the output at which the long-run average cost curve becomes horizontal.
- the output at which the long-run average cost curve becomes vertical.

9. A firm's variable costs
- directly reflect the price of the company's outputs.
- are dependent upon the level of fixed costs in the long-run.
- are always equal to the company's total average costs in the long-run.
- are determined by quantity of output produced by a firm.
- decrease as output increases.

10. A firm's average fixed costs
- are determined by dividing the total fixed cost by the amount produced.
- are always larger than variable costs in the short- and long-run.
- are the same no matter what quantity the firm produces.
- are equal to zero only when the level of production is also zero.
- always increase as output increases.

11. If the marginal product of labor is increasing,
- output is increasing at an increasing rate.
- output is increasing at a constant rate.
- output is increasing at a decreasing rate.
- output is decreasing.
- There is not enough information to answer the question.

12. If the firm is experiencing diminishing returns to labor (in the short run), total output is _____, and marginal cost is _____.
- increasing, increasing
- increasing, decreasing
- decreasing, increasing
- decreasing, decreasing
- increasing, constant

3. Long Answer Questions

13. 14. 15. 16. 17.

Suppose you are thinking of starting your own mountain bike company, "Dirt Lovers." You calculate that your initial costs for your plant and equipment will be $10,000. The wage is $15 per hour, and each bike requires $50 of materials. You determine that producing 5 bikes will require 15 hours of labor, 25 bikes

will require 100 hours of labor, and 50 bikes will require 250 hours of labor. For each level of production, 5, 25, and 50 bikes, calculate the following: Total Fixed Cost, Average Fixed Cost, Total Variable Cost, Average Variable Cost, Average Total Cost, and Marginal Cost.

2. In the production of a new airplane, workers often "learn by doing." With each plane they build, the workers determine a more efficient method for building the next plane. What does this imply about the long-run average cost curve and the economies of scale facing the aerospace industry? Explain.

Notes

THE MARKET STRUCTURE AND THE FACTORS MARKET AND COMPETITION

We have pointed out that the theory of supply and demand is a theory of price in **competitive** markets. In this chapter, we draw on the previous two chapters and show how the supply curve is determined by price competition and profit maximization.

We have now built up enough background to study what economists call "industrial organization" — that is, the study of industries, including their organization and **structure,** how they **conduct** business, how they respond to change and evolve, and how efficiently they **perform.**

Economists believe all these things are interrelated, so this is sometimes called a **structure/conduct/performance** approach.

After you complete this chapter, you should be able to:

- What is a market?
- What are the different type of market structures?
- What is competition?
- Conduct and performance of these markets
- Theories of the firm : maximizing & non-maximising

Market Structures

- Degree of competition in the industry
- High levels of competition – Perfect Competition
- Limited Competition – Monopoly
- Degrees of competition in between

Market Structures

Market Structure

- Determinants of market structure
  - Freedom of entry and exit
  - Nature of the product – homogenous (identical) differentiated?
  - Control over supply/output
  - Control over price
  - Barriers to entry

Market Structures

- Type of market structure influences how a firm behaves:
  - Pricing
  - Supply
  - Barriers to Entry
  - Efficiency
  - Competition

Market Structure

- Perfect Competition:
  - Free entry and exit to industry
  - Homogenous product – identical so no consumer preference
  - Large number of buyers and sellers – no individual seller can influence price
  - Sellers are price takers – have to accept the market price
  - Perfect information available to buyers and sellers

Market Structure

- **What happens in a competitive environment?**
  - New idea? – firm makes short term abnormal profit
  - Other firms enter the industry to take advantage of abnormal profit
  - Supply increases – price falls
  - Long run – normal profit made
  - Choice for consumer
  - Price sufficient for normal profit to be made but no more!

Market Structure

- Examples of perfect competition:
  - Financial markets – stock exchange, currency markets, bond markets?
  - Agriculture?
- To what extent?

Market Structure

- **Imperfect or Monopolistic Competition**
  - Many buyers and sellers
  - Products differentiated
  - Relatively free entry and exit
  - Each firm may have a tiny 'monopoly' because of the differentiation of their product
  - Firm has some control over price
  - **Examples** – restaurants, professions – solicitors etc, building firms– plasterers, plumbers etc

Market Structure

- **Advantages of Perfect Competition:**
- High degree of competition helps allocate resources to most efficient use
- Price = marginal costs
- Normal profit made in the long run
- Firms operate at maximum efficiency
- Consumers benefit

Market Structure

- **Oligopoly – Competition amongst the few**
  - Industry dominated by small number of large firms
  - Many firms may make up the industry
  - High barriers to entry
  - Products could be highly differentiated– branding or homogenous
  - Non – price competition
  - Price stability within the market - kinked demand curve?
  - Potential for collusion?
  - Abnormal profits
  - High degree of interdependence between firms

Market Structure

- **Examples of oligopolistic structures:**
  - Supermarkets
  - Banking industry
  - Chemicals
  - Oil
  - Medicinal drugs
  - Broadcasting

Market Structure

- **Monopoly:**
- Pure monopoly – industry is the firm!
- Actual monopoly – where firm has >25% market share
- Natural Monopoly – high fixed costs – gas, electricity, water, telecommunications, rail

Market Structure

- Measuring Oligopoly:
- **Concentration ratio** – the proportion of market share accounted for by top X number of firms:
  - E.g. 5 firm concentration ratio of 80% - means top 5 five firms account for 80% of market share
  - 3 firm CR of 72% - top 3 firms account for 72% of market share

Market Structure

- **Monopoly:**
  - High barriers to entry
  - Firm controls price OR output/supply
  - Abnormal profits in long run
  - Possibility of price discrimination
  - Consumer choice limited
  - Prices in excess of MC

Market Structure

- **Duopoly:**
- Industry dominated by two large firms
- Possibility of price leader emerging – rival will follow price leaders pricing decisions
- High barriers to entry
- Abnormal profits likely

Market Structure

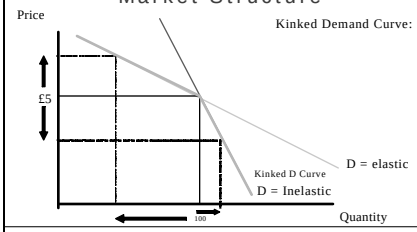
- **Advantages and disadvantages of monopoly:**
- **Advantages:**
  - May be appropriate if natural monopoly
  - Encourages R&D
  - Encourages Innovation
  - Development of some products not likely without some guarantee of monopoly in production.
  - Economies of Scale can be gained – consumer may benefit

Market Structure

- **Disadvantages:**
  - Exploitation of consumer – higher prices,
  - Potential for supply to be limited- less choice
  - Potential for inefficiency –

    X-inefficiency – complacency over controls on costs

Market Structure



Notes

LESSON 18:
PERFECT COMPETITION

As usual, the first step will be some terminology. Economists in general recognize four major types of market **structures** (plus a larger number of subtypes):

- Perfect Competition
- Monopoly
- Oligopoly
- Monopolistic competition

For now, I will just define the first of the four, and then use concepts related to "Perfect Competition" to define the other three.

18.1 What is a Market?

These are forms of markets. We might pause for a moment to think about just what a market is. **A market consists of all the (potential) buyers and sellers of a particular good or service.**

In order for a market to exist, though, these potential buyers and sellers must have some way to communicate offers to buy and sell with one another. One possibility is for them to come together and yell at one another (as at the New York Stock Exchange). In traditional societies, craftsmen in a particular trade may all be located on the same street, so that customers know where to go to buy.

Of course, many modern markets make use of a wide range of electronic communication methods, as do NASDAQ and the international currency markets.

Thus, it is natural to identify a market with the place where traders come together (as in the case of a stock market) or the means by which they communicate. But the market consists of the people willing to buy and sell.

18.2 Perfect Competition

What many economists call "P Competition" is an idealized structure of an industry in which price competition is dominant — in fact the only form of competition possible. The terminology "Perfect Competition" is quite common but not quite universal. The term "Pure Competition" is also sometimes used. I will use the term "P-Competition," where the P can stand for perfect, pure, or price competition — whichever you like.

A P-Competitive structure is defined by four characteristics. For an industry to have a P-competitive structure, it must have all four of these characteristics:

- Many buyers and sellers
- A homogenous product
- Sufficient knowledge
- Free entry

These are all characteristics that favor price competition. Each of these characteristics will be explained in turn.

18.2.1 Many Buyers and Sellers

The idea is that the sellers and buyers are small relative to the size of the market, so that no one of them can "fix the price." If there are "many small sellers," it makes it much harder for any seller or any group of sellers to "rig the price." Similarly, if there are "many small buyers," there is little opportunity for buyers to "rig the price" in their own favor.

Each seller reasons as follows: "If I try to charge a price above the market price, my customers will know that they can get a better price from my competitors. My own share of the market is so small that all of my customers will be able to buy what they want from the competition — and I won't have any customers left!" Thus, the seller treats the price as being given and determined by "the forces of the market" independently of her own output.

How many sellers? How small? There is no absolute answer to that question; but there must be enough sellers and they must each be small enough so that each regards the price as being determined by the market, so that none of the sellers sees any opportunity to push the price up by cutting back on his or her output.

Similarly, there must be enough buyers, and each small enough, that each one treats the price as being determined by the market, and beyond her or his own ability to influence.

These conditions encourage the buyers and sellers in the market not even to try to control the price, but instead to compete against one another whenever quantity supplied differs from quantity demanded, driving the price toward the equilibrium of supply and demand.

18.2.2 Homogeneity

If the product (or service) of one seller differed significantly from that of another seller, then each seller would probably be able to retain at least some of the customers, even at a very high price. These would be the customers who just prefer this seller's product (or service) to that of someone else. The assumption of homogenous products serves to rule that out.

But this assumption should not be taken too literally. No two potatoes are exactly alike. We are not assuming that the goods are alike: only that the goods produced by one supplier are good substitutes for those offered by another seller. Thus, the potatoes don't need to be just alike — provided that, on the average, Farmer Jones' potatoes are just as good as Farmer Green's.

This is especially important with respect to services. It would be hard to prove that two haircuts are just alike! But so long as the haircuts supplied by one barber are substitutable for the haircuts supplied by another — and their conversation is about equally amusing — then the "homogenous products" assumption is fulfilled.

"Homogenous products" means all suppliers sell products that are perfect substitutes. If different sellers sold different products, then customers might be reluctant to switch suppliers when one supplier raises the price. They might stick with the supplier even at the higher price, because, even at the higher price, they like the product of that firm better than the product of another firm. By ruling this out, the homogeneity of products encourages price competition.

18.2.3 Knowledge

Some versions of the "perfectly competitive" structure include "perfect knowledge" as one of its characteristics — but, of course, "perfect knowledge" never exists in reality.

Perfect information is a little less clear than the other assumptions — we can hardly assume that people know everything there is to know! In practice, what is important is that each buyer and seller knows all about her or his opportunities to make deals, that is, knows the terms on which other market participants will buy and sell. Remember what we said in the paragraph on "many small sellers:" a seller would assume that her or his customers would know if the competition were selling more cheaply. If the customers didn't know that they had alternatives, then even a very small seller might get away with pushing the price up, without losing many customers. Thus, the "perfect information" assumption complements the other assumptions. The assumptions that there are many small buyers and many small sellers, and the assumption of free entry, all mean that buyers and sellers have many alternatives of potential buyers and sellers to choose among. The assumption of sufficient information says that they know what those alternatives are.

Traders need to know quite a bit to compete effectively in markets. They need to know the terms on which other people are offering goods and services, or offering to buy; the quality of the goods and services offered, and enough about costs to judge whether the trade is profitable or not. This is what I mean by "sufficient knowledge"

18.2.4 Free Entry

Remember Adam Smith's concept of the "natural price:" when the price of beer is high, so that brewing is especially profitable, people will enter the brewing trade and their competition with the established breweries will force the price down toward the "natural" price. As Smith was aware, in the long run the entry of new competition — or the exist of unprofitable firms from the industry, to go into other trades — is one of the most important aspects of competition and is thus one of the four characteristics of the P-competitive structure.

Free entry means that new companies can set up in business to compete with established companies whenever the new competitors feel that the profits are high enough to justify the investment. This is, first and foremost, a legal condition. That is, in a "perfectly competitive" market there are no government restrictions on the entry of new competition. This legal status is often called by the French phrase "laissez faire," meaning "let them make (whatever they want to make for sale)." But it could also be a practical condition. For example, if no-one could set up in business without enormous capital investments, that might prove an effective limit on the entry of new competition — especially if, for some reason, the capital cannot be raised by borrowing or issuing shares.

Now you can very well sum up the four characteristics of P-Competition :

1. Many Small Sellers

The more sellers there are, the more substitutes the consumer has

2. Homogenous Product

When the product is homogenous, then the substitutes are "perfect substitutes."

3. Sufficient Knowledge

When customers know the prices offered by other sellers, they will be better able to switch — increasing elasticity further.

4. Free Entry

In the long run, companies may even enter the market to provide still more substitutes

18.3 Other Market Forms

The other three market structure models can be defined in terms of the ways in which they deviate from the characteristics of P-Competition.

In a **Monopoly** there is just one seller of a good or service for which there is no close substitute.

In an **oligopoly** there are two or more, but only a few firms.

In **Monopolistic Competition**, the products are not homogenous but are "differentiated."

We don't have a standard model for "insufficient knowledge" but, at least in some cases, that seems to work similarly to "product differentiation."

18.4 The Competitive Firm

Our next step is to explore the operation of a firm in a P-Competitive industry. To be specific, what does the demand curve for the individual firm look like?

You have already noticed, some time back, that the individual firm's demand curve is different from that of the industry, and is more elastic. This is because substitutes increase elasticity, and the customer of the firm has many good substitutes for that firm's output — namely, the output of other firms in the industry.

Now lets move to the firms statistics in P-Competitive market.

18.5 Firm Demand in P-Competition

Since a P-Competitive structure is an idealization of these tendencies, we say that the demand curve for a P-Competitive firm is **infinitely** elastic.

In fact the demand curve for a P-Competitive Firm is a **horizontal line corresponding to the going price.**

And that makes sense, because the price in a P-Competitive market is determined by supply and demand — not by the seller or the buyer. Conversely, so far as the seller or the buyer is concerned, the price must be a given, since it is determined by supply and demand. The seller has no control over the price,

and to say that the seller has no control control over the price is to say that the price is given — a constant, a horizontal line — from the point of view of the seller.

Economists sometimes express this by saying that the price is "parametric," meaning that while it may change from time to time, it does not change in response to the firm's output decision.

18.6 Firm Supply and Demand

Happily — but not by coincidence! — all of our examples of profit maximization to date are based on the assumption of given prices. Thus, we already know that the supply curve of a P-Competitive firm is the firm's marginal cost curve.

Thus, marginal cost = price is the same as quantity supplied = quantity demanded for the individual firm.

When marginal cost = price for each firm in the industry, we have quantity supplied = quantity demanded in the industry as a whole.

Here's a picture:

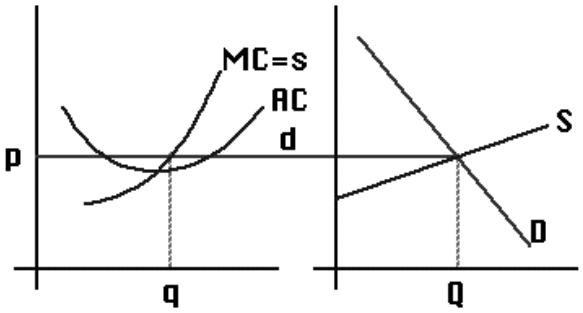


Figure 18.1

In the figure, the lower case q, s and d refer to output, supply and demand from the point of view of the individual firm, respectively, and the capital S, D, and Q are for the industry as a whole. Cost and supply curves are shown in red and demand curves in green. Price (per unit sold) is the same from all points of view.

18.7 Profits and Entry

We notice that, in the picture just shown, the firm is making an "economic profit." All costs, explicit and implicit, are included in the firm's Average Cost curve. In particular, Average Cost includes the opportunity cost of capital investment — so another way of putting it is that investors in this industry are making more than their best alternative investment in any other industry.

These profit opportunities will attract new firms into the industry. With "free entry," the (short run) supply curve of the industry shifts to the right, causing the price to drop until the economic profits are eliminated.

This process of entry and price change is known as the "long run equilibrium process" and it continues until "long run equilibrium" is attained. Here is a picture of the firm and industry in "long run equilibrium:"

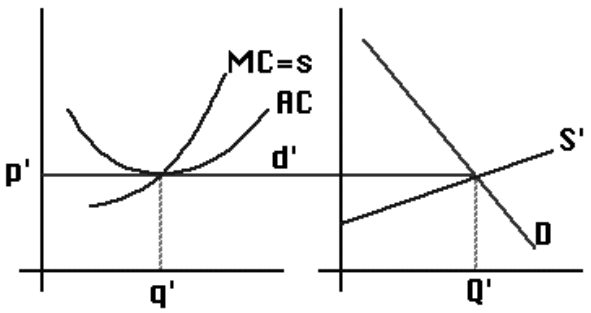


Figure 18.2

The new price, quantity, firm demand and short run supply are indicated by primes — p', q'. Q', d', S'. We see that, at a slightly lower price, the individual firm is lower on the MC curve and produces a little less, but since there are more firms in the industry, the industry as a whole produces more.

18.8 P-Competitive Equilibrium as an Ideal

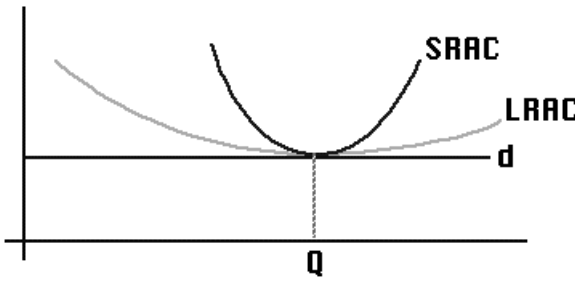


Figure 18. 3

We notice something else about the long-run equilibrium of a P-Competitive industry: each firm chooses the plant and equipment and productive capacity that gives the lowest average cost overall. This is shown by the above figure.

To see what this means, we might ask the following hypothetical question:

- If an industry is to produce a certain amount of output, how should the output it be divided up among the different firms?
- More specifically, how many firms should share that production assignment?

If there are very many firms, then each will be producing at a very small scale. They will not be taking advantage of the economies of scale, and cost per unit will be high. On the other hand, if there are very few firms, each will be producing on a very large scale, and suffering from diseconomies of scale, so, again, unit costs would be high. It would be best to balance the disadvantages of too large scale against the disadvantages of too small scale, and have just enough firms in the industry so that each is at the bottom of its average cost curve. The **total cost of producing that output is then at a minimum.**

What we see is that the equilibrium in a P-competitive industry does just that. That is one reason why economists often think of P-competition as an ideal.

Remember the assumption behind this whole argument! The assumption is that the long run average cost curve is u-shaped as shown. There may be some industries for which that is not true, and the argument would not be applicable to those industries.

18.9 Points About Long Run Equilibrium

We have seen that profits will lead to the entry of new firms into a P-competitive industry. This also works in the opposite direction: if firms in the industry were taking losses, supply in the industry will decrease. Firms in the industry might continue to produce in the short run, despite the losses. Remember that a profit-maximizing firm will continue to produce, in the short run, so long as it can cover its variable costs. However, in the long run, firms will drop out of the industry, if they continue to lose money. Thus, the supply curve of the unprofitable industry will shift to the left. But that in turn means prices will rise, and the long run equilibrium comes where the price is equal to average cost, as shown in Figure 18.2.

This discussion could apply to any economic activity to which there is "free entry." Economic profits — profits over and above the opportunity cost of capital — will attract new entrants. Returns less than the opportunity cost of capital will cause firms to get out of the industry. This will continue until the return to capital in that activity is the same as the opportunity cost of invested capital, that is, until profits are zero.

We might call this principle "the Entry Principle." It says that

In the long run, with free entry, returns to invested capital in an industry are just enough to offset the opportunity cost. When there are economic profits or losses, entry into the industry or exit of firms from it will shift the industry supply until economic profits are zero.

18.10 Long Run Supply 1

We can use these principles to explore the long run supply curve of the industry. Remember, in the short run, the capital plant and productive capacity of the industry is given, and the industry is made up of a certain group of firms. But in the long run, all of these things are variable. By definition, capital plant is variable in the long run — in fact, all inputs are variable. And the number and identity of the firms in the industry is variable. As we have just seen, economic profits will bring more firms into the group, and losses will result in the exit of firms. Alfred Marshall expressed this by saying that a competitive industry is like a forest, and the firms are like the trees. A forest does not grow by having bigger trees, but by having more trees. And similarly, a competitive industry grows or shrinks primarily by having more or fewer firms.

Let's consider a simple case. The simplifying assumptions are

1. All firms have identical cost curves
2. The cost curves are u-shaped as shown in Figure 18. 3
3. The cost curves remain stationary as the number of firms in the industry changes

Now, let's look at Figure 18. 4. We start out in long run equilibrium with 10,000 firms producing a total output of Q_1. Short run supply and demand curves are not shown. The average and marginal cost curves shown represent approximately those for the 10,000th firm with respect to its contribution to industry output, after the other 9999 firms have produced their parts. (These cost curves are very exaggerated. Drawn to scale they would be invisible).

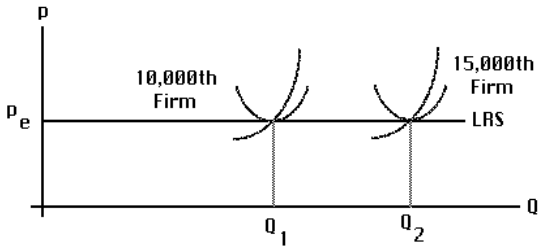


Figure 18.4

Now suppose there is an increase in demand (not shown) so that the price rises above p_e. The existing 10,000 firms will enjoy economic profits. These profits will attract new firms, so the number of firms in the industry will grow, shifting the short run industry supply (not shown) to the right and thus depressing the price back toward p_e. When will these new entries stop? Only when economic profits are back to zero. We suppose, for the example, that this happens when there are 15,000 firms in the industry, producing Q_2 of output. As the (exaggerated) cost curves for the 15,000th firm show, the price will be back at p_e.

What this example shows us is that

The long run supply curve in this case is a horizontal line corresponding to the bottom of the average cost curve of a firm in the industry.

This is a bit of a surprise. We have been thinking of supply curves as being upward sloping — but as we noted, there are some exceptions. In the long run, supply may or may not be upward sloping.

But first, let's define the long run supply curve a little more carefully.

Long Run Supply

For each industry output, the long run supply curve shows the lowest price at which that output can be produced, so that the price covers all costs including the opportunity cost of invested capital.

Thus the long run supply curve is a boundary — the boundary between profitable and unprofitable prices, given industry output.

We have seen that, on the simplifying assumptions:

1. All firms have identical cost curves
2. The cost curves are u-shaped
3. The cost curves remain stationary as the number of firms in the industry changes

the Long Run Supply curve (LRS) is a horizontal line. In economics, this special case is known as a **constant cost industry,** for reasons that are probably pretty obvious. But these simplifying assumptions can't always be applied. Assumption 3. is the tricky one. For example, agricultural industries would pretty clearly be exceptions to it.

Suppose, for example, that the demand for wheat increases, so wheat farming becomes profitable. Then more farmers switch from producing other crops to producing wheat. But that increases the demand for the best wheat land, so the rental cost of that land increases. New farms will have to pay the higher rental cost or make use of land that is less well suited to producing wheat. Either way, the new firms will have higher costs. Thus, the cost curves will shift upward as the number of wheat farms increases, and the new long run equilibrium price will be higher. Agricultural industries are not constant cost industries, but increasing cost industries.

For an increasing cost industry, the long run supply curve is upward sloping. Some economists believe there are also decreasing-cost industries, with downward-sloping long run supply curves. However, we will leave these complications for a course at a more advanced level.

Instead, let's look at another example of long run supply in a constant-cost industry: computer software.

I don't really know for sure that computer software is a constant cost industry. One problem is that computer programs are not identical, so we would have to measure the output of the industry in some uniform units. Perhaps, to a rough approximation, the output can be measured in lines of code. Of course, not all lines are equal — some lines are wasted and some are inspired — but this measure may work OK on the average, and, in fact, businessmen do use lines of code as a measure of programmer output.

Let's assume, for the sake of the example, that software output can be measured on the average by lines of code and that the industry is a constant-cost industry. The example is illustrated by Figure 18. 5:

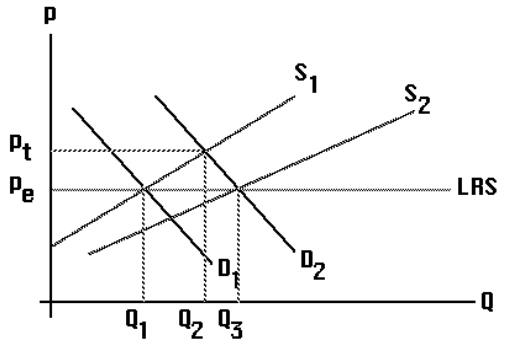


Figure 18.5

In the diagram, the long run supply of computer software is the gray line LRS. At the beginning, we have long run equilibrium at p_e and Q_1. Demand is D_1 and short run supply is S_1.

Now a breakthrough in computer hardware, a complementary good, increases the usefulness of computer software and so increases the demand for software. (The invention of desk-top computers pretty clearly had this effect). In the short run, the price of computer software rises to p_t with Q_2 lines of software produced. At this price, software is a profitable industry — the price is above the long run supply curve, which, by definition, is the boundary between profitable and unprofitable prices. Thus, there will be entry into the software industry (and more programmers and software engineers will be trained, an investment in human capital) so that the short run supply curve shifts to the right. The shift continues until the new long run equilibrium is reached at price p_e and production Q_3, with short run supply S_2.

This two-stage adjustment process is characteristic of industries characterized by free entry and supply-and-demand pricing.

18.11 Summary

The conduct and performance of an industry will depend to some extent on its structure. Among four major types of structures recognized by economists we have focused on the P-Competitive — sometimes called Perfectly or Purely Competitive — structure as the one corresponding most closely to the supply and demand model. This structure can be described by four basic characteristics:

- Many buyers and sellers
- A homogenous product
- Sufficient knowledge
- Free entry

All of these characteristics push an industry toward predominant **price** competition, so P-competitive could also stand for "price-competitive." In the short run, the P-competitive industry's supply curve is its marginal cost curve. In the long run, entry of new firms and exit of firms already in the industry will lead to a price corresponding to average cost, inclusive of the opportunity costs of capital and other resources. In other words, there are no "economic" profits. We also have found that the P-competitive model defines a kind of ideal in which rational self-interest leads to an allocation of resources in which given quantities of outputs are produced by enterprises of an efficient cost-minimizing scale. This remarkable finding is one modern counterpart to Adam Smith's conception of the "invisible hand."

Now when you are aware of perfect competition lets move to imperfect competitions.

LESSON 19: MONOPOLY

In this chapter we consider the workings of an industry in which there is no possibility of competition among sellers: a monopoly.

19.1 What is a Monopoly?

We have seen that P-Competition has some remarkable results. As we have seen, P-competition defines an ideal market structure in two senses: 1) in P-competition, price competition dominates all other forms of competition and forces the price to the supply-and-demand equilibrium, and 2) at that price and the corresponding output, marginal benefit is equal to marginal cost, so the allocation of resources is efficient — net benefits are maximized.

But it seems unlikely that all industries are P-competitive, even approximately. We expect to find P-competition in industries that (among other characteristics) have many small sellers. But we observed some industries (cable TV, for example) in which there is only one seller in a particular local market, and other industries in a whole spectrum from one to few to many sellers. P-competition is the many-small-sellers extreme of that spectrum.

Now we consider the opposite extreme: a monopoly. By definition, a monopoly is the only seller of a product for which there is no close substitute. It is an industry in which there is only one firm — or., conversely, a firm that has the whole industry to itself.

Once you know what is a monopoly, you need to know the Causes of Monopoly

Most economists regard monopoly as an exceptional case in a modern economy. In an economy populated by alert profit-seekers, it seems that any profitable monopoly would quickly attract competitors. For a monopoly to be stable, there must be some "barrier to entry." The assumption of free entry into the industry must not apply. Thus, we ask what might create the exception — what might "cause" a monopoly, what the "barrier to entry" might be.

Most texts give four causes of monopoly, which I will also give and add a fifth.

- patents and other forms of intellectual property
- control of an input resource
- government
- decreasing cost
- crime

We will discuss these five causes in turn.

19.1.1 Patents and Other Forms of Intellectual Property

Patent law is designed to increase the incentive to invent new methods of production and new goods. The inventor is granted a temporary monopoly on the use of the invention. The idea is that the patent makes the invention more profitable, during the term of the patent, and that these profits encourage inventors and so increase the rate of technical progress.

For example, the Polaroid company has owned the basic patents on instant cameras. When the Kodak company produced instant cameras in competition with Polaroid, a court found that this violated Polaroid's patent rights, and Kodak had to cease and desist and pay a penalty to Polaroid.

Other forms of "intellectual property" include copyrights on books and works of art and such, trade-marks, and trade secrets. Copyrights and trade-marks probably do not create monopolies in and of themselves. There may be close substitutes for copyrighted books, and close or even perfect substitutes can be offered for trade-marked goods, provided they do not falsify the trade-mark. However, it is possible that trade secrets might create monopolies. The formula for Coca-Cola, for example, is a trade secret. While Coca-Cola probably is not a monopoly, this is a matter of degree — Coca-Cola is a distinctive product. Whether other colas are close substitutes or not we leave to the judgment of the reader.

19.1.2 Control of an Input Resource

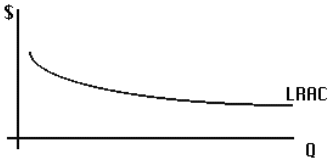
Products which require a natural-resource input may be monopolized if one supplier can get control of all known supplies of the natural resource. For example, at one time all known supplies of nickel were controlled by a single company. Aluminum ore, too, was at one time controlled by a single supplier.

19.1.3 Government Grants of Monopoly

Monopolies can be created by legislation. Historically, this has been an important source of monopolies as, for example, a monarch might grant a monopoly of wine to a favorite. In the modern world, governments may still encourage monopoly in many countries of the world, though this seems less common as time goes on.

19.1.4 Decreasing Costs

Monopolies can come about because there are decreasing costs (increasing returns to scale) in the long run. In such a case, the long run average cost slopes downward, as shown in the picture:



Increasing Returns to Scale

In such a case, the largest producer can undersell the rest, and still make a bigger profit. Therefore, in an industry in which there are increasing returns to scale, we would not be surprised to find a monopoly, in the absence of any other causes. Such a case is called a "natural monopoly."

We will consider this case in some detail later in the chapter.

19.1.5 Crime

While economists usually limit themselves to discussion of legal activities, it is clear that criminal coercion can prevent competition and so create monopolies. This is probably most common in activities that are anyway illegal, such as gambling, which often seem to be local monopolies. (It's hard to be sure, for obvious reasons, and probably pretty changeable, too).

Once established in illegal activities, criminals may use their profits and means of coercion to monopolize businesses that are legal, in principle, such as small scale lending. A possible example is so-called "loan sharking." The "loan shark" makes risky loans, which is legal in itself, but limits his risk by using the threat of violence (which is not legal) to limit the risk and assure that the loan and interest are paid. High interest rates will be charged, and these may be illegal. When there is a high risk of default, loans will be supplied only at very high interest rates, legal or not. But, if the loan shark uses coercive threats to maintain a monopoly of these risky loans, the rate of interest may be even higher than the risk of default requires, because of a monopoly mark-up

19.2 Monopoly Demand

The demand curve for a monopoly is different from that of a P-Competitive firm.

In P-competitive industry, we have to distinguish between the industry demand and the demand for the output of an individual firm, which are quite different. As we recall, the P-competitive firm has a horizontal demand curve. More general, if there are two or more firms in an industry, we have to distinguish between the industry demand and the firm demand. But in a monopoly — an industry with only one firm — there is no such distinction.

The demand curve for the monopoly is the demand curve for the industry — since the monopoly controls the output of the entire industry — and the industry demand curve is downward sloping. So the monopoly's demand curve is downward sloping, That means the monopoly can push the price up by limiting output. If the monopoly cuts back on its output, it can move up the industry demand curve to a higher price.

To illustrate what this means, let's "tell a story" about monopoly. In this story the monopoly will be created by legislation.

The story begins with a competitive industry consisting of many firms. As usual we will call it the "widget" industry. Economists often use the word "widget" meaning "some small good or service, not specific." (This came from a 1950's musical comedy, "How to Succeed in Business Without Really Trying.") You won't be far off if you think of it as a small manufactured good.

We'll need to make a few simplifying assumptions. Assumption: Each firm operates under constant costs in the long run. We recall that with constant costs in the long run, each firm's long run average cost is a horizontal line. We need two more facts about constant costs in the long run, and one more assumption.

Fact: with constant long run costs, the firm's long run marginal cost is also a horizontal line, identical with the LRAC curve.

Assumption: the firm cost curves remain the same as the number of firms in the industry increases.

Fact: when the assumption is true, the long run average cost, marginal cost, and supply curve of the industry are also the same horizontal line.

With these assumptions, the competitive industry's supply curve is a horizontal line. The discussion would be a little more complex in general, allowing for more complicated supply curves, but the overall results would be the same.

Here is a picture of long run supply and demand under these assumptions. In the picture, it is assumed that the constant average cost for the widget industry is $40 per unit, so the long run supply curve is the horizontal red line at $40, while the industry demand is the downward sloping green line.

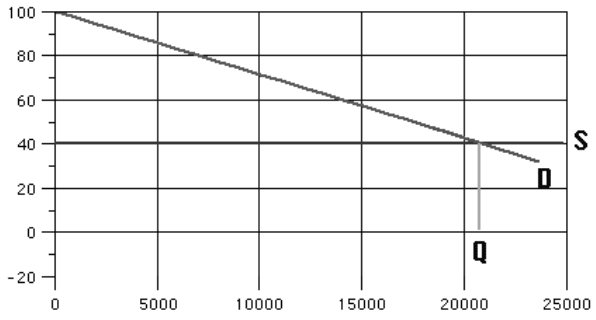


Figure 19.1

As we see, the long run equilibrium output for the widget industry is 21,000 units of output at a price of $40.

Here, in Table 19.1, are the numbers for the example.

Table 19.1

Price	Quantity
100	0
97	1000
94	2000
89	4000
83	6000
77	8000
71	10000
66	12000
60	14000
54	16000
49	18000
43	20000
37	22000
31	24000
26	26000
20	28000
14	30000

Now we go on to the next stage of our story.
The national legislature passes a law limiting competition in the widget industry. All the small firms in the industry are consolidated into one large corporation. The owners of the old competitive firms are issued shares in the corporation proportionate to their ownership in the old firms. The corporation is now the only firm allowed to supply the widget industry.

The board of directors of the widget industry now meet to consider their policy, and they aim at maximizing profits. How will they go about this?

In general, the logic of monopoly profit maximization is very much like the logic of the other "maximization" questions we have dealt with. We want to use the marginal-cost, marginal-benefit logic to solve it.

But it **is** different from the P-competitive firm, because the price is no longer fixed, from the point of view of the monopoly.

Of course, the monopoly will charge "all the market will bear," that is, will choose a price and output on the demand curve. But that doesn't tell us much. Starting at the competitive equilibrium price of $40, the monopoly will raise its price to increase its profits, but its quantity sold will drop with each price increase. Eventually, its profits will begin to decline because the lost sales offset the higher prices. How far up will the monopoly push the price?

To solve that problem, we need one more concept: marginal revenue.

19.3 Marginal Revenue

We can define marginal revenue by a formula that should be familiar by now, at least in its broad outlines.

$$\text{Marginal Revenue} = \frac{\Delta R}{\Delta Q}$$

where R is revenue (that is, price times quantity sold) and Q is the quantity sold. As usual, this is an approximative formula, and the smaller the change in Q the better the approximation. We can interpret marginal revenue as (approximately) the increase in total revenue as a result of selling one more unit of output. Here's an example of calculation of the approximation: suppose output increases from 10000 to 11000 and revenue increases from 754286 to 714286. Then we have

$$\frac{\Delta R}{\Delta Q} = \frac{754286-714286}{11000-10000} = \frac{40000}{1000} = 40$$

Thus, between 10,000 and 11,000 units of output, the marginal revenue is approximately $40.

19.3.1 Monopoly, P-Competition, and Marginal Revenue

Here is a difference between monopoly and P-competition. For the P-competitive firm, the marginal revenue is the same as the price, since each unit sold will add the price to revenue. For the monopoly, it is different. In order to sell one more unit, the monopoly has to drop its price a bit. The additional unit sold will add something to revenue, but the cut in price will decrease the revenue from the units the monopoly could have sold at the old price, without cutting. So the net addition to revenue will be less than the price at which the additional unit is sold, and could even be negative — the lost revenue from the price cut could be more than the price for which the additional unit is sold.

For example, suppose the monopoly is selling 11,000 widgets at $69 each. This gives a total revenue of $759,000. In order to increase sales to 13,000, the monopoly has to reduce its price to $63. The price of $63 applies to the first 11,000 units as well as the remaining 2000 — all units of output are sold at the same market price. So 11,000 units at $63 per unit yields only $693,000. This is more than made up by the $126,000 earned from selling 2000 more units at $63, leaving a total revenue of $819,000 — an increase in revenue relative to the $759,000 the monopoly started with, but the increase is not in proportion to the additional sales.

19.3.2 Marginal Revenue Example: Table

Here is a table of the output, price and marginal revenue in our numerical example for this chapter. Notice that the marginal revenue drops much faster than the price, and in fact is negative when the price is $40 or more.

Table 19.2

Output	Price	Marginal Revenue
0	0	97
1000	97	89
3000	91	77
5000	86	66
7000	80	54
9000	74	43
11000	69	31
13000	63	20
15000	57	9
17000	51	-3
19000	46	-14

19.3.3 A Picture of Demand and Marginal Revenue

Here is a picture of demand and marginal revenue for our example, based on the data in the previous table. Demand is shown in the darker green, and marginal revenue in the lighter.

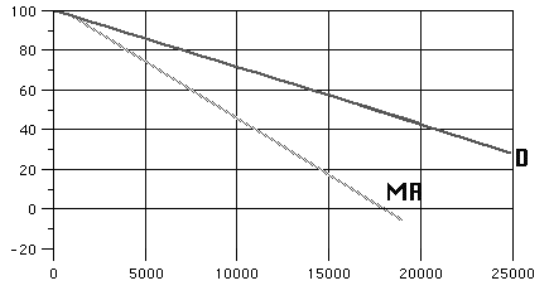


Figure 19.2

19.4 Monopoly Profit Maximization

The rule for monopoly profit maximization will come as no surprise. It is

MR=MC

That is, the rule says that the monopoly should increase output up to the level where the marginal cost curve intersects the marginal revenue curve, in order to maximize its profits. The price charged is the corresponding price on the demand curve. Notice that this is a two-stage analysis:

- at the first stage, the marginal cost, shown in red, and the marginal revenue, shown in light green determine the output. Profit maximizing output is the output at which they intersect, shown by the gold line.
- at the second stage, the output and the demand curve determine the price. Trace up the vertical gold line to the dark green demand curve, and that's the profit-maximizing price.

The diagram is a little more complex. Here it is:

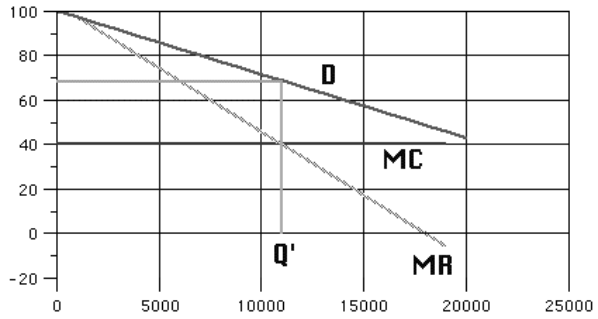


Figure 19.3

The output that corresponds to maximum profits is Q', which is 10,500 widgets, and the monopoly price is $70 per widget.

19.4.1 Monopoly Output Restriction

In the diagram, the monopoly maximizes its profits by selling Q'.

Now, let's get back to our story. Recall, the widget industry had been monopolized by an act of the legislature. Before it was monopolized, it had sold 21,000 units at a price of $40 — to the right where demand crosses the marginal cost line. Now, we see the monopoly selling much less. In fact it will sell just half what the competitive industry sold — 10,500 units, at a price of $70.

(It will not be this simple as a rule. Remember, we have made a lot of simplifying assumptions to get here. What we can be sure of in general is that a profit-maximizing monopoly will sell less than the supply-demand output, at a higher price).

19.4.1 Monopoly Profits

In our example, the widget industry started out in long run equilibrium, with zero economic profits. Once the monopoly has cut back to its long run equilibrium, at a higher price, it will have positive economic profits. In the picture below, profits are shown by the shaded blue area. The amount of monopoly profit in this simple example is $30 (the mark-up over $40 of cost per unit) times 10,500 widgets sold, or $315,000.

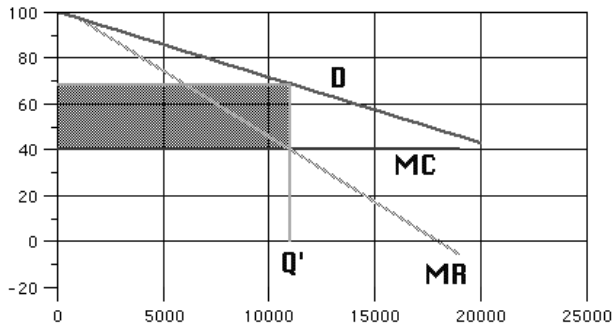


Figure 19.4

19.5 Monopoly Inefficiency

The restriction of output by the monopoly is inefficient. This inefficiency is shown in the following figure:

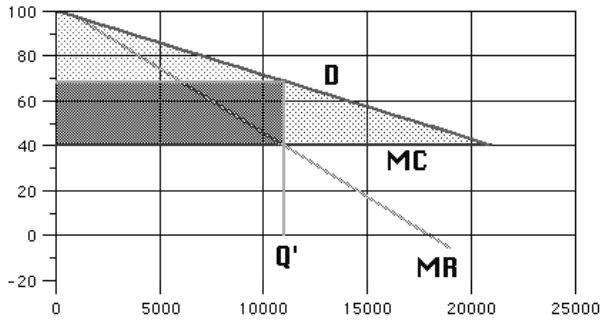


Figure 19.5

We can explain the inefficiency of monopoly by using the concept of consumers' surplus, using the diagram. There are three areas, the lightly shaded area above the profit rectangle, the lightly shaded area to its right, and the profit rectangle itself. Before the industry is monopolized, consumers buy 21,000 widgets at $40 per widget and their consumers' surplus is the sum of the three areas. After the monopolization, the consumers buy 11,500 widgets at $70, and their consumers' surplus is the area of the upper triangle.

Let's add up the benefits of monopolization. After monopolization, the net benefits from widget production have two components: the profit rectangle plus the upper consumers' surplus triangle. But the opportunity cost of monopolization is the consumers' surplus the consumers would have enjoyed if they had continued to buy at $40 — the sum of the three areas. Thus, the benefits of monopolization are less than the costs, and the difference — the excess cost — is measured by the area of the (red) triangle to the right.

This loss of consumers' surplus is called the "deadweight loss" (meaning the monopoly profits are not enough to offset it) or the "welfare triangle" and is a measure of the waste due to monopoly restriction of output. In this very simplified example, it is half of the monopoly profits.

19.6 Complications in the Theory of Monopoly

But there are still some complications we have not taken account of. So far we have been assuming that the monopoly has the same cost conditions that a competitive industry would have. The complications arise when the monopoly and a competitive industry cannot have the same cost conditions.

This could happen for two reasons.

The monopoly, lacking the spur of competition, wastes resources so that its cost curves are above those of a competitive industry. The term for this is "X-Inefficiency."

This is a deviation from the neoclassical assumption of absolute rationality, but perhaps a very important realistic deviation. A good deal of observational evidence suggests that different firms can have quite different costs, in what seem to be the same circumstances. Competition, driving prices down, will tend to eliminate this problem — by eliminating the high-cost firms. In the absence of any competition at all, we would expect costs to be higher than they would be in a competitive industry. There is some evidence that this is true.

There are economies of scale, so the monopoly, operating on a larger scale, can achieve lower costs.

This is the case of natural monopoly. We will explore it in a bit more detail.

19.7 "Natural" Monopoly

"Natural" monopoly creates a dilemma for neoclassical economics and (perhaps) for market economies.

Here is a picture-example of "natural" monopoly. The example assumes that there is one indivisible cost, but that once it is paid, the firm can produce an unlimited amount at a constant marginal cost. Thus, the Long Run Marginal Cost is horizontal, but the Long run Average Cost is downward- sloping

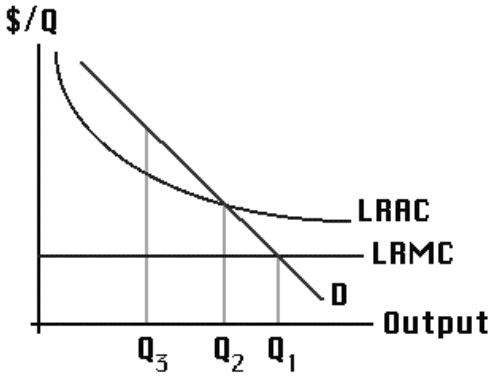


Figure 19.6

The dilemma is that output Q_1, where MC=price, is still the efficient output. But at that output, the monopoly cannot cover its total costs. On the other hand, a profit-maximizing monopoly will produce much less, at Q_3, which covers costs but it inefficient.

In different countries and at different times, governments have dealt with this problem in three primary ways:

- Government Ownership
- Regulation
- Deregulation

19.7.1 Government Ownership

Government ownership has been a pretty common response outside the United States, and there are cases of municipal ownership in the US.

Government ownership could, in principle, solve the problem, since the government could operate the monopoly efficiently, charging a price equal to marginal cost, and cover the losses out of tax revenues.

In practice, however, government monopolies usually seem to have been operated as "cash cows" for the government, and that's not a solution to the problem of high monopoly prices! It has been quite common around the world (for example) for public telephone monopolies to raise the price of telephone service to pay the deficit of the postal system. Poor telephone service at a high price is the predictable result.

In recent years, many of these government monopolies have been privatized.

19.7.2 Regulation

In the United States, for most of the 20th Century, the most common response has been regulation. In this system, a private monopoly is recognized and protected as such, on the condition that it keep its price down below the profit-maximizing level. The monopoly would be allowed to earn a "fair rate of return."

Over the years, this was more and more interpreted as meaning that the monopoly would operate at Q_2, where the price just covers average cost. This is less efficient than Q_1, but better than Q_3. However, there are some other complications that led economists and regulators to question this interpretation by the 1960's. In recent years, the trend has been away from regulation.

19.7.3 Deregulation

Natural monopolies are complex businesses, with different lines of business and different cost conditions for the different lines of business, and changing technology may change the cost curves, making more competition possible. The telephone industry provides some examples. The Bell monopoly of the 1970's offered both long distance and local service, as well as some other lines of business. Microwave technology and other technical developments were making it possible for smaller firms to compete with Bell in long distance service. But, so long as Bell remained under "natural monopoly" regulations, all its lines of business were interdependent and had to be regulated in complicated ways.

This led some economists to argue that, even in natural monopoly conditions, it is best to rely more on market forces and less on government. Under the influence of those economists, US natural monopolies have increasingly been "deregulated." This began with the Jimmy Carter administration (1977-80) and has been continued since by all administrations, Republican or Democratic.

Deregulation has not meant that all regulations were eliminated, but their scope has been cut back a great deal. At a minimum, the companies have had more freedom to set their own prices,

while in most cases they are no longer protected from new competition.

19.7.4 Deregulation as Market-Based Regulation

It is not clear that "deregulation" has really reduced the scope of regulation of "natural monopolies." In practice, what emerged in the 1990's might be better described as market-based regulation. The nature of the regulations have changed in ways that are designed to encourage, rather than restricting, price competition.

This is illustrated by a recent court decision. Michael Weinstein, a columnist for the New York Times, writes (January 28, 1999, p. C2) that the supreme court in January upheld Federal Communications Commission regulations that had been challenged by the local telephone service providers in some parts of the United States. The regulations were designed to encourage new companies to enter the local markets to compete with the Bell Telephone Companies and other local providers. Free entry, remember, is a key characteristic of price-competitive markets. "The problem was that no company was in a position to provide local service using only its own equipment. Entrants would need to use, at least initially, some or all of their of the Bell network, including wires into homes, switching equipment and operator services." Weinstein quotes Gene Kimmelman, a Consumers' Union representative, as saying that the FCC had imposed "smart rules that compel the Bell companies to compete on fair terms with their rivals." The rules required that the local providers lease their physical facilities to the new entrants and limit the price they can charge. "... under the commission's rules, the Baby Bells would charge enough to cover legitimate costs — the costs of a low-cost provider — and not one penny more." Similar rules have been imposed on local providers of electrical power in some areas, including the Philadelphia region.

The local service providers had argued that they should not be forced to help their competitors by allowing them to use the physical facilities which are the property, not of the new entrants, but of the established local providers. They argued that this was against the spirit of deregulation and perhaps against the constitutional protection of private property. However, the FCC argued that a failure to require the local providers to rent their facilities at a controlled price "posed the danger that the Bells would exploit the 1996 act to dominate telecommunications markets, making a mockery of Congress' will."

Once again we see the dilemma of natural monopoly. It may be that the new sort of regulations, designed to make the telecommunications and other "public utility" markets function more like price-competitive markets, will be more effective in the long run than the old sort, that tolerated the monopoly so long as it did not raise the price too high. But the idea that we can get rid of regulation and rely on the spontaneous forces of competition seems as far away as ever

19.8 Summary on Monopoly

Monopoly provides an important example of an exception to the Fundamental Principle of Microeconomics, in that there is not enough competition to push the price down to the supply-demand level. Indeed there is only one seller.

We have seen that a profit-maximizing monopoly will

- produce less than a comparable P-competitive industry
- charge a higher price
- this output restriction is inefficient

However, if there are economies of scale, P-competition may simply not be possible, and the extreme case of "natural monopoly" leaves us with a choice of the least of three evils: public ownership, regulation, or deregulation.

Now when you know both the extremes, let's study the stuff in between.

Notes

LESSON 20:
IMPERFECT COMPETITION

As I have said earlier, that we will use the concepts related to perfect competition to define other market structures. Now let's start with the imperfect competition

20.1 Imperfect Competition

In the previous chapters, we have identified P-competition as an ideal, in that the P-competitive structure encourages strong price competition, which in turn leads to price and output at the supply-and-demand equilibrium levels. In appropriate circumstances, the supply-and-demand equilibrium is efficient, as the Fundamental Principle of Microeconomics points out.

For an industry to have a P-competitive structure, it must have all four of these characteristics:

- Many buyers and sellers
- A homogenous product
- Sufficient knowledge
- Free entry

Some industries (such as financial service industries and many agricultural industries) seem to approximate the P-competitive structure nearly enough that there is little doubt that the P-competitive theory can be applied to them. A few industries are equally clearly exceptions to which the P-competitive theory cannot be applied: monopolies, which we studied in the last chapter. There seem to be a considerable number of industries that don't clearly fit either extreme. The discussion of these cases in the 1920's and 1930's led to a classification of industries and markets into four major types:

- "Perfect Competition"
- Monopoly
- Oligopoly
- Monopolistic competition

We now need to say what we can about the last two of these types. **"Oligopoly" and "Monopolistic Competition" are often lumped together as "Imperfect Competition."**

Here is a table to illustrate the the four characteristics of "Pure Competition" and how the four market forms differ.

	Sellers	Product	Knowledge	Entry
Pure Competition	Many	Homogenous	Sufficient	Free
Monopoly	One	?	?	None
Oligopoly	Few	?	?	None or limited
Monopolistic Competition	Many	Differentiated	?	Free

20.2 Forms of Imperfect Competition

The two recognized forms of imperfect competition are defined by the ways in which they deviate from the four characteristics of P-competitive markets:

Oligopoly

The term "oligopoly" comes from Greek roots meaning "few sellers," and that is the way that oligopoly differs both from P-competition and monopoly — there is more than one seller, but not "very many." Of course, this is a vague conception — the vagueness is unavoidable in the use of the relative term "few" — and economists have debated on the exact boundaries between "few" and "many. For the small number of sellers to be stable, there presumably must be some "barriers to entry" of new competitors.

Monopolistic Competition

In monopolistic competition the products sold by the different firms in the industry group are not homogenous but differentiated. Thus, each firm has a "monopoly" of its own product. But it is not a true monopoly, such as we considered in the last chapter, because the differentiated products are "close substitutes." Once again, we have a certain vagueness here: how close is a "close substitute?" But once again, the vagueness is in the facts, not in the discussion: the products of real firms may be more or less close substitutes in different cases. For monopolistic competition, however, entry is free.

All of this vagueness is a bother, but the thing to keep in mind is that these are broad umbrella terms, each of which includes a range of possibilities, and that may overlap to some extent. For example, some oligopolies sell differentiated products. Rather than try to discuss these types separately, we will consider some of the characteristics that may be observed in imperfectly competitive markets in general, and that be more or less important in different cases.

20.3 Implications of Imperfect Competition

The main significance of the four characteristics of P-Competitive structure is that they are conducive to price competition. When those characteristics are missing, we may see

- increased nonprice competition
- decreased price competition.

We will explore these possibilities in turn. After exporing the kinds of nonrpice competition, we will explore monopolistic competition, a market structure associated with a great deal of nonprice competition, then go on to discuss oligopoly, in which reduced price competition is the central issue.

20.4 Other Forms of Competition

In these other two market structures, which together are often called "imperfectly competitive" structures, there may just be less competition, or there may be other forms of competition, "nonprice competition."

Nonprice competition includes

- competition in the characteristics of the good
  - differentiation
  - quality competition
- Advertising
  - informational
  - persuasive

20.5 Product Differentiation

Products are differentiated when the products of different companies are not perfect substitutes — instead, "every company has a monopoly of its own product." Nevertheless, companies may compete by changing the characteristics of the product they sell. The idea is not necessarily to make a better product than the competitor, just different — to appeal to a different "market niche."

Again, economists (on the whole) regard this form of competition as a mixed bag:

- It increases variety, thus increasing the range of consumer choice, which is good, but
- It divides up the market, leading to higher prices and costs

According to the traditional ideas on "imperfect competition" developed in the first half of the twentieth century, this form of competition is especially common in "monopolistic competition." In fact it is part of the definition of monopolistic competition. But is also observed in many oligopolies.

20.6 Competition in Quality

In some industries, there is a very hot competition to introduce a product that is superior to rival products. This form of competition has a good reputation but it can certainly be overdone and may be overrated.

Competition in quality lends itself to "horse races," in which only the winner gets any profit at all, and recent research indicates that these "winner take all" competitions tend to lead to overinvestment and waste of resources. The constant upgrading of computer software may be an example. We will take a closer look at "horse race" competition in a later chapter. In any event, to much of a good thing — that is, spending too much on a good thing — can always change it into a bad thing.

On the other hand, quality improvement makes consumers better off in an obvious and probably pretty major way, so competition in quality is **at worst** a mixed bag.

20.7 Informational Advertising

In a P-competitive market, there is no advertising, because there is no need to advertise — the firm can sell its profit-maximizing output at the market price, so why should it spend on advertising? But in the real world, advertising is a very major competitive strategy.

When the aim of the advertising is to give people information about the availability, characteristics and prices of goods, we call it "informational advertising." This sort of advertising increases the consumer's range of choice and may improve the quality of the decisions consumers make. It can also contribute to the effectiveness of price competition, so (in some markets) it may complement price competition and bring the market closer to the supply-and-demand outcome than it otherwise would be.

The purpose of cutting your price is to attract more customers, after all. If nobody knows about the price cut, it will be less effective, so why bother? Conversely, an advertised price cut can be the most powerful form of price competition.

Some "professional ethics" laws have prohibited physicians and lawyers from advertising, on the grounds that such advertising would be inconsistent with professionalism. In recent years, as part of the trend toward increased reliance on markets in modern economies, those laws have been repealed in some jurisdictions. The repeal does seem to lead to increased price competition.

All the same, like other forms of nonprice competition, economists regard informational advertising as something of a mixed bag:

- It gives people information they might lack, which is good, but
- It can be overdone, spending too much on advertising and driving prices up, which is not so good.

20.8 Persuasive Advertising

The purpose of persuasive advertising is to shift the utility functions of the customers, thus to shift their demand curves in favor of the good being advertised. Since we judge consumers' benefits in terms of **given** consumer utility functions, it is hard to say if consumers benefit or lose — so persuasive advertising looks like spending money without making consumers any better off, and many economists regard that as a negative.

We expect that both forms of advertising will be especially common in oligopolies and whenever products are differentiated, as in monopolistic competition. Monopolies, too, may find it profitable to advertise.

20.9 Summary

While the various forms of nonprice competition each seems to be a mixed bag — some good, some bad consequences — many economists feel that, on the whole, more competition in all these categories tends to be better than less competition. This is especially likely where the competition results in an increased range of consumer choice. Common sense tells us that choices "keep the suppliers honest," and economic theory tells us that even when the suppliers are as honest as they could be, more choices can make people better off, but never worse off — so nonprice competition, even though it has costs, can have pretty substantial benefits that more than balance them.

Let me tell you a very interesting example of garbage pickup. Garbage pickup is a rather interesting industry, in that we can find just about all forms of economic organization in this field:

Government Organization
In many localities, the garbage is hauled by public employees

Privatization
There are a considerable number of communities in which garbage is hauled by private monopolies under a contract with the government.

Free competition
You must have seen some people picking up garbage from road side.

Autarky
Autarky means each person must provide the service for himself or herself, and there are rural areas in which there is indeed no garbage pickup service at all.

Now let us switch over to types of imperfect competition one by one. First we will take up Monopolistic Competition.

Notes

LESSON 21:
MONOPOLISTIC COMPETITION

21.1 What is Monopolistic Competition

In discussing industries that are neither monopolies nor p-competitive, economists have tended to begin from the four characteristics of a p-competitive industry. We recall that those characteristics are:

- Many buyers and sellers
- A homogenous product
- Sufficient knowledge
- Free entry

Competition can be "imperfect" in an industry if the industry deviates from any one of the four. Thus, if there are just a few firms (but more than one), deviating from the first characteristic, the industry is said to be an "oligopoly." Since the nineteen-twenties, economists have also discussed the situation when an "industry" deviates only in the second characteristic. This is called "monopolistic competition," and we have "monopolistic competition" when a group of firms sell closely related, but not homogenous products. Instead, the products are said to be "differentiated products." Thus, the characteristics of "monopolistic competition" are:

- Many buyers and sellers
- Differentiated products
- Sufficient knowledge
- Free entry

To say that products are differentiated is to say that the products may be (more or less) good substitutes, but they are not perfect substitutes. For an example of a monopolistically competitive "industry" we may think of the hairdressing industry. There are many hairdressers in the country, and most hairdressing firms are quite small. There is free entry and it is at least possible that people know enough about their hairdressing options so that the "sufficient knowledge" condition is fulfilled. But the products of different hairdressers are not perfect substitutes. At the very least, their services are differentiated by location. A hairdresser in Center City Philadelphia is not a perfect substitute for a hairdresser in the suburbs — although they may be good substitutes from the point of view of a customer who lives in the suburbs but works in Center City. Hairdressers' services may be differentiated in other ways as well. Their styles may be different; the decor of the salon may be different, and that may make a difference for some customers; and even the quality of the conversation may make a difference. A very good friend of mine changed hairdressers because her old hairdresser was an outspoken Republican. My friend said that she just couldn't take it any more without answering back — and it's not a good idea to get into a controversy with one's haircutter!

21.1.1 Product Differentiation

On the previous page, the word "industry" was put in quotes, when it referred to a group of firms with product differentiation. That's because the boundaries of the industry become much more vague when we talk about product differentiation. A hairdresser in Center City Philadelphia and another in a Philadelphia suburb may be pretty close substitutes — but the Philadelphia hairdresser's service will be a very poor substitute for the services of a hairdresser in Seattle! Are they in the same industry? Or should we think of hairdressing industries as localized, so that Philadelphia hairdressing is a different industry than Seattle hairdressing? And, too: barbers may cut women's hair, and hairdressers may cut men's hair. Are hairdressers and barbers part of the same industry, or different industries? There really is no final answer to this question, and some economists have avoided any reference to industries in dealing with monopolistic competition. Instead they talk about "product groups." A product group is a group of firms selling products that are "good," but not necessarily "perfect" substitutes. And, of course, a product group is not unique, since it depends on how "good" we require the substitutes to be, so there will be broader and narrower product groups. Coke and Pepsi are both members of the product group "cola drinks," while Coke, Dr. Pepper, Sprite and Squirt are members of the broader product group "carbonated soft drinks."

This illustrates another point. Product differentiation is characteristic of monopolistic competition, but not limited to monopolistic competition. Oligopolies, too, may have product differentiation. Cola drinks would probably be thought of as a differentiated oligopoly, an oligopoly product group, rather than a monopolistically competitive group.

And what about "free entry?" For monopolistic competition, that means free entry into the "product group." Again, let's think of hairdressers as the example. If a hairdresser is especially successful with a Seattle-punk style at a certain location in Center City Philadelphia, there is nothing to prevent other hairdressers from setting up at a nearby location, and cutting in a similar style. In that sense, there is "free entry" into the product group. In general, when one monopolistically competitive firm is quite profitable, we may expect that other firms will set up in business producing similar products, and established firms may change the characteristics of the products they produce, to make those products more similar to the successful one. In that sense, there is free entry into the monopolistically competitive product group.

21.1.1.1 The Short Run

In the short run, then, the monopolistically competitive firm faces limited competition. There are other firms that sell products that are good, but not perfect, substitutes for the firm's own product. In the words of British economist Joan

Robinson, every firm has a monopoly of its own product. When the product is differentiated, that means the firm has some monopoly power — maybe not much, if the competing products are close substitutes, but some monopoly power, and that means we must use the monopoly analysis, as if Figure 1 below.

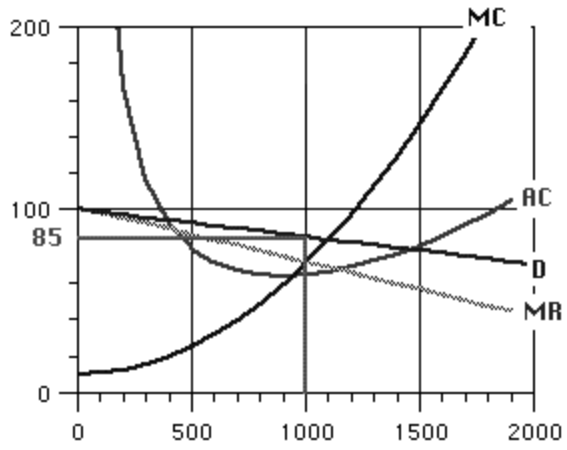


Figure 21.1: Monopolistic Competition

We see that, as usual in monopoly analysis, the marginal revenue is less than the price. The firm will set its output so as to make marginal cost equal to marginal revenue, and charge the corresponding price on the demand curve, so that in this example, the monopoly sells 1000 units of output (per week, perhaps) for a price of $85 per unit.

But this is just a short run situation. We see that the price is greater than the average cost (which is $74 per unit, in this case) giving a profit of $11,000 per week. We remember too that this is economic profit — net of all implicit as well as explicit costs — so this profitable performance will attract new competition in the long run. What that means is that new firms will set up, and existing firms will change their products, so that there will be more, and closer, substitutes in the long run. That will shift the demand for this firm's profits downward, and perhaps cause the cost curves to shift upward as well, squeezing the profit margins.

21.1.1.2 The Long Run

In monopolistic competition, when one firm or product variety is profitable, it will attract more competition — more substitutes and closer substitutes for the profitable product type. Thus, demand will shift downward and (perhaps) costs will increase. This will go on as long as the firm and its product type remain profitable. A new "long run equilibrium" is reached when (economic) profits have been eliminated. This is shown in Figure 2:

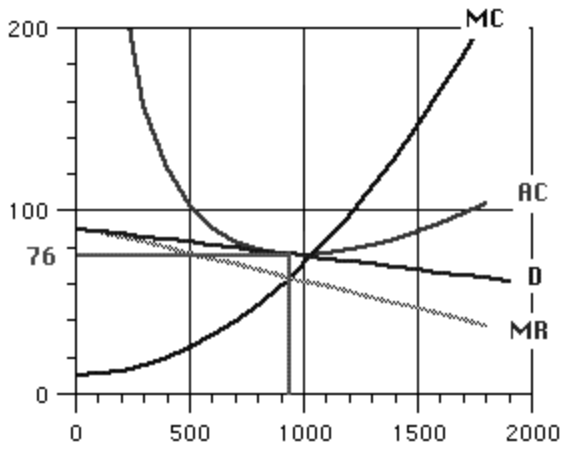


Figure 21.2

In this example, the firm can break even by selling 935 units of output at a price of $76 per unit. The profit — zero — is the greatest profit the firm can make, so profit is being maximized (as usual) with the output that makes MC=MR.

Zero (economic) profit is also the condition for long run equilibrium in a p-competitive industry. But this equilibrium is not the ideal that the long run equilibrium in a p-competitive industry is. Many economists feel that the long run equilibrium in a monopolistic industry has some problems:

Inefficiency

Notice that, either in the long run or in the short, the price is greater than marginal cost. But the condition for efficient production is that price is equal to marginal cost. Thus, an individual firm's output is less that would be efficient, according to the traditional standard.

Excess Capacity

We see that, in the long run, the firm is not producing at the bottom of its long run average cost curve. Instead, it is operating on a scale that is smaller and less efficient — the firm has a capacity to produce more at a lower average cost. To put it a little differently, each firm is serving a market that is too small, and there are too many firms, so that the product group as a whole has the capacity to serve more customers than there are — excess capacity.

Advertising and Nonprice Competition

A firm in a p-competitive industry will not advertise at all. Why should they? The p-competitive firm can sell all it wants to sell, without cutting its price, so why spend money to get more customers? But the monopolistically competitive firm cannot sell all it wants without cutting its price, and advertising to get more customers may be more profitable than cutting price. Thus, economists expect to see monopolistic competition associated with advertising. Moreover, advertising seems to go along with differentiated products, and as a profitable firm attracts more competition, with more substitutes and closer substitutes for their product, the firm may feel that it needs to spend more on advertising. (That's why the cost curve could be

higher in a long run equilibrium). In this context, advertising is seen as wasteful.

Of course, all of this is controversial. Some economists have been quite critical of the idea of monopolistic competition from the start. Here are some responses the critics might make to these points.

Inefficiency

While the hypothetical monopolistically competitive firm does operate inefficiently, it doesn't miss efficiency by much. The deviation of marginal cost from marginal price, and of the firm's production and price from the efficient, p-competitive quantities could be only a few percent — less than we can detect in practice. Thus, despite the differentiation of products, the p-competitive theory may be a "good enough" approximation, especially in the long run.

Excess Capacity

Again, for concreteness, let's think of hairdressing as a typical instance of "monopolistic competition." What this is telling us is that if some of the existing hairdressing enterprises were combined, so that there would be fewer hairdressers each serving a larger market, they could serve that market at a lower cost and price. Perhaps; but some consumers would lose out, since they would have to go further from their homes or offices to find the nearest hairdresser. More generally, getting rid of "excess capacity" means sacrificing variety — and that's a loss. Whose favorite is to be eliminated?

Advertising and Nonprice Competition

Actually, advertising is common in most industries, however, competitive — as a disequilibrium event. When price is a little above equilibrium, it makes sense for a competitive firm to include advertising it its competitive mix; but when price competition brings the price down to its equilibrium level, sellers no longer have any reason to compete for buyers — either by price cuts or advertising or any other way. In the real world, a competitive industry is always reacting to changing events, always moving toward the equilibrium — but it may never stay there for very long. And this advertising is useful, in keeping consumers up to date about their opportunities. So it is not clear that monopolistically competitive advertising is something to be concerned about.

As we have seen before, economic theory favors price competition, while nonprice competition — by advertising and other means — is often seen as a mixed bag of good and bad. But some economists claim that monopolistic competition promotes a particularly unfortunate kind of nonprice competition. Here's the idea:

21.1.2 Increasing Product Differentiation

A monopolistically competitive firm faces competition from other products that are good substitutes for its own product type. One way that it might be able to improve its profit margins is by changing its product type so that the other products are less substitutable for it. This is "increasing the differentiation of the product" or "(further) differentiating the product," and may be accomplished by creating marketing, engineering redesign, or other means.

How does "increasing the differentiation of the product" work in the model of monopolistic competition? Thinking back to the section on elasticity, we recall

The more and the closer substitutes there are for a product, the more elastic the demand for that product is.

So, if the effort to differentiate the product is successful, the elasticity of demand for the product will be decreased. In turn, we recall

A less elastic demand is correlated with a steeper demand curve.

So we can visualize a more differentiated product as having a steeper demand curve. That's the idea behind Figure 21.3:

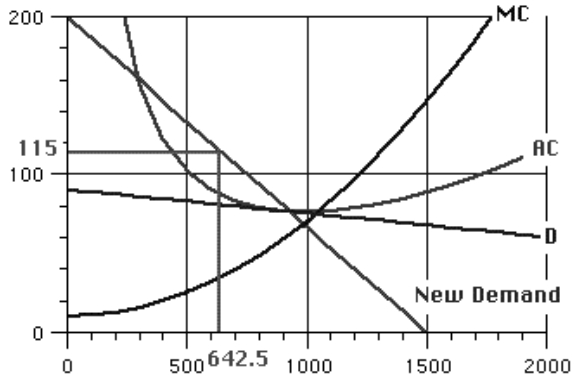


Figure 21.3

In the figure, the firm has succeeded in further differentiating its product, (starting from the long-run equilibrium we saw before) without losing any of their customers. This substitutes the lighter green New Demand curve for the old demand curve D. As a result, referring to the new marginal revenue curve (which is not shown, to keep this complicated diagram from being any more complicated) the firm will maximize profits by selling 642.5 units at $110 each for profits somewhat over $70,000 at an average cost of $90 per unit. That compares with zero profits in the long run equilibrium on demand curve D — not bad!

But, of course, it is still a short run gain. In the long run, the new product type will attract new competition, and profits will again be eroded. That's life in a competitive business. Profits in the short run beats no profits at all.

All the same, let's take a look at the long run.

21.1.2.1 Increasing Product Differentiation : Long Run

If the increasing product differentiation is successful in increasing profits in the long run, it will attract new competition until the economic profits are wiped out in the new long run equilibrium. Since all of the firms are trying to increase the differentiation of their products, the competing products are not closer substitutes, but just because of increasing numbers of firms, demand decreases and profits disappear. Here is the way the firm's new long run equilibrium would look:

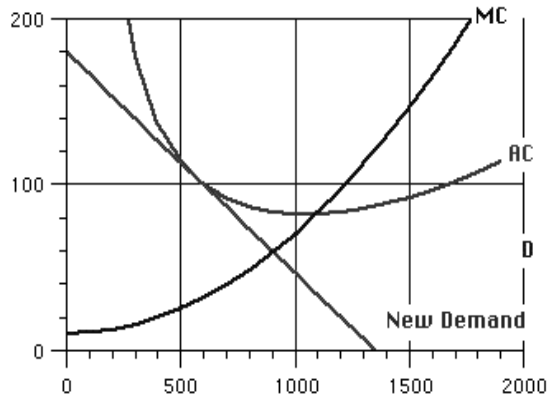


Figure 21.4

Now we see zero profits on the new demand curve, with sales of 610 units at a price of $99 per unit. Comparing the two long run equilibria, we see that in this case, nonprice competition has increased the price and cost from $76 per unit to $99 per unit, while production has been cut back from 935 units to 610 units. This doesn't look very good for monopolistic competition.

21.2 The Controversy on Monopolistic Competition

It sometimes happens in the reasonable dialog of economics that issues are raised that cannot be fully resolved to everybody's satisfaction. That has been the case with the theory of monopolistic competition. Some of the great economists of the 1920's and 1930's began the study of monopolistic competition. American economist Edward Chamberlin and British economists Joan Robinson and Abba Lerner all made important contributions to the theory, while British economist Roy Harrod and American economist George Stigler criticized it. As we have just seen, increasing differentiation of products could lead to nonprice competition with increasing prices and costs. Critics argued that a firm that was unwise enough to adopt this price-raising strategy would soon be undercut by a low-cost producer before very long; in other words, that a high-price situation could not be a long-run equilibrium. The critics also held that the monopolistic competition theory was too imprecise. While it may not be very easy to see, in this simple discussion, the p-competitive theory could be restated in very precise ways. The theory of monopolistic competition was harder to restate in precise terms. By about 1970, it seemed as if the critics had won. But during the 1980's and 1990's, monopolistic competition theory had made something of a comeback. Product differentiation and variety seemed to important to leave out of economic theory, and economists found ways to make this idea quite precise. It also seemed especially important in international trade: when both the United States and Germany import automobiles to one another, it must mean that automobiles are a differentiated product, and that the American cars are different from the German cars. The problem is that we still really do not know what that implies for efficiency. The theorists of the 1990's tend to put a great deal of stress on the tendency of nonprice competition to encourage innovation and the introduction of new products, rather than any tendency to raise prices. The discussion is still going on.

What we are pretty sure of is that product variety is important. But we have a lot to learn about how the market system creates product variety, and whether it creates too much, too little, or just about enough. Perhaps one of the readers of this book will discover the answer to that question.

21.3 Decreased Price Competition

In a P-competitive industry, in the long run, economic profits will be zero and prices determined by supply and demand. In a monopoly, prices are higher and the industry profit is at a maximum. How high will the profits and prices of an imperfectly competitive industry be? This issue has been discussed by economists for over a hundred years, and is still not entirely resolved!

Of course, the answer may depend somewhat on the structure of the industry we are looking at — remember, there is a wide range of imperfectly competitive structures. Here, we do need to distinguish between monopolistic competition and oligopoly.

For monopolistic competition, there is free entry, so (as in P-competition) economic profits will be zero in the long run. As long as there are positive economic profits, new competition will be attracted into the industry group, and therefore positive economic profits will not be stable in a monopolistically competitive industry. However, some economists believe that a monopolistically competitive industry may have high prices because nonprice competition pushes up costs.

For oligopoly, we will have to consider several hypotheses.

21.3 Oligopoly Prices

Economists have been discussing oligopoly prices and profits since the 1840's. In principle, the oligopoly's profits could never be higher than those of a monopoly — since the monopoly chooses the price that maximizes industry profits. A price higher than the monopoly price would just reduce profits (by driving away too many customers), so we wouldn't expect to see a price above the monopoly price. At the other extreme, in the long run, no industry — and in particular no oligopoly — could be stable with prices and profits lower than those of a P-competitive industry in long-run equilibrium: zero economic profits and the supply-and-demand equilibrium price. But where in this range will the oligopoly's prices settle?

There are four major hypotheses about oligopoly pricing:

1. The oligopoly firms will conspire and collaborate to charge the monopoly price and get monopoly profits.
2. The oligopoly firms will compete on price so that the price and profits will be the same as those of a P-competitive industry.
3. The oligopoly price and profits will be somewhere between the monopoly and competitive ends of the scale.
4. Oligopoly prices and profits are "indeterminate." That is, they may be anything within the range, and are unpredictable.

Unfortunately, theoretical reasoning has not been able to determine which of these hypotheses is most likely. Economists have developed models that lead to each of these four results. Instead of relying on theory, we will have to try to discover the answer by observation.

Hypothesis 3 — that oligopoly price and profits will be somewhere between the monopoly and competitive ends of the scale — seems to be the commonsense guess, and in this case commonsense seems to be right. Since the 1950's there have been many studies in the fields of "industrial organization" and "econometrics" (economic statistics) to try to resolve this question. While there is room for some controversy, the weight of the evidence seems to favor commonsense hypothesis 3.

We can go a little further. We say that an industry is more concentrated when it is more dominated by a few large firms — in other words, when its structure comes closer to the monopoly end of the scale. One rough way to measure the concentration of an industry is to compute the portion of the industry's sales revenue that is earned by the biggest four (or three or eight) firms in the industry. This is called the four-firm concentration ratio, and the bigger the four-firm concentration ratio, the more concentrated and monopoly-likethe industry is. There are other, more complicated ways to measure industry concentration that take into account all of the firms in the industry, rather than just the biggest four; but we will find industries rank about the same for concentration however we measure it.

Common-sense suggests that the more "concentrated" the oligopoly is — that is, the fewer and bigger the firms are, so that it more nearly resembles a monopoly — the nearer monopoly profits and prices the industry will come. And, once again, the evidence of observation agrees with common sense. Here, again, there is some possible controversy, but the weight of evidence is that more concentrated industries do have somewhat higher profits (counting interest payments as a cost). So we may think of a spectrum of industries, with competitive industries at one end and monopolies at the other, and with oligopolies falling at the points in between. that's not very precise, but it seems to fit the facts pretty well.

21.3 Oligopoly Pricing Strategy

All the same, those who believe in the "indeterminate" theory have a pretty good case. It seems that traditional microeconomic theory just doesn't have an answer to the question of oligopoly pricing.

In the 1930's, it began to seem that the problem was that oligopoly pricing decisions are strategic decisions. That isn't so for P-competition. P-competitive firms don't decide on a price — they just decide how much to sell at the price determined by supply and demand. Nor are monopoly pricing decisions strategic, at least in the same sense. Since the monopoly has (by definition) no rivals, it does not have to consider how its rivals will respond to its decision. But that's the essence of a strategic decision, and oligopolists' decisions are strategic in that sense: they have to consider the reactions of their rivals (one another).

The difficulty of resolving the question of strategic price decisions within traditional microeonomic theory led some economists, around 1940, to consider a new theory of strategic decisions that was being developed by the great mathematician John von Neumann. It was called Game Theory. Over the years, game theory has become increasingly important as an approach to microeconomic questions in general — not just in oligopoly pricing. Accordingly, we will now go on to a quick survey of game theory, before concluding this chapter.

21.4 Game Theory Basics

Game theory is a distinct and interdisciplinary approach to the study of strategic behavior. The disciplines most involved in game theory are mathematics, economics and the other social and behavioral sciences. Game theory (like computational theory and so many other contributions) was founded by the great mathematician John von Neumann. The first important book was The Theory of Games and Economic Behavior, *which von Neumann wrote in collaboration with the great mathematical economist, Oskar Morgenstern. Certainly Morgenstern brought ideas from neoclassical economics into the partnership, but von Neumann, too, was well aware of them and had made other contributions to neoclassical economics.*

A Scientific Metaphor

Since the work of John von Neumann, "games" have been a scientific metaphor for a much wider range of human interactions in which the outcomes depend on the interactive strategies of two or more persons, who have opposed or at best mixed motives. Among the issues discussed in game theory are

1. What does it mean to choose strategies "rationally" when outcomes depend on the strategies chosen by others and when information is incomplete?
2. In "games" that allow mutual gain (or mutual loss) is it "rational" to cooperate to realize the mutual gain (or avoid the mutual loss) or is it "rational" to act aggressively in seeking individual gain regardless of mutual gain or loss?
3. If the answers to 2) are "sometimes," in what circumstances is aggression rational and in what circumstances is cooperation rational?
4. In particular, do ongoing relationships differ from one-off encounters in this connection?
5. Can moral rules of cooperation emerge spontaneously from the interactions of rational egoists?
6. How does real human behavior correspond to "rational" behavior in these cases?
7. If it differs, in what direction? Are people more cooperative than would be "rational?" More aggressive? Both?

Thus, among the "games" studied by game theory are

- Bankruptcy
- Barbarians at the Gate
- Battle of the Networks
- Caveat Emptor
- Conscription
- Coordination
- Escape and Evasion
- Frogs Call for Mates
- Hawk versus Dove
- Mutually Assured Destruction
- Majority Rule
- Market Niche

- Mutual Defense
- Prisoner's Dilemma
- Subsidized Small Business
- Tragedy of the Commons
- Ultimatum
- Video System Coordination

(This list is extracted from an index of games discussed in Roy Gardner, *Games for Business and Economics*)

Rationality

The key link between neoclassical economics and game theory was and is rationality. Neoclassical economics is based on the assumption that human beings are absolutely rational in their economic choices. Specifically, the assumption is that each person maximizes her or his rewards — profits, incomes, or subjective benefits — in the circumstances that she or he faces. This hypothesis serves a double purpose in the study of the allocation of resources. First, it narrows the range of possibilities somewhat. Absolutely rational behavior is more predictable than irrational behavior. Second, it provides a criterion for evaluation of the efficiency of an economic system. If the system leads to a reduction in the rewards coming to some people, without producing more than compensating rewards to others (costs greater than benefits, broadly) then something is wrong. Pollution, the overexploitation of fisheries, and inadequate resources committed to research can all be examples of this.

In neoclassical economics, the rational individual faces a specific system of institutions, including property rights, money, and highly competitive markets. These are among the "circumstances" that the person takes into account in maximizing rewards. The implications of property rights, a money economy and ideally competitive markets is that the individual needs not consider her or his interactions with other individuals. She or he needs consider only his or her own situation and the "conditions of the market." But this leads to two problems. First, it limits the range of the theory. Where-ever competition is restricted (but there is no monopoly), or property rights are not fully defined, consensus neoclassical economic theory is inapplicable, and neoclassical economics has never produced a generally accepted extension of the theory to cover these cases. Decisions taken outside the money economy were also problematic for neoclassical economics.

Game theory was intended to confront just this problem: to provide a theory of economic and strategic behavior when people interact directly, rather than "through the market." In game theory, "games" have always been a metaphor for more serious interactions in human society. Game theory may be about poker and baseball, but it is not about chess, and it is about such serious interactions as market competition, arms races and environmental pollution. But game theory addresses the serious interactions using the metaphor of a game: in these serious interactions, as in games, the individual's choice is essentially a choice of a strategy, and the outcome of the interaction depends on the strategies chosen by each of the participants. On this interpretation, a study of games may indeed tell us something about serious interactions. But how much?

In neoclassical economic theory, to choose rationally is to maximize one's rewards. From one point of view, this is a problem in mathematics: choose the activity that maximizes rewards in given circumstances. Thus we may think of rational economic choices as the "solution" to a problem of mathematics. In game theory, the case is more complex, since the outcome depends not only on my own strategies and the "market conditions," but also directly on the strategies chosen by others, but we may still think of the rational choice of strategies as a mathematical problem — maximize the rewards of a group of interacting decision makers — and so we again speak of the rational outcome as the "solution" to the game.

21.4 The Prisoners' Dilemma

Recent developments in game theory, especially the award of the Nobel Memorial Prize in 1994 to three game theorists and the death of A. W. Tucker, in January, 1995, at 89, have renewed the memory of its beginnings. Although the history of game theory can be traced back earlier, the key period for the emergence of game theory was the decade of the 1940's. The publication of The Theory of Games and Economic Behavior was a particularly important step, of course. But in some ways, Tucker's invention of the Prisoners' Dilemma example was even more important. This example, which can be set out in one page, could be the most influential one page in the social sciences in the latter half of the twentieth century.

This remarkable innovation did not come out in a research paper, but in a classroom. As S. J. Hagenmayer wrote in the Philadelphia Inquirer ("Albert W. Tucker, 89, Famed Mathematician," Thursday, Feb. 2, 1995, p.. B7) " In 1950, while addressing an audience of psychologists at Stanford University, where he was a visiting professor, Mr. Tucker created the Prisoners' Dilemma to illustrate the difficulty of analyzing" certain kinds of games. "Mr. Tucker's simple explanation has since given rise to a vast body of literature in subjects as diverse as philosophy, ethics, biology, sociology, political science, economics, and, of course, game theory."

The Game

Tucker began with a little story, like this: two burglars, Bob and Al, are captured near the scene of a burglary and are given the "third degree" separately by the police. Each has to choose whether or not to confess and implicate the other. If neither man confesses, then both will serve one year on a charge of carrying a concealed weapon. If each confesses and implicates the other, both will go to prison for 10 years. However, if one burglar confesses and implicates the other, and the other burglar does not confess, the one who has collaborated with the police will go free, while the other burglar will go to prison for 20 years on the maximum charge.

The strategies in this case are: confess or don't confess. They payoffs (penalties, actually) are the sentences served. We can express all this compactly in a "payoff table" of a kind that has become pretty standard in game theory. Here is the payoff table for the Prisoners' Dilemma game:

Table 1

		Al	
		confess	don't
Bob	confess	10,10	0,20
	don't	20,0	1,1

The table is read like this: Each prisoner chooses one of the two strategies. In effect, Al chooses a column and Bob chooses a row. The two numbers in each cell tell the outcomes for the two prisoners when the corresponding pair of strategies is chosen. The number to the left of the comma tells the payoff to the person who chooses the rows (Bob) while the number to the right of the column tells the payoff to the person who chooses the columns (Al). Thus (reading down the first column) if they both confess, each gets 10 years, but if Al confesses and Bob does not, Bob gets 20 and Al goes free.

So: how to solve this game? What strategies are "rational" if both men want to minimize the time they spend in jail? Al might reason as follows: "Two things can happen: Bob can confess or Bob can keep quiet. Suppose Bob confesses. Then I get 20 years if I don't confess, 10 years if I do, so in that case it's best to confess. On the other hand, if Bob doesn't confess, and I don't either, I get a year; but in that case, if I confess I can go free. Either way, it's best if I confess. Therefore, I'll confess."

But Bob can and presumably will reason in the same way — so that they both confess and go to prison for 10 years each. Yet, if they had acted "irrationally," and kept quiet, they each could have gotten off with one year each.

Dominant Strategies

What has happened here is that the two prisoners have fallen into something called a "dominant strategy equilibrium."

DEFINITION Dominant Strategy: Let an individual player in a game evaluate separately each of the strategy combinations he may face, and, for each combination, choose from his own strategies the one that gives the best payoff. If the same strategy is chosen for each of the different combinations of strategies the player might face, that strategy is called a "dominant strategy" for that player in that game.

DEFINITION Dominant Strategy Equilibrium: If, in a game, each player has a dominant strategy, and each player plays the dominant strategy, then that combination of (dominant) strategies and the corresponding payoffs are said to constitute the dominant strategy equilibrium for that game.

In the Prisoners' Dilemma game, to confess is a dominant strategy, and when both prisoners confess, that is a dominant strategy equilibrium.

Issues With Respect to the Prisoners' Dilemma

This remarkable result — that individually rational action results in both persons being made worse off in terms of their own self-interested purposes — is what has made the wide impact in modern social science. For there are many interactions in the modern world that seem very much like that, from arms races through road congestion and pollution to the depletion of fisheries and the overexploitation of some subsurface water resources. These are all quite different interactions in detail, but are interactions in which (we suppose) individually rational action leads to inferior results for each person, and the Prisoners' Dilemma suggests something of what is going on in each of them. That is the source of its power.

Having said that, we must also admit candidly that the Prisoners' Dilemma is a very simplified and abstract — if you will, "unrealistic" — conception of many of these interactions. A number of critical issues can be raised with the Prisoners' Dilemma, and each of these issues has been the basis of a large scholarly literature:

- The Prisoners' Dilemma is a two-person game, but many of the applications of the idea are really many-person interactions.
- We have assumed that there is no communication between the two prisoners. If they could communicate and commit themselves to coordinated strategies, we would expect a quite different outcome.
- In the Prisoners' Dilemma, the two prisoners interact only once. Repetition of the interactions might lead to quite different results.
- Compelling as the reasoning is that leads to the dominant strategy equilibrium may be, it is not the only way this problem might be reasoned out. Perhaps it is not really the most rational answer after all.

We will consider some of these points in what follows.

21.5 Oligopoly prices and "Solutions" to Pricing Games

We are concerned in this chapter about oligopoly pricing and market strategy. Here is an example. It is a very simplified model of price competition. Like Augustin Cournot (writing in the 1840's) we will think of two companies that sell mineral water. Each company has a fixed cost of $5000 per period, regardless whether they sell anything or not. We will call the companies Perrier and Apollinaris, just to take two names at random.

The two companies are competing for the same market and each firm must choose a high price ($2 per bottle) or a low price ($1 per bottle). Here are the rules of the game:

1. At a price of $2, 5000 bottles can be sold for a total revenue of $10000.
2. At a price of $1, 10000 bottles can be sold for a total revenue of $10000.
3. If both companies charge the same price, they split the sales evenly between them.
4. If one company charges a higher price, the company with the lower price sells the whole amount and the company with the higher price sells nothing.
5. Payoffs are profits — revenue minus the $5000 fixed cost.

Here is the payoff table for these two companies

Table 2

		Perrier	
		price = $1	price = $2
Apollinaris	price = $1	0,0	5000,-5000
	price = $2	-5000,5000	0,0

Once again, as in the Prisoners' Dilemma, each company has a strong rationale to choose one strategy — and in this case it is a price cut. For example, Appolinaris might reason "Either Perrier will cut to $1 or it will not. If it does, then I had better cut, too — otherwise I'll lose all my customers and lose $5000. On the other hand, if Perrier doesn't cut, I'm still better off to cut, since I'll take their customers away and get a profit of $5000." Thus, the price cut is a dominant strategy.

But this is, of course, a very simplified — even unrealistic — conception of price competition. Let's look at a more complicated, perhaps more realistic pricing example:

Another Price Competition Example

Following a long tradition in economics, we will think of two companies selling "widgets" at a price of one, two, or three dollars per widget. the payoffs are profits — after allowing for costs of all kinds — and are shown in Table 5-1. The general idea behind the example is that the company that charges a lower price will get more customers and thus, within limits, more profits than the high-price competitor. (This example follows one by Warren Nutter).

Table 3

		Acme Widgets		
		price = $1	price = $2	price = $3
Wiley Widgets	price = $1	0,0	50,-10	40,-20
	price = $2	-10,50	20,20	90,10
	price = $3	-20,40	10,90	50,50

Unlike the mineral-water example (and more realistically), industry profits in this example depend on the price and thus on the strategies chosen by the rivals. Profits may add up to 100, 20, 40, or zero, depending on the strategies that the two competitors choose. We can also see fairly easily that there is no dominant strategy equilibrium. Widgeon company can reason as follows: if Acme were to choose a price of 3, then Widgeon's best price is 2, but otherwise Widgeon's best price is 1 — neither is dominant.

Nash Equilibrium

We will need another, broader concept of equilibrium if we are to do anything with this game. The concept we need is called the Nash Equilibrium, after Nobel Laureate (in economics) and mathematician John Nash. Nash, a student of Tucker's, contributed several key concepts to game theory around 1950. The Nash Equilibrium conception was one of these, and is probably the most widely used "solution concept" in game theory.

DEFINITION: Nash Equilibrium If there is a set of strategies with the property that no player can benefit by changing her strategy while the other players keep their strategies unchanged, then that set of strategies and the corresponding payoffs constitute the Nash Equilibrium.

Let's apply that definition to the widget-selling game. First, for example, we can see that the strategy pair p=3 for each player (bottom right) is not a Nash-equilibrium. From that pair, each competitor can benefit by cutting price, if the other player keeps her strategy unchanged. Or consider the bottom middle — Widgeon charges $3 but Acme charges $2. From that pair, Widgeon benefits by cutting to $1. In this way, we can eliminate any strategy pair except the upper left, at which both competitors charge $1.

We see that the Nash equilibrium in the widget-selling game is a low-price, zero-profit equilibrium. Many economists believe that result is descriptive of real, highly competitive markets — although there is, of course, a great deal about this example that is still "unrealistic."

Let's go back and take a look at that dominant-strategy equilibrium in Table 4-2. We will see that it, too, is a Nash-Equilibrium. (Check it out). Also, look again at the dominant-strategy equilibrium in the Prisoners' Dilemma. It, too, is a Nash-Equilibrium. In fact, any dominant strategy equilibrium is also a Nash Equilibrium. The Nash equilibrium is an extension of the dominant strategy equilibrium.

21.6 Cooperative Games

These three examples have the following things in common:

1. In each case, there are two decision-makers.
2. The payoffs to each decision-maker depends on both decisions.
3. Both decision-makers seek their own interests.
4. Each chooses in isolation from the other, taking the other decision as given.
5. As a result, both have relatively bad outcomes — long prison terms or zero profits.

The Prisoners' Dilemma has been influential throughout the social sciences, in the second half of the 1900's, because it offers such a vivid illustration of how this can happen: rational and self-interested decision-makers, choosing their strategies in isolation from one another, find that the strategies interact so that they both have bad outcomes.

In application to the problem of oligopoly pricing, the examples given so far seem to give strong support to the second hypothesis of oligopoly pricing, the hypothesis that

oligopoly prices will be the same as those in a P-competitive market: zero profits. But that's not really so clear.

The key assumption in these examples is assumption 4 — that each chooses in isolation from the other, taking the other decision as given. But is it really rational for them to do so? In the Prisoners' Dilemma game, the isolation is imposed by the rules of the game — the Prisoners have been isolated by the Police, and have no choice in the matter. But the oligopolists could, in principle, get together, agree on a common strategy, and share out the gains from it among themselves. They wouldn't be taking one anothers' strategies as given. Instead, they would be coordinating their strategies.

Of course, antitrust laws are designed to make such a price-fixing agreement illegal. But we haven't always had antitrust laws — they were enacted because many people believed that businessmen were collaborating to fix high prices. And even now, there may be ways to get around the law.

When the decision-makers in a "game" get together, agree on a common strategy, and share out the gains from it among themselves, the agreement they come to is called a "cooperative solution" to the game. The examples we have looked at so far are "noncooperative solutions."

It appears that we cannot rule out the possibility of a cooperative solution to the oligopoly pricing game, so we need to look a bit at the cooperative alternative in game theory.

21.6.1 A Basic Cooperative Game

It's not very hard to find an example of a cooperative solution to game. In fact, we have been talking about cooperative solutions all through this course. Buying and selling is a cooperative game in which the buyer and the seller are the two "players" and the price they agree upon is their common strategy. Here's a numerical example to illustrate what I mean.

We suppose that Joey has a bicycle. Joey would rather have a game machine than a bicycle, and he could buy a game machine for $80, but Joey doesn't have any money. We express this by saying that Joey values his bicycle at $80. Mikey has $100 and no bicycle, and would rather have a bicycle than anything else he can buy for $100. We express this by saying that Mikey values a bicycle at $100.

The strategies available to Joey and Mikey are to give or to keep. That is, Joey can give his bicycle to Mikey or keep it, and Mikey can give some of this money to Joey or keep it all. It is suggested that Mikey give Joey $90 and that Joey give Mikey the bicycle. This is what we call "exchange." Here are the payoffs:

Table 4

		Joey	
		give	keep
Mikey	give	110,90	10,170
	keep	190,10	100,80

EXPLANATION: At the upper left, Mikey has a bicycle he values at $100, plus $10 extra, while Joey has a game machine he values at $80, plus an extra $10. At the lower left, Mikey has the bicycle he values at $100, plus $100 extra. At the upper left, Joey has a game machine and a bike, each of which he values at $80, plus $10 extra, and Mikey is left with only $10. At the lower right, they simply have what they begin with — Mikey $100 and Joey a bike.

If we think of this as a noncooperative game, it is much like a Prisoners' Dilemma. To keep is a dominant strategy and keep, keep is a dominant strategy equilibrium. However, give, give makes both better off. Being children, they may distrust one another and fail to make the exchange that will make them better off. But market societies have a range of institutions that allow adults to commit themselves to mutually beneficial transactions. Thus, we would expect a cooperative solution, and we suspect that it would be the one in the upper left.

Thus, we can see that cooperative solutions are not uncommon in a market society. On the contrary! They are the essence of a market system! Thus, we cannot rule out the possibility that oligopolists will arrive at a cooperative solution, and charge a monopoly price and get monopoly profits.

21.7 The Oligopoly Problem

It seems that game theory doesn't solve the oligopoly problem after all. There are at least two kinds of solutions to the problem of oligopoly pricing — cooperative and noncooperative — and we can't rule either one of them out.

Actually, it's a bit worse than that. In each of the two categories, there is actually more than one sort of solution, depending on how we approach the problem! So game theory really isn't a "solution" to the problem of oligopoly pricing at all!

That had become pretty clear to economists by the 1960's, and many economists lost interest in game theory. But despite its failure on this specific point, game theory has proved to be a powerful tool of economic thinking, so that it has become more influential since the 1960's, culminating in the Nobel Prize for three game theorists (including John Nash, who invented the Nash-Equilibrium) in 1994.

And it isn't just simply a failure in the analysis of oligopoly prices — not really. In a comic strip a few years ago, a schoolboy was discussing his grades with his guardian. The schoolboy said that he had gotten an "A" on an algebra test. The guardian said, "Well, then, you must understand algebra." "No," the schoolboy said, "but I'm confused at a much higher level." We economists can say that about oligopoly pricing.

Sometimes it's important to be confused at a higher level. We now understand not only that oligopoly pricing is a hard problem, but something of why it is a hard problem. The pricing examples we have seen here give some insight about the reason why price competition — when it does occur — is so powerful in bringing prices down to the lowest stable level. And we can apply the same methods to a range of other problems, both related to imperfect competition and in other fields of economics.

Let's try applying it to advertising, for example.

21.7.1 Advertising

Some economists and others believe that persuasive advertising may be largely wasteful. Here is an example that illustrates what they have in mind. We consider two whiskey manufacturers, one selling Black George brand, and the other selling Wild Chicken brand whiskey. If they advertise, they can sell more whiskey, and the larger market will give them both more revenues — but the cost of advertising more than offsets the increased revenues, so that profits are slightly less. On the other hand, if only one advertises, he takes most of the customers and gets a large profit. Here is the payoff table:

Table 5

		Black George	
		advertise	don't
Wild Chicken	advertise	40,40	110,10
	don't	10,110	50,50

Once again, we see an outcome like the Prisoners' Dilemma: if each seller takes the other seller's decision as given, then they both advertise, and their profits are lower as a result. Again, as in the Prisoners' Dilemma example, we can contrast a cooperative solution with this noncooperative solution. A cooperative solution to the advertising game would be one in which the whiskey-sellers coordinate their strategy for mutual benefit.

We get these Prisoners'-Dilemma-like results because of the way the payoffs are interrelated. Other payoff assumptions might lead to a quite different analysis. We should stress that this is an hypothesis, not a fact. Some economists believe that advertising is often wasteful, and explain their opinion by this Prisoners'-Dilemma-like sort of reasoning. But, so far as evidence is concerned, the jury is still out. We don't know whether advertising is a "Prisoners' Dilemma" or not.

The value of game theory, in a case such as this, is not that it answers the questions but that it enables us to ask them more precisely and clearly. That's helpful, and is a step on the way to getting the answer.

21.8 Summary

This chapter has undertaken to survey the industries that don't seem to fit into either the "P-competitive" or "monopoly" category: oligopolies and monopolistically competitive industries, or, taking both together, imperfectly competitive industries. These industries deviate from one or another of the four characteristics of P-competition, but also involve some competition among two or more firms selling close substitutes, if not the same products.

The two major implications of imperfect competition, by comparison with P-competition, are

Increased Nonprice Competition

Unlike price competition, nonprice competition may be costly and may or may not make consumers better off all in all. Nonprice competition is a mixed bag, but most economists are persuaded that, on the whole, more competition among sellers is better than less.

Decreased Price Competition

Price competition favors customers and promotes efficiency, so decreased price competition is clearly a bad aspect of imperfect competition. But how much will price competition be reduced? Turning to game theory, we learn that there is no one "rational" answer to the choice of price strategies in oligopoly, so the question cannot be given a precise answer. But there is some evidence that the intensity of price competition increases as the number of firms gets larger and their size gets smaller, relative to the industry.

These problems mean that the decisions of imperfectly competitive firms are strategic in a sense that monopoly decisions and the decisions of P-competitive firms are not. Accordingly, we have followed modern economics (and mathematics and other social sciences) in digressing a bit on an important modern theory of strategic choices called "game theory." Here, too, we find not clear answers but a range of possibilities: the "solutions" to or equilibria in "games" may be cooperative or noncooperative. The noncooperative games sometimes lead to results that nobody wants — such as low prices and zero economic profits in an oligopoly — but we also have to consider the possibility of a cooperative solution with monopoly prices and profits.

Imperfect competition remains a controversial area in economics. Some economists would argue that imperfect competition is the rule, rather than the exception, and they conclude that the "Fundamental Principle of Microeconomics" — however fundamental for theory — has little application in the real world. Other economists would argue that, even if imperfect competition is pretty wide-spread, the deviations from supply-demand pricing in the real world are small, minor and temporary, so that "supply and demand" remains our best guide to prices and outputs in our market economy, with a very few obvious exceptions. This latter view has become more widespread and influential over the past 30 years, and that change has contributed to the political climate that led to deregulation in the years since 1977.

LESSON 22:
THE ROLE OF GOVERNMENT IN THE MARKET ECONOMY

22.1 The Rationale for Regulation

1. Efficiency
2. Regulation
3. Equity

Economic Considerations

1. Market Failure
2. **Externalities:** differences between private and social costs or benefits. Failure by incentive. This is a situation where social costs and benefits differ considerably from the private costs and values of producers and consumers.

   a. negative: a cost of producing, marketing, or consuming a product that is not borne by the product's producers or consumers. Environmental pollution is the best known example.

   b. Positive: a benefit of production, marketing, or consumption that is not reflected in the product pricing structure and, hence does not accrue to the products producers or consumers. The rapid acceptance of technology in some industries which has led to reduced costs for suppliers and hence consumers.
3. **Political Considerations -** play an important role in the design of regulatory policies:

   a. preservation of consumer choice

   b. limit concentration of economic and political power.

   c. Important political considerations lead to the argument compelling power for government to be in the marketplace

22.2 Regulatory Response to Incentive Failures

Operating Right Grants

1. Example: FCC
2. may be effective but it is imprecise
3. The cost of inefficiency

Patent Grants

1. Government grants exclusive right to produce, use or sell an invention or innovation for a limited period of time (17 years in U.S.).
2. a legal monopoly - goal is to achieve the benefits of both monopoly and competition in the field of research and development.
3. Firms cannot use patents to unfairly monopolize or otherwise limit competition (e.g., Xerox).

Subsidies

1. provided to private business firms
2. indirect - construction grants which benefit trucking
3. direct - agricultural payment-in-kind programs; special tax treatments, and government-provided low-cost financing.

Tax Policies

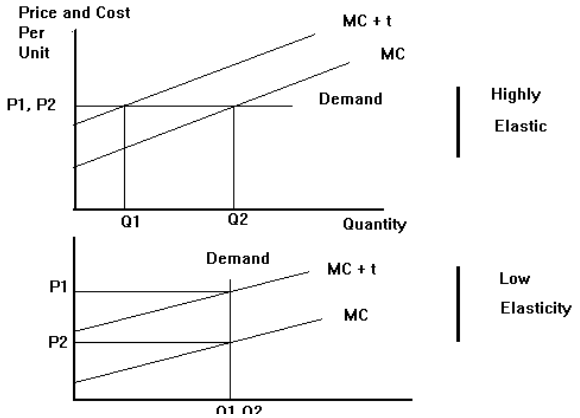
1. includes both regular tax payments and fines or penalties that may be assessed intermittently.
2. Differentiate from subsidy and grant programs - subsidies infer the right to do something (pollute); taxes as a penalty assert societies right.

Operating Controls

1. Standards - are designed to limit undesirable behavior by compelling certain actions while prohibiting others. Limits on auto emissions as an example; fuel efficiency standards is another. In another sector — wage and price controls to curb inflation.
2. Relies on non-monetary incentives

22.3 Who Pays for Regulation?

1. Depends on the elasticity of demand for the final products of affected firms
   - **a.** tax incidence vs. tax burden. The point of tax collection versus the issue of who really pays. For example, the polluting foundry may pay the pollution tax (incidence) but the cost may be passed onto consumers (burden).
   - **b.** Highly elastic product demand places the burden of regulation-induced cost increases on producers. Production must be cut from Q1 to Q2
   - **c.** Low elasticity of product demand allows producers to raise prices from P1 to P2. Consumers bear the burden of regulation-induced cost increases.



2. Problem of Underproduction

Utility Price Regulation

1. unrestricted monopoly
2. reduced dollar profit
3. lower return on investment

Problems in the Utility Industry

1. Pricing problems
2. Output level problems
3. Inefficiency
4. Investment level
5. Regulatory lag and political influence
6. Cost of regulation

22.4 Antitrust Policy

Designed to promote competition and prevent unwarranted monopoly. The laws seek to improve economic efficiency by enhancing consumer sovereignty and the impartiality or resource allocation while limiting concentrations in both economic and political power.

Sherman Act of 1890

1. forbade contracts, combinations, or conspiracies in restraint of trade.
2. too vague.

Clayton Act of 1914

1. addressed problems of mergers, interlocking directorates, price discrimination, and tying contracts.
2. focused on price discrimination

22.5 The Regulatory Mess

1. Cost of regulation
2. The size-efficiency problem
3. Capture Theory: an economic theory which suggests that industry seeks regulation to limit competition and to obtain government subsidies.
4. The deregulation movement

22.6 Price Controls

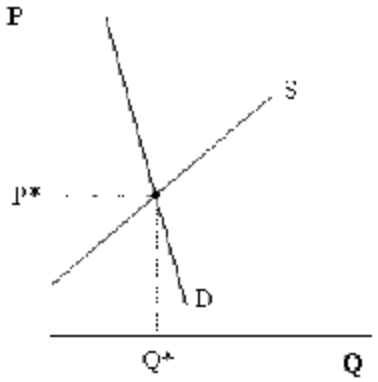
When markets are free to choose a price and quantity, the result is an equilibrium. Prices become the mechanism that directs demanders and suppliers toward this point. If, in the opinion of government decisionmakers, the resulting equilibrium price is too high or too low, then the government may intervene in the market by imposing price controls. Those controls are then enacted to prevent prices from falling into unacceptable ranges. Of course, normative concerns like fairness are arbitrary and represent something that government must define. That is, government must decide exactly how high is too high or how far a price can fall before it's too low.

Price controls are important because they can alter the behavior within a market. Prices act as an incentive to buy and sell. For example, if government deems a price too high, then it is possible that the government may restrict the price from rising above a certain point by placing a price ceiling on the good sold in this market. The price ceiling serves as a price maximum.

Similarly, if the price was too low, then the government could impose a price floor (price minimum).

The point here is that if prices cannot reach what would be the equilibrium, then a gap will emerge between the quantity demanded and quantity supplied. In turn, a lack of equality between the quantity demanded and supplied causes other events to occur. Although these secondary effects are indirectly related to the price control, they may be of a sufficient magnitude and therefore deserving of our attention.

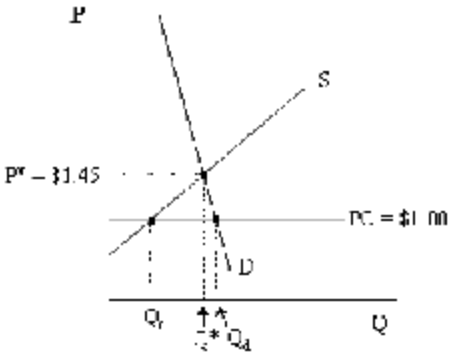
Let's consider an example where a price ceiling is imposed on a specific market. Our market will be the market for regular unleaded gasoline (87 octane) and we'll assume that this market can be described by the demand and supply model in the following graph (note the steep slope of demand implies something about how people respond to changes in the price of regular unleaded gas):



In this setting, Q* units of regular unleaded gasoline, which we'll assume is 9.5 units (we'll assume that each unit is one hundred thousand gallons), are sold at a price of P*, which we'll also assume is $1.45 per gallon.

Suppose consumers lobby the government, asking for intervention in this market. Their claim is that the high price of regular unleaded gasoline hurts middle and lower income individuals more than higher income drivers because rich people use cars that require a higher octane that what's provided in regular unleaded. Let's assume that these people are successful and that the government places a price ceiling of $1 on regular unleaded gas (but no ceiling on the more premium brands). Demanders and suppliers can still agree to transact at any price they want, but only as long as that price is below the ceiling.

We can illustrate this ceiling on the previous graph as follows:



Because the price cannot legally rise above $1, demanders and suppliers must interact at this new price (not the price that would otherwise be the equilibrium price). We determine how much suppliers would provide at this price by looking at the quantity supplied (Q_s) at the $1 price. Let's suppose that the resulting quantity supplied is 5.25 units (i.e. 525,000 gallons of regular unleaded gasoline). Similarly, if we consider the quantity demanded (Q_d) at $1, then we observe 1 million gallons being demanded. Of course, demanders will not be able to get this one million gallons, because only 525,000 gallons are being supplied. Therefore, the price ceiling leads to a shortage of 475,000 gallons. The change in quantity demanded and supplied, and resulting shortage, is a direct effect of the price ceiling.

Note also that there are fewer gallons being exchanged in this market. That is, not only does the price ceiling cause a 475,000 gallon shortage to occur, but there are also 425,000 fewer gallons of regular unleaded being sold. Many of the drivers who formerly purchased regular unleaded gas will have to turn to an alternative fuel (or fuel source) because they cannot buy the gasoline they used to buy before the price ceiling.

The price ceiling may also have some indirect effects on this market. Whereas the price ceiling may directly lead to a shortage, the shortage may (in turn) have an effect on the behavior of demanders and suppliers.

Several outcomes are possible, but let's consider the short run possibilities first. We'll define the short run as a period of time where demanders and suppliers face at least some adjustment constraints on their behavior. For instance, it's difficult for suppliers to just leave the market in a short time because existing firms not only must first liquidate their assets, pay off creditors, etc., but also might want to wait and see if the market changes after a while (e.g. perhaps the government might change its mind about the price ceiling). Consequently, we think of exiting the market as something firms would do in the long run rather than the short run.

Some of the possible short run adjustments might include changes in the purchasing patterns of demanders, or selling patterns of suppliers. Demanders need gas because of the shortage, so perhaps they'll turn to buying one of the more expensive, higher octane gasolines. If demanders cannot afford these higher priced brands, then they may also drive further to buy regular unleaded gas in markets where there is no shortage (or price ceiling). It's also possible that suppliers may decide to impose certain fees on the sale of regular unleaded. For example, a 25 cents per gallon "pumping charge" might be imposed on every gallon sold during busy times of the day.

In the long run, suppliers might adjust by shifting toward supplying more of these higher octane gasolines. Suppliers might expand their ability to supply higher octane gas by increasing their ability to store higher octane gas (and buying smaller storage tanks for regular unleaded gas). It is also possible that some gasoline suppliers will not sell gasoline any more. Some may leave the market, some may shift over to providing mechanic services instead. Overall, there will be a decrease in the number of suppliers of regular unleaded gas. When the number of suppliers decreases, the supply curve shifts left and the shortage only grows.

22.7 Price Ceilings – A Welfare Economics

One way in which the central authority may regulate an industry is by controlling the market price. For example, one type of price control is a price ceiling (where the government sets an upper bound on the market price). Price ceilings set below the equilibrium price cause shortages. With a shortage, it is necessary to determine how the product will be allocated. This handout illustrates that the size of deadweight loss can vary with the allocation rule.

Assume a that there's a perfectly competitive market, where consumers buy (at most) one unit. The demand and supply curves are:

Demand: $D(P) = 100 - P$ Supply: $S(P) = 10 + .25P$

The equilibrium price and quantity become: $P^* = \$80$, $Q^* = 20$. Assume that a price ceiling (Pc) is set at $40, causing a shortage of 50. What are the welfare effects of using different allocation rules? We'll consider three.

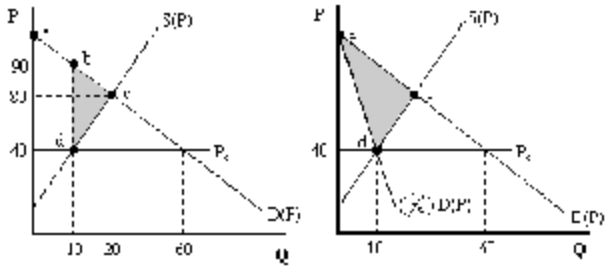


Figure 1 Figure 2

Rule 1. Product is Allocated According to People's Willingness to pay for it

In this regime, we allow consumers to purchase the good on the basis of their willingness to pay. By lining consumers up, according to their willingness to pay, we have the market demand curve illustrated in Figure 1. Consumers with the highest willingness to pay are allowed to purchase the good first. Here, only the consumers along line segment "ab" can buy the good. Comparing the surplus achieved under competition with the surplus achieved under Rule 1, we see a decrease. This decrease in surplus corresponds with the (lost) surplus derived from consumers along line segment "bc". Consequently, deadweight loss becomes the shaded area "bcd".

Rule 2. Product is Allocated by the Lottery Method

In this regime, the government allocates the good randomly by giving consumers the right to purchase the good if their name is selected from a lottery. Recall that each consumer buys (at most) one unit of the good. At $P_c = \$40$, 60 units are demanded and only 10 supplied. Therefore, each consumer has a 1/6 chance of purchasing a unit. Since only 1/6 of the $D(P_c)$ consumers can purchase the good, our allocation rule implies that the resulting consumer surplus will be 1/6th of the total area below the demand curve and above the price of $40. In Figure 2, by adding the allocation rule [(1/6)D(P)], we can see

this area of consumer surplus as lying below line segment "ad" and above the ceiling price.

Consequently, the deadweight loss resulting from Rule 2 corresponds with the shaded area "acd". This loss arises because of two factors: (a) the decrease in output associated with the price ceiling (equal to area "bcd" in Figure 1); and (b) the loss in welfare associated with improper allocation (equal to the difference between areas "acd" and "bcd"). Regarding the second factor, the allocation is improper because a limited number of units are allocated to consumers who place a lower value on the good (i.e. those consumers associated with lower points on the demand curve). Those who want the good the most, in terms of how much they're willing to pay for the good, won't necessarily get the good.

If a black market developed and consumers were allowed to resell the good, the deadweight loss in Figure 2 would approach the size of the deadweight loss in Figure 1. When a low value consumer resells the product to a higher value consumer, the low value consumer derives "rents" from the sale. In the end, if only the higher value consumers hold the good (i.e. those consumers along line segment "ab" in Figure 1), then the resulting deadweight loss is the same as that of the previous example (shown by Figure 1).

Rule 3. Product is Allocated by the Good's Suppliers

When suppliers allocate the good, consumers compete against one another to acquire the product. This competition may occur in a variety of ways, but we expect that the high value consumers will be willing to pay extra for the good - up to the difference between their respective reservation price and the ceiling price. There are several ways in which they may do so. One is to wait in line. In this case, consumers will consider their opportunity cost of time and how long they must wait in line to purchase the good. If the opportunity cost of waiting in line (for each consumer) corresponds with lost wages of $10 per hour, and consumer A's reservation price is $80, then consumer A will wait in line no more than 4 hours to purchase the product. Since time is a real resource, giving up time to wait in line represents an additional welfare loss to society. For example, time spent waiting in line could have been spent providing labor in the production of goods and services.

It is also possible that consumer A might bribe the supplier, in order to acquire the product. No bribe will be greater than the difference between a consumer's reservation price and ceiling price of $40. Since monetary bribes involve exchanging a pecuniary resource, there is no additional welfare loss to society other than that associated with area bcd in Figure 1. The reason is that a monetary bribe is simply a financial transfer from consumer to supplier.

Notes

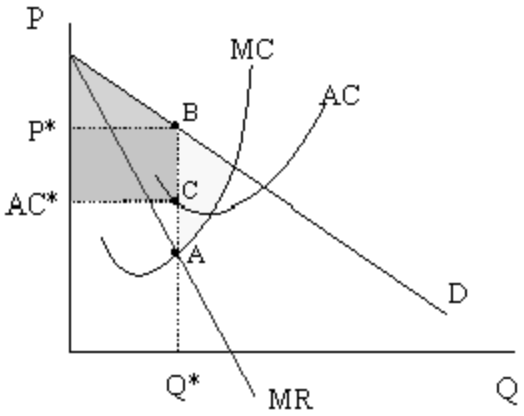
LESSON 23:
PRICING STRATEGY

23.1 Pricing Strategies for the Monopolist

When firms can set their own price, then there are a variety of strategies that each firm may follow. Naturally, if a firm is profit maximizing, then the strategy chosen will be that which brings in the most (economic) profit. Three of these approaches, linear pricing, price discrimination and the two part tariff, are discussed below.

1. **One price for all units sold.** In economics circles, this approach is referred to as linear pricing and is the most commonly discussed approach in the microeconomics course. A firm will adjust the quantity of output it supplies until finding a point where the marginal revenue associated with selling that quantity is equal to the marginal cost of producing that quantity. That is, the firm will produce where MR = MC.

When following this approach, the firm will then charge a specific price that applies to each unit sold. If we consider this in the context of a monopolist choosing an output level, then we have the graph below.



The monopolist finds where MR = MC, which occurs at pt A. Directly below pt A, we see the monopolist's output (Q*). Going up from pt A to pt B, and then left to the Price axis, we get the monopolist's price (P*). Profit (blue area) is the difference between the price and average cost associated with supplying Q* units of this good. AC* represents the monopolist's average cost of supplying Q* units, which comes from where the dotted line (up from Q*) hits the AC curve at pt C.

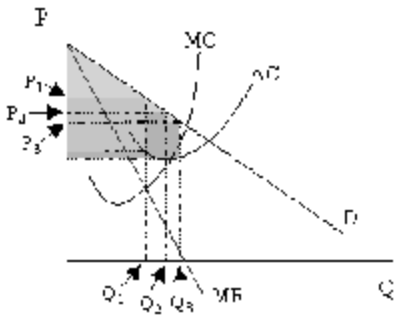
Just as with profits, we also see that there is some consumer surplus (orange area above the price and below demand) and deadweight loss (yellow area to the right of Q*, between demand and marginal cost). The deadweight loss arises because the firm produces an inefficient amount of output. That is, the firm is not producing where P = MC, which is considered the efficient amount of output.

Here, the firm charges one price (P*) for all units sold. If two consumers purchase the same item, but at a different price, then the difference in price corresponds with what must be the different cost of supplying these two consumers. For example, if consumer A buys one unit of this good at P*, and consumer B buys one unit for P** (where P* > P**), then it must be true that the price of B's unit arose from a shift in the MR or MC curve. It's also possible that there was a transportation cost (e.g. shipping and handling) tacked onto A's price that wasn't reflected in the graph.

2. **Different prices for different consumers.** This approach is referred to as price discrimination, and, unlike the first approach, corresponds with differences in what consumers are willing to pay, not as the result of changes in demand or differences in supply cost.

For example, entertainment providers often charge different prices to students, the general public, seniors, etc., even though the cost of supplying that entertainment to each consumer is identical. Profit maximizing movie theatres, carnivals, etc., realize that some consumers are willing to pay more, or are able to pay more, for entertainment than other consumers. The problem is that demand-related consumer information is not readily available to suppliers. Ticket booth operators at the movies could try to ask consumers about their willingness to pay for tickets, but they aren't likely to get much information because consumers have no incentive to tell the truth about their preferences. As a result, movie theatres must consider other variables, variables that are easily observed like age, income or time of purchase, which are likely to be correlated with willingness to pay.

Assume that there is only one form of entertainment in town, and that the movie theatre is a monopoly provider of this entertainment. Once the movie theatre manager determines what variables are correlated with willingness to pay, the manager will charge different prices to each of the different identifiable groups. The result is something like what the graph illustrates below.



Rather than charge just P_1 (which corresponds with P* in the one price for all units sold approach), the theatre owner can charge prices P_2 and P_3 to seniors and youth (under 18) respectively. Both of these prices will attract additional sales, which would never have occurred when the theatre charged P*. Because these groups can buy tickets at a price that is above AC, the theatre can exact additional profits (given as the pink and greenish areas to the right of the original blue profit box corresponding with P_1).

Note that the market (if this theatre is a monopolist) generates additional consumer surplus (orange areas above each price) and dramatically decreases the deadweight loss (still dark green, to the right of Q_3 and between demand and MC) that occurred when only one price was charged for all units sold. Deadweight loss became smaller because more units were sold. Therefore, this approach accomplishes three important objectives: the firm receives additional profits, consumer surplus has increased, and the market output is closer to what we (society) consider an efficient level of output (i.e. where P = MC for the last unit sold).

Only one problem remains for price discriminating firms like this theatre owner. If general admission consumers are able to buy tickets at youth prices, then the whole pricing strategy could fall apart. That is, the theatre owner could possibly become worse off by attempting price discrimination instead of charging one price for all units sold. To make price discrimination successful then, the movie theatre must prevent the resale of movie tickets between the different consumer groups. One way to do this might be to color-code the tickets, so that adults couldn't have their kids buy tickets for the whole family to get adults and kids in at youth prices.

Price Discrimination, A Numerical Example

Assume that a local concert provider has a monopoly over the provision of heavy metal rock concerts. The firm has estimated that its market demand curve can be drawn from the following equation (where P = price and Q = quantity demanded):

$P = 130 – 2Q$

Marginal revenue (MR), in this case, would be $MR = 130 - 4Q$. The firm's average and marginal costs are constant, in that the AC and MC equations are both always equal to $40. These equations appear as follows:

$AC = 10$

$MC = 10$

If the firm was to charge one price for every ticket it sells, the demand, MR, MC and AC curves inform us that the firm will sell 30 tickets at a price of $70 per unit and make economics profits of $1800 (you may want to verify this on your own).

Let's assume that the firm has enough information on its market to utilize a price discrimination pricing strategy. To be a price discriminating monopolist, this firm must do two things:

1. Separate consumers into different groups, based on differences in their maximum willingness to pay for the firm's product
2. Prevent the resale of the firm's product between these different groups

Let's further assume then, that the firm has determined that younger consumers (under 50) are willing to pay up to $80 per ticket for an upcoming heavy metal band concert and that older consumers (50 or older) are willing to pay up to $30 per ticket to *rock on* at this concert. If the firm uses price discrimination, based on age differences, how do we measure the effects of this pricing strategy on profits?

If the monopolist sets a price of $80, then we calculate the number sold by plugging P = 80 into the market demand equation and solving for Q.

$P = 130 – 2Q$

$80 = 130 – 2Q$

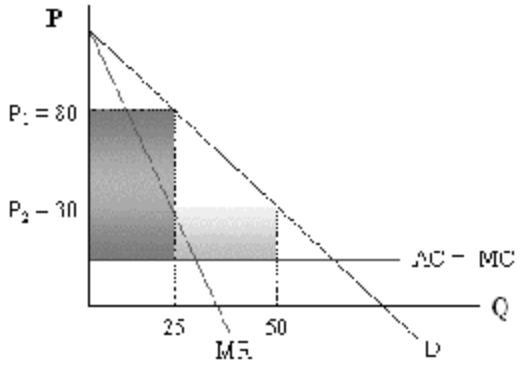
$Q = 25$

If the firm sets a price of $30, then we can similarly calculate the number that would be sold at P = 30.

$30 = 130 – 2Q$

$Q = 50$

Of course, if there are two prices charged, we want to consider the additional sales that occur because of the lower price - which is the difference between the two quantities (25 and 50). Therefore, as the graph shows below, the $80 price will result in 25 tickets being sold to the younger group whereas charging the older consumers a price of $30 will cause the overall sales to increase to 50 tickets (i.e. an 25 additional tickets are sold).



Profits (p) are measured as the net revenue generated from the sales of this good at the two prices given above (where the "1" subscript denotes the younger group, and the "2" subscript corresponds with the older group). The pink area corresponds with the profits derived from sales to group 1, and the green area corresponds with the profits derived from sales to group 2.

p = "Pink Area" + "Green Area"

$p = (P_1 – AC_1)Q_1 + (P_2 – AC_2)(Q_2 - Q_1)$

$p = (80 – 10)25 + (30 – 10)(50 - 25)$

$p = 2250$

The firm can obviously make more profits now than what would have been attained in the pre-price discrimination system.

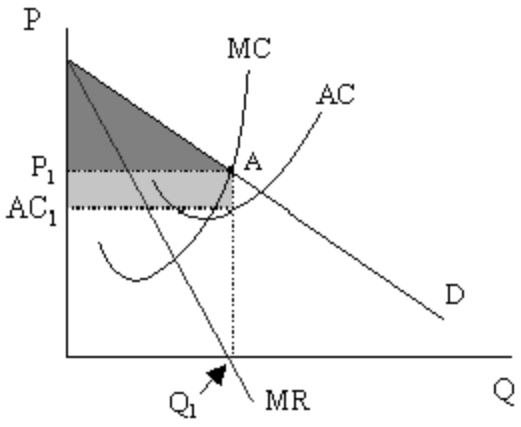
Note that we could have chosen a different pair of prices to work with, and that our profits from price discrimination could go up or down, depending on which prices we chose. For example, prices of $90 and $10 would yield total profits of $1600; whereas prices of $75 and $30 would yield total profits of $2287.50.

3. **Set up a "club" and charge one price for all units.** This approach is referred to as the two-part tariff. Consumers first pay a flat (fixed) fee which effectively gives them the "right" to buy as many units of a good as they want at a given price. The flat fee has many names. Sometimes it's called a membership fee, at other times it's a hookup fee, and in some situations the fee is called an entry fee. In each case, however, all consumers will pay the same fee regardless of whether they end up buying anything (thereafter) or not. Once the fee is paid and the consumer is effectively *inside the store*, each consumer can purchase varying amounts of whatever is being sold.

This pricing strategy has many examples. The State Fair charges an entry fee, and then specific prices for each ride. Some discount stores (e.g. Sam's Club) ask shoppers to first pay a membership fee, before going inside the store to purchase various products at their discounted prices.

If the individual prices of each good exceed the good's average cost, the firm will make profit. Additional profits can also be brought in, however, with the fixed fee. The amount of fixed fee revenue collected depends on the available consumer surplus. As the firm sets a lower price for the units supplied, the consumer surplus becomes larger. Larger consumer surplus makes it possible for the firm to collect more revenue by either charging a larger membership fee or by adding members.

How might this approach work if the theatre manager decided to use this pricing approach instead of either of the previous two approaches? Again, we can see this with the use of another graph.



The manager could charge an entry fee, something that allows movie-goers to enter the theatre and buy their tickets. Because consumers are willing to pay an entry fee that is no greater than their potential consumer surplus, the movie theatre realizes that it is possible to collect up to the amount of the aggregate consumer surplus from this market (the dark green area above the price P_1). The lower the price, the greater the market consumer surplus and the higher the (potential) entry fee revenues. Therefore, we should expect the movie theatre to set low ticket prices as an inducement to get consumers to pay the entry fee. While lower prices would not bring monopoly-like profits, the entry fee revenues could potentially raise profits above those attained by the other monopoly's one price for all units sold strategy.

In the graph, the movie theatre sets a price that is equal to marginal cost (to do so requires setting the price where MC cross the demand curve at pt. A). This price brings forth profits that are represented by the blue profit box. If the entry fee is set high enough, the movie theatre can add the entire green consumer surplus area to total profits also.

Similar to the price discrimination approach, we see that the two-part tariff pricing strategy may lead to a big reduction in deadweight loss and concurrent increase in output. Consequently, we can see that the monopolist may not be as inefficient as first believed with the one price for all units sold approach. We should concede as well, however, that consumer surplus may potentially disappear with the deadweight loss if the firm can successfully set prices that are equal to each consumer's maximum willingness to pay (in the case of price discrimination) or set a fixed fee that allows the firm to secure all of each consumer's consumer surplus.

Two-Part Tariff, A Numerical Example

Suppose the campus bookstore has a monopoly over the supply of textbooks. The bookstore hires someone to estimate their (market) demand curve and receives the following information (where P = price and Q = quantity demanded):

P = 100 - 1.5Q

Marginal revenue (MR), in this case, would be MR = 100 - 3Q. The firm buys all of its books from a book publisher at $40 per book, making the bookstore's average and marginal cost (AC and MC, respectively) always equal to $40. The bookstore's AC and MC equations would be:

AC = 40

MC = 40

If the firm was to charge one price for book it sells, the demand, MR, MC and AC curves help us in determining that the bookstore will sell 20 books at a price of $70 per book. Economic profits would be $600 (you may want to verify this on your own).

Let's assume that the bookstore owner hears about two-part tariffs and would like to implement this pricing strategy. Students are asked to pay a cover charge, just to enter the store, and may then buy all the textbooks they want at some pre-determined price.

The lower the textbook price, the more consumers save. More specifically, the lower the price, the greater the consumer surplus. The bookstore knows that the two-part tariff pricing approach allows them to recover any lost profits (from lower prices) by raising the cover charge, so the firm will adjust the cover charge and textbook price to a point where profits are as high as possible.

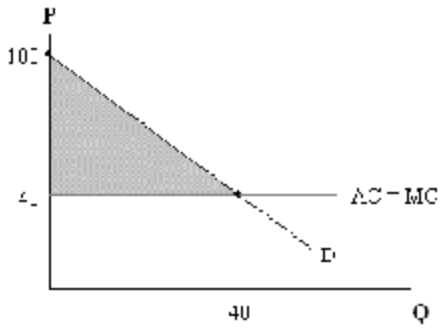
Given the demand for economics textbooks, the bookstore decides on a price of $40 per book. That is, the bookstore decides to sell the textbooks at cost. To determine how many books are sold at this price, we take the demand curve and plug the price of $40 in for P before solving for the quantity sold (Q).

P = 100 - 1.5Q

40 = 100 - 1.5Q

Q = 40

In the absence of any cover charge, this would allow consumers to obtain (overall) consumer surplus of $1200. This is illustrated in the graph below, where the blue area represents consumer surplus.



Because consumer surplus is the area of the triangle bordered by the demand curve, price and vertical axis, we can calculate the area of this triangle as:

CS = 0.5("base" x "height")

CS = 0.5(Q* x ["y-intercept" - P*])

CS = 0.5(40 x [100 – 40])

CS = 1200

Because we don't have enough specific information about the various consumers making up this market, including how many consumers there will be, we can only make guesses at the cover charge.

For example, suppose there are 30 students who think they can save money by paying the cover charge to enter the bookstore. If the bookstore sets the cover charge at $25 per person and there are 30 students willing to enter, then the firm can earn total profits (i.e. profits from booksales + revenues from the cover charge) of $750.

23.2 Price Discrimination: A Summary

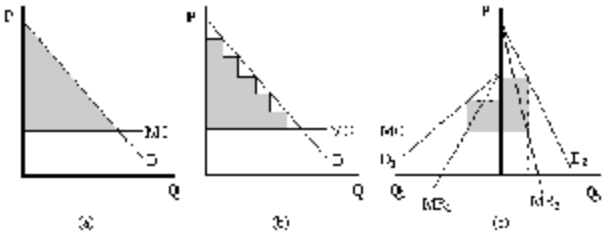
Discussions of firm pricing behavior often assume that a firm will charge the same price to all consumers. In reality, we find examples like theatres who charge different prices to students, the general public, seniors, etc. - even though the cost of supplying "entertainment" to each of these consumer types is the same. This corresponds with a practice known as price discrimination.

What is price discrimination? The standard discussion of price discrimination centers on the following brief definition: "Price discrimination is the sale (or purchase) of different units of a good or service at price differentials not directly corresponding to differences in supply cost." (Scherer and Ross, 1990)

How do firms conduct price discrimination? Price discrimination is founded on a firm's ability to distinguish amongst buyers, based on their varying demand characteristics for a particular product. The more a firm is able to do so, the more *perfect* the degree of price discrimination.

Three conditions must exist to enable a firm to profitably price discriminate: (a) the firm must have market power, (b) the firm must be able to distinguish among buyers on the basis of their demand-related characteristics (e.g. demand elasticity or reservation price), and (c) the firm must be able to constrain resale between buyers with high and low reservation prices (or demand elasticities).

There are three degrees of price discrimination (illustrated below): (a) **first degree** (perfect), where firms charge each consumer their reservation price for the good; (b) **second degree**, where firms charge "blocks" of consumers their reservation price for the good; and (c) **third degree**, where firms divide consumers into two or more submarkets, each with its own demand curve, and independently maximize profits in each submarket.



What types of price discrimination are found in practice? There are three main classes, each with differing intra-type examples: *personal discrimination*, which is based on differences among individual consumers; *group discrimination*, where intergroup differences are the distinguishing factor; and *product discrimination*, where different products are priced in a discriminating manner.

Here are some examples of each type of price discrimination (from Scherer and Ross, 1990):

Personal Discrimination

1. **Haggle-every-time:** each transaction is a separately negociated bargain. Examples: Middle Eastern bazaars, and new/used car sales.
2. **Size-up-their-income:** wealthier (individual) customers are expected to possess more inelastic demand and are charged more than less affluent consumers. Examples: legal and medical services.
3. **Measure-the-use:** customers who use a product more are charged a higher price that is not proportional to any difference in costs. Example: Xerox machine rental charges.

Group Discrimination

1. **Dump-the-surplus:** goods in excess supply are exported at reduced prices, to prevent depressing domestic monopoly prices. Example: export market dumping (e.g. televisions, computer chips, etc.)

2. **Promote-new-customers:** new customers are offered lower prices than existing customers to develop new brand loyalty. Examples: newspapers and magazines.
3. **Keep-them-loyal:** special discounts are given to high volume buyers or prized customers. Example: frequent flier programs.
4. **Sort-them-by-time-value:** coupons which involve a time commitment for redemption are given to customers. Those who redeem these coupons are presumed to have a lower opportunity cost of time, which corresponds with a lower reservation price. Examples: mail-in rebates, and newspaper coupons.
5. **Divide-them-by-elasticity:** separating customers on the basis of belonging to a particular group, when there is an expectation that the demand elasticity or reservation price will vary among each group. Examples: business vs. tourist rates on travel, and student vs. general admission prices for entertainment.

Product Discrimination

1. **Appeal-to-the-classes:** pricing higher quality products to achieve larger markups than with lower quality products. Examples: cloth vs. paperbound books, and luxury vs. mid-size economy cars.
2. **Make-them-pay-for-the-label:** charging higher prices for (homogeneous) goods, based on name recognition. Examples: Name-brand vs. generic aspirin, salt, etc.
3. **Clear-the-stock:** clearance sale prices are charged on certain items when inventories need to be reduced, with the hope that these lower prices will induce purchases by customers with tight budgets. Example: Macy's, or other high-end store clearance sales.
4. **Switch-them-to-off-peak-times:** for goods and services with varying time-consumption patterns, lower prices are charged during off-peak periods. Examples: hotel and motel rates, and long distance telephone rates.
5. **Skimming:** setting high introductory prices that are designed to exploit customers eager to buy a new product. Example: introductory automobile prices.

Notes

LESSON 24:
ACQUIRING AND USING MARKET POWERS

24.1 A Review of Perfect Competition and Monopoly

It is typical in microeconomic analysis to discuss perfect competition and monopoly. This handout is a basic review of these two extreme cases of market structure, but it is also intends to motivate the further investigation of some related topics.

Perfect Competition

When defining what is meant by a "competitive market" one usually thinks of many firms, with each firm charging a low price. Each firm's price will be high enough to ensure a "fair return" but also low enough to keep other firms from outselling them.

A perfectly competitive industry is characterized in the short run as one with many small firms, each selling a homogeneous (standardized) product. In the long run, a perfectly competitive industry has no barriers to entry or exit. That is, firms may enter and exit the industry as necessary.

What do these conditions directly imply about perfectly competitive industries?

1. Selling a homogeneous product leads to one price for all firms.
2. Many small firms implies that each firm produces for a small segment of the overall market, so small in fact that no single firm can affect the market price. It is the behavior of the market as a whole which determines the market price. In turn, the market price serves as the demand curve for each firm.
3. Free entry into or exit from an industry implies that the industry's firms will make zero economic profits in the long run.

What do these conditions indirectly imply about perfectly competitive industries?

1. Since firms cannot affect the market price, their marginal revenue will always equal the market price.
2. Firms adjust their output levels in order to maximize profits. They produce where their marginal cost (of producing a specific output level) equals the market price. This implies further that the firm's marginal cost curve is its supply curve as well.
3. Once the market price and output levels are set, we find that total economic surplus is at a maximum. This results from the fact that each firm is producing where P = MC, which implies allocative efficiency.

Monopoly

In the popular board game Monopoly, the object is to eliminate one's competitors by buying up their property. That is, to become a monopolist. En route to this end, as a player buys up more and more property, their "profits" rise. In both the game and popular culture, monopolies are often characterized as high profit firms, made rich no doubt after charging exorbitant prices to hapless widows and children.

By definition, a monopolized industry is an industry inhabited by only one firm. In the long run, a monopolistic industry has high barriers to entry and exit. Entry and exit are made difficult by either natural causes (e.g. cost conditions or trade secrets) or more artificial ones (e.g. patents or government involvement).

What do these conditions directly imply about monopoly?

1. With one firm meeting all of market demand, that one firm determines its own market price.
2. When firms may not freely enter or exit an industry, the economic profits of the industry's firms are greater than or equal to zero in the long run. The actual profit level will depend upon the firm's average costs.

What do these conditions indirectly imply about monopoly?

1. Since a monopolist can set its own price, that price will exceed the monopolist's marginal revenue.
2. A monopolist maximizes profits by producing where her marginal cost equals her marginal revenue.
3. Since the firm's price will be greater than its marginal cost, total economic surplus is not at a maximum, implying allocative inefficiency.

Further Questions

While instructive in a theoretical context, these extreme cases aren't often observed in the real world. We may then want to ask some deeper, more empirically driven questions. Here are some examples:

1. Since perfectly competitive firms are unable to affect the market price and since monopolists operate without competition, we see that there is no strategic interaction among firms within perfect competitive or monopoly markets.
   a. How would we model strategic interaction within a market?
   b. What effect does strategic interaction have on prices and output levels?
2. Real firms compete in ways other than just setting a price or output level. For example, in the short run, firms may differentiate their product from that sold by a competitor or firms may compete through innovation, affecting their rate of output growth over the long run.
   a. How do firms differentiate their products, and how does that product differentiation affect market outcomes?
   b. While some market structures may lead to allocative inefficiency, is it possible that through innovation these

same markets may have higher rates of innovation and achieve higher rates of output growth?

3. Our previous analysis says little about how monopoly is achieved.

a. Should it matter how firms come to dominate a market?

b. What happens in monopoly markets when barriers are not so high as to prevent entry forever?

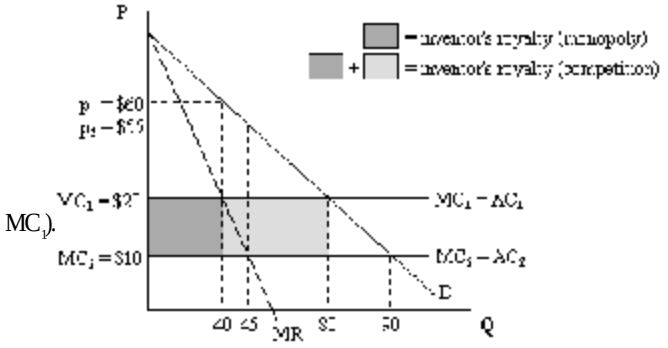
24.2 Innovation and Market Structure

One of the industries is a monopoly (industry A), the other is perfectly competitive (industry B). What is the incentive to innovate in these two industries? Assume that the industries face identical demand and cost curves:

- Industry demand: P = 100 - Q (monopolist's MR = 100 - 2Q)
- Marginal Cost: MC = AC = $20
- An inventor designs a way for each industry to lower unit costs by $10 per unit

A. Incentive to innovate = size of the inventor's royalty

Suppose that our inventor can choose between selling the innovation to industry A or B. Whoever pays the biggest royalty gets the innovation. How much are they willing to pay? The answer is, no more than the change in their costs (i.e. $10 per unit). Therefore, we'll set the royalty fee at $10 per unit. The innovation would lower MC in both markets (to MC_2), but the royalty fee would bring MC back to its pre-innovation level (to MC_1).



The inventor would collect a larger royalty from the competitive industry than the monopoly. The monopolist would pay $400, and the competitive industry would pay $800. This higher royalty from the competitive industry has been interpreted to imply a greater incentive to innovate with perfect competition than with monopoly.

Some researchers argue that it is misleading to define the incentive to innovate as the size of the royalty paid to the inventor because this approach is biased against monopolies. Another way to define the incentive to innovate is in terms of that innovation's effect on industry profits.

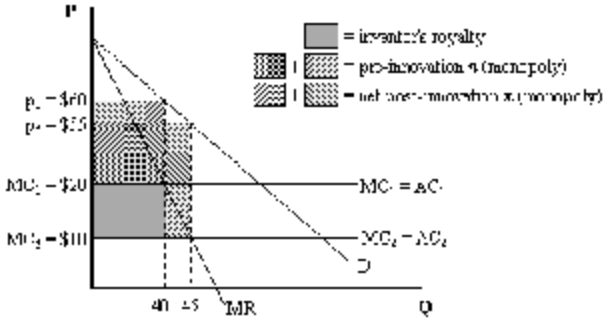
B. Incentive to innovate = change in industry profits

Instead of assuming that an inventor can sell this innovation to the highest bidder, assume that the royalty paid is the same between industries A and B. To do this, assume that industries A and B the same marginal revenue and cost curves (rather than the same demand and cost curves as we did above).

Suppose there's a two-step process to this example: (a) industry and inventor get together and decide the amount of the royalty, and (b) given the royalty, the industry determines how much to produce. Set the royalty at $400 again. Taking this $400 payment as given allows each industry to operate along a new, lower MC curve. In both industries, output rises from 40 to 45.

What happens to profits in industry B? In the competitive industry, price falls to marginal and average cost - implying zero (gross) profits for the industry, both before and after the innovation is introduced. Taking the $400 royalty out of this amount means a $400 loss. Consequently, the net change in profits is -$400.

What happens to profits in industry A? The monopolist's pre-innovation profits are $1600. After the innovation, MC declines. Output rises to 45 units and the price falls to $55. The monopolist's post-innovation (gross) profits become $2025. After paying the royalty, the monopolist's net change in profits is $25. This implies a stronger incentive to conduct innovation in monopoly than in perfect competition.



24.3 The Measurement of Market Power

Two firms produce a homogeneous product at constant costs and maximize their profits by optimally setting their output levels. There is some degree of strategic interaction between these firms, measureable to some extent by looking at the ability of each firm to mark up its price over marginal cost. The firms face the same market demand curve (where $q_1 + q_2$ represents the market's total output):

$P = 100 - (q_1 + q_2)$

Firm 1 maximizes its profits by setting its own marginal revenue equal to its marginal cost. Their marginal revenue would be: $MR = 100 - 2q_1 - q_2$. Rather than complete this exercise, however, let's consider what happens if we rearrange firm 1's MR a little (firm 2's MR would be the same as firm 1's except that we would have to swap the subscripts on the q's).

Another way of writing firm 1's MR is: $P - q_1(1 + a_1)$

The most important term in this expression is the "a" term. This term measures the change is called the *conjectural variation* of firm 1, and it reflects the degree of interdependence mentioned above. More technically, it is a measure of how changes in firm 1's output affect firm 2's output choice: Dq_2 / Dq_1. We can assume that $Dq_2 / Dq_1 = Dq_1 / Dq_2$, which means that both

firms hold the same conjectural variation (allowing us to drop the subscript on this term).

Now, let's look again at firm 1's profit maximization process. Set MR = MC:

$P - q_1(1 + a) = MC$

which can be rearranged as follows:

$P - MC = q_1(1 + a)$

Divide both sides by P, and multiply the righthand side (RHS) by Q/Q. The result is:

$$\frac{P-MC}{P} = \left(\frac{q_1}{Q}\right)\left(\frac{1}{\varepsilon_d}\right)(1+\alpha)$$

When all the firms in the industry have the same MC, they produce the same level of output (in this example, each firm's MC is constant at the same amount). That being the case here, we can re-express the q_1/Q term as $1/n$ (where n = the number of firms, which in this example is 2):

$$\frac{P-MC}{P} = \left(\frac{1}{2}\right)\left(\frac{1}{\varepsilon_d}\right)(1+\alpha)$$

On the lefthand side of the equation, we have firm 1's price-cost margin (PCM). Another name given to this expression is the Lerner index. As the price elevates above marginal cost, the firm makes steadily higher profits. On the RHS are three reasons for variations in firm 1's PCM: (a) changes in firm 1's market share (q_1/Q or $1/2$); (b) changes in the market's elasticity of demand (e_d), or (c) changes in the firm's conjectural variation.

From this Last Equation, three Interesting Cases Emerge

Case 1: a = -1. If firm 1 increases their output, and firm 2 decreases their output by the same amount, we have Bertrand competition. This condition implies that changes in either firm's output do not affect the market price, which also occurs in perfectly competitive markets. These firms have no market power and their PCM is zero.

Case 2: a = 0. If firm 1 increases their output, and firm 2 is passive - which means that they make no changes in their output level - there is Cournot competition. In this setting price will exceed marginal cost, which signals the presence of some market power.

Case 3: a = 1. If firm 1 increases their output, and firm 2 increases their output by the same amount as firm 1, as though the firms were acting in unison, then we have perfect collusion. The PCM reaches its greatest level as the firms act as though they were a single firm. Perfect collusion may be explicit, by the forming of a cartel, or implicit, as firms collude tacitly through observation and reaction. A more general way of stating this case is by saying that it occurs when a = n - 1.

24.4 Vertical Product Differentiation

Let's begin by assuming the following:

- 2 firms, each produces a differentiated good at zero cost. Firm 1 produces good H, and firm 2 produces good L.
- Firm 1 has a higher quality product than firm 2. If k_i (where i = H or L) represents quality, then this assumption implies that: $k_H > k_L > 0$
- 100 consumers with identical incomes, E, but different tastes, q_i (i = 1,...,100), that are uniformly distributed between zero and one hundred.
- Each consumer must decide whether to buy (at most) one unit from firm 1, one unit from firm 2, or buy nothing.

Utility functions determine what each consumer does:

- if consumer x buys nothing: $U_x(0,E) = E$
- if consumer x buys good H: $U_x(k_H,E) = (E - p_H) + q_x k_H$
- if consumer x buys good L: $U_x(k_L,E) = (E - p_L) + q_x k_L$

1. There is one consumer who is indifferent between buying nothing and buying the lower quality product of firm 2. This person is identified by their taste parameter, q_z. Their utility from buying nothing is equal to their utility from buying good L:

$E = (E - p_L) + q_z k_L$

"Solving this equation for q_z, gives us the "location" of this consumer:

$$\theta_z = \frac{p_L}{k_L}$$

2. There is another consumer who is indifferent between buying good L and good H. This person is identified by their taste parameter: q_w.

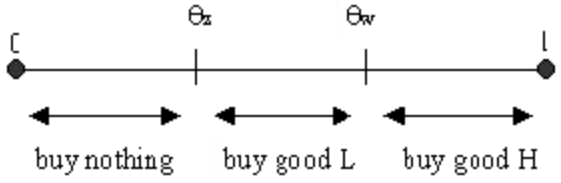
Their utility from buying good L is equal to their utility from buying good H:

$(E - p_L) + q_w k_L = (E - p_H) + q_w k_H$

"Solving this equation for q_w, we get the "location" of this consumer also:

$$\theta_w = \frac{p_H - p_L}{k_H - k_L}$$

"Think about these consumers as standing in a line, from the lowest q to the highest. By including the position of our two "indifferent consumers", we can get an idea of the "market share" for these goods.



From that information, we can get these demand functions (since the line above measures market share, we have to multiply by the number of consumers to get the demands):

$$D_H(p_H,p_L) = (1-\theta_w)100 = \left(1-\frac{p_H-p_L}{k_H-k_L}\right)100$$

$$D_L(p_H,p_L) = (\theta_w-\theta_z)100 = \left(\frac{p_H-p_L}{k_H-k_L}-\frac{p_L}{k_L}\right)100$$

Notice that if $p_H = p_L$, everybody buys good H and nobody buys good L. That verifies we are working with a model of vertical differentiation.

Profit Maximization

Each firm maximizes its profits (interdependently). That means each firm must equate their own MR with MC (remember that MC is zero), and then we solve "simultaneously" for their different prices or output levels (here it's prices).

Below, we walk through the profit maximization process, to the stage where we obtain the equilibrium prices and output levels for firms 1 and 2, as well as their equilibrium profits.

$\pi_H = p_H D_H(p_H,p_L)$

$\pi_H = p_H\left(\left(1-\frac{p_H-p_L}{k_H-k_L}\right)100\right)$

$\frac{d\pi_H}{dp_H} = 100\left(1-\frac{2p_H-p_L}{k_H-k_L}\right) = 0$

$p_H = \left(\frac{k_H-k_L}{2}\right)+\frac{1}{2}p_L$

$p_H^* = \frac{2k_H(k_H-k_L)}{(4k_H-k_L)}$

$D_H^* = \frac{400(k_H-k_L)(k_H)^2}{(4k_H-k_L)}$

$\pi_H^* = \frac{800(k_H)^3(k_H-k_L)^2}{(4k_H-k_L)^3}$

$\pi_L = p_L D_L(p_H,p_L)$

$\pi_L = p_L\left(\left(\frac{p_H-p_L}{k_H-k_L}-\frac{p_L}{k_L}\right)100\right)$

$\frac{d\pi_L}{dp_L} = 100\left(\frac{p_H k_L-2p_L k_H}{(k_H-k_L)k_L}\right) = 0$

$p_L = \frac{1}{2}\cdot\frac{k_L}{k_H}p_H$

$p_L^* = \frac{k_L(k_H-k_L)}{(4k_H-k_L)}$

$D_L^* = \frac{100(k_H-k_L)(k_H k_L)}{(4k_H-k_L)}$

$\pi_L^* = \frac{100k_H(k_L)^2(k_H-k_L)^2}{(4k_H-k_L)^3}$

What should we notice about these results?

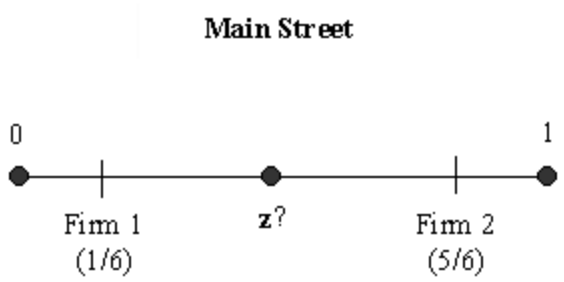
- Firm 1 charges a higher price than firm 2.
- Even though this is Bertrand competition, both firms have P > MC
  (as long as $k_H > k_L > 0$).
- Firm 1 always has a profit-incentive to increase the quality of their good, firm 2 does not have that same incentive. For firm 2, an increase in quality is bad for profits when the degree of differentiation is small (if $8k_H^2 - 15k_Hk_L + k_L^2 < 0$, then p_L falls with an increase in k_L).

24.5 Horizontal Product Differentiation

We begin by assuming:

- 2 firms, each sells the same good at zero cost but from different locations along Main Street. Firm 1 sells good 1, and firm 2 sells good 2.
- · Main Street is a mile long (from one end of town to the other). Firm 1 sells good 1 at their shop, located 1/6 of a mile from where Main Street begins, and firm 2 sells good 2 at their shop, located 1/6 of a mile from where Main Street ends.
- Consumers are uniformly distributed along Main Street, with their location being determined by where they live. Consumers incur a cost in walking to either shop. That cost equals (t x c), where t = distance travelled and c = cost per unit of distance.
- Consumers may buy (at most) one unit from firm 1 or one unit from firm 2.

Somewhere along Main Street is a consumer who is indifferent between walking to firm 1's shop or firm 2's shop. This person is identified by their address, which we'll call **z**. We don't know where z is actually located (hence the question mark below), but once we do it's possible to know who buys from which firm. Here's what this would look like visually.



If this consumer is indifferent between buying good 1 and good 2, then it's because they'd pay the same price for either good. Including the cost of walking to a particular firm, the effective price is p + (t x c). We can figure out the value of t by calculating the distance between z and each firm.

- The distance between **z** and firm 1 is (z - 1/6)
- The distance between **z** and firm 2 is (5/6 - z)

Let's substitute these values in for t, and set the effective prices of each firm's good equal to one another: $p_1 + (z - 1/6)c = p_2 + (5/6 - z)c$

Solving this equation for z, we get the location of the indifferent consumer:

$$z = \frac{1}{2}+\frac{p_2-p_1}{2c}$$

Everybody between z and the beginning of town will shop with firm 1, and everybody at the other end of town will shop with firm 2. Therefore, their demands correspond with the following (under the added assumption that $p_2 - c < p_1 < p_2 + c$):

$$D_1(p_1,p_2) = \frac{1}{2}+\frac{p_2-p_1}{2c}$$

$$D_2(p_1,p_2) = 1-\left(\frac{1}{2}+\frac{p_2-p_1}{2c}\right)$$

Profit Maximization

Each firm maximizes its profits (interdependently). That means each firm must equate their own MR with MC (remember that MC is zero), and then we solve "simultaneously" for their different prices or output levels (here it's prices).

Below, we walk through the profit maximization process, to the stage where we obtain the equilibrium prices and output levels for firms 1 and 2, as well as their equilibrium profits.

$\pi_1 = p_1 D_1(p_1, p_2)$ $\qquad$ $\pi_2 = p_2 D_2(p_1, p_2)$

$\pi_1 = p_1\left(\frac{1}{2} + \frac{p_2 - p_1}{2c}\right)$ $\qquad$ $\pi_2 = p_2\left(\frac{1}{2} - \frac{p_2 - p_1}{2c}\right)$

$\frac{d\pi_1}{dp_1} = \frac{-(2p_1 - c - p_2)}{2c} = 0$ $\qquad$ $\frac{d\pi_2}{dp_2} = \frac{-(2p_2 - c - p_1)}{2c} = 0$

$p_1 = \frac{p_2 + c}{2}$ $\qquad$ $p_2 = \frac{p_1 + c}{2}$

$p_1^* = c$ $\qquad$ $p_2^* = c$

$D_1^* = \frac{1}{2}$ $\qquad$ $D_2^* = \frac{1}{2}$

$\pi_1^* = \frac{c}{2}$ $\qquad$ $\pi_2^* = \frac{c}{2}$

What should we notice about these results?

1. Firms 1 and 2 charge the same price (since the goods are homogeneous).
2. Even though this is Bertrand competition, both firms have P > MC
   (as long as c > 0).
3. Both firms have an incentive to change locations. For example, if either firm moved to a point that is 1/2 mile from the beginning of town (i.e. the same location as z), after reworking the equations above we'd find that that firm's profits would increase from c/2 to 50c/81.

24.6 Collusion and Competition within a 2 firm industry

Suppose a market exists where there are only two firms. Assume that the firms (called A and B) are symmetric, which means that they have identical costs. Assume also that they produce a homogeneous (identical) product. The market demand and firm A and B's costs are:

Market Demand: $P = 100 - q_A - q_B$ (P = market price, q = output)

There are two possibilities: (1) they maximize their own profits, or (2) they maximize their collective (joint) profits. We'll consider both.

1. If each firm maximizes its own profits.

This is saying that the firms are engaged in (typical) competitive behavior. To maximize their own profits, the firms must produce where their own marginal revenue is equal to their marginal cost. Each firm's MR and MC are:

Firm A's MR $\quad MR_A = 100 - 2q_A - q_B$

Firm B's MR $\quad MR_B = 100 - q_A - 2q_B$

Firm A's MC and AC $\quad MC_A = AC_A = 20$

Firm B's MC and AC $\quad MC_B = AC_B = 20$

A. Find equilibrium output for each firm.

	Firm A	**Firm B**
1. Set MR = MC	$100 - 2q_A - q_B = 20$	$100 - q_A - 2q_B = 20$
2. Solve for (own) q	$q_A = 40 - 0.5q_B$	$q_B = 40 - 0.5q_A$
3. Plug q_A equation into q_B equation		$q_B = 40 - 0.5(40 - 0.5q_B)$
4. Solve for q_B*		$q_B^* = 80/3$
5. Plug q_B* into q_A equation	$q_A = 40 - 0.5(80/3)$	
6. Solve for q_A*	$q_A^* = 80/3$	

B. Find the equilibrium market price.

1. Plug q_A* and q_B* into the market demand equation $\quad P = 100 - (80/3) - (80/3)$
2. Solve for P* $\quad P^* = \$140/3$

"C. Find each firm's equilibrium profits.

	Firm A	**Firm B**
1. Find (P - AC)q* for each firm	$((140/3) - 20)(80/3)$	$((140/3) - 20)(80/3)$
2. Calculate profits	$_A^* = \$711.11$	$_B^* = \$711.11$

2. If the two firms maximize their collective profits.

This is says that the firms are colluding and may have formed a cartel. It is as though the firms merged into a single firm (forming a monopoly). The cartel sets the (market) marginal revenue equal to their marginal cost. The cartel's MR and MC are:

Cartel's MR $\quad MR_C = 100 - 2Q$

Cartel's MC and AC $\quad MC_C = AC_C = 20$

A. Find equilibrium output and price for the cartel.

	Firm A	**Firm B**
1. Find each firm's output share	$q_A^* = 40/2 = 20$	$q_B^* = 40/2 = 20$
1. Find (P - AC)q* for each firm	(60 - 20)20	(60 - 20)20
2. Calculate profits	$\pi_A^* = \$800$	$\pi_B^* = \$800$

Summary: the firms each produce 20 units, charge a common price of $40 and make $800 in profits. With a lack of competition, we see prices rising and output falling.

If the firms operate independently, maximizing their own profits, they make profits of $711.11. However, if they collude, then they can make profits of $800. Direct collusion is illegal in the United States, but even if it were not there are reasons why many cartels would not remain intact for long.

3. Collusion and the incentive to cheat

When firms form an agreement to fix prices (i.e. collude) there is always some incentive to cheat on the agreement. To understand why, consider the following.

Suppose firm A decides to increase their output slightly, to 80/3 units (i.e. the amount they would produce if maximizing their own profits). This would have no effect on their own AC, but would change the market price. Firm A would produce 80/3 units while B would continue producing their share of the cartel output, so the price would be: P = 100 - (80/3) - 20 = 160/3.

Profits for firm A become: $p_A^* = ((160/3) - 20)(80/3) = \$888.89$

While firm B's profits are: $p_B^* = ((160/3) - 20)(20) = \$666.67$

Firm A obviously benefits, at firm B's expense, from cheating on the collusive agreement. Of course, once firm B realizes that firm A is doing this, then firm B should increase their output too. The result is that the cartel falls apart and both firms end up producing their original (competitive) output levels. This problem illustrates a classic conflict between maximizing one's own welfare - in this case, profits - vs. that of the group.

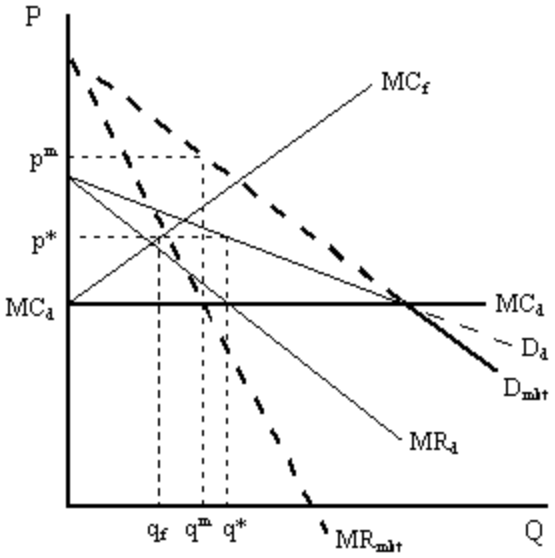
24.7 Strategic Entry Barriers : The Dominant Firm and Limit Pricing

In some highly concentrated industries, a single ("dominant") firm serves a majority of the market and a group of smaller ("fringe") firms supply the rest. Martin (1994) summarizes the difference between a monopolist and a dominant firm as follows:

"A dominant firm differs from a monopolist in one important respect. The only constraint on the monopolist's behavior is the market demand curve: if the monopolist raises price, some customers will leave the market. Like the monopolist, the dominant firm is large enough to recognize that a price increase will drive some customers from the market. But the dominant firm faces a problem that the monopolist does not: the possibility that a price increase will induce some customers to begin to buy from firms in the fringe of small competitors. That dominant firm, in other words, must take into account the reaction of its fringe competitors."

To best understand how the dominant firm may attempt to prevent entry by strategically setting its price, we need to examine how the dominant firm and its competitive fringe determine output. In the graph (below), we derive the residual demand curve - facing the dominant firm (D_d) - from the MC, or supply curve, of the fringe (MC_f) and market demand (D_{mkt}). That is, we find the "new" demand curve faced by the dominant firm when the market is shared with a competitive fringe.

As a monopolist, the dominant firm would charge p^m and produce q^m.



With entry by the fringe, the dominant firm now faces a residual demand curve, rather than the market demand curve. Notice that if the market price falls below the point where the residual demand curve, D_d, crosses the market demand curve, D_{mkt}, the dominant firm is (once again) a monopolist. This point corresponds with where MC_f crosses the vertical axis (where the fringe produces zero output).

Since the dominant firm chooses to produce where current profits are maximized (i.e. where $MR_d = MC_d$) the price falls to p* and output responds by rising to q*. The competitive fringe induces the dominant firm to exercise some restraint in price setting. As long as this restraint is present, the market price will remain lower. At the price p*, consumers will be willing to buy more than q* units and so we find the fringe producing as well. The fringe firm(s) will produce at q_f (where $p^* = MC_f$).

As the dominant firm confronts the entry of the fringe, what is the proper course of action? One answer concerns the strategic use of pricing. There are several approaches: static limit pricing and dynamic limit pricing. Let's examine these individually.

Static Limit pricing. One option is that the firm sets a price that prevents the fringe from entering the market (i.e. causes $q_f = 0$). The dominant firm could do this by lowering the price - from p^m to MC_d - causing the dominant firm's economic profits to fall to zero. One problem with static limit pricing concerns whether a rational firm would ever engage in non-profit maximizing behavior. That is, the firm could make higher (current) profits by setting a different price. Naturally, the dominant firm is concerned with future profits as well, and so the answer depends upon both current and future profits.

Dynamic Limit pricing. Another option is one where the firm considers what is called the present discounted value of the stream of profits it receives over time. There are two ways of viewing this approach. Each way concerns how the firm is able to "gaze into the future." Suppose the firm has myopic

foresight, and can only view each period as it occurs. In this event, they will take entry as given and maximize current profits. In the graph above, this leads to the price p*. As entry continues to expand the size of the fringe, the fringe "supply" curve becomes flatter. A flatter fringe supply curve causes a flatter residual demand curve for the dominant firm. Thus, the dominant firm's price will fall as the fringe expands.

If the dominant firm has perfect foresight, then they will maximize the present discounted value of profits. Likewise, the firm takes entry as given and allows the fringe to expand over time. Essentially, the difference is that the myopic foresight price will exceed the perfect foresight price. In both cases, however, the price falls over each successive period until reaching the limit price (here, note that the limit price is MC_d - the perfectly competitive price).

Either way, whether there is perfect or myopic foresight, the dominant firm reacts passively to the expansion of the fringe. The dominant firm seeks only to maximize their profits in some sense (either the present discounted value or current value per period). As a result, this model becomes a way of predicting movements in price and market share for industries like the auto or steel industry where a dominant firm or set of firms face competition from smaller entrants.

Notes

LESSON 25:
NATURAL MONOPOLY

25.1 Natural Monopolies and Pricing Policy

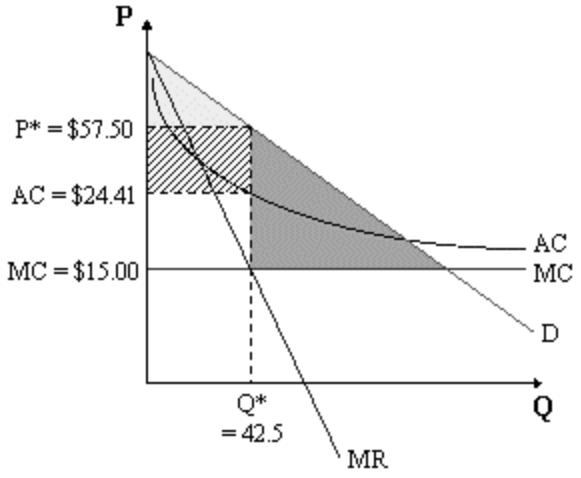
Assume that a certain natural monopolist has the following demand and cost related curves:

Demand:	P = 100 - Q
Marginal Revenue:	MR = 100 - 2Q
Average Cost:	AC = 15 + (400/Q)
Marginal Cost:	MC = 15

Why is this a Natural Monopoly?

The answer stems from the monopolist's natural (cost-related) barriers to entry. The relative position of the AC and MC curves give the natural monopolist a cost advantage over its competition. Taking a closer look at these equations, you'll see that AC is always going to be greater than MC. Remembering the relationship between marginal and average values, AC will be declining as long as MC is below it. In general then, for a natural monopoly, AC is said to decrease (as Q increases) through "some relevant range of market output".

On a graph, it looks like this:



We'll calculate the values for P and Q* below, and also explain the meaning of the shaded areas.*

If allowed to decide herself, how much will this natural monopolist produce, and at what price?

If allowed to set her own output and price, this natural monopolist will produce where MR = MC:

Set MR = MC, and solve for Q*

100 - 2Q = 15

Q* = 42.5

Find price by plugging Q* into the demand equation:

P = 100 - (42.5) = 57.5

Therefore: Q* = 42.5 and P* = $57.50

Suppose we also want to find the monopolist's profits. To do that, we use the formula (P - AC)Q. Before plugging things into this equation though, we must find AC. The value for AC is found by plugging Q* into the AC equation to get AC = $24.41 (i.e. AC = 15 + 400/42.5).

Profits are then calculated as:

p = [$57.50 - $24.41]42.5 = $1406.33

To make these kind of profits (the area represented on the graph by the striped rectangle), the monopolist sets a price exceeding what might occur within a more competitive market. This high price makes consumer surplus (shaded yellow in the graph) rather small.

One big problem with this result is that since the natural monopolist produces less output than what is possible under perfect competition, there is some deadweight loss (shaded blue on the graph) — which represents the value of output not produced as a result of P > MC.

To get rid of the DWL, a government regulator might step in and force the monopolist to set its price at marginal cost.

1. Marginal Cost Pricing

When the regulating agency forces this firm to set its price at marginal cost, we have what is called marginal cost pricing. In this case, that means setting P = $15.

How much will the firm produce when P = MC?

Set Demand, or P, equal to MC and solve for Q*

100 - Q = 15

Q* = 85

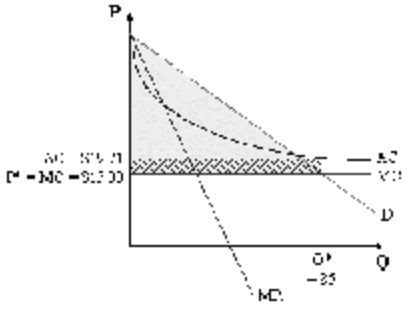
What are the firm's profits when P = MC?

If P* = $15 and Q* = 85, then AC = $19.71

Profits become:

p = [$15.00 - $19.71]85 = -$400.35

which represents a loss. On the graph below, these values and the areas for consumer surplus and profits are illustrated. Notice that the area of consumer surplus overlaps that corresponding with profit (loss), and that there is no deadweight loss since P = MC.



Since the firm is making a loss, it needs to consider the future. That is, should the monopolist stay in this industry if, over the long run, the best it can ever do each year is make some type of loss? The answer would obviously be no, and so if the price were set at $15, the firm would eventually exit the industry.

The whole point of government involvement here relates to the fact that regulators wanted to make things more efficient (in terms of allocative efficiency). However, achieving this particular type of efficiency causes the firm to eventually exit the industry — leaving consumers with nothing. Therefore, to prevent the firm from leaving, our regulator must also allow the monopolist to cover her losses. One way to do this is by subsidizing the monopolist the amount of her loss ($400.35). Another way is to give up on the idea of producing where P = MC.

2. Average Cost Pricing

One possibility is that the government regulator might want to allow the firm to charge a slightly higher price, but make zero economic profit. This is accomplished when P = AC, an approach that is called Average Cost pricing.

Figuring out (algebraically) what the price will be is a bit more involved than what we did above. To algebraically find the price that would equal average cost, we first set Demand (Price) equal to Average Cost (AC), then solve for Q* and lastly plug Q* into the Demand equation to get P*:

Set P = AC:

$100 - Q = (400/Q) + 15$

Rearrange this equation to get:

$Q^2 - 85Q + 400 = 0$

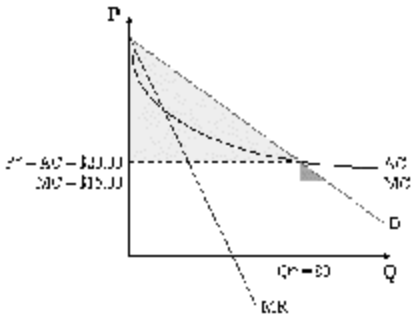
Using the quadratic formula, we can solve for Q:

$Q^* = 80$

Plugging Q* into the Demand equation, we can solve for P*

$P^* = 100 - (80) = \$20$

Consequently, setting P = AC means setting P* = $20, and getting 80 units of output. In the graph below, these values are given, as are the corresponding shaded areas for consumer surplus and deadweight loss (remember that profits are zero here since P = AC, but there will be some deadweight loss since P > MC).



Of course, the problem here is that while the natural monopolist is able to make zero profits, thereby ensuring that the firm will stay in business, some deadweight loss reoccurs — the very thing that government involvement was trying to eliminate.

Another potential problem with government imposing this type of (average cost) pricing is that it may create an incentive for the firm to inflate its fixed costs. This is called overcapitalization, because the firm may overinvest in capital equipment since it is, in a sense, guaranteed a "normal return".

What we find is that, when charging a single price to all consumers, a natural monopolist's costs force us to choose between allocative efficiency and allowing the firm a fair return on its investment (without subsidizing it). A final approach involves using two different "prices", what we'll call here a two part tariff.

3. Two-part Tariff

Suppose the regulator forces our monopolist to sell every unit of output at $15 (i.e. P = MC), but also allows her to charge a fixed (flat) fee that all consumers must pay before buying this product at $15.

In other words, the natural monopoly is allowed to charge something we could call an admittance fee. This fee establishes who is in the market. Those consumers who pay the fee are subsequently allowed to buy as much product as they want at $15 per unit (the MC price).

Before this extra fee, a price of $15 caused the monopolist to lose $400 in profits. But with it, the fixed fee allows the monopolist to recoup those losses (by setting the fixed fee = 400/N, where N is the number of consumers who want to purchase the firm's product).

For example, if there are 50 consumers who want to buy this natural monopolist's product, then the firm should be allowed to charge a flat fee of $8. Doing so, allows the firm to produce 85 units of output and make zero economic profit. This means that the government regulators get what they want — no deadweight loss — and the firm gets what she wants — a fair return on her investment (which is what we interpret zero economic profit to imply).

25.2 Multiple Products and Natural Monopoly

With a single product natural monopoly, having economies of scale implies having a natural monopoly. Is this true with a multi-product monopoly as well? Consider the following example.

Suppose that Firm A produces two different goods: products 1 and 2. If the firm decides to produce both products at the same time, then its cost function looks like this:

$C(q_1,q_2) = q_1 + q_2 + (q_1q_2)^{1/3}$

Another option is to produce both products separately (e.g. Firm A could break down into two smaller firms, each specializing in the production of product 1 or 2). If the firm produced the products separately, specializing in the production of product 1, then its cost function becomes:

$C(q_1) = q_1$

On the other hand, if Firm A produced only product 2, then its cost function would be:

$C(q_2) = q_2$

1. Does the Firm have Economies of Scale?

If the firm can double its output without doubling its costs, then it has economies of scale.

a. If the firm produces 8 units of each product, its costs would be $20

b. If the firm's costs doubled, they would be $40

c. If the firm doubled each product's output, its costs would be $38.35

Firm A has economies of scale since it can double its output without doubling its costs.

2. Does the Firm have Economies of Scope?

If Firm A can produce both products (simultaneously) more inexpensively than it could as two separate companies (where each company produces one product, rather than both), then the firm has economies of scope.

a. If the firm produces 8 units of each product, its costs would be $20

b. If the firm produced only 8 units of product 1 (or product 2), its costs would be $8 (the total cost for the two smaller companies would be $16)

Firm A does not have economies of scope, because the total cost of producing products 1 and 2 separately ($16) is less than the total cost of producing them jointly ($20).

Conclusion: when a monopolist produces only one product, then having economies of scale implies having a natural monopoly. When a monopolist produces more than one product, then having economies of scale is not enough to imply having a natural monopoly. It takes both economies of scale and scope in this case. In more technical terms, the meaning of additivity is expanded to include "scope" when the monopoly produces multiple products.

Notes

LESSON 26:
PROFIT MAXIMIZATION AND THE PERFECTLY COMPETITIVE FIRM

Our first look at firm behavior comes within the context of perfect competition. What comes below is a step by step explanation of how perfectly competitive firms maximize their profits, both algebraically and graphically, and a discussion of our result.

Remember that, in perfectly competitive markets, no individual firm has any influence over the market price (since there are many firms and each is a small player in the overall market). Since each firm's product is identical to that of other firms (i.e. products are homogeneous), all firms face the same price.

While firms cannot individually influence the market price through their actions, they can collectively. Therefore, our starting point will be the market demand and supply curves. These are the same demand and supply curves from the earlier material on Consumer Theory (i.e. they do all the same tricks, like demand shifting when there's a change in income, that those other demand and supply curves did).

(Market Demand) $P = 100 - .078Q_d$

(Market Supply) $P = .02Q_s + 2$

Solving for equilibrium price and quantity, we get: P*= $22 and Q*= 1000 units. These values represent the price that each firm will charge and the total number of units that will be produced overall.

A typical firm within this market has the following costs:

(Total Costs) $TC = q^2 + 2q + 100$

(Average Costs) $AC = q + 2 + (100/q)$

(Marginal Costs) $MC = 2q + 2$

Let's note a few things about the first two equations before proceeding. In the TC equation, $q^2 + 2q$ represents the firm's variable costs and 100 represents the fixed costs. The AC equation is obtained by dividing the TC equation by q. This means that, in the AC equation, q + 2 are the average variable costs and 100/q are the average fixed costs.

1. Given these costs, how much should the Firm Produce?

The firm will always produce where the MC of a certain level of output equals the market price. That is, the firm will adjust its output level until P = MC. To find this output level, we set the MC equation equal to the equilibrium price:

P* = MC

22 = 2q + 2

q = 10

The firm will maximize its profits by producing 10 units. It is possible to characterize this firm and market level information with the following pair of demand and supply graphs. The graph on the right represents the market, while the graph on the left represents the firm.

The equilibrium price corresponds with where the market demand (D_M) intersects the market supply (S). The firm accepts this price and decides how much to produce. This occurs where the firm's marginal cost curve (MC) crosses the firm's demand curve (D_f). Note that the firm's demand curve is a horizontal line at the equilibrium price of $22.

Another way to see whether the firm is maximizing profits is to assume that our P = MC rule isn't true. Suppose that the firm decides to test this rule by varying its output. If profits decline as we move away from where q = 10 (e.g. as we move between 8 and 12 units), then profits must be maximized in the row where P = MC.

P	q	MC	AC	Profits
22	8	18	22.5	- 4.0
22	9	20	22.1	- 0.9
22	10	22	22	0
22	11	24	22.1	- 1.1
22	12	26	22.3	- 3.6

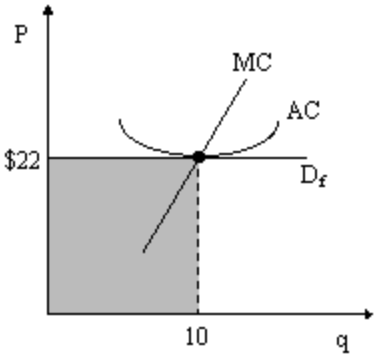
As the table makes clear, profits reach their highest level when the firm produces 10 units. Although it is true that the price equals **both** marginal and average cost in this row, this is only coincidence right now (in the short run). Profit maximization only necessitates that P = MC.

2. How do we Calculate the firm's Profits?

To find the firm's profits, we take one of two approaches (where TR = total revenue, which is (P x q)):

TR - TC approach	**P - AC approach**
Profit = TR - TC	Profit = (P - AC)q
Profit = (22 x 10) - [$(10)^2$ + 2(10) + 100]	Profit = (22 - [(10) + 2 + (100/(10))]) x (10)
Profit = 220 - 220	Profit = (0) x 10
Profit = 0	Profit = 0

The result is that this firm produces 10 units and makes zero economic profit. Graphically, we find this result by comparing P and AC. Recall that P comes from the action of the market (as a whole), and it is represented by the horizontal demand curve D_f. AC is found by: (a) locating the firm's output level, (b) tracing a dotted line from this output level to the AC curve, and (c) from the point where the dotted line hits AC - go left, over to the vertical axis.



In the graph above, both P and AC are the same. We find TR by multiplying P and q, and TC by multiplying AC and q. By this method, the firm's TR and TC are represented by the same shaded area on the graph.

3. Why Would the Firm Produce if it makes Zero Profit? One way to answer this question is by seeing what happens if the firm shuts down. Then we'll compare the profits (or losses) under the two situations: producing vs. shut down. Recall that the firm has fixed costs of $100. Assume that these fixed costs are all sunk (i.e. non-recoverable). If so, shutting down will cost the firm its $100 in sunk costs. This is worse than making zero profits, so the firm will produce.

Supposing that the fixed costs are all recoverable, then the firm would be indifferent between producing and shutting down since both situations would involve making zero profit. In a lot of introductory economic analysis, however, fixed costs are implicitly assumed to be 100% sunk.

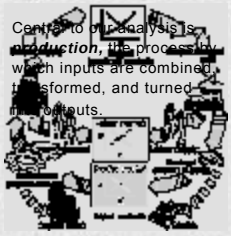
The important thing to remember here is that these profits are economic profits, not accounting profits. To see why this is important, consider how economic profits and accounting profits are calculated:

Economic profit = Actual revenue - (Actual costs + Opportunity costs)

Accounting profit = Actual revenue - Actual costs

While zero accounting profit would be undesirable, zero economic profit is not. A person could work all day to make $1 in accounting profits and be very unhappy since that person could probably do better in some other money-making activity (i.e. the next best alternative occupation). By including opportunity cost, economic profit accounts for things like the value of one's time in producing a good or service.

Production



Central to our analysis is ***production***, the process by which inputs are combined, transformed, and turned into outputs.

What Is A *Firm*?

- A ***firm*** is an organization that comes into being when a person or a group of people decides to produce a good or service to meet a perceived demand. Most firms exist to make a profit.
- Production is not limited to firms.
- Many important differences exist between firms.

Perfect Competition

Perfect competition is an industry structure in which there are:

- ***many firms***, each small relative to the industry,
- producing virtually ***identical products*** and
- in which ***no*** firm is large enough to have any ***control over prices***.
- In perfectly competitive industries, new competitors can ***freely enter and exit*** the market.

Homogeneous Products

- ***Homogeneous products*** are undifferentiated products; products that are identical to, or indistinguishable from, one another.
- In a perfectly competitive market, individual firms are ***price-takers***. Firms have no control over price; price is determined by the interaction of market supply and demand.

Profits and Economic Costs

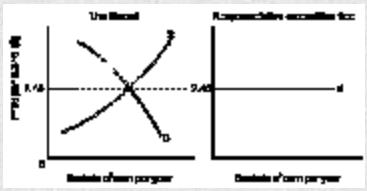
- ***Profit (economic profit)*** is the difference between total revenue and total economic cost.

economic profit ? total revenue? total economic cost

- ***Total revenue*** is the amount received from the sale of the product:

$$(q \; x \; P)$$

Demand Facing a Single Firm in a Perfectly Competitive Market

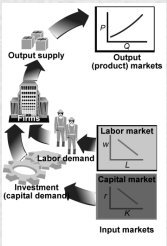


- The perfectly competitive firm faces a ***perfectly elastic demand*** curve for its product.

Profits and Economic Costs

- ***Total cost (total economic cost)*** is the total of
  1. Out of pocket costs,
  2. Normal rate of return on capital, and
  3. Opportunity cost of each factor of production.

The Behavior of Profit-Maximizing Firms



- The three decisions that all firms must make include:

1. How much output to supply
2. Which production technology to use
3. How much of each input to demand

Profits and Economic Costs

- The ***rate of return,*** often referred to as the yield of the investment, is the annual flow of net income generated by an investment expressed as a percentage of the total investment.

Profits and Economic Costs

- The ***normal rate of return*** is a rate of return on capital that is just sufficient to keep owners and investors satisfied.
  - For relatively risk-free firms, the normal rate of return be nearly the same as the interest rate on risk-free government bonds.

Short-Run Versus Long-Run Decisions

- The ***short run*** is a period of time for which two conditions hold:
  1. The firm is operating under a fixed scale (or fixed factor) of production, and
  2. Firms can neither enter nor exit the industry.

Profits and Economic Costs

- ***Out-of-pocket costs*** are sometimes referred to as ***explicit costs*** or ***accounting costs***.
- ***Economic costs***, often referred to as ***implicit cots***, include the full opportunity cost of every input.

Short-Run Versus Long-Run Decisions

- The ***long run*** is a period of time for which there are no fixed factors of production. Firms can increase or decrease scale of operation, and new firms can enter and existing firms can exit the industry.

Calculating Total Revenue, Total Cost, and Profit

Initial Investment:	**$20,000**
Market Interest Rate Available:	**.10 or 10%**
Total Revenue (3,000 belts x $10 each)	**$30,000**
Costs	
Belts from supplier	$15,000
Labor Cost	14,000
Normal return/opportunity cost of capital ($20,000 x .10)	2,000
Total Cost	**$31,000**
Profit = total revenue ? total cost	**? $ 1,000[a]**

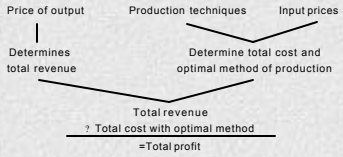
[a]There is a loss of $1,000.

The Bases of Decisions

- The fundamental things to know with the objective of maximizing profit are:

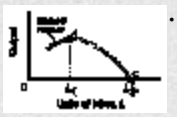
1.	2.	3.
The market price of the output	The techniques of production that are available	The prices of inputs

Determining the Optimal Method of Production



- The ***optimal method of production*** is the method that minimizes cost.

Marginal Product



- ***Marginal product*** is the additional output that can be produced by adding one more unit of a specific input, *ceteris paribus*.

$$\text{marginal product of labor} = \frac{\text{change in total product}}{\text{change in units of labor used}}$$

The Production Process

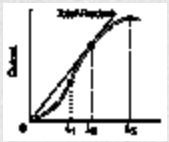
- ***Production technology*** refers to the quantitative relationship between inputs and outputs.
- A ***labor-intensive technology*** relies heavily on human labor instead of capital.
- A ***capital-intensive technology*** relies heavily on capital instead of human labor.

The Law of Diminishing Marginal Returns

- The ***law of diminishing marginal returns*** states that:

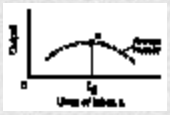
  When additional units of a variable input are added to fixed inputs, the marginal product of the variable input declines.

The Production Function



- The ***production function*** or ***total product function*** is a numerical or mathematical expression of a relationship between inputs and outputs. It shows units of total product as a function of units of inputs.

Average Product



- ***Average product*** is the average amount produced by each unit of a variable factor of production.

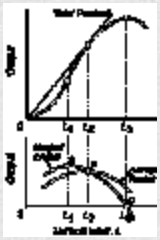
$$\text{average product of labor} = \frac{\text{total product}}{\text{total units of labor}}$$

Production Function for Sandwiches

Production Function

(1) LABOR UNITS (EMPLOYEES)	(2) TOTAL PRODUCT (SANDWICHES PER HOUR)	(3) MARGINAL PRODUCT OF LABOR	(4) AVERAGE PRODUCT OF LABOR
0	0	?	?
1	10	10	10.0
2	25	15	12.5
3	35	10	11.7
4	40	5	10.0
5	42	2	8.4
6	42	0	7.0

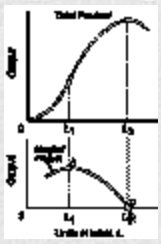
Total, Average, and Marginal Product



Remember that:

- As long as marginal product rises, average product rises.
- When average product is maximum, marginal product equals average product.
- When average product falls, marginal product is less than average product.

Total, Average, and Marginal Product



- Marginal product is the slope of the total product function.
- At point A, the slope of the total product function is highest; thus, marginal product is highest.
- At point C, total product is maximum, the slope of the total product function is zero, and marginal product intersects the horizontal axis.

Production Functions with Two Variable Factors of Production

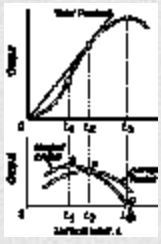
- In many production processes, inputs work together and are viewed as complementary.
  - For example, increases in capital usage lead to increases in the productivity of labor.

Inputs Required to Produce 100 Diapers Using Alternative Technologies

TECHNOLOGY	UNITS OF CAPITAL (K)	UNITS OF LABOR (L)
A	2	10
B	3	6
C	4	4
D	6	3
E	10	2

- Given the technologies available, the cost-minimizing choice depends on input prices.

Total, Average, and Marginal Product



- When average product is maximum, average product and marginal product are equal.
- Then, average product falls to the left and right of point B.

Production Functions with Two Variable Factors of Production

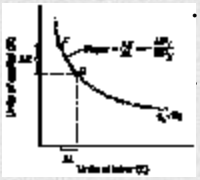
Cost-Minimizing Choice Among Alternative Technologies (100 Diapers)

(1) TECHNOLOGY	(2) UNITS OF CAPITAL (K)	(3) UNITS OF LABOR (L)	(4) COST WHEN $P_K = \$1$ $P_L = \$1$	(5) COST WHEN $P_K = \$5$ $P_L = \$1$
A	2	10	$12	$52
B	3	6	9	33
C	4	4	8	24
D	6	3	9	21
E	10	2	12	20

Review Terms and Concepts

Accounting costs	Marginal product
Average product	Normal rate of return
Capital-intensive technology	Optimal method of production
Economic costs	Out-of-pocket costs
Economic profit	Perfect competition
Explicit costs	production
Firm	Production function or total product function
Homogeneous products	Production technology
Implicit costs	Profit (economic profit)
Labor-intensive technology	Short run
Law of diminishing returns	Total cost (total economic cost)
Long run	Total revenue

Appendix: Isoquants and Isocosts



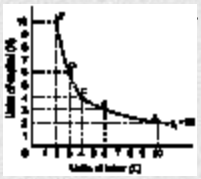
- Along an isoquant:

$$? K ? MP_K \; ? \; ? \; ? L ? MP_L$$

The slope of an isoquant is called the ***marginal rate of technical substitution***.

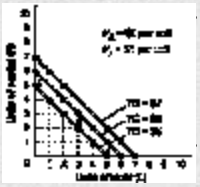
$$\frac{? K}{? L} \; ? \; ? \frac{MP_L}{MP_K}$$

Appendix: Isoquants and Isocosts



An ***isoquant*** is a graph that shows all the combinations of capital and labor that can be used to produce a given amount of output.

Appendix: Isoquants and Isocosts

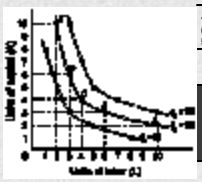


An ***isocost line*** is a graph that shows all the combinations of capital and labor that are available for a given total cost.

The equation of the isocost line is:

$$P_K \; ? K \; ? \; P_L \; ? L$$

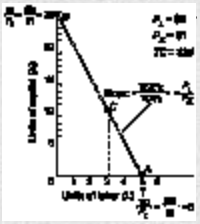
Appendix: Isoquants and Isocosts



Alternative Combinations of Capital (K) and Labor (L) Required to Produce 50, 100, and 150 Units of Output

	$q_1 = 50$		$q_2 = 100$		$q_3 = 150$	
	K	*L*	*K*	*L*	*K*	*L*
A	1	8	2	10	3	10
B	2	5	3	6	4	7
C	3	3	4	4	5	5
D	5	2	6	3	7	4
E	8	1	10	2	10	3

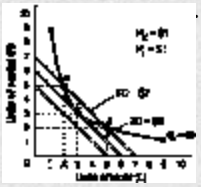
Appendix: Isoquants and Isocosts



Slope of the isocost line:

$$\frac{? K}{? L} \; ? \; ? \frac{TC / P_K}{TC / P_L} \; ? \; ? \frac{P_L}{P_K}$$

Appendix: Isoquants and Isocosts



By setting the slopes of the isoquant and isocost curves equal to each other,

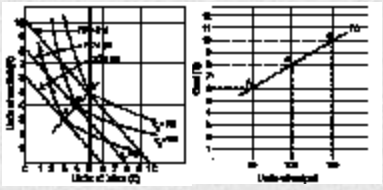
$$\frac{MP_L}{MP_K} ? \frac{P_L}{P_K}$$

we derive the firm's ***cost-minimizing equilibrium condition*** is found

$$\frac{MP_L}{P_L} ? \frac{MP_K}{P_K}$$

Appendix: Isoquants and Isocosts

- Plotting a series of cost-minimizing combinations of inputs (at points A, B, and C), yields a ***cost curve***.



Notes

THE MARKET STRUCTURE AND THE FACTORS MARKET FACTOR MARKET

In the chapter on Supply and Demand we limited our attention to the supply and demand for consumer goods, to keep things a little simpler. In this chapter we extend the supply and demand approach to cover markets for resources, or, in the economist's traditional jargon, factors of production. We will discuss markets for

- labor
- land
- natural resources
- capital

People demand consumer goods for the direct benefits of consuming them. That's not true of labor, land and capital. These factors of production are demanded in order to use them in producing consumer goods. In other words, the demand for factors of production is a "derived demand" — that is, it is derived from the demand for the consumer goods. When a firm demands resources in order to produce potatoes (for example) the firm's demand for resources will depend on two things: the demand for potatoes, and the productivity of the resources in producing potatoes. To be more precise - drawing on ideas from preceding Chapter - it will depend on the market price of potatoes and on the marginal productivity of the resources in producing potatoes.

We will take up where we left off in Chapter 2 & 3, with the demand for the human resource: labor.

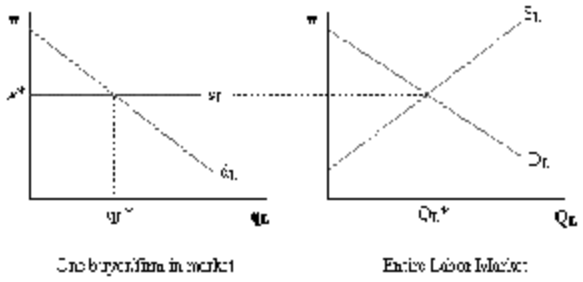
Notes

LESSON 27:
DEMAND AND SUPPLY IN FACTOR MARKET

27.1 Demand and Supply in Competitive Factor Markets

Factor markets are much like output markets, but with one important difference. The demand for factors is a derived demand. For example, a firm's demand for labor depends upon how much output the firm will produce. More output leads to a greater demand for laborers, less output leads to a lower demand for laborers.

Consider the case where there is one variable factor called labor and that this labor is bought and sold in a competitive labor market. The use of the term competitive designates a situation that is similar to what we would find in a competitive output market. In this case, each buyer of labor is one of many buyers in the market. The labor market determines the wage facing individual buyers, who purchase as many workers as needed at the given wage rate. As a result, the supply of labor facing any individual buyer is a horizontal line at the going wage rate. The graph below shows this.



The quantity of labor hired by each firm depends on where the marginal benefit of each hired factor equals the cost of hiring that factor. The marginal benefit (MB) is given by how each extra laborer affects the firm's revenues, whereas the marginal cost (MC) of each additional laborer is how each laborer affects the firm's total costs. In equation form, MB and MC are given as:

MB = DTR/ DL

MC = DTC/ DL

Because firms can hire as many workers as needed at the existing wage rate, we know that the second equation (which describes the supply of labor) can be rewritten as MC = w.

The first equation (which gives us the demand for labor) can be rewritten as: MB = (DTR/ DQ)(DQ/ DL). That is, the marginal benefit of hiring additional workers is the product of a firm's marginal revenue and marginal product of labor.

Note that marginal revenue is determined by the firm's output decision in the output market, while marginal product is determined by the firm's hiring decision in the factor market. Because a firm's marginal revenue curve depends on the market structure of the output market, the firm's demand for labor will depend on the market structure in the output market as well. For example, if a firm operates in a perfectly competitive output market, then the firm's marginal revenue is equal to the market price. If a firm is a monopolist in the output market, then the firm's marginal revenue is less than the price.

Let's compare the demand for labor between a perfectly competitive industry and a monopoly that operate in identical output markets. Assume that both industries draw from the same labor market (e.g. the unskilled labor market) and that there are many other firms doing the same thing.

The demand for labor by firms in any industry is the product of marginal revenue and the marginal product of labor. Perfectly competitive firms have a marginal revenue that equals the market price (i.e. MR = P). If MR = P, then we can rewrite the demand for labor by firms in a perfectly competitive industry as ($P \times MP_L$). This term can be refered to as the marginal value product of labor (MVP_L).

Monopolists produce where their marginal revenue is less than the market price. Therefore, we cannot rewrite the monopolist's demand for labor as ($P \times MP_L$). The monopolist will have a demand for labor equal to ($MR \times MP_L$), which we call the marginal revenue product of labor (MRP_L).

27.2 Monopsonist Factor Market

In output markets with only one seller, there is monopoly. When there is only one buyer in a factor market, we have monopsony. A key difference between monopsonistic and competitive factor markets is how total costs change with the purchase of additional factors.

For example, in a competitive labor market, firms may purchase as many laborers as needed at the market wage because each firm's labor supply curve is horizontal. Because a monopsonist is the only buyer of labor in a particular labor market, the monopsonist faces the entire (positively sloped) market labor supply curve. The difference is clearly important when considering the marginal cost of hiring each additional worker.

When facing a positively sloped market supply curve, the firm's total costs change with the hiring of additional workers as follows:

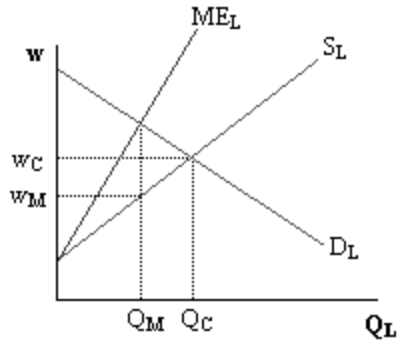
DTC = wDL + (Dw)L

If we divide both sides of the equation by DL and then rearrange, we have:

DTC/ DL = w + (Dw/ DL)L

This equation is refered to as the marginal expenditure of labor. Note that the marginal expenditure of labor (or marginal cost of hiring additional workers) is greater than the wage here. Therefore, the marginal expenditure curve is steeper than the market labor supply curve.

In addition, because the monopsonist hires a quantity of labor that coincides with where the marginal cost of hiring each additional worker equals the marginal benefit of each additional worker, the firm hires at the point where the firm's demand for labor equals the marginal expenditure for labor. This is illustrated on the graph below.



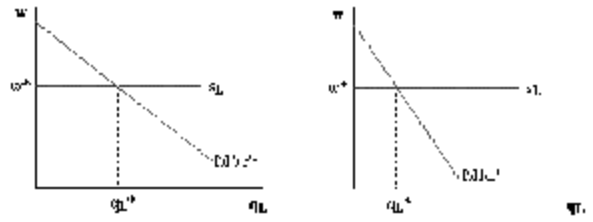
On the graph, the marginal expenditure equals the firm's demand for labor at the quantity Q_M. To get this quantity of workers to supply labor, the firm must pay a wage that comes from the labor supply curve. On the graph, that wage is w_M.

Contrary to the monopsonistic market, hiring in a competitive labor market occurs where the demand for labor equals the supply of labor. To provide a comparison with a similar monopsonistic labor market, the competitive market equilibrium values are included in the graph as well (Q_C and w_C). Note that a monopsonist would hire less labor and pay a lower wage than what would result in competitive labor market with similar demand and supply conditions.

27.3 Different Factor Markets

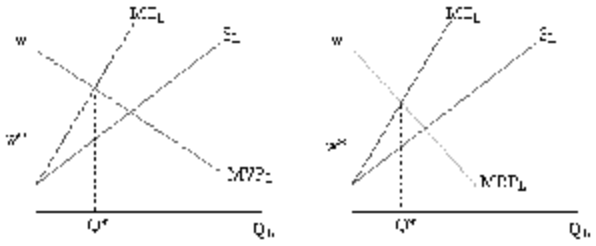
In the short run, let's assume that labor is the only variable factor hired by different firms. Equilibrium in the market depends on two things. First, we must know whether the firms buying labor are competitive or monopolistic in their respective output markets. Second, we must know whether the labor market itself is competitive or monopsonistic. These different situations are illustrated below.

Assume that the labor market is competitive, so that firms hire where demand equals supply. In this situation, firms face a horizontal labor supply curve. If the firm is perfectly competitive in its output market, then its labor demand curve is called the marginal value product (MVP). If the firm is a monopolist in its output market, then its labor demand curve is called the marginal revenue product (MRP).



The graph (above) on the left represents the perfectly competitive firm buying labor from this competitive labor market, whereas the graph on the right represents the monopolistic firm buying labor from this competitive labor market. In each case, labor is hired and paid a wage that corresponds with the point where the demand for labor equals the supply of labor.

If the labor market involves monopsony, then firms hire where their marginal expenditure curve crosses their demand for labor curve and pay a wage that corresponds with the labor supply curve. Depending on whether the firm is perfectly competitive in its output market or monopolistic, the firm's demand for labor is either called the marginal value product curve (left graph below) or the marginal revenue product curve (right graph below).

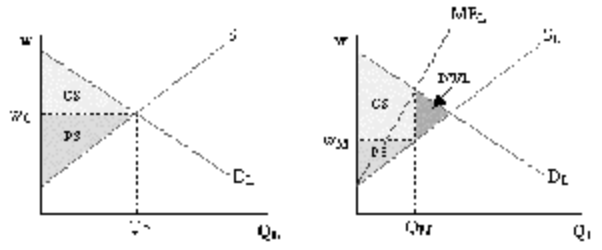


In both situations, and in the absence of any unionization, the wage is determined by where the vertical dotted line (coming down from the intersection of either MVP_L and ME_L to Q* on the left graph or from MRP_L and ME_L to Q* on the right graph) crosses the labor supply curve. The wage in each case is given as w*.

27.4 Factor Markets and Surplus

Just as in output markets, factor markets give rise to consumer and producer surplus. We can define these surpluses in the same manner as with output markets. Consumer surplus in a factor market is the difference between maximum acceptable factor price and the actual factor price for all units purchased. Producer surplus becomes the difference between the actual factor price and the minimum acceptable factor price for all units purchased.

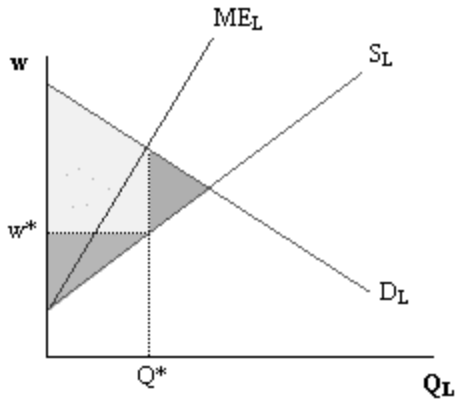
Let's consider a world where there is only one variable factor (labor). In a competitive labor market, firms hire where their labor demand equals the existing labor supply. Considering the labor market as a whole, this implies an equilibrium such as that found in the left-side graph below. Consumer surplus is the greenish area bordered by w_C and D_L - all the way out to Q_C. Producer surplus is the purplish area bordered by w_C and S_L - all the way out to Q_C. These surpluses are labelled as CS and PS respectively.



If the labor market is monopsonistic, then the monopsonist hires labor to the point where ME_L crosses D_L. The wage (w_M) comes from the point on S_L that is directly above Q_M.

Because ME_L is above S_L, the monopsonist hires less labor than would be hired in a more competitive labor market. As a result, the surpluses change in size. Consumer surplus extends down into former producer surplus, whereas producer surplus shrinks. If labor is not hired at the point where demand equals supply, then we also have some deadweight loss (given in the graph as the tan area called DWL).

Although the existence of DWL implies an inefficient allocation of labor, the size of this area can differ across otherwise similar labor markets. This is because DWL results from the substitution that occurs within the demand or supply side of the market. For example, the graph below shows that an increasingly inelastic labor demand curve decreases the size of DWL. As the labor demand curve becomes steeper, we realize that the labor from other labor markets is an increasingly poorer substitute for the labor in this market. At the extreme, where D_L is perfectly inelastic (i.e. D_L is vertical), there would be no DWL at all.



Notes

TUTORIAL 6

1. Multiple Choice Questions

Top of Form

1. A perfectly competitive firm is operating at an output where price is greater than marginal cost. Therefore:

- the firm should produce more to maximize profit.
- the firm should produce less to maximize profit.
- the firm is making a profit.
- the firm is taking a loss.
- the firm should shut down.

2. If your new business (in which you have invested $50,000, which you previously kept in savings, earning 10% interest) earns an accounting profit of $30,000 in the first year, and you had previously been employed as a ditch-digger, earning $10,000 per year, your economic profit is:

- -30,000.
- -15,000.
- $0.
- $15,000.
- $20,000.

3. Fill out Table 1, and use it to answer the following three questions.

Table1

Q	***VC***	***TC***	***MC***
0	0	50	—
1	10		
2	30		
3	60		
4	100		
5	150		

If the price of the product is $40, at what output does this firm maximize profit?

- 1
- 2
- 3
- 4
- 5

4. Refer to Table 1. If the price of the product falls to $20, what should this firm do?

- produce 0 (shut down) because profit is less than zero
- reduce production to 1
- reduce production to 2
- reduce production to 3
- do nothing; the choice is still optimal

5. You've been hired by an unprofitable firm to determine whether it should shut down its operation. The firm currently uses 70 workers to produce 300 units of output per day. The daily wage (per worker) is $100, and the price of the firm's output is $30. The cost of the other variable inputs is $500 per day. Although you don't know the firm's fixed costs, you do know that they are high enough that the firm's total cost exceeds total revenue. What is the best recommendation you can make, based on the information that you have?

- exit the industry in the long run
- shut down in the short run
- continue to produce in the short run
- continue to produce in the long run
- not enough information to tell

6. Which of the following conditions exist in long-run competitive equilibrium?

- P = LRAC
- Individual firms operate at the most efficient scale of plant.
- The level of output produced coincides with the minimum point on the LRAC curve.
- All of the above.

7. Assume that a perfectly competitive industry is in long run equilibrium. If demand increases, which of the following will occur?

- market price will increase.
- firms will produce more output.
- firms will increase their profits.
- the industry will not be in long-run equilibrium.
- All of the above are correct.

8. Think about a graph showing per-unit cost and revenue numbers for a competitive firm in the short run. When computing the amount of profit obtained from any level of output produced in this graph, which of the curves below is NOT necessary?

- The firm's demand curve.
- The SATC curve.
- The MC curve.
- None of the above. All of these three curves are necessary to estimate the amount of profit.

9. Think about a graph showing per-unit cost and revenue numbers for a competitive firm in the short run. To determine what level of output the firm should produce in order to maximize profit, one of the curves below can be discarded. Which one?

- The firm's demand curve, either D_0 or D_1.
- The SATC curve.
- The MC curve.
- None of the above. All of these curves are necessary.

10.The firm's short-run supply curve is:

- the marginal cost curve.
- the marginal cost curve above average variable cost.
- the marginal cost curve above average total cost.
- the total cost curve.
- the total revenue curve.

11.Refer to Figure 1. At a price of $15, how much should the firm produce?

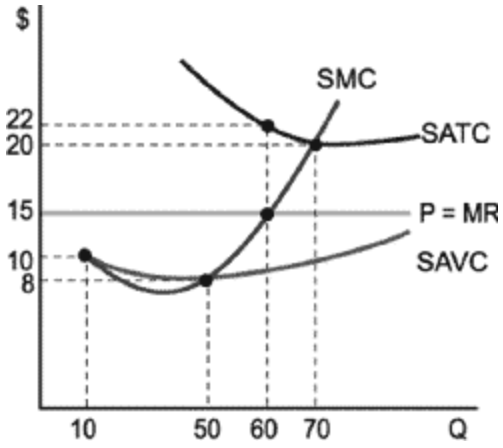


Figure 1

- 0
- 10
- 50
- 60
- 70

12.Refer to Figure 1. What is the lowest price at which this firm will produce anything?

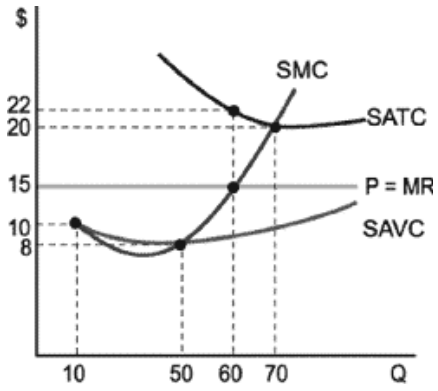


Figure 1

- $8
- $10
- $15
- $20
- $22

13.Refer to Figure 1. What is the lowest price at which this firm can break even?

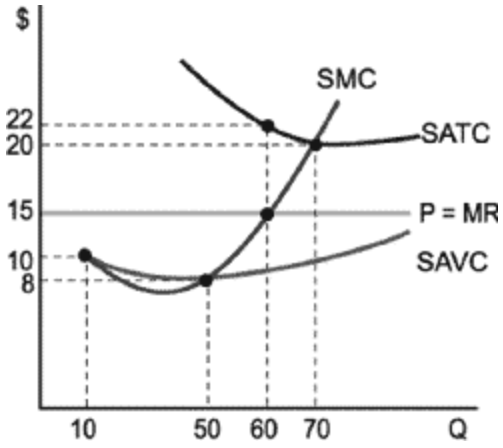


Figure 1

- $8
- $10
- $15
- $20
- $22

14.When a firm expands its scale of operations, and such expansion leads to lower cost per unit, the firm faces:

- Constant returns to scale.
- Decreasing returns to scale.
- Increasing returns to scale.
- Diminishing returns.

15.A firm exhibits economies of scale:

- Along the decreasing portion of the long-run average cost curve (LRAC).
- Along the increasing portion of the long-run average cost curve (LRAC).
- Exactly where LRAC reaches its minimum point.
- Anywhere along the LRAC, as long as increasing the scale of operations does not affect cost per unit.

16.If the fast-food industry is known to be a constant cost industry, which of the following best explains the effect of an increase in demand?

- In the short run, prices will rise, inducing firms to enter, and causing the price to return to the original price in the long run.
- In the short run, prices will rise, inducing firms to enter, and causing the price to fall but remain higher than the original price in the long run.

- In the short run, prices will rise, inducing firms to enter, and causing the price to fall below the original price in the long run.
- In the short run, prices will fall, inducing firms to exit, and causing the price to return to the original price in the long run.
- In the short run, prices will fall, inducing firms to exit, and causing the price to fall below the original price in the long run.

17. When the long run supply curve is horizontal,

- The industry is a constant-cost industry.
- The industry is a decreasing-cost industry.
- The industry is an increasing-cost industry.
- The industry experiences external economies.

18. Why is the long-run supply curve expected to be flatter than the short-run supply curve?

- because entry of new firms drives up input costs in the long run
- because firms experience diminishing returns in the short run but not the long run
- because demand is more likely to increase in the short run
- because firms cannot adjust inputs in the long run
- because economic profit must be zero in the long run

19. Decreasing-cost industries occur when:

- input prices rise as the industry expands output.
- input prices remain constant when the industry expands output.
- input prices fall when the industry expands output.
- the long-run supply curve shifts leftward.
- the long-run supply curve shifts rightward.

20. In the long run, entry will occur in a perfectly competitive market:

- as long as economic profits are greater than zero.
- as long as there are no barriers to entry.
- as long as demand is increasing.
- as long as price is greater than short-run average variable cost.
- all of the above

2. Long Answer Questions

1. Explain both the potential benefits and potential costs of allowing mergers. Under what circumstances is the Department of Justice most likely to allow mergers? Under what circumstances are mergers most likely to be disallowed?
2. How did the deregulation of air travel affect the price of air travel?

3. Case Study - I

Auto Insurers and Price Discrimination

LEAD STORY-DATELINE: The Wall Street Journal**, Wednesday April 7, 2004.**

Automobile insurers have practiced price discrimination for years. Good student discounts, multiple policy discounts, and discounts based on driving record are all ways to segment the market and practice price discrimination. Now some insurance companies are experimenting with 5% to 10% occupation discounts.

Farmers Insurance Group offers residents in some states discounts if they are police officers, firefighters, nurses, scientists, engineers, or medical doctors. Other insurers have extended these discounts to include teachers. Many of these discounts are pilot programs that will be extended if successful.

One of the primary reasons the programs have yet to be expanded is the problem of occupation verification. For now, many rely on customer honesty. Other companies do spot checks. And ironically, while many of these programs offer discounts to medical doctors, this occupation is the second highest accident-prone profession calculated using number of accidents per 1,000 individuals. Students rank at the top with more accidents than any other group.

Thinking About the Future!

Now that you are aware of price discrimination, I suspect you can identify several more examples. Look around and see what other businesses practice price discrimination and see how they accomplish this task. For example, think about how banks segment customers into groups that receive different prices.

Talking It Over And Thinking It Through!

1. What is price discrimination?
2. Why would a firm practice price discrimination?
3. Provide some other examples of price discrimination.

4. Case Study - II

Will Andersen Conviction Benefit Enron's Shareholders?

LEAD STORY-DATELINE: Wall Street Journal**, June 19, 2002.**

Arthur Andersen LLP's conviction looks like it could provide ammunition to Enron Corp. investors suing the accounting firm. Not only did the government's recently concluded criminal trial provide damning testimony about Andersen's activities, the guilty verdict will be admissible in civil cases. This particular judgment, however, will not help Enron shareholders get their hands on any cash. Since Andersen will cease auditing publicly owned clients by August 31, the firm's financial picture is murky. This means plaintiffs' prospects for financial recovery in their civil cases increasingly will ride on their ability to hold Enron's banks and law firms liable. Under federal securities law, however, plaintiffs will have to prove the firms had intent to defraud or were recklessly indifferent.

Another potential source of recovery for Anderson and for the plaintiffs might come from Andersen's overseas affiliates which are required by their contracts to pay 1.5 times last year's revenue to break away. To date, June 2002, few of these payments have been made. Still, the Enron shareholder claims are "contingent," meaning that it hasn't yet been determined how much, if anything, Andersen will owe on them. If Andersen files for bankruptcy-court protection before closing its doors, litigants would have to take their claims to bankruptcy court, then stand in line with other unsecured creditors for a piece of the remain-

ing cash. One advantage for the plaintiffs, the burden of proof in civil claims is far lower than that in criminal cases, such as the government's successful action against Andersen. The lawsuit on behalf of Enron employees alleges that Enron and its bankers conspired to defraud Enron workers out of their retirement money. If proven, J.P. Morgan & Co., Citigroup Inc., Credit Suisse First Boston, and Vinson & Elkins would be welcome sources of collection for potential damages.

The limits to limited-liability partnerships, designed to shield its partners from the liabilities of the firm, have never been tested. Most legal beagles believe LLP shields partners from malpractice claims, but don't extend to partners implicated in wrongdoing. What does all this have to do with economics? We will explore those implications as we talk it over and think it through. The potential damage to the economy and to future economic policies is far-reaching.

Thinking About the Future!

Frank Savage was a board member of Enron and of the investment firm Alliance Capital Management, which until recently was Enron's largest institutional investor. Alliance was dangerously close to last in selling Enron stock. Alliance bought large blocks of the stock on August 15, 2001-the day after CEO Jeffrey Skilling resigned-and continued to buy even after Enron's October 22 announcement that it was under investigation by the SEC. By the time Alliance sold its 43 million shares of Enron stock, it had lost hundreds of millions of dollars for its investors, including $334 million from the Florida state pension fund. Governor Jeb Bush and state officials sued Alliance for negligence. As a board member of both companies, Savage deserves special recognition for poor business judgment or worse. Nell Minow, a corporate watchdog running the Web site *The Corporate Library*, and others are developing ratings systems for board members based on factors such as attendance and prior performance. In this way, board members such as Savage may be shamed into early retirement from corporate boards. Ironically, however, the impetus for reform may be money. Taking a plank from the idea of "market efficiency," the insurance industry is eager for director rankings. This industry foots the bill when boards like Enron's fail. In the future, with Adam Smith-styled markets back to work, companies with lousy boards will pay hefty premiums for directors' and officers' insurance. Markets work, but not to benefit everyone equally. That's the rub, or is it?

Talking it Over and Thinking it Through!

1. Why is it economically important to protect shareholders rights and provide restitution to shareholders when their losses are due to fraud?
2. Adam Smith's invisible hand of the market - which translated the pursuit of self-interest into public benefit - recently has been interpreted to mean that our self-interest in the stock market's performance would benefit the public. Are there different consequences for greed than for Adam Smith's concept of self-interest? Why or why not?
3. Sometimes markets are described as if they operate independently of other institutions such as property rights and the judicial system. What lesson can be learned from the Andersen/Enron case regarding this seeming independence?
4. During the past two decades, the Securities and Exchange Commission (SEC) has been very lax in its oversight of corporations and of the markets. Do you believe that it should become more vigilant since the Andersen/Enron case? Why or why not?
5. What was the essential reason for the establishment of limited liability partnerships (LLPs)?

Notes

LESSON 28:
LABOUR

1. Profit Maximization and The Demand for Labor

As we learned in Previous chapter, the way to maximize profits then is to hire enough labor so that

VMP=wage

where the wage is the price of labor, per hour or week or year as the case may be, and VMP stands for the Value of the Marginal Product. In turn the VMP is defined as follows:

VMP=p*MP

where p is the price of output and MP is the marginal productivity of labor in units of output. We may think of the VMP as the marginal productivity of labor in money terms.

Using the example of producing potatos, p would be the market price of potatos, MP would be the marginal productivity of labor in the potato industry. We recall the definition of

marginal productivity of labor:

the additional output as a result of adding one unit of labor, with all other inputs held steady and ceteris paribus.

Thus, the value of the marginal productivity (in the potato example) is the market value of the additional potatos produced by one additional worker in the potato industry.

Let's use these ideas to visualize the demand for labor. On this page, we will make one simplifying assumption: the price of the output (the price per bushel of potatos, for example) is held constant as the wage varies for an individual firm. (We'll see in the next page what difference the simplifying assumption makes).

We can visualize the VMP and the wage as in the following figure:

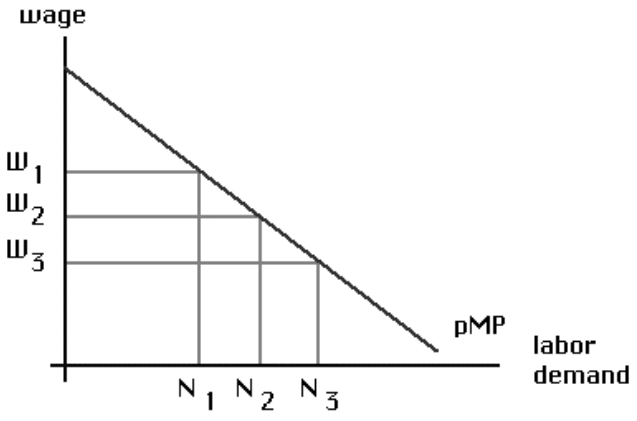


Figure 1: VMP = wage

As the wage drops from W_1 to W_2 and then to W_3, we see the demand for labor increasing from N_1 to N_2 and then to N_3. The firm is moving down the curve of diminishing marginal productivity, and the cheaper labor is, the more sense it makes to move further down that curve, and the further down the curve the profit-maximizing labor input is.

What this is telling us is that (with the output price constant) the VMP curve is the firm's demand curve for labor. Remember the definition of a demand curve: it tells us, for each price, what will be the quantity demanded. The price of labor is the wage, and the VMP curve tells us, for each wage, what is the quantity of labor demanded.

a. A Complication

But, in general, the price of the output will not remain steady as the wage changes. Here's the reason: if labor is cheaper, to the industry as a whole, that will shift the industry supply curve to the right. That will lead to lower prices for industry output. This leads to the slightly more complicated diagram in figure 2:

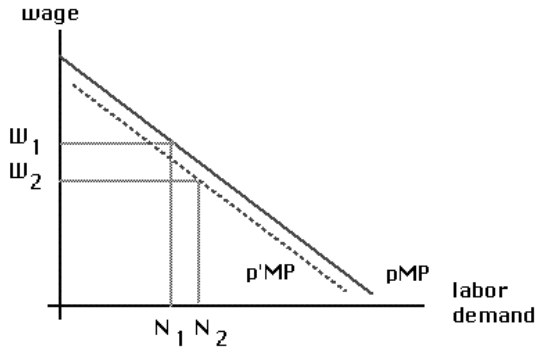


Figure 2: Changing Output Prices

As the wage drops from W_1 to W_2, the price of industry output drops from p to p', shifting the pMP curve to the left. The new pMP curve is shown by the dotted green line. This leads to a new profit-maximizing labor input at N_2. This means that the industry demand curve of labor is actually a little steeper than the VMP curve for a constant output price

b. Labor Supply

The supply and demand approach is based on both — supply and demand. So we need to consider the supply as well as the demand for labor.

Any individual worker's supply of labor will depend on her or his opportunity for income from sources other than labor and on her or his preferences between leisure and earning income. When we look at the supply of labor from the point of view of the economy as a whole, this can lead to some surprises. Some economists argue that the labor supply curve could slope backward, for at least a part of its range. This is shown in figure 3, below.

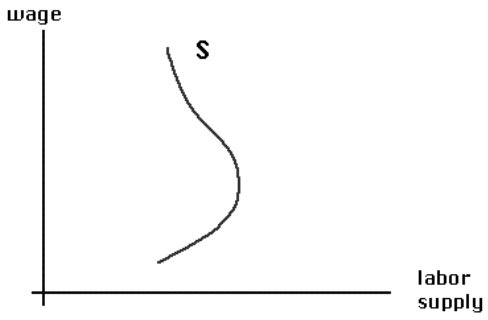


Figure 3: Supply of Labor to the Economy as a Whole — Perhaps

Here is the idea: when people earn a higher wage per hour, they can earn more income for working the same or even fewer hours. When people can "have it all," they often choose to do just that — have more of all the good things. Thus, when wages per hour of labor rise, people are getting better off, and eventually they decide to take some of their increased potential income in the form of leisure rather than money. That means the quantity of labor supplied — in hours per year — is less, and the labor supply curve slopes backward (as shown) at the higher wage levels.

When we look at the supply of labor to a particular industry, we don't need to worry about this. For example, when wages paid by the potato growing industry are low, most people will find that they can make more money in other industries, and so they don't supply labor to the potato industry. When potato grower wages rise, some of those people will find that they now can earn more in the potato industry than in their alternatives, and will switch their labor supply into the potato industry. Thus, the supply curve of labor to the potato industry will be upward sloping. Since the supply of labor to a particular industry is dominated by this switching-back-and-forth from other industries, the supply curve of labor to an individual industry will usually be upward sloping, From now on, we will show them as upward sloping.

c. Equilibrium in the Market for Labor

Now we can put the supply and demand for labor together and discuss an equilibrium. This is shown in Figure 4, below. Recall, the price of labor is the wage, shown by w, and the quantity demanded is the number of labor-hours employed, shown by N. The demand for labor is the pMP, the price of output times the marginal productivity of labor in units of output. As usual, the equilibrium price (wage) and the equilibrium quantity demanded and supplied (employment) are at the point where the supply and demand curve intersect.

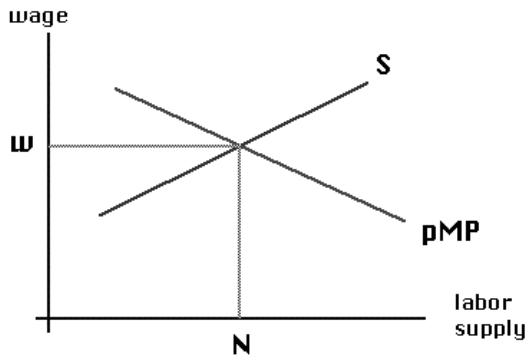


Figure 4: Supply and Demand for Labor in an Industry

"The proof of the pudding is in the eating," so let's look at an application.

2. Labor Market with a Minimum Wage Law

One important application is to analyze the effects of a minimum wage law. This is shown in Figure 5, below. In the figure, the minimum wage is at W, above the equilibrium wage. As a result, employers will demand and hire only N_d labor-hours, less than would be hired at the equilibrium wage. On the other hand, N_s labor hours are supplied, and we have an excess supply.

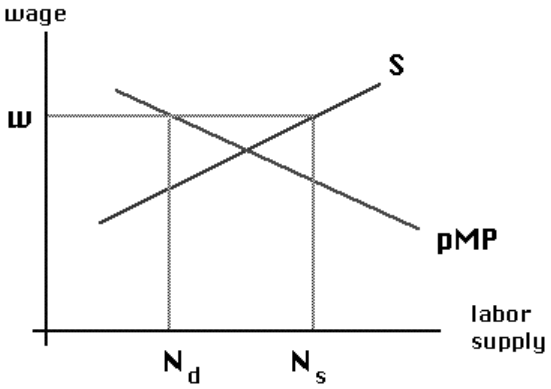


Figure 5: Supply and Demand for Labor with a Minimum Wage Law

Workers still employed under the minimum wage law are presumably better off, but there are workers offering N_s-N_d labor hours who cannot find jobs in the industries covered by the minimum wage. What are they to do? They might

a. Shift into industries with equilibrium wages above the minimum wage.

But most will not be able to do this — if they could get jobs at higher than minimum wages, they probably would have done it already.

b. Shift into industries that pay less than the minimum wage but are not covered by the minimum wage law.

Over time, the number of such industries has decreased, but there are still some. They will be working for lower wages and be worse off in this case.

c. Become self-employed in some very small enterprise.

Again, they will presumably obtain less income and be worse off — otherwise, we would suppose that they would have shifted before the minimum wage law was enacted.

d. Drop out of the labor force entirely.

Some may retire, or rely on the income of spouses or relatives, while some may drop out of the legal labor force to engage in illegal "hustling" for an income.

e. Become unemployed.

Unemployment is really a macroeconomic concept, but in the simplest terms people who are looking for a job and not finding one are said to be unemployed.

Many economists believe that a portion of them will become unemployed. In any case, this analysis leads the majority of economists to believe that minimum wage laws are a poor policy. Presumably they are intended to help wage-earners, but at least some wage-earners are worse off as a result of the minimum wage laws. While there has been some controversy in all this, and the controversy has been renewed in the 'nineties, it has not shaken the predominant feeling of economists that minimum wage laws have some very undesirable side-effects.

3. Criticism of the Supply-and-Demand Approach to Labor Markets

Not all economists would accept the supply-and-demand analysis of labor markets, employment and wages. There are several criticisms of the supply-and-demand model of labor markets and the John Bates Clark model of the business firm. Each of the criticisms could lead to an alternative analysis of labor markets. Since this is not a text of labor economics, we will just mention the criticisms, but not attempt to sketch the alternative analyses. Critics mention the following points, among others:

- Wages and income distribution may be influenced or determined by bargaining power.
  - Labor unions can create bargaining power for employees
  - Employers may limit wages because of limited competition for employees and by means of wage discrimination.
- Perceptions of fairness may influence wages and working conditions.
- Productivity itself may be influenced by wages and working conditions.
- The Marxist view is that employment is a social relationship based on exploitation, and that wages and labor cannot be understood except in those terms.

4. Criticisms Related to Bargaining Power

Some critics stress the importance of bargaining power in influencing wages and employment, shifting them away from the supply-and-demand equilibrium or replacing supply and demand completely as the determinant of wages. Bargaining power may be exercised by employers or employees or both:

- Employers may influence the wage by restricting their hiring. When the wage is below the supply-and-demand equilibrium, employers find it profitable to hire more workers **at the going wage,** but hiring the additional workers could force wages higher (moving upward along the supply curve of labor), leaving the employers worse off on net. Conversely, the employers as a group may have higher profits with less labor, but a lower wage — lower on the supply curve of labor. Strong competition for employees, in which each employer offers higher wages to get more employees according to their marginal productivity, would eliminate the low wages and lead the labor market to the supply-and-demand equilibrium. However, if employers can avoid this competition for labor, they can keep wages down and profits higher. Some critics believe this is possible because labor markets are typically segmented by skill, experience and location. As a result of this segmentation, they feel many labor markets are dominated by one buyer (called **monopsony**) or a few buyers (called **oligopsony**).
- Employers may also keep their wage costs down by **discrimination,** paying different wages to different workers and, as nearly as possible, paying each one the minimum necessary. Some employees may accept lower wages than others because their alternatives are limited by gender, race, age, disability, or unpredictable personal characteristics. Having worse alternatives, they may accept lower pay and thus allow employers to expand their work forces without moving up the labor supply curve and paying higher wages overall. Again, strong competition for labor would tend to limit (perhaps not eliminate) wage discrimination.
- On the employees' side, **labor unions** can favor the employees' bargaining power (again) by limiting competition for jobs among the employees and potential employees. Where this is very successful, it leads to **collective bargaining** between employers and employees. That is, the union or a group of unions, and the employer or a group of employers together negotiate a contract that determines wages and conditions for a whole group of employees, sometimes a whole industry or group of industries. This factor may be exaggerated — only about 20% of the American labor force is unionized — but unions are more important in some other industrialized countries and collective bargaining settlements may influence the wages and conditions for non-unionized employees. Also, a group of employees may find means of limiting their competition in the absence of unions.
- In the absence of any union, when wages are determined by **individual bargaining**, the bargaining power of the individual workers may be an important influence on wages. This is especially likely when employers practice wage discrimination. However, the average bargaining power of the individual worker may be increased or decreased by general conditions in society, such as the distribution of political power and — of course! — the overall balance of supply and demand for labor.

These criticisms lead a minority of economists to question the analysis of minimum wage laws given in an earlier page. In that discussion, a minimum wage law leads to a decrease in employment as the market moves up the demand curve for labor. But in some of the models based on bargaining power, employ-

ment is not on the labor demand curve. In that case, the impact of a minimum wage law on unemployment is unpredictable.

5. Perceived Fairness

Fairness is, of course, normative economics — to say that something is fair is to say that "it ought to be." But a minority of economists believe that perceptions of fairness may have an influence on wages paid. This is especially likely if bargaining (either collective or individual) plays a role in determining wages. If workers see a low wage offer as "unfair," they may be more likely to resist it, making it more difficult for employers to insist on it. And employers may be more likely to agree to a wage demand if they perceive it as "fair."

Thus, **perceptions** of normative economic concepts, such as fairness, can have an impact on "what is," positive economics. But perceptions are changeable and may depend on wider social attitudes and customs. For example, it appears that employers are less concerned with "fairness," and more determined to maximize profits by reducing wage costs, in recent years than they had been in the "good old days." In recent years social attitudes have been more accepting toward this sort of competition, and it may be that this accepting attitude has changed the behavior of business. But it may be that only the rhetoric has changed, and that employers are more open about their decisions than they were when social attitudes were less favorable to "profit maximization."

Unfortunately, we **know** very little about the impact of perceived fairness, but we cannot rule it out as a factor in wage bargaining.

6. Effects of Wages and Working Conditions on Productivity

The supply-and-demand approach as we have seen it here assumes that the marginal productivity curve remains unchanged as wages and working conditions change. But there is some evidence that a change in wages or working conditions can shift the marginal productivity curve upward or downward, changing the relationship between the number of units of labor employed and the marginal productivity of labor. In particular, a cut in wages may shift the marginal (and average) productivity downward.

- In very poor countries, lower wages may lead to reduced nutrition and worse health for the employees, and thus to lower productivity.
- In richer countries, a wage above the employee's alternative (that is, above the supply curve) can limit the turnover of the work force. Turnover is costly in itself, and with less turnover the employees have more opportunity to learn to work well together, increasing productivity.
- Wages above the supply curve can also increase the incentive to work hard and to "work smart." If wages are only on the supply curve, then the worker has less to lose if she or he is dismissed on grounds of not working hard enough.
- The perceived fairness of a wage above the supply curve may also make the employees more willing to work hard and to "work smart."

Now, suppose that an employer cuts the wages (starting above the supply curve of labor) in the hope of cutting overall costs and so increasing profits. Let's call that the **direct cost effect.** At the same time, the lower wages shift the marginal and average productivity downward. This increases costs, offsetting the wage cut. Let's call that the **productivity effect.** Profits may either increase or decrease, depending on whether the direct cost effect is bigger than the productivity effect. Some economists argue as follows: when wages are very high, the direct cost effect will be greater than the productivity effect; but as wages drop the productivity effect will increase and the direct cost effect will decline, so that there is a particular wage (above the supply curve of labor) that gives maximum profits. This is called the **"efficiency wage."**

If market wages are often efficiency wages, then labor markets could behave quite differently than the John Bates Clark supply and demand approach suggests. But what is the evidence?

7. Evidence on Productivity Effects

We have little direct evidence on the efficiency wage hypothesis, but there is a good deal of evidence on related issues.

- Profit sharing is a form of labor compensation in which the employees get higher wages when company profits are higher. The idea is that profit sharing increases the incentive to work harder and "work smarter," and thus increases profits. On the whole, the studies confirm this, showing that there is at least some scope for increasing profits through the productivity effect.
- Studies of unionization provide evidence that, on the whole, unionized companies have higher labor productivity than non-unionized companies (except, of course, when the union is on strike!) This is a surprising result, and there has been some controversy about how to explain it. One possible explanation is that the higher productivity comes from the productivity effect of better wages (and working conditions) the unions were able to obtain. The productivity effect does not seem to increase profits in this case, however. Apparently the increase in productivity is just about balanced out by the increase in wages.
- A **cooperative enterprise** is an enterprise in which the directors and officers are elected directly by the employees. The employees may be the owners of the enterprise or the enterprise may be owned by some nonprofit, public, or philanthropic agency. Such enterprises have existed in various parts of the world (including the United States, Britain, and other "capitalist" countries) for over 150 years. Studies of cooperative enterprises show that they generally do better than investor-controlled or government enterprises in labor productivity, although they can have difficulty raising investment capital. Presumably they are benefiting from a large productivity effect, although money wages may not be the main reason for it — because of their difficulties raising investment capital, their wages are not necessarily high.
- Many other enterprises are more or less compromises between investor-controlled and cooperative forms. For example, the employees may elect a certain proportion of the Board of Directors. This is called **Codetermination.** In

Germany all of the large corporations are required to have half of the members of their Boards of Directors elected by the employees. This is called **Codetermination with Parity.** International comparisons are difficult, but the German economy has certainly been a strong performer, with high labor productivity, very high wages, and a very strong export balance. Comparisons of different enterprises within countries are easier. Studies of this kind have compared enterprises that come nearer the cooperative end of the spectrum with other enterprises that are nearer the investor-controlled end of the spectrum and, allowing for some differences in circumstances, the more nearly cooperative enterprises have had higher labor productivity on the average.

Thus, there is some evidence that productivity effects can be important, but the evidence also suggests that the organization of the enterprise is at least as important as the money wage level in creating productivity effects.

8. Marxist Labor Economics

For Marxism, of course, labor and employment are the central concepts of social science. However, a Marxist would reject not only the supply-and-demand approach to labor markets, but many of the neoclassical criticisms and alternative models of "labor markets."

We can only sketch a few concepts of Marxism here. From the Marxist point of view, only labor is productive, so that all production is attributable to labor. Marxist ideas were developed before the marginal productivity approach was known, and so marginal productivity plays no role in the Marxist approach. Many Marxists would see the marginal productivity approach as relevant only in the short run; by contrast, the Marxist analysis is a long period analysis.

Also, of course, Marxism accepts the labor theory of value. This is the meaning of the phrase "only labor is productive:" specifically, only labor produces **value.**

Marxism says that, in the long run, the price of any commodity is the same as its average cost. Cost and price are measured in labor value, so this is the same as saying that, in the long run, the price is equal to the value. Since labor is a commodity in a capitalist system, the price of labor — the wage — will be equal to the average cost of production of **labor.** Since the productivity of labor is greater than its average cost of production, labor produces some surplus that is not paid out as wages, even in the long run. In turn, the cost of production of labor is itself measured in labor value, so we are saying that labor creates more value than is required to produce the labor. The difference, in terms of labor value, is called "surplus value." Surplus value is the source of profits in a capitalist system

Thus, on the Marxist view, labor is "exploited" through the wage employment system, which transfers the surplus value of labor to the capitalist class. In the short run, this can be offset or exaggerated by such factors as monopoly power, temporary scarcities of investment funds, machinery or natural resources, or political interference in the marketplace. In this context, marginal productivity may play its short-run role. The labor values define a **typical** capitalist system, in that prices in any real period tend toward the labor values. If, hypothetically, a capitalist system were to exist for an indefinite period without any new disturbing factors, the prices would eventually equal the labor values. However, no capitalist system could exist for an indefinite time — the same tendencies will produce a revolution and lead to some new kind of system. Nevertheless, Marxists regard the surplus-value theory as the correct general explanation of profits and the most useful intellectual tool for understanding wages, labor and employment in a capitalist system.

9. The Marginal Productivity Approach in General

So far, we have applied the marginal productivity approach to analyze the demand for labor, in the Neoclassical tradition. However, the marginal productivity corresponds to the demand for any input. In general, we can define the

marginal productivity of any input as

the additional output as a result of adding one unit more unit of that input, with all other inputs held steady.

In algebraic terms, that is approximately

$$MP_{input} = \frac{\Delta Output}{\Delta Input}$$

that is, the increase in output divided by the corresponding increase in the input, while the other inputs do not vary.

We can then define the value of the marginal product for any input as

$VMP_{input} = p * MP_{input}$

where p stands for the price of the output, as before. In general, we may think of the VMP as the marginal productivity of the input in money terms. The rule for maximization of profits, for each input, is to increase the use of the input until

$VMP_{input} = price_{input}$

Now let's apply that to the land input in next lesson.

Notes

LESSON 29:
LAND, CAPITAL AND NATURAL RESOURCES

1. The Marginal Productivity Approach in General

So far, we have applied the marginal productivity approach to analyze the demand for labor, in the Neoclassical tradition. However, the marginal productivity corresponds to the demand for any input. In general, we can define the **marginal productivity** of any input as the additional output as a result of adding one unit more unit of that input, with all other inputs held steady.

In algebraic terms, that is approximately

$$MP_{input} = \frac{\Delta Output}{\Delta Input}$$

that is, the increase in output divided by the corresponding increase in the input, while the other inputs do not vary.

We can then define the value of the marginal product for any input as

$$VMP_{input} = p * MP_{input}$$

where p stands for the price of the output, as before. In general, we may think of the VMP as the marginal productivity of the input in money terms. The rule for maximization of profits, for each input, is to increase the use of the input until

$$VMP_{input} = price_{input}$$

Now let's apply that to the land input.

2. Land

Now, let's apply the marginal productivity approach to land. We may think of a potato farmer who is considering renting additional land to farm. How much land? Of course, that will depend on the rent per acre — the price of land.

Using the general formulae for marginal productivity and the value of the marginal product from the previous page, we can define the marginal productivity of land as

$$MP_{land} = \frac{\Delta Output}{\Delta Land}$$

that is, the increase in output (measured in bushels of potatos) divided by the corresponding increase in the number of acres of land used, while the other inputs do not vary.

The value of the marginal product of land will be

$$VMP_{land} = p * MP_{land}$$

where p, again, stands for the price of the output — a bushel of potatos, in this case.

Continuing the example of producing potatos, the value of the marginal productivity (in the potato example) is the market value of the additional potatos produced on one additional acre of land. The farmer will increase the number of acres of land he rents until

$$VMP_{land} = price_{land}$$

And so the value of the marginal product of land is the demand curve for land of a standard quality.

3. Differential Rent

However, land is not a homogenous resource, and that important complication cannot be skipped over. It is the basis of the theory of rent, first proposed by English economist David Ricardo, and still considered the correct theory of rent by just about all economists.

Some land is more fertile than other kinds of land, or more profitable because it is closer to markets; and some land is more suitable to one kind of crop than another. These differences in fertility will be reflected in the marginal productivities and therefore in the demands for the different kinds of land. Let us think of a very small economy with just three kinds of land: good, fair, and bad. There are just 10,000 acres of each kind of land. The supply and demand for each kind of land is shown in Figure 6 below:

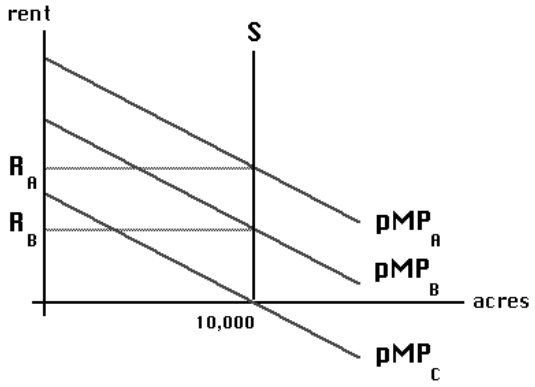


Figure 1: Marginal Productivity of Land with Different Fertilities

The demand for good land is pMP_A, for fair land pMP_B, and for bad land pMP_C. The supply of land of a particular quality is always a vertical line, because "they're not making any more of it" — the supply of land cannot be increased no matter how high the price. Since there are 10,000 acres of each sort of land, the three kinds of land have identical supply curves, all shown by the vertical line at S.

In a supply-and-demand equilibrium, then, the rent per acre of good land will be R_A. For fair land it will be R_B, and for bad land zero. The bad land in this example is what Ricardo called "marginal" land — good enough to be cultivated, but only if it can be had free of rent.

Thus, the rent on fair land is just enough to offset its greater productivity relative to the marginal land. Similarly, the rent on

good land is just enough to offset its productivity advantage over marginal land. If the rent of good land were any lower than that, no-one would want to use fair or marginal land, but all would compete for the limited supply of good land — forcing the rent on the good land up until it is large enough to offset the productivity advantage of that good land. Similarly, the difference in rent between the good and fair land is just enough to offset the productivity differential between them. This is called the "differential" theory of rent — that the rent of any land is just large enough to offset its differential productivity relative to marginal land. To stress the basis of land rent, it is often called differential rent.

This idea — that rent is based on differential productivity, which is given by nature and not the result of the landowner's action — is what led the American social activist, Henry George, to propose that land rent ought to be largely confiscated by taxation.

4. Natural Resources

Natural resources include such things as standing forests and schools of fish, deposits of petroleum, copper ore and gold, and similar biological and mineral resources. Traditionally, these resources are lumped with land in economics. One reason for this tradition is that the market prices of all natural resources include an element of differential rent.

The case is simplest for such renewable natural resources as second-growth forests. Essentially, woodland is one crop that may be grown on the land, and if the land is best suited for woodland, its rent will be based on its value in growing timber. The productivity of land in this sort of use will also depend on the **cost of extraction.** If the land is very hilly or swampy, then it may be difficult to use machinery, or more costly machinery may be needed — raising the cost of harvesting the timber. With natural resources in general, cost of extraction is inversely related to productivity and therefore rent.

For mineral resources such as petroleum and gold (and old-growth forests), no renewal is possible in an economically meaningful time frame. Nevertheless, the supply-and-demand price of natural resources will include a component of differential rent. The cost of extraction of mineral resources will vary. For example, petroleum at great depth or under deep water will cost much more to drill and operate the well than shallow, dry-land oil deposits. Thus the shallow, dry-land deposits with lower costs of extraction are more productive. In general the cheaper deposits will be extracted first. But oil or minerals from costly deposits must sell for the same as those from deposits that are cheap to extract, so the ones from deposits with low costs of extraction will be sold at over their cost, the difference being differential rent.

5. Capital

The third of the classical "factors of production" is capital. The earliest generation of economists did not think of capital as being an independent factor. But as the nineteenth century wore on, it was increasingly clear that capital (and specifically, mechanization) is a key factor in modern production. It was Nassau Senior who pointed out that capital investment in and of itself would increase productivity. Senior wrote "That the powers of Labor, and of the other instruments that produce wealth, may be indefinitely increased by using their Products as the means of further production."

As we know, capital is more than just machinery, but it may be helpful to think in terms of a specific kind of machine. We may think of a tractor to be used on the potato farm. As we increase the number of machines in use, with the same amount of land and labor, output will increase, but at a decreasing rate. Capital, like the other inputs, is subject to diminishing returns.Once again, we will focus on the "marginal productivity" of the machines. On the other hand, the costs of using the machinery will also increase as the number of machines increases.

The machines will gradually wear out and will have to be replaced.

It is customary to deduct wear and tear from the output, so that the total and marginal productivity are **net of wear and tear.** But, for practical applications, we should remember that wear and tear is a real cost and must be taken into account.

The resources "tied up" in the machine have an opportunity cost.

The money laid out to buy the machines pays for the resources used in producing the machines. The money (and resources) will be recovered only gradually, using the machines to produce goods and services. However the money (and the resources) could be used for other purposes. For example, the investor might instead have bought a vineyard, which would produce a crop of wine grapes every year. It will not make sense to invest in the machine unless the net revenue from using the machine to produce goods and services is worth at least as much as the wine grapes. Similarly, the investor might instead have leant his money to someone else, to finance either production or consumption expenditures. It will not make sense to invest in the machine unless the net revenue from using the machine to produce goods and services is worth at least as much as the interest the investor could get on the loan.

Of course, capital includes many kinds of producers' goods, from tractors and other machines through grapevines and orchards and many intangible assets. What they all have in common is the opportunity cost corresponding to the interest rate. With that in mind, we identify the price of capital as the interest rate. The demand for capital is the marginal productivity of capital (net of wear and tear) times the price of output.

As for the supply of capital, that will depend on the decisions made by savers. Many economists believe that an increase in interest rates will result in an increase in saving and so in the quantity of capital supplied, giving an upward sloping supply curve of capital. However, for capital as for labor, it is logically possible that the supply curve (in the economy as a whole) could be backward sloping. For an individual industry, however, the supply of capital will probably be horizontal and correspond to the opportunity cost of capital in other industries.

In the next page, we see a picture of the supply and demand for capital as many economists understand them.

6. Demand for Capital

Here is a diagram the demand for capital by an individual firm as it is sketched in the previous page. We assume that the firm uses a given quantity of labor and land and that the quantity of capital used varies. The quantity of capital used (measured in dollars' worth) is marked off on the horizontal axis. On the vertical axis is the rate of interest, which we understand as the price of capital.

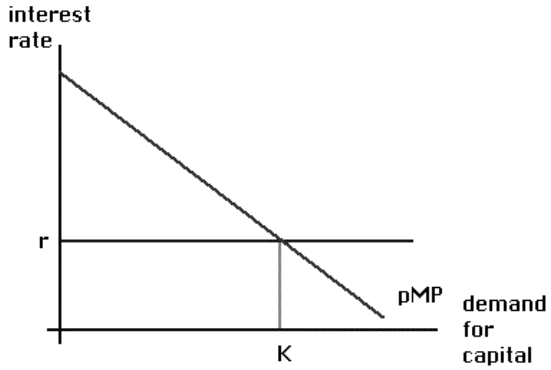


Figure 2: Marginal Net Productivity as the Demand for Capital

In the diagram, the horizontal red line, corresponding to the market interest rate, is the supply curve of capital to the firm. The green line is the demand for capital, that is, the marginal productivity of capital (net of the cost of wear and tear of specific capital goods) times the price of the output — the value of the marginal product of capital. The profit-maximizing demand for capital for the firm is shown by K. That is, it will be profitable to expand the capital stock of the company until diminishing returns reduces the value of the marginal product to r, the market rate of interest, at K.

Some economists criticize this approach to the supply and demand for capital on the grounds that "capital" really consists of many different kinds of capital goods and cannot be expressed as a single amount of "capital." If we accept that argument, we would have to think of a marginal productivity curve and demand for each respective capital good, measure the capital goods in natural units (number of machines, number of grapevines, and so on), and treat the price in a little more complicated way. But, for microeconomics, the results would be pretty much the same: the demand for a capital good, like the demand for any input, depends on marginal productivity.

Now let's look at an important application of the marginal productivity approach in the economics of factors of production.

Notes

LESSON 30:
INCOME AND WEALTH

1. Income Distribution and Labor Demand

As we observed in previous lessons, the marginal productivity approach originated with John Bates Clark. Clark was especially concerned with the division of income between labor and property, and the John Bates Clark model provides us with a visualization of income distribution. Let's look at that in a little more detail.

We will use the marginal productivity approach to the demand for labor as the basis of our discussion. Here is another look at the key picture.

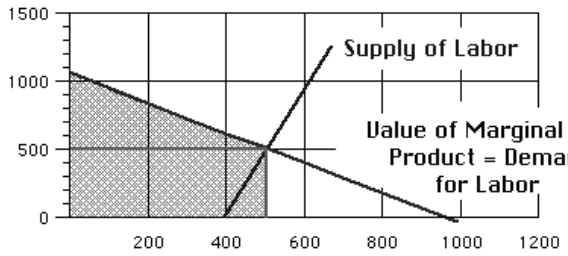


Figure 1: Wages and Profit

Notice the shaded area between the VMP curve and the price (wage) line. Remember, the area of the shaded triangle is the total amount of payments for profits, interest, and rent — in other words, everything the firm pays out for factors of production other than labor. We have seen that the demand for land and capital can be visualized in quite different ways, but there is no conflict there — the same thing can look different when we look at it from different points of view. When we want to visualize the demand for land or for capital individually, then we will use marginal productivity diagrams as we did in the last few pages. But it is equally correct to think of the payments to land and capital as what is left over after the wage bill has been paid. That's the point of view we will take in the rest of this chapter.

2. The Labor Market and the Distribution of Income

In the idealized market society that John Bates Clark envisioned, the wage is determined by the supply and demand for labor. The demand for labor is the Value of the Marginal Product of labor in the society as a whole. The supply of labor is determined by the population and the preferences of workers with respect to more income and consumption versus more holidays and shorter work hours. Here is a picture to illustrate the idea:

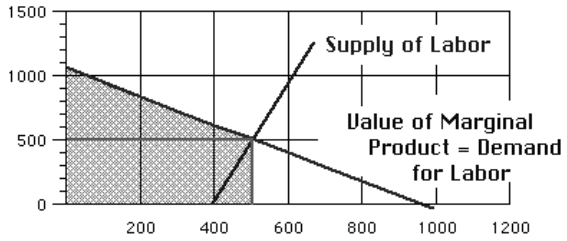


Figure 2: Division of Income Between Work and Property

The equilibrium is shown by the orange lines: the equilibrium wage is $500 per labor week, as in our earlier examples, and the equilibrium quantity of labor hired is something less than 500 units of labor. The total income of the workers is shown by the rectangular area shaded light green, and the total income to proprietors is the triangular area shaded light pink.

3. Functional and Personal Distribution

This John Bates Clark theory of the distribution of income illustrates what economists call the "functional distribution of income." Economists distinguish two concepts of the distribution of income:

The functional distribution of income

The functional distribution of income is the distribution between groups in society who own different factors of production, i.e. the proportion of income going to employees, landowners, and owners of capital respectively.

The personal distribution of income

The personal distribution of income is the distribution of income among individuals, families or households, regardless of what factors of production they own.

Of course, the personal distribution depends partly on the functional distribution, but it also depends on who owns what. In other words, the words of British radical economist Joan Robinson, the functional distribution tells us the distribution of income among the factors of production, but the personal distribution depends on the distribution of the "factors among the chaps."

This is important if we are concerned about inequality in the distribution of income. Generally, income from land and capital is much less equally distributed than income from labor. In the idealized John Bates Clark world, income from labor would be absolutely equally distributed among those who work the same hours, because labor is assumed to be homogenous. But that's a simplifying assumption. In the real world, people who have better skills or stronger bodies may be able to earn a higher wage. This makes for some inequality in labor income. But the ownership of land and capital is much more unequally distrib-

uted than the capacity to earn wages, so income from property is much more unequally distributed among persons.

4. Just Distribution of Income

Thus, the John Bates Clark model provides us with a theory about "what is" the **functional** distribution of income between labor and property-owners. Clark went still further and claimed that this distribution would be just.

Of course, justice is a matter of normative economics, not positive economics, since saying that something is just implies that "it ought to be." How could Clark claim that payment according to the marginal product is just? Clark was applying a value judgment that it is just to pay people according to what they contribute to the product of society. He claimed that the marginal product of labor is the appropriate measure of what the individual worker contributes, so it would be just that the worker be paid her or his value marginal product. The rent, interest, and profits, incomes to property owners, would correspond to the marginal product of the land and capital, according to Clark, so their payment would also be just.

But, of course, this is controversial. Not everyone would accept the idea that people ought to be paid according to their contribution. The Reasonable Dialog perspective is especially important here. We can easily make two opposite mistakes: first, to suppose that this value judgment is obviously true — if we agree with it in the first place! — or that it is a matter of arbitrary opinion, without any basis in reason at all. Here are some criticisms of the value judgment that "it is just to pay people according to what they contribute to the product of society:"

What people contribute depends on the opportunities they have to contribute. For example, if a person cannot find a job, despite his best attempts, his failure to contribute to society is unavoidable, and it is unjust that he be deprived of an income on that account.

Clark would probably respond that in the competitive society he envisioned this would be impossible, since all obstacles to finding employment would be removed. The idea behind the supply and demand model is that everyone who is willing to supply labor finds a matching demand — at the equilibrium wage.

What people contribute to production is mostly a result of factors they themselves had nothing to do with. For example, very few of us could contribute anything much without relying on the methods, technology and experience we have taken from the generations that preceded us. Some of our output may also be the result of gifts of nature (or of God) that we are fortunate to have. Why should we be rewarded for being lucky enough to have access to these resources of natural gifts and technology and historic experience?

I'm not sure how Clark would respond to this one. He would probably say that natural gifts really don't differ very much, and that in any case the resources of natural gifts and technology and experience wouldn't produce anything if they were not propelled by human effort, so that the product really is proportionate to the effort, and that effort should be rewarded.

Just what is a contribution? Followers of the American activist Henry George, for example, say that land is not a contribution on the part of the landowner, since land is a free gift of nature. Marxists and other labor radicals would go even further, saying that only work is a contribution to the social product, so that workers are entitled (individually or as a group) to 100% of the product.

Clark would probably dismiss this as a confusion, saying that the Marxist misses the point of marginal analysis. But that response could be hasty. The real difference seems to be in the definition of terms: what really counts as a "contribution?"

Another unspoken assumption in the John Bates Clark version of normative economics is that the distribution of property is just. No-one would say that it is just for a thief to be paid for contributing the property he has stolen. Income to property-owners cannot be just unless the owner has a just claim to own the property. Not everyone would agree that capitalist property ownership is just. Marxists say that all capitalist property originates from exploitation, or from actual plunder, so they would say the distribution and ownership of property in a capitalist market system is itself not just, so no distribution of income to property could be just. Clark's model doesn't address this question.

So there is plenty of room for controversy about the normative economics of income distribution. But even if we do not agree with it, we must recognize Clark's contribution of a precise and logically constructed account of one view in normative economics. Because of its precision and clarity, it lends itself to more constructive discussion, both from a critical and supportive point of view.

However, we will not be able to explore the controversy much further in our book, but instead will return to positive economics — the description and explanation of what is. As we have seen, the John Bates Clark theory is mostly a theory of the functional distribution of income, while normative economics is also concerned with the personal distribution of income. Let's look a little more carefully at equality and inequality in the personal distribution of income.

5. Visualizing the Personal Distribution of Income

We can visualize the personal distribution of income using a graphic presentation called the Lorenz curve. Here is a Lorenz curve for the American economy in 1994.

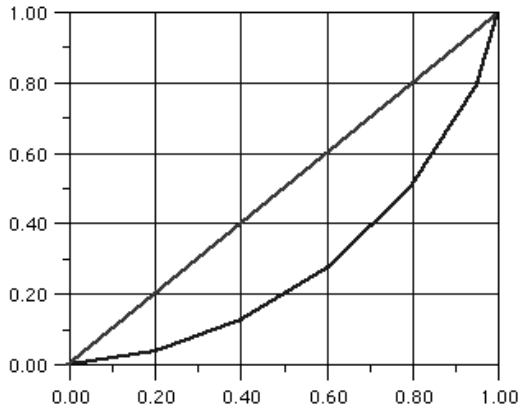


Figure 3 Lorenz Curve

The Lorenz curve is the dark red curve in Figure 8 above. On the horizontal axis we have the proportion of the population, and on the vertical axis we have the proportion of total income earned by the corresponding fraction of the population. Thus, the figure tells us that the poorest 20% of the population earned only 3.6% of the income, the poorest 40% of the population earned 12.5 percent of income, and so on. similarly, the least well off 95% of the population earned 78.8% of the income, leaving 21.2% for the richest 5% of income for the richest 5%.

If income were distributed with perfect equality, then the poorest 20% of the population would earn exactly 20% of income, the poorest 40% of the population would earn 40% of income, and the richest 5% of the population would earn exactly 5% of income. As a result, the Lorenz curve for an equal distribution would be a forty-five degree diagonal line, shown by the green line in Figure 8. The curvature of the Lorenz curve, as it droops below the 45 degree line, shows the inequality of the income distribution.

The data for these examples were supplied by the census bureau.

6. Measuring Inequality in Income Distribution

The Lorenz Curve construction also gives us a rough measure of the amount of inequality in the income distribution. The measure is called the **Gini Coefficient**. Computation of the Gini Coefficient is illustrated by Figure 4below.

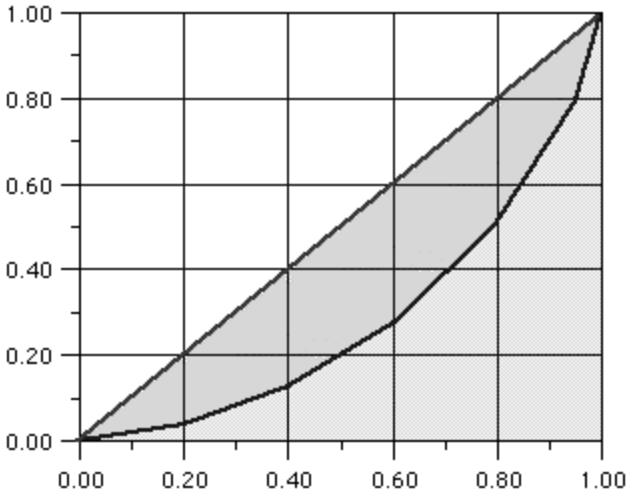


Figure 4. Lorenz Curve and Gini Coefficient

To compute the Gini Coefficient, we first measure the area between the Lorenz Curve and the 45 degree equality line. This area is divided by the entire area below the 45 degree line (which is always exactly one half). The quotient is the Gini coefficient, a measure of inequality. In other words, the Gini coefficient is the area shaded in pink divided by the total of the areas shaded in pink and light blue-green.

For a perfectly equal distribution, there would be no area between the 45 degree line and the Lorenz curve — a Gini coefficient of zero. For complete inequality, in which only one person has any income (if that were possible) the Lorenz curve would coincide with the straight lines at the lower and right boundaries of the curve, so the Gini coefficient would be one. Real economies have some, but not complete inequality, so the Gini coefficients for real economic systems are between zero and one.

The Gini coefficient for the United States in 1994 (according to the Census Bureau) was 0.456.

7. Changing Income Inequality

Income inequality in the United States has increased in recent decades, and this has recently become a concern to some economists and others interested in economic and political issues. Lorenz curves for the United States are shown in Figure 5 below.

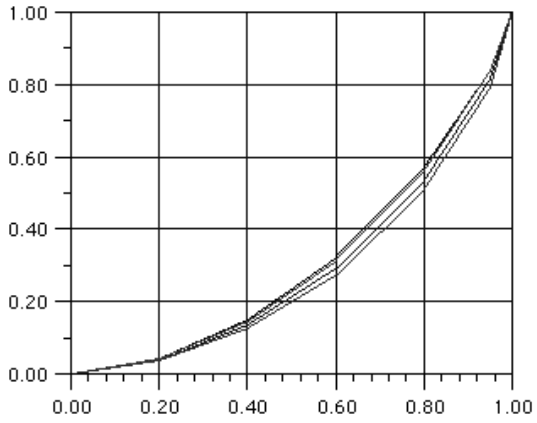


Figure 5. Lorenz Curves of the United States

The four Lorenz curves are for 1970, 1980, 1990, and 1994, and are color-coded as follows:

Table 1

years	1970	1980	1990	1994
colors	black	green	blue	red
GiniCoefficients	0.394	0.403	0.428	0.456

We observe a steady, and perhaps accelerating, tendency toward increased inequality over this period. This can also be observed when we look at the Gini Coefficients for the four years, shown in the third row of the table. We see that the over-all increase is about sixteen percent, or one-sixth.

8. Explaining the Changes

How are we to explain these changes in terms of the functional distribution of income? There are at least three possible explanations:

- The distribution of wage income may have become more unequal.
- The distribution of income from property may have become more unequal.
- The functional distribution may have shifted, so that property incomes are a larger fraction of total incomes.

The last one, a shift of the functional income distribution so that property income is a larger fraction of the total, would make the personal distribution of income more unequal because property incomes are the more unequal component.

Increasing the weight on the more unequal component will increase overall inequality.

When we look at the record in detail, it seems that both the first and the last of the three things seems to have happened. We have to distinguish between the markets for **unskilled** and **skilled** labor. This is a little more complex that the John Bates Clark model, which thinks in terms of homogenous labor, but the same principles can be applied without much difficulty. One explanation that fits the facts pretty well is that there has been a big decrease in the demand for **unskilled** labor. It's easy to see how that would cause an increase in the inequality of wage income. Unskilled workers were already being paid less, then a drop in the demand for their labor depresses their wages, and they make even less, relative to skilled labor. The John Bates Clark model helps out by showing us how this would also tend to shift the functional distribution of income.

Take a look at Figure 6:

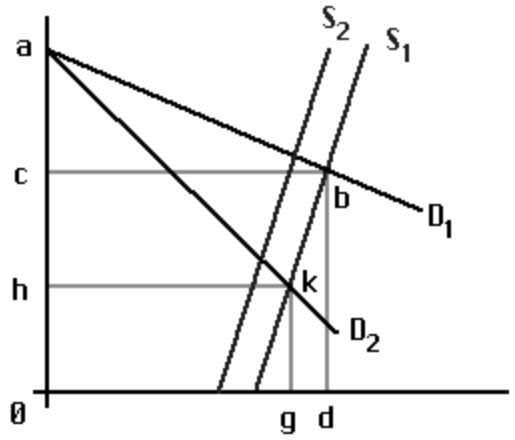


Figure 6: A Decrease in the Demand for Unskilled Labor

The demand for unskilled labor is, again, the value of the marginal product. Before the decrease in demand, the supply and demand for unskilled labor are S_1 and D_1, and the market wage is c. The unskilled workers earn an amount visualized by the rectangle cbd0, and profits on their work amount to the triangle abc. Now there is a big decrease in the demand for unsilled labor, to D_2, partially offset by a slight decrease in the supply, to S_2. The new wage for unskilled labor is h, much lower than before, as we anticipated. Now, in addition, unsilled workers are earning a much smaller share — hkg0 by comparison with profits akh. Since the share of unskilled workers is less, if nothing has changed the share of skilled workers, the overall share of labor in the functional distribution of income will have declined.

9. The Functional Distribution, Again

We can check that explanation against the facts. Did the shares in income in fact shift away from labor in the last 25 years?

It seems that they did. Here are some data on labor incomes and property incomes. Labor incomes include all wages and other labor compensation, while property incomes include interest, rent, and dividends, as reported by the census bureau. The incomes are before taxes and do not include government subsidies, either. The following table shows the fraction of the total labor and property incomes that were labor incomes, in the second row, and property incomes in the third row. the fourth row shows the breakdown between labor and property incomes as a "pie chart."

Table 2

year	1970	1980	1990	1994
labor incomes	0.86	0.83	0.80	0.82
property incomes	0.14	0.17	0.20	0.18
				

In the "pie charts," of course, the labor share is represented by the blue shaded "pie" and the share of property income by the pinkish shaded "piece of the pie." We see the share of property income getting consistently larger through the first two periods. That's consistent with the explanation that greater inequality is a result of a shift of the functional distribution of income toward a greater proportion of property incomes. In 1994, however, we see a slight shift back toward a larger share for labor incomes, although the inequality continued to increase from 1990 to 1994. It would seem that other factors became more important in the four years 1990-1994.

What the evidence suggests is that shifts in both the functional distribution of income and the inequality of wages are factors in the increase in inequality since 1970. In general, we will find the John Bates Clark model is a good starting point, but only gives us part of the story.

10. Chapter Summary

In this chapter we have explored the supply and demand for factors of production, land, labor, and capital. The discussion complements our discussions in Chapter 3 of the supply and demand for consumer products. The basic ideas for this discussion are the marginal productivity of the input and the "value of the marginal product," that is, marginal productivity times the price.

- We began with a discussion of the supply and demand for labor.
  - The demand for labor is identified with the value of the marginal product" of labor.
  - The supply of labor depends on the population and on its preferences between earning more income and enjoying more leisure.
  - The price of labor is the wage, and we see an equilibrium if the wage is just high enough so that the quantity of labor demanded is equal to the quantity supplied.
  - This leads to an application to minimum wage regulations — and the conclusion that such regulations are likely to put some employees out of work.
  - Some economists doubt that the supply-and-demand approach really works in labor markets, and propose alternative approaches. But in many cases the alternative

approaches incorporate marginal productivity along with other influences on wages and employment.

- We next discussed the supply and demand for land and natural resources.
  - Here again, demand is identified with the value of the marginal product.
  - The supply, however, is given by nature.
  - But different parcels of land have different fertility, so that demanders have to pay "differential rent."
- The supply and demand for capital were next discussed.
  - We think of the interest rate as the price of capital.
  - The demand is again identified with the value of the marginal product of capital.

We then applied the marginal productivity approach to explain the distribution of income and its trends in recent years. The marginal productivity approach gives us a theory of the functional distribution of income, which is one aspect of the distribution among persons. We find that the recent trends in income distribution do "make sense" when we apply the marginal productivity approach to them. But is that the only explanation? The criticisms of markets for labor can be applied also to the marginal productivity approach as a whole, and to this application, so there may be other aspects, or even a completely different approach, that will make better sense.

However, we have come to the limit of what this book can say about the marginal productivity approach and the alternatives to it. Students who are interested in alternative approaches, or a deeper discussion of this approach, will find it in more advanced textbooks

Factor Markets

Factor Market

- a market for a factor of production.
- example: The market for construction workers brings together the buyers and sellers of construction workers' services.

Factor of Production

- used to produce some output.
- also called an **input** or a **productive resource**.
- examples: labor, machinery, raw materials, land

Derived Demand

The demand for an input is derived from the demand for the output that the input helps produce.

Note

A firm might be a perfect competitor in the product market and might not be a perfect competitor in the factor market, or vice versa.

Example: The local water company is the only water company in the area. It is one of many employers who hire accountants.

This firm is **not** a perfect competitor in the product market (water market).

Four Possibilities for a Firm

- Perfect competitor in the product market, and perfect competitor in the factor market.
- Perfect competitor in the product market, but **not** a perfect competitor in the factor market.
- **Not** a perfect competitor in the product market, but a perfect competitor in the factor market.
- **Not** a perfect competitor in the product market, and **not** a perfect competitor in the factor market.

Example: The local water company is the only water company in the area. It is one of many employers who hire accountants.

This firm is **not** a perfect competitor in the product market (water market).

It may be a perfect competitor in the factor market (market for accountants).

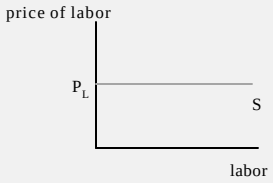
Example: The local water company is the only water company in the area. It is one of many employers who hire accountants.

Example: A small mill town is owned by a textile company. The company is the only employer in town.

Example: A small mill town is owned by a textile company. The company is the only employer in town.

This firm may be a perfect competitor in the product market (textile market).

The Supply Curve of Labor to a Firm that is a Perfect Competitor in the Labor Market



Example: A small mill town is owned by a textile company. The company is the only employer in town.

This firm may be a perfect competitor in the product market (textile market).

It is **not** a perfect competitor in the factor market (labor market).

Factor Market Terms

Price-Taking in the Factor Market

Just as a firm in a perfectly competitive product market takes the price of the product as given, a firm in a perfectly competitive factor market takes the price of the factor as given.

The firm can hire as much of the input as it wants at the going input price.

So, the supply curve of the input to the firm is a horizontal line at the input price.

Marginal Resource Cost (MRC)

the change in total cost that results from the employment of an additional unit of an input.

$$MRC = \Delta TC / \Delta L$$

Marginal Physical Product (MPP) or Marginal Product (MP)

the change in the quantity of output that results from the employment of an additional unit of an input.

$$MPP = \Delta Q / \Delta L$$

Alternative formula for MRP

$$MRP = \frac{\Delta TR}{\Delta L} = \frac{\Delta TR}{\Delta L}\frac{\Delta Q}{\Delta Q} = \frac{\Delta TR}{\Delta Q}\frac{\Delta Q}{\Delta L} = MR \cdot MPP$$

So, **$MRP = MR \cdot MPP$**

Marginal Revenue Product (MRP)

the change in total revenue that results from the employment of an additional unit of an input.

$$MRP = \Delta TR / \Delta L$$

Value of the Marginal Product (VMP)

the price of the output multiplied by the marginal physical product of the input.

$$VMP = P \cdot MPP$$

What is the difference between the MPP and MRP?

Suppose your company produces chairs.

The MPP tells how many more **chairs** you can make if you hire another worker.

The MRP tells how much more **revenue** you can make from the additional chairs produced by the additional worker

Sometimes MRP = VMP, but not always.

Recall: If a firm is a **perfect competitor in the product market**, marginal revenue is equal to the price of the output (MR = P).

Then, MRP = MR · MPP
= P · MPP
= VMP

So, MRP = VMP

for a firm that is a perfect competitor in the product market.

Example: A firm sells its shirts in a perfectly competitive product market for $10 each.

L	Q
0	0
10	70
20	130
30	180
40	220
50	250
60	270
70	280

Recall: If a firm is a **not** perfect competitor in the product market, marginal revenue is less than the price of the output (MR < P).

Since MRP = MR · MPP
and VMP = P · MPP,
MRP < VMP,

for a firm that is **not** a perfect competitor in the product market.

Example: A firm sells its shirts in a perfectly competitive product market for $10 each.

L	Q	MPP=?Q/? L
0	0	---
10	70	7
20	130	6
30	180	5
40	220	4
50	250	3
60	270	2
70	280	1

If a firm is perfectly competitive in the product market, then MRP = VMP.

If a firm is not perfectly competitive in the product market, then MRP < VMP.

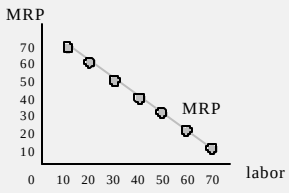
Example: A firm sells its shirts in a perfectly competitive product market for $10 each.

L	Q	MPP=?Q/? L	TR=PQ
0	0	---	0
10	70	7	700
20	130	6	1300
30	180	5	1800
40	220	4	2200
50	250	3	2500
60	270	2	2700
70	280	1	2800

Example: A firm sells its shirts in a perfectly competitive product market for $10 each.

L	Q	MPP=?Q/?L	TR=PQ	MR =?TR/?Q
0	0	---	0	---
10	70	7	700	10
20	130	6	1300	10
30	180	5	1800	10
40	220	4	2200	10
50	250	3	2500	10
60	270	2	2700	10
70	280	1	2800	10

Plotting points we have a graph of the MRP curve.



Example: A firm sells its shirts in a perfectly competitive product market for $10 each.

L	Q	MPP=?Q/?L	TR=PQ	MR =?TR/?Q	MRP=?TR/?L MRP= MR. MPP
0	0	---	0	---	---
10	70	7	700	10	70
20	130	6	1300	10	60
30	180	5	1800	10	50
40	220	4	2200	10	40
50	250	3	2500	10	30
60	270	2	2700	10	20
70	280	1	2800	10	10

When should you employ more of an input?

MRP > MRC	employ more input
MRP < MRC	cut back employment
MRP = MRC	profit-maximizing employment level

Focusing on the first and last columns of the previous table, we have the MRP schedule.

L	MRP
0	---
10	70
20	60
30	50
40	40
50	30
60	20
70	10

profit-maximizing condition for input usage:

MRP = MRC

MRC in a Perfectly Competitive Labor Market

Each time a firm hires another unit of labor, its cost increases by the price of the labor (P_L).

So for a firm in a perfectly competitive labor market, MRC = P_L .

(If a firm is not in a perfectly competitive labor market, this is not true.)

Remember we have been trying to determine what the demand curve for labor looks like for this firm.

All of our demand curve points have been points on the MRP curve.

The demand curve for labor by the firm is just (the downward sloping part of) the MRP curve.

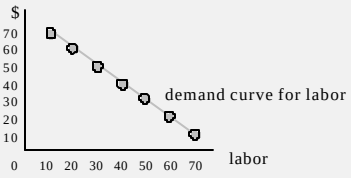
Suppose the firm in the example we considered earlier is also perfectly competitive in the labor market.

So the MRC is the same as the price of labor or the market wage.

Let's see what the demand curve for labor is for this firm.

What we need to know is how many workers will be hired at various wage levels.

A Firm's Demand Curve for Labor



Remember: You hire workers as long as they add at least as much to revenues as to cost.

L	MRP
0	---
10	70
20	60
30	50
40	40
50	30
60	20
70	10

Suppose the market wage is $70. How many workers will you hire?
10
Suppose the market wage is $60. How many workers will you hire?
20
Suppose the market wage is $50. How many workers will you hire?
30
Suppose the market wage is $40. How many workers will you hire?
40

TUTORIAL 7

1. Case Study

The New Workforce

Lead Story-dateline: The Economist, August 25-31, 2001.

Of all the big developed countries, America now has the smallest proportion of factory workers in its labor force. Before World War I, there was not even a word for people who made their living other than by manual work. The term "service worker" was coined around 1920, but it has turned out to be rather misleading. These days, fewer than half of all non-manual workers are service workers. The only fast-growing group in the workforce in America, and in every other developed country, are "knowledge workers" and "knowledge technologists" - people whose jobs require formal and advanced schooling. They now account for a full third of the American workforce, outnumbering factory workers by two to one. In another 20 years or so, they are likely to make up close to two-fifths of the workforce of all rich countries. The term "knowledge industries," "knowledge work," and "knowledge worker" are only 40 years old, coined around 1960. Knowledge workers, collectively, are the new capitalists who own the means of production, and through their stakes in pension funds and mutual funds, they have become majority shareholders and owners of many large businesses in the knowledge society, and, therefore, are capitalists in the old sense of the word. Knowledge workers see themselves as equal to those who retain their services: as "professionals" rather than as "employees." The knowledge society is a society of seniors and juniors rather than of bosses and subordinates.

Thinking About the Future!

The upward mobility of the knowledge society comes with the psychological pressures and emotional traumas of the rat race. These pressures create hostility to learning. America, Britain, Japan, and France are allowing their schools to become viciously competitive. The fact that this has happened over such a short time - 30 to 40 years - indicates how much the fear of failure has already permeated the knowledge society. Given this competitive struggle, successful knowledge workers "plateau" in their 40s. Knowledge workers, therefore, need to offset this by developing an alternative non-competitive life and community of their own for personal contributions and achievement.

Talking It Over And Thinking It Through!

1. What are the implications of the shift toward knowledge workers for the role of women in the labor force?
2. What is the difference between knowledge workers and knowledge technologists?
3. How does knowledge compare with traditional skills and means of production?

ESSENTIAL PRINCIPLES

What are the Essential Principles of Economics?

Here are the Essential Principles of Economics I have tried to explain, illustrate and apply in this book. There are ten major principles. That's a nice round number. But I have kept the number of main principles small mainly by putting some pretty important topics among the 25 to 35 subordinate topics (depending on how you count). Links are given, but some of the linked pages may be hard to understand out of context.

Division of Labor

I put the division of labor first mainly because Adam Smith did, arguing that division of labor is the key cause of improving standards of living. Modern economics doesn't do much with the concept of division of labor, but two closely related concepts are important:

Returns to Scale

Returns to scale may be increasing, constant or decreasing. Increasing returns to scale is the case that leads to special results, and division of labor is one cause (arguably the main cause) of increasing returns to scale.

Virtuous Circles in Economic Growth

For Smith, a major consequence of division of labor and resulting increasing productivity was a "virtuous circle" of continuing growth. Modern "virtuous circle" theories have more dimensions, but division of labor and (resulting?) increasing returns to scale are among them.

Opportunity Cost

The idea is that anything you must give up in order to carry out a particular decision is a cost of that decision. This concept is applied again and again throughout modern economics. If (God forbid) you were to learn only one of the Principles of Economics thoroughly, this should be the one.

Scarcity

According to modern economics, scarcity exists whenever there is an opportunity cost, that is, where-ever a meaningful choice has to be made.

Production Possibility Frontier

The production possibility frontier is the diagrammatic representation of scarcity in production.

Comparative Advantage

A very important principle in itself, and a key to understanding of international trade the principle of comparative advantage is at the same time an application of the opportunity cost principle to trade.

Discounting of Investment Returns

Another application of the opportunity cost principle that is very important in itself, this one tells us how to handle opportunities that come at different times.

The Equimarginal Principle

This is the diagnostic principle for economic efficiency. It has wide applications in modern economics. Two of the most important are key principles of economics in themselves:

The Fundamental Principle of Microeconomics

This principle describes the circumstances under which market outcomes are efficient, and

The Externality Principle

describes some important circumstances in which they are not. Of course, the equimarginal principle is founded on

Marginal Analysis

Also an important principle in itself and very widely applied in modern economics. There is no major topic in microeconomics (I believe) that does not apply marginal analysis and opportunity cost. The link shown above is the marginal analysis of productivity, but marginal analysis also has applications to cost, revenue, consumers' utility and benefits, and more.

Market Equilibrium

The market equilibrium model could be broken down into several principles — the definitions of supply, demand, quantity supplied and demanded and equilibrium, at least — but these all complement one another so strongly that there is not much profit in taking them separately. However, there are many applications and at least four important subsidiary principles:

Elasticity and Revenue

These ideas are a key to understanding how market changes transform society.

The Entry Principle

This tells us that, when entry into a field of activity is free, profits (beyond opportunity costs) will be eliminated by increasing competition. This has a somewhat different significance depending on whether competition is "perfect" or monopolistic.

Cobweb Adjustment

This might give the explanations when the market does not move smoothly to equilibrium, but overshoots.

Competition vs. Monopoly

Why economists tend to think highly of competition, and lowly of monopoly.

Diminishing Returns

Perhaps the best-known of major economic principles, the Principle of Diminishing Returns is much more reliable in short-run than in long-run applications, so the Long Run / Short Run dichotomy is an important subsidiary principle. Modern economists think of diminishing returns mainly in marginal terms, so marginal analysis and the equimarginal principle are closely associated.

Game Equilibria

Game theory allows strategy to be part of the story. One result is that we have to allow for several kinds of equilibria. We have

- Noncooperative equilibrium
  - Prisoners' Dilemma (dominant strategy) equilibria which are
  - Nash (best response) equilibria, (but not all Nash equilibria are dominant strategy equilibria), and
- Cooperative equilibria

  And that's just the beginning. The main applications in this book, and traditionally, are in the study of
- oligopoly.

Measurement Principles

Economics is multidimensional, and that creates some difficulties in measuring things like production, incomes, and price levels. Some of the problems can be solved more or less fully.

Value Added and Double Counting

One for which we have a pretty complete solution is the problem of double counting: the solution is, use value added.

"Real" Values and Index Numbers

Since we measure production and related quantities in dollar terms, we have to correct for inflation. Index numbers are a pretty good workable solution, but there are some problems and criticisms.

Measurement of Inequality

Another issue is that the "average income" may not mean very much, because nobody is average and income is unequally distributed. Even if we cannot correct for that (what would that mean?) we can get a rough measure of the relative inequality and see where it is going.

Medium of Exchange

Money is whatever is generally acceptable as a medium of exchange. That means a bank, or similar institution, can literally create money, so long as people trust the bank enough to accept its paper as a medium of exchange. We might call this magical fact the Fiduciary Principle.

Income-Expenditure Equilibrium

Like the market equilibrium principle, but even moreso, this model pulls together a number of subsidiary principles that complement one another and together constitute the "Keynesian" theory of aggregate demand. The implications of this theory are less controversial than the word "Keynesian" is — controversy has to do more with the details than the applications. Among the subsidiary principles are

- Coordination Failure
- The income-consumption relationship
- The Multiplier
- Unplanned inventory investment
- Fiscal Policy
- The Marginal Efficiency of Investment
- The influence of money on interest
- Real Money Balances
- Monetary Policy

The Surprise Principle

People respond differently to the same stimuli if the stimuli come as a surprise than they would if the stimuli do not come as a surprise. This new economic principle plays the key role with respect to aggregate supply that "Income-Expenditure Equilibrium" plays with respect to aggregate demand.

Rational Expectations

People don't want too many unpleasant surprises. If they use the information available to them efficiently, then they won't be surprised in the same way very often. This can lead to

Policy Ineffectiveness

But it is hard to reconcile this way of thinking with the apparent

Permanence

of many economic changes, especially those in unemployment. These suggest that the economy has a high degree of

Path Dependence,

and that would put the independence of aggregate supply into some doubt.

GLOSSARY

Glossary For Microeconomic Chapters

Adjustment process Path by which a system moves from a position of isequilibrium to one of equilibrium.

Adverse selection A problem in insurance markets caused by asymmetric information. Insurance tends to be purchased most by those who know that they are high risks.

Aggregation A method of simplifying theory by combining many markets into a large, composite market.

Allocation The determination of what goods and services will be produced from available resources. Allocation can be done with markets or with hierarchy.

Antitrust policy Policy that makes companies act in a competitive manner by breaking up companies that are monopolies, prohibiting mergers that would increase market power, and finding and fining companies that collude to establish higher prices.

Arbitrage Simultaneously buying in a cheap market and selling in an expensive one.

Budget Constraint A line that separates outcomes that are affordable from outcomes that are not affordable. Occasionally called a consumption-possibilities frontier.

Change in demand A shift in the demand curve.

Change in quantity demanded A change in the amount people buy because a change in price moves them along a stationary demand curve.

Change in quantity supplied A change in the amount sellers sell because a change in price moves them along a stationary supply curve.

Change in supply A shift in the supply curve.

Circular flow model A simplified picture of a market economy showing the flow of products from businesses to households and the flow of resources from households to businesses.

Consumer Sovereignty In a market economy, it is ultimately the wants of the consumers, not the preferences of the producers, that determine what goods and services are produced.

Consumers' surplus The difference between the maximum a buyer would pay and the actual price.

Consumption-possibilities frontier See budget constraint.

Contingent behavior Behavior that exists when each person's actions depend on what he expects others to do.

Cross-price elasticity (cross-elasticity) A measure of whether goods are substitutes or complements.

Demand curve The relationship between price and the amount of a product people want to buy.

Derived demand The demand for a resource depends on, or is derived from, the demand for the things that the resouce helps produce.

Disequilibrium A condition that exists when a system is not at rest and has a tendency to change.

Dollar voting An explanation of how a market economy determines what goods are produced, made with an analogy to the political process of voting.

Duopoly A market in which there are two sellers.

Economic efficiency A situation in which value is maximized. Given resources, technology, and preferences, no changes will increase value. Also called Pareto optimality.

Economic inefficiency A situation in which there is potential value that no one captures. Given resources, technology, and preferences, there is some change which will improve the well-being on one individual without harming anyone else.

Economic Rent See producer surplus.

Elasticity A measure of responsiveness.

Entrepreneur An individual who creates new a new organization, market, or product, usually in the quest for profit. Entrepreneurs are innovators.

Equilibrium A condition that exists when a system is at rest and has no further tendency to change.

Externality A cost or benefit that a decision maker passes on to a third party. Pollution is an example of a negative externality.

Excise tax A sales tax on a specific item.

Fixed cost Cost that does not change as output changes.

Free rider Person who does not pay for good or service because there is no way to exclude those who do not pay from using the good or service.

Game theory An analysis of interactions in which the outcome a person faces depends not only on his strategy of action, but also on the strategies of others.

Human capital Peopleís assets in the form of investment in themselves.

Income elasticity of demand A measure of the responsiveness of people's purchases to changes in income.

Indifference curve In a graphical representation of a utility function; an indifference curve plots all combinations of goods and services which provide the same utility. Occasionally called an isoutility curve.

Inferior good A good that people buy less frequently if their incomes rise.

Invisible hand A phrase that expresses the belief that the best interests of a society can be served when individual consumers and producers compete to achieve their own private interests.

Isoquant In a graphical representation of a production function, an isoutility curve shows all ways of producing a specificlevel of output.

Law of demand The principle that there is an inverse relationship between the price of a good and the quantity that buyers are willing to purchase.

Law of diminishing returns Adding more of one input while holding other inputs constant eventually results in smaller and smaller increases in added output.

Lorenz Curve A graphical way to illustrate the equality or inequality of the distribution of income.

Marginal cost The change in total cost caused by a one-unit change in an activity, or the slope of the total cost curve. In the case of a business, the change in total cost is caused by a change in output.

Marginal rate of substitution The ratio at which people will trade good B for good A.

Marginal rate of transformation Slope of the production-possibilities frontier, which shows how much of good B must be given up to produce more of good A.

Marginal resource cost The change in total cost caused by a one-unit change in an input.

Marginal revenue The change in total revenue resulting from a change in sales; the slope of the total cost curve.

Marginal revenue product The change in total revenue resulting from a one-unit change in an input.

Market failure A situation in which a market yields a result that is economically inefficient, that is, there is value that is not captured.

Maximization principle The rule that net benefits are maximized when marginal benefit equals marginal cost.

Monopoly An industry with only one seller.

Monopolistic competition An industry that has easy entry and exit, but in which sellers are price searchers.

Monopsony A market with only one buyer.

Moral hazard An insurance problem; when the cost of a disaster is reduced with insurance, people have less incentive to avoid the disaster.

Negative-sum game In terms of game theory, an interaction in which losses exceed winnings.

Normative analysis An analysis based on a judgement about what is desirable and what is undesirable.

Oligopoly An industry in which there are few sellers.

Paradox of Value The puzzle of why essential items such as water are cheap while frivolous items such as diamonds are expensive. The paradox is easily resolved when one understands the difference between total value and marginal value.

Pareto optimality See Economic efficiency.

Positive analysis An analysis limited to statements about the actual consequences of an event or policy, with no judgement about whether the consequences are desirable or not.

Positive-sum game In terms of game theory, an interaction in winnings exceed losses.

Present value Money in the future is less valuable than an equivalent amount of money now because money in the future gives fewer options. The comparison of money in different time periods is made with a present value computation.

Price ceiling Legally established maximum price a seller can charge.

Price-discrimination Charging different prices for the same good or service.

Price floor Legally established minimum price a seller can be paid.

Price elasticity of demand Measures how much consumers respond to a change in price in their buying decisions

Price elasticity of supply The same formula as price elasticity of demand, except that the quantity in the formula refers to the quantity that sellers will sell.

Price searcher A seller (buyer) who can influence price by the amount he sells (buys).

Price taker A seller (buyer) who has no control over price, but sells (buys) at the given price.

Principal-agent problem The potential conflict of interest when a person (the principal) has someone (the agent) acting on his behalf.

Prisonersí dilemma A game theory illustration that self-interest can lead to group disaster.

Problem of the commons Refers to the absence of any automatic mechanism or incentive to prevent the overuse and depletion of commonly-held resources.

Producers' surplus The difference between the lowest price a producer will accept and the actual price. Also called economic rent.

Production function The mathematical way of stating that output depends on inputs.

Production-possibilities frontier Frontier that separates outcomes that are possible for an individual (or a group) to produce from those that cannot be produced.

Profit seeking Efforts to obtain value through exchange by providing a good or service that others consider valuable. (See rent seeking for the alternative.)

Progressive tax A tax that charges a higher percentage of income as income rises.

Proportional tax A tax that charges the same percentage of income, regardless of the size of income.

Public good A good or service that, once produced, has two properties: Benefits are available to all and there is no way to bar people who do not pay (free riders) from consuming the good or service.

Quota Limit on the quantity of a good that may be imported in any time period.

Regressive tax A tax that charges a lower percentage of income as income rises.

Rent seeking Efforts to obtain value through transfer without providing anything in return.

Scarcity The condition in which human wants exceed the available supply of goods, time, and resources. In a world without scarcity, there would be no economics.

Shortage The market condition existing when quantity demanded exceeds quantity supplied. Generally an increase in price will eliminate a shortage.

Speculation Attempting to buy when the price is low and sell when it is high.

Sunk cost Cost that cannot be recovered.

Supply curve The relationship between the quantity sellers want to sell during some time period (quantity supplied) and price.

Surplus The market condition existing where the quantity supplied is greater than the quantity demanded. Generally a decrease in price will eliminate a surplus.

Tariff Excise tax on imported goods.

Tax incidence Taxes can be shifted from those who write the check to the government to others. The study of tax incidence is the study of who ultimately bears the burden of the tax.

Tournament theory An analysis of conditions under which small differences in ability result in large differences in reward.

Utility An abstract variable, indicating goal-attainment or want-satisfaction.

Utility function A mathematical way of saying that utility depends on consumption of goods and services.

Zero-sum game An interaction in which the sum of winnings and losses equals zero.

REFERENCE MATERIAL

Text Books

1. Microeconomics (5th Edition)
by Robert S. Pindyck (Author), Daniel L. Rubinfeld (Author) (Hardcover — July 27, 2000)

2. Price Theory and Applications
by Steven E. Landsburg (Hardcover — July 20, 2001)

3. Financial Accounting : In an Economic Context
by Jamie Pratt (Author) (Hardcover — February 22, 2002)

4. Managerial Economics (5th Edition)
by Edwin Mansfield (Editor), et al (Hardcover — February 2002)

5. Microeconomic Theory
by Andreu Mas-Colell, et al (Hardcover — June 1995)

6. Microeconomics: An Integrated Approach
by David Besanko (Author), Ronald R. Braeutigam (Author) (Hardcover — September 6, 2001)

7. Managerial Economics: Applications, Strategy and Tactics
by James R. McGuigan, et al (Hardcover — March 7, 2001)

8. Microeconomic Theory: Basic Principles and Extensions
by Walter Nicholson (Hardcover — October 26, 2001)

9. Foundations of Microeconomics
by Robin Bade (Author), Michael Parkin (Author) (Paperback — July 12, 2001)

10. Auction Theory
by Vijay Krishna (Author) (Hardcover — March 1, 2002)

11. Introduction to the Mathematics of Financial Derivatives
by Salih Neftci (Author), Salih N. Neftci (Hardcover — April 2000)

12. Microeconomic Analysis
by Hal R. Varian (Hardcover — March 1992)

13. Health Economics and Policy
by James W. Henderson (Paperback — June 27, 2001)

14. Intermediate Microeconomics
by Hal R. Varian (Textbook Binding — February 1, 1999)

15. Economics, Organization and Management, The
by Paul Milgrom (Author), John Roberts (Author) (Hardcover — February 1, 1992)

16. Applied Microeconomics
by Edwin Mansfield (Hardcover — January 1997)

17. Forecasting : Methods and Applications
by Spyros G. Makridakis (Author), et al (Hardcover — December 1997)

18. Managerial Economics
by William F. Samuelson (Author), Stephen G. Marks (Author) (Hardcover — April 19, 2002)

19. Price Theory
by Milton Friedman (Hardcover — June 1, 1976)

20. Principles of Microeconomics
by N. Gregory Mankiw (Paperback — June 6, 2000)

Net Support

Introductory Economics

Cybernomics: A Semi-Interactive, Almost Multimedia Way to Learn Economics

An online principles text by Robert E. Schenk (Saint Joseph's College, IN)

The Economics Net-TextBook by Ted Black (University of Maryland)

This online text provides useful simulation and review material on an extensive collection of micro, macro, and international economics topics.

Essential Macroeconomic Principles by John Bouman (Howard Community College)

An online macroeconomics text available in either Adobe pdf or Microsoft Word format.

Essential Miacroeconomic Principles by John Bouman (Howard Community College)

An online microeconomics text available in either Adobe pdf or Microsoft Word format.

Essential Principles of Economics: A Hypermedia Approach by Roger A. McCain (Drexel University)

The most extensive attempt at an online economics text using hypermedia.

Human Society and the Global Economy by Kit Sims Taylor (Bellevue Community College)

This site contains an online version of a survey text in economics that relies on an institutionalist/Post-Keynesian approach. The text may be viewed online or *Word 7.0* copies of the chapters may be downloaded.

A Pedestrian's Guide to Economics by Orley Amos (Oklahoma State University)

An online expanded version of his *Economic Literacy* book. This book is designed to provide an intuitive explanation of economic concepts to the general public (and struggling principles students).

Introductory Microeconomics

Elements of Economics: The People's Introduction to Economic Theory by Richard Ruble (University of Virginia)

This work is a draft of an introductory/intermediate level microeconomics text.

Intermediate Microeconomics

Price Theory: An Intermediate Text, by David Friedman

An online copy of the text that was published by SouthWestern College Publishing.

Elements of Economics: The People's Introduction to Economic Theory by Richard Ruble (University of Virginia)

This work is a draft of an introductory/intermediate level microeconomics text.

International Economics

International Trade Theory & Policy Analysis by Steven M. Suranovic (George Washington University)

This web site contains links to a free html version of the text as well as a pay-per-view version in Adobe Acrobat format. The free version does not contain answer keys.

Econometrics

Handbook of Econometrics, vol. 1-4

Elsevier has now placed the first 4 volumes of the *Handbook of Econometrics* online in Adobe pdf format.

Structural Analysis of Discrete Data and Econometric Applications, edited by Charles F. Manski and Daniel L. McFadden

The complete text of this classic 1981 MIT Press volume is available in either Adobe PDF or Postscript format.

Urban Travel Demand: A Behavioral Analysis by Tom Domencich and Daniel L. McFadden

The entire text of this 1975 North-Holland volume is available at this site in both Adobe PDF and Postscript formats.

Game Theory

Game Theory Evolving by Herbert Gintis

This text provides a dynamic and evolutionary approach to game theory. This text will remain online until its publication by Princeton University Press in 2000.

Strategy and Conflict: An Introductory Sketch of Game Theory by Roger A. McCain

This online text contains a brief, nontechnical solid introduction to game theory concepts and their application to economics.

Law and Economics

Law's Order: What Economics Has to Do With Law and Why It Matters" by David Friedman

This online copy of a 2000 Princeton University Press text provides a superb introduction to the relationships between law and economics. Topics included in this text include economic analyses of: crime, externalities, marriage, fertility, divorce, the value of life, contract law, tort law, and many other topics.

Public Choice

Understanding Democracy: An Introduction to Public Choice by J. Patrick Gunning

This text provides a basic introduction to public choice theory. He asks that users of this text provide him with feedback and suggestions.

Production Economics

Production Economics: A Dual Approach to Theory and Applications. Volume I: The Theory of Production, edited by Melvyn Fuss and Daniel L. McFadden

The entire text of this 1981 North Holland volume is available online in either Adobe pdf or Postscript formt.

Public Finance

Public Finance: Government Revenues and Expenditures in the United States Economy by Randall G. Holcombe

This site provides the complete text (in either .pdf or Word format) of the Public Finance text published by West in 1996 (currently out of print).

EIILM UNIVERSITY
SIKKIM

Jorethang, District Namchi, Sikkim- 737121, India
www.eiilmuniversity.ac.in