The Theoretical Essentials of Intermediate Microeconomics

second edition

Donald W. Katzner
University of Massachusetts/Amherst

This book is published by the University of Massachusetts Amherst Libraries

https://openpublishing.library.umass.edu/

DOI: 10.7275/dcby-t473

ISBN: 9781945764325

Preface

Over the past 50 years or so, if textbooks are any guide, intermediate microeconomics courses have expanded their purview by adding topics such as uncertainty, game theory and behavioral economics. They have also increasingly identified theoretical constructions as techniques for solving microeconomic problems and greatly expanded their focus on applications to real economic issues. What seems to have become somewhat lost in all of this is the imparting of (i) an understanding of the power and significance of theoretical argument and, perhaps more importantly, (ii) an appreciation of one of the central aims of microeconomic theory itself, namely, to provide an explanation of how a real, capitalist microeconomy might operate. That is, with respect to the latter and in the slightly modified words of K.J. Arrow and F.H. Hahn,[1] how is it that a decentralized economy motivated by self-interest and guided by price signals could be compatible with a coherent distribution of resources and productive outputs? Why wouldn't there be chaos? The standard answer to these questions and the sought-after coherence provided by economists is usually some version of the Walrasian model of the microeconomy, a structure that is often seen as part of the theoretical basis for capitalism as it currently appears in much of economic reality today. The belief that there is coherence of this form in the real microeconomic world is held not only by economists but by many non-economists as well. It seems, therefore, that this should be an important part of any intermediate microeconomics course.

This de-emphasis of theoretical argument and the coherence issue has manifested itself in several ways. First, theoretical arguments do not stand out as much as they might. Second, there are a number of theoretical matters that are not given the attention they deserve. For example, the role of assumptions, what happens to the theoretical constructions those assumptions generate if the latter are weakened or discarded, and the relation of those assumptions or their implications or the structures they generate to the real microeconomy, tend to be largely ignored. Third, even in chapters devoted to discussions of general equilibrium found in many textbooks, the focus is on how the individual pieces of the microeconomy (consumers, firm, and markets) interact with each other. The fact that these pieces fit together to form a significant unified totality that can be thought to say something about functioning of the real microeconomy as a whole receives scant emphasis and is likely to be missed by students. Fourth, important methodological matters such as what it means to theoretically explain real economic phenomena like a consumer's or firm's behavior or the operation of a market or of a microeconomy, that is, how such explanation actually works to explain, how explanation in economics differs from explanation in the physical sciences, and the fact that certain theoretical explanations are actually accepted by economists and the general public as providing the reason for an observed economic phenomenon, are not given much attention. And lastly, the ways the general. real microeconomy can be evaluated in order to determine if action should be taken to make improvements in actual microeconomic outcomes often does not stand out as much as it could.

These points are the main focus of the following 25 chapters. I have tried to present and bring to the forefront a clear and uncluttered discussion centering attention narrowly on the basic

[1] *General competitive Analysis* (San Francisco: Holden Day, 1971), p. vii.

theoretical ideas, arguments, and interrelationships that constitute the traditional Walrasian explanation of the workings of the perfectly competitive microeconomy, along with a few examples of market failures to account for more realistic microeconomic circumstances. Many details and asides that appear in standard textbooks are omitted. Discussion is set in the context of a world with two final goods, two persons, two firms, and two inputs or resources,[2] and is often phrased in reference to graphs. There is no game theory and no discussions of uncertainty and behavioral economics. Nor are there any applications. Only a few examples, usually abstract, illustrate theoretical ideas. In an Appendix, notes supplemental to the 25 chapters provide summaries of scattered material and several detailed explorations not contained in the chapters themselves so as to avoid interrupting the thread of argument. This volume is not intended as a replacement for a complete, standard intermediate economics textbook. Its only purpose is to allow an uninterrupted and distraction-free concentration on the fundamental theoretical pieces of intermediate microeconomic theory and on the coherence and other issues described above.

The prerequisites necessary for fully understanding these chapters are a university-level introductory course in microeconomics and a university-level first semester of calculus. For better or worse, in today's environment it has become impossible to avoid the use of mathematics in presenting much of theoretical microeconomic argument. But in the present volume, mathematics is used more to represent concepts as a language for communication than for solving problems or as a substitute for verbal argument – although the derivative derivation of first- and second-order conditions for the maximization or minimization of functions of single variables is occasionally used to determine economic outcomes. And the second chapter below provides a summary of the main mathematical ideas that are employed.

Finally, I would like to thank Michael Ash for encouraging me to pursue this project. Thanks are also due to him and Iryna Bobukh for reading the entire manuscript and making helpful comments and suggestions, and to Erin and Todd Katzner for assistance with formatting. It should aso be noted that the present, second edition of this book differs from the first only in that it corrects errors, fills in gaps in reasoning, and adds clarity to arguments.

[2] In this context, the vision of the microeconomy presented in these notes is succinctly summarized in D.W. Katzner, *The Walrasian Vision of the Microeconomy*: *An Elementary Exposition of the Structure of Modern General Equilibrium Theory* (Ann Arbor: University of Michigan Press, 1989).

Table of Contents

Chapter 1
Introduction

Why is it that when an individual wants to buy a rug, there are various rugs available in various stores for him/her to buy? Why is it that if a person wants to work, there are generally jobs available for which he/she can apply? The answers to these kinds of questions all have to do with the functioning of what is called the real microeconomy.

The purpose of this volume is to present an explanation of how a real (capitalist) microeconomy might operate, examine a specialized evaluation of its outcomes, and consider some alternative possibilities. Formally, the real microeconomy of a country or society consists of the collection of all consumers, all firms, all markets, and all of their economic interactions. It is geared toward the production of commodities and services from a distribution of its factors or resources – land[1], labor, capital,[2] and enterprise[3] – among firms. The distribution of its production output among consumers is, in part, the outcome of microeconomic activity.

The first step is to describe what is being explained. However, the actual microeconomy is so large and complex that it is way beyond the capacity of human beings to fully describe. That is, there are many different consumers, each with different goals and preferences among commodities. There are many different firms, each with different aims and means of producing even the same commodities. There are many different varieties of the same commodity and of the same kinds of land, labor, and capital, each sold at possibly different prices. And lastly, there are many different markets for every commodity and every factor, each with its own characteristics. Due to this immense variety of elements and the intricate and entangled relationships among them, it is necessary to abstract from reality to obtain a simplified version of the microeconomy that is capable of being handled.

The method of abstraction employed here is to regard all consumer ways of determining what they will buy and sell as the same and all firm procedures for determining what outputs they will produce and sell along with the inputs they will buy for producing those outputs as identical.[4] The method also condenses all varieties and prices of the same commodity or resource into one commodity or resource and one price, and all markets for the same good into a single market. Thus the consumers, firms, goods, and markets in this abstraction are mental constructs that do not exist in reality. In addition, consumers will be taken to own all of the economy's resources, and the presence of governments, foreign economies, and banking systems in the real microeconomy will be ignored. The workings of this abstract microeconomy are pictured in the schematic, circular-flow diagram of Figure 1-1. In that diagram, consumers sell all resources to firms and receive the payments of rent for land, wages for labor, what will be referred to as

[1] The term 'land' includes such things as iron ore in the ground and fish in lakes and streams.

[2] 'Capital' refers to physical items like machines and factories used to repeatedly produce many units of output.

[3] 'Enterprise' will be taken to mean the willingness and ability to assume the risk of creating and running a business concern.

[4] That is, in the case of consumers, each individual can buy or sell different things. But all consumers make buying and selling decisions using the same method. Similarly, different firms can produce different products with different inputs. But they, too, all make their producing, selling, and buying decisions by in the same manner.

returns for capital, and profit for enterprise. They use all of the payments they receive to buy commodities produced by firms. Firms, on the other hand, produce capital and noncapital commodities for consumers and other firms in the microeconomy using all of the revenue obtained from the sale of their output to buy from consumers and other firms the inputs needed to produce that output. All goods (which move along the outer circular lines in the diagram) and payments (which move along the inner circular lines) flow through markets. The commodity or final goods markets are represented in the upper half of the diagram, the resource or factor markets in the lower half of the diagram. Intermediate commodity markets in which firms buy and sell to each other fall within the box labeled Firms.[5] The buyers and sellers in each market are indicated in Table 1-1. When drawing attention to the seller's side of a market for a produced good, the market is often referred to as an industry.

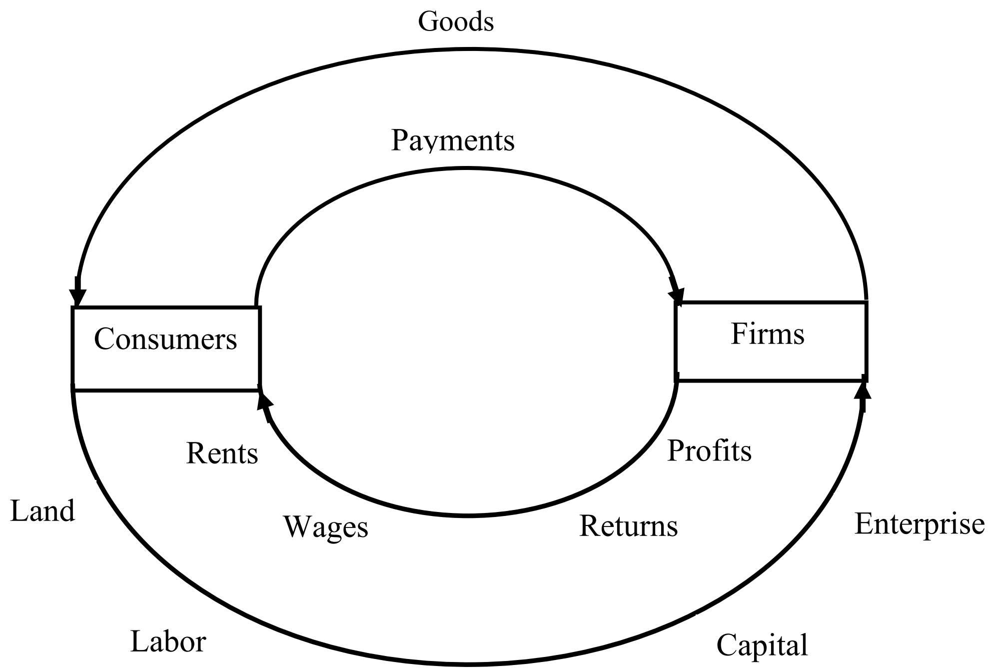


Figure 1-1

Table 1-1

Market	Buyers	Sellers
Final Commodities	Consumers	Firms
Intermediate commodities	Firms	Firms
Factors	Firms	Consumers

[5] Then possibility of consumers buying and selling to each other (in the box labeled Consumers) is not considered although consumers trading among themselves is briefly discussed in Chapter 18.

To understand how the real microeconomy operates, an explanation will first be developed of how the abstract or simplified microeconomy of Figure 1-1 and Table 1-1works to produce final goods for consumers. The explanation will consist of four parts:

1. An explanation of consumer buying and selling behavior in each market.
2. An explanation of firm buying and selling behavior in each market.
3. How markets operate to distribute commodities to consumers and factors to firms.
4. How all of these elements fit together to form the (simplified) microeconomy.

This explanation will then be applied (as an approximation) in understanding the functioning of the real microeconomy.

The method used to explain and understand the way the simplified microeconomy operates is to construct a model. A model of something, call it A, is a construct having enough in common with A so that insight into A can be obtained by studying the construct.[6]

> **Example:**[7] Suppose you were asked to explain how a particular clock with observable rotating hands works, but you are not permitted to remove its cover and look inside. One way to do this is to obtain the necessary parts and construct a physical model of the clock that duplicates its observable behavior: Your model would have the same numbers in the same relative places and hands that rotate in the same way and at the same speed as the clock you are trying to explain. You can then open up your model to display how it works and say that the clock works like my model. Because the workings of the clock cannot be seen, the workings of your model are unlikely to be identical to those of the clock. But you still have one possible explanation of how the clock works. Moreover, there are many models that will produce the same observable behavior of the clock. And so there are many possible explanations of how the clock works.

Recognizing that much of what goes on behind purchases and sales in the real microeconomy cannot be seen, a model whose purpose is similar to that of the clock example will be constructed to explain the buying and selling behavior of consumers and firms in all types of markets, how markets operate, and how all of this fits together to form a unified and coherent whole, in other words, how the simplified microeconomy works. That model will then be used as an explanation of the workings of the real microeconomy.[8] Unlike the model of the clock, which is a physical creation constructed with tangible parts, the present model is a mental formulation built from assumptions and structures derived from them. In this model:

1. Consumer tastes or preferences among commodities and the technology for producing

---

[6] What is considered to be "enough in common with A" depends on the circumstances relating to the model's construction, such as the use to which the model is to be put and the purpose of the analysis of which the model is a part

[7] A. Einstein and L. Infeld, *The Evolution of Physics* (New York: Simon & Schuster, 1938), p. 33.

[8] Like the model of the clock, this provides only one possible explanation of how the real microeconomy operates.

output available to firms are taken to be specified and fixed.

2. Individuals and firms decide how much to buy and sell through a process of maximization.

3. Markets are taken to be perfectly competitive in that they have
   a) a large number of small buyers and sellers, each sufficiently small and there are a sufficiently large number of them so that no buyer or seller can have any influence on market price.
   b) a standardized (homogeneous) product,
   c) free entry into and exit from them, and
   d) perfect information in that all buyers and sellers have the same information about products and prices in them.

4. Market forces operate to equilibrate demand and supply.

This model is often referred to as Walrasian after its creator Léon Walras. It is also called a general-equilibrium model because all of its variable elements are permitted to interact simultaneously. When one or more variable element is held fixed while those remaining are allowed to vary, the model is reduced to what is called a partial-equilibrium model.

While this method of analysis and explanation applied to the real microeconomy is similar in many ways to that employed in the physical sciences, there are at least two major differences. First, the real microeconomic world is much more fluid and subject to change than the real physical world. Physical laws such as the Law of Gravity generally apply to all time – past, present, and future – and across all space throughout the universe. But what might pass for a laws in economics often has no such constancy. For example, general characteristics of human preferences, what in subsequent chapters will be seen as the foundation for the general properties of consumer buying and selling behavior, can vary over time and be different in different places. A change in non-economic circumstances such as an illness, the birth of a child, or moving to a different climate can affect such preference characteristics. Thus, although an aspect of a specific individual's economic behavior might appear as a law at one time and in a particular place, it need not apply at a different time or in a different place.

Second, economists do not have a laboratory with sufficient controls to be able to regard the results of a test of the validity of their theoretical conclusions about reality with the same confidence as physicists and chemists. To determine the speed of falling objects in a vacuum, the physicist can generally ensure that the (near-vacuum) environment in which objects are actually dropped does not change significantly from one drop to the next, and thereby conclude with reasonable confidence that all objects, such as a metal ball and a feather, fall at the same rate. But laboratory experiments in economics involve eliciting responses from human beings to stimuli in laboratory-created circumstances. And the latter are incapable of accurately duplicating actual real-life situations. Of course, participants can be threatened with the deprivation of monetary gain that might accrue from their participation in the experiment. But they cannot be threatened with the loss of their real-life incomes or their real-life jobs. Unlike decision-making in real-life circumstances, decisions made in response to laboratory stimuli have no consequences or impact on the participant's life outside of the laboratory apart from whatever small monetary gain is obtained. In this sense, taking action (e.g., buying or selling) in real-life situations is different from indicating what action one might take in a laboratory. For example, attitudes towards purchasing a commodity may be different in the two situations. The results of such experiments, then, cannot carry the same force as results of experiments in the physical sciences.

It will take the next 16 chapters to build the Walrasian model described above. Although more general aspects will often be provided, analytical parts will be set in a world of two persons, two consumer goods, two firms and two resources or inputs (labor and capital). Upon completion of this construction, what it suggests about the "welfare" derived from the consumption commodities produced and distributed to individuals by the real microeconomy will be explored. Finally, adjustments in parts of the model will be introduced to account for more realistic situations (outside of the realm of perfect competition) that involve what are called market failures.

Chapter 2
Mathematics Used in this Book

In this volume, mathematics is used mostly as a language for communication rather than as a means for solving problems. The main ideas to be employed are as follows:

A. Functions of a single variable.

In general, a function is defined as consisting of two sets, a domain and a range, together with a rule. The rule, call it f, assigns to each element in the domain a unique element in the range. Formally, the function is written as

$$y = f(x),$$

where x, the independent variable, varies over the domain and y, the dependent variable, varies over the range.

There are three ways of making functions specific. First is the equation form, for example, $y = 2x + 1$ with domain, say, $\{x: x \geq 0\}$. Second is the tabular form, where specified values for x are listed on the right and the corresponding unique values for y generated by $y = 2x + 1$ appear on the left as in Table 2-1:

Table 2-1

y	x
1	0
3	1
5	2

The third is the geometric form of Figure 2-1 secured by plotting the numbers of Table 2-1 on a graph with x on the horizontal axis and y on the vertical, and given the present specification of the function as $y = 2x + 1$, connecting the points so obtained with a straight line:

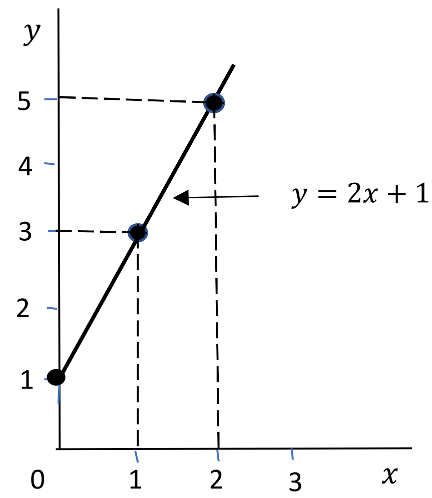


Figure 2-1

For purposes of this book, it will be convenient to regard all three representations of $y = 2x + 1$ as equivalent.

The first-order derivative of f at x^0 is defined as

$$f'(x^0) = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x^0 + \Delta x) - f(x^0)}{\Delta x},$$

where $\Delta x = x' - x^0$ and $\Delta y = y' - y^0$. The geometric depiction of the derivative of f at x^0 is the slope of the graph of f at x^0, that is, the slope of the red straight-line tangent to the curve at x^0, as pictured at point A in Figure 2-2. In this diagram the graph of f is drawn as the curved

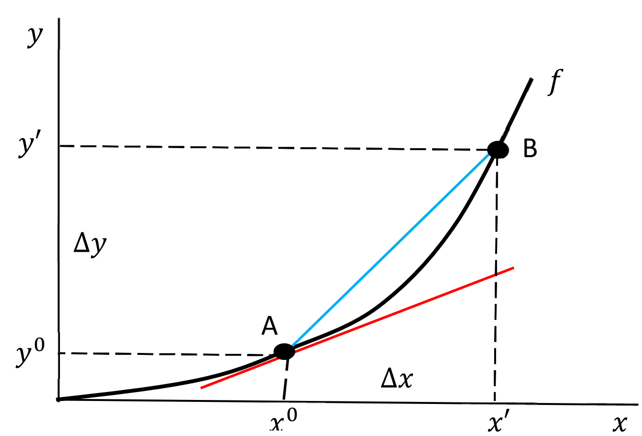


Figure 2-2

line emanating from the origin. The expression

$$\frac{\Delta y}{\Delta x} = \frac{f(x^0 + \Delta x) - f(x^0)}{\Delta x} \tag{2.1}$$

represents the slope of the blue straight line connecting points A and B. That slope is an approximation of the slope of the curved line at A. The approximation gets better and better as $x' \to x^0$ or, equivalently, as $\Delta x \to 0$.

In Economics, derivatives are often called marginals, such as marginal cost or marginal utility. It will often be convenient in subsequent chapters to make no distinction between the derivative of f at x^0 (or the slope of its graph at x^0) and approximations of it. That is, for example, the marginal cost at x^0 denoted by, say $MC(x^0)$, may be thought of either in derivative or in approximate terms.

In approximate form, derivatives or marginals can be calculated from tables like that of Table 2-1 even if the numbers of the table correspond to points on a curved line rather than the straight line of $y = 2x + 1$. In Table 2-1, if x represents, say, output and y the total cost of producing that output, then the approximate derivative at $x = 1$ or the marginal cost of producing the second unit of output when already producing the first is calculated using equation (2.1) as ${}^{\Delta y}/_{\Delta x} = 2$, where $\Delta x = 2 - 1 = 1$ and $\Delta y = 5 - 3 = 2$. In this case, at any point on the graph the approximate derivative is always the same as the actual derivative. If the numbers of a table related to a function whose graph is the curved line in Figure 2-2 where, say, $x^0 = 1$, $x' = 2$, $y^0 = 1$, and $y' = 6$, then the calculation ${}^{\Delta y}/_{\Delta x} = 5$ yields the slope of the blue line in Figure 2-2 as an approximation of the derivative or slope of the curve at x^0 (that is, the slope of the red line).

The second-order derivative of f at x^0, written $f''(x^0)$, is the derivative of $f'(x)$ at x^0.

If a function is known to have a unique maximum or minimum value at an x^0 somewhere in its domain, and if it has continuous first- and second-order derivatives everywhere, then that x^0 can be found by solving the first-order condition

$$f'(x^0) = 0.$$

To determine whether $f(x^0)$ is a maximum or minimum, it is sufficient to look at the sign of the second-order derivative at x^0. Thus

$$f''(x^0) \begin{cases} < 0 & \text{indicates a maximum,} \\ > 0 & \text{indicates a minimum.} \end{cases}$$

B. Functions of two variables.

The general definition of a function of two variables is the same as that of a single variable except that now the elements of the domain contain pairs of numbers as in (x, y) and the function is written as, say, $z = f(x, y)$. As an illustration, take f to be

$$z = (xy)^{1/4} \tag{2.2}$$

on the domain $\{(x, y): x \geq 0 \text{ and } y \geq 0\}$. The equivalent tabular form would have three columns instead of two, and the 3-dimensional graph drawn in perspective in the two-dimensional plane of the paper would look like a portion of a bowl on its side with its bottom point located at the origin as in Figure 2-3:

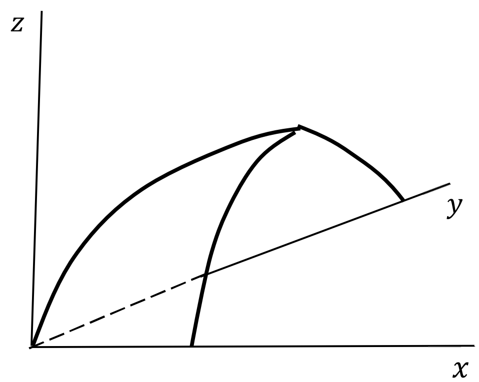


Figure 2-3

Derivatives of these functions will be obtained by holding x or y fixed at, say x^0 or y^0 respectively, thereby reducing f to a function of a single variable, and applying the above definition of derivative to the reduced function. The first-order derivative at x^0 with respect to x holding y fixed at y^0 is denoted by $f_x(x^0, y^0)$; that at y^0 with respect to y holding x fixed at x^0 by $f_y(x^0, y^0)$. Thus, for example,

$$f_x(x^0, y^0) = \lim_{\Delta x \to 0} \frac{f(x^0 + \Delta x, y^0) - f(x^0, y^0)}{\Delta x},$$

where, as before, $\Delta x = x' - x^0$.

Geometrically, the 3-dimensional graph of a function of two variables can be reduced to a two-dimensional graph by slicing a plane through the former and focusing on the graph that appears in that plane. The resulting two-dimensional graph can be thought of as the geometric picture of a function of a single variable. Three methods of doing this, illustrated below with respect to the 3-dimensional graph of Figure 2-3 above, will be employed in subsequent chapters. In each case, the curve at issue in the left-hand 3-dimensional diagram appears in red and is reproduced in red in the relevant two-dimensional coordinate system in the right-hand diagram.

i. In Figure 2-4, a plane at z^0 parallel to the x-y plane cuts through the 3-dimensional graph of Figure 2-3. The curve in the z^0 plane is projected vertically onto the x-y plane. All points on the projected curve in the x-y plane have the same function value z^0. That curve can be described mathematically as $\{(x, y)\colon\ z^0 = f(x, y)\}$ or as the graph of a function

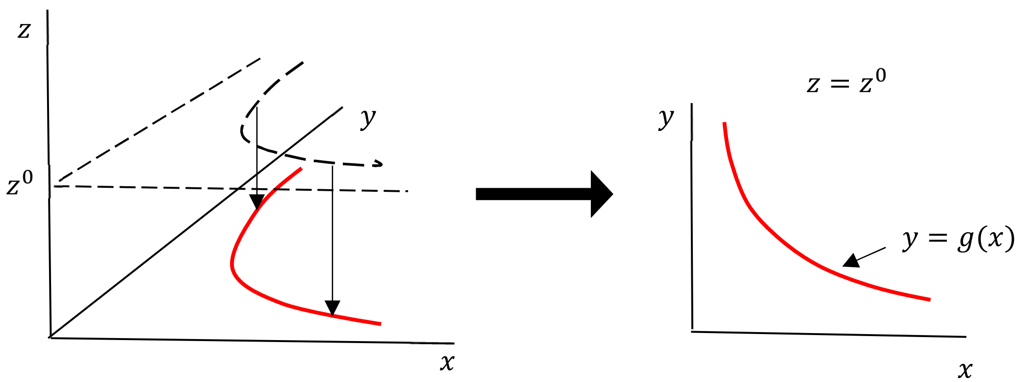


Figure 2-4

written as, say, $y = g(x)$ and obtained by solving the equation $z^0 = f(x, y)$ for y as a function of x. In terms of equation (2.2), $z^0 = f(x, y) = (xy)^{1/4}$ and solving for y gives

$$y = g(x) = \frac{(z^0)^4}{x}.$$

ii. In Figure 2-5, a plane at y^0 parallel to the x-z plane cuts through the 3-dimensional graph of Figure 2-3. The curve in the y^0 plane is the graph of the function of the single variable $z = f(x, y^0)$. From equation (2.2), the latter becomes $z = f(x, y^0) = (xy^0)^{1/4}$.

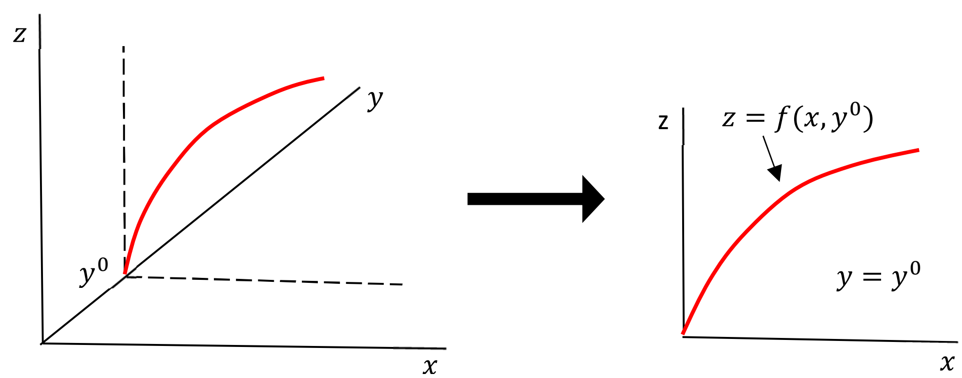


Figure 2-5

iii. In Figure 2-6, a plane at x^0 parallel to the y-z plane cuts through the 3-dimensional graph of Figure 2-3. The curve in the x^0 plane is the graph of the function of the single variable $z = f(x^0, y)$ or, with respect to equation (2.2), $z = f(x^0, y) = (x^0 y)^{1/4}$.

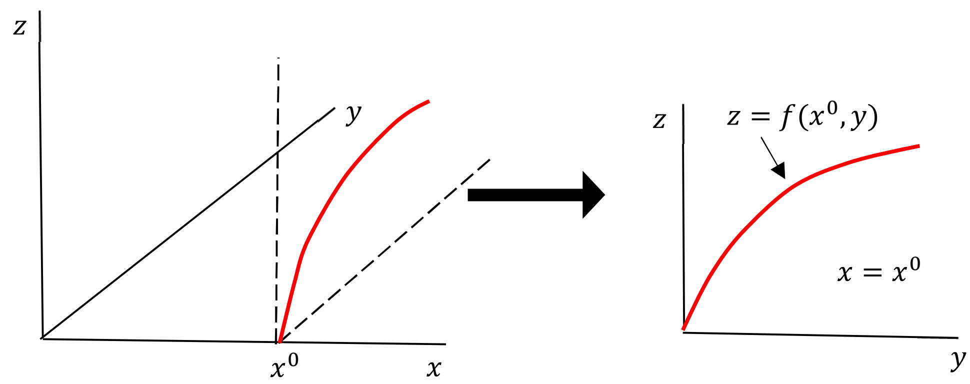


Figure 2-6

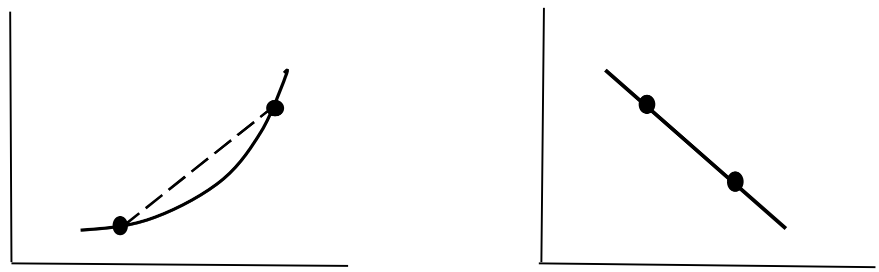
C. Convexity and Concavity.

A set *S* is convex if the straight-line segment connecting **any** two points of *S* lies entirely in S as in the left-hand diagram of Figure 2-7. In the right-hand diagram, the straight line

Figure 2-7

falls outside of the set for some pairs of points in *S*. That is why the set is not convex.

A function f is convex if the straight-line segments connecting **any** two points on its graph lies on or above the graph as in both diagrams of Figure 2-8. It is strictly convex if **all** such line segments lie entirely above the graph except for their end points. Only the function whose graph appears in the left-hand diagram of Figure 2-6 is strictly convex. The set of all points lying above the graph of a convex function is convex.

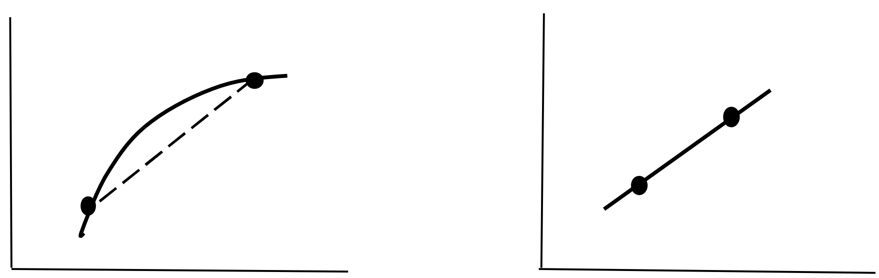


Convex and Strictly Convex

Convex, Concave, and Linear

Figure 2-8

A function f is concave if the straight-line segments connecting **any** two points on its graph lies on or below the graph as pictured in Figure 2-9 and the right-hand diagram of Figure 2-8. It is strictly concave if **all** such line segments lie entirely below the graph except for their end points as in only the left-hand diagram of Figure 2-9. The set of all points lying below the graph of a concave function is convex.



Concave and Strictly Concave

Concave, Convex, and Linear

Figure 2-9

The graphs of convex or concave functions can be linear or contain straight-line segments. This is not possible for graphs of strictly concave or strictly convex functions.

A function is linear if and only if it is both concave and convex.

As can be seen in the left-hand diagrams of Figures 2-8 and 2-9, if the second-order derivative of a function is everywhere positive, that function is strictly convex. If the second-order derivative is everywhere negative, the function is strictly concave. It is the second-order derivative that determines the strictly-convex or strictly-concave shape of the curve.

The function whose graph appears in Figure 2-10 has strictly convex and strictly concave parts but is neither strictly concave nor strictly convex.

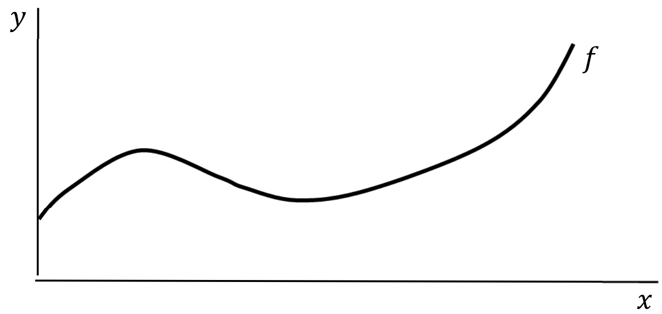
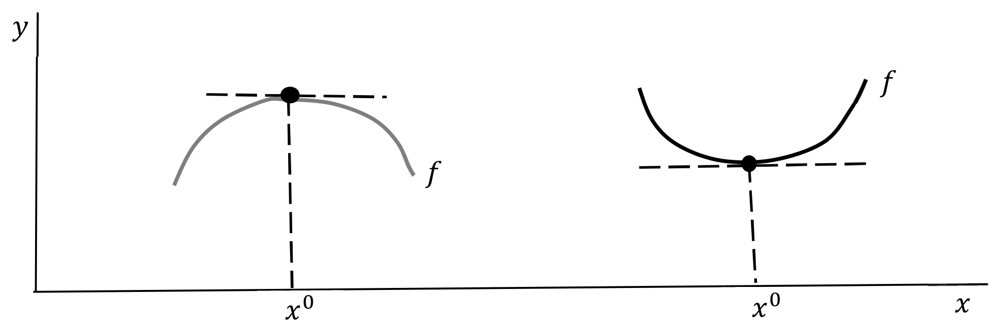


Figure 2-10

The reason for introducing the ideas of strictly concave and strictly convex functions is that, in this book, their use will sometimes avoid the second-order maximization and minimization conditions in derivative terms as described above. This simplifies matters considerably, especially when dealing with "constrained" maximization and minimization and functions of two variables. To illustrate how this works in the simplest context of a function of a single variable, consider the following schematic: To say that $y = f(x)$ has a unique maximum or minimum at x^0:

	<u>Maximum</u>	<u>Minimum</u>
	$f(x^0) > f(x)$ for $x \neq x^0$	$f(x^0) < f(x)$ for $x \neq x^0$



First-order condition:	$f'(x^0) = 0$	$f'(x^0) = 0$
Second-order condition:	$f''(x^0) < 0$	$f''(x^0) > 0$
Assumption to replace the second-order derivative condition:	f is strictly concave	f is strictly convex

Thus, combining the first-order condition with the statement that the functions is strictly concave or strictly convex is sufficient to determine if the function has, respectively, a maximum or minimum at x^0.

The strictness of concave and convex functions is important here. If only concavity and convexity were assumed without strictness, the graph of a function for which a maximizing or minimizing value of x is sought could appear as the straight line in Figure 2-11 below. In this case all values of x both maximize and minimize f. Suppose, then, that one wants to explain,

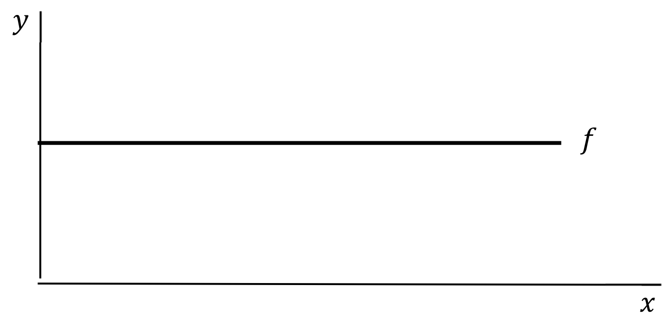


Figure 2-11

say, the amount of output a firm produces at some moment or period of time as the outcome of maximizing a profit function. Since the firm produces only one amount of output, the appropriate function has to determine, upon maximization, a unique output value. Functions with graphs pictured in Figure 2-11 cannot do that. Therefore, when the maximization or minimization of a concave or convex function is intended to pick out a unique value in its domain, strictness has to be assumed.

Chapter 3
The Operation of Markets:
Demand and Supply

The first step taken here in the construction of a model to explain the functioning of the microeconomy is to focus on the operation of an isolated market in the context of the abstract world described in Chapter 1. Because it is a mental construct, neither the model as a whole nor any of its parts exist in the real microeconomic world. To be concrete, focus on the market for a commodity in which buyers are consumers and sellers are firms. The demand and supply sides of that market are expressed in terms of market demand and supply functions. To obtain those functions, present discussion begins with the general definition of demand functions for individual consumers. These will be explained and derived from consumer preferences in later chapters. To further simplify, consider a world with only two persons 1 and 2, and two consumer goods x and y. The same symbol is used to denote a good and quantities of it. The following notation is used:

x_1 – quantities of good x demanded by person 1. Similarly for y_1, x_2, and y_2.
p_x, p_y – prices of good x and good y.
m_1, m_2 – incomes of persons 1 and 2 (the amount of money he/she has to spend).

The demand functions h^1 and g^1 for, respectively, goods *x* and *y* of person 1, and the corresponding functions h^2 and g^2 for person 2 are written as

$$x_1 = h^1(p_x, p_y, m_1) \qquad x_2 = h^2(p_x, p_y, m_2)$$
$$y_1 = g^1(p_x, p_y, m_1) \qquad y_2 = g^2(p_x, p_y, m_2)$$

where h^i and g^i, for $i = 1{,}2$, are the symbols (names) identifying the functions and x_1, say, becomes the quantity of good x demanded by person 1. These functions, defined in the standard mathematical way for all positive prices and incomes, indicate the quantities the induvial demands of the two goods at the two price and income values.

The individual demand curve is obtained by fixing the "other price" and income in the demand function and graphing the function of the single variable that remains. For example, the demand curve for good *x* of person 1 is found by setting $p_y = p_y^0$ and $m_1 = m_1^0$ in h^1 and graphing $x_1 = h^1(p_x, p_y^0, m_1^0)$ to obtain the curve labeled h^1 and illustrated in Figure 3-1. Note the reversal of mathematical convention in the labeling of the axes: The independent variable p_x appears on the vertical axis while the dependent variable x_1 appears on the horizontal.

There are a number of ideas in relation to demand curves that should be noted:

1. The "law of demand" is the statement that demand curves slope downward and to the right. This is not a law is the sense that the term is used in the physical sciences (recall Chapter 1). As will be seen in Chapter 7, theoretically the law

does not appear to hold up as a law. Empirically it almost always seems to hold, although there have been observations of upward sloping demand curves. Goods with upward sloping demand curves are often called Giffen goods.

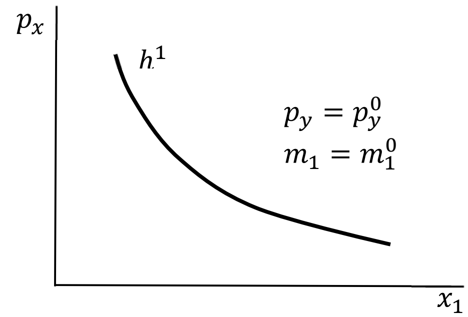


Figure 3-1

2. The phrase, "changes in quantity demanded" indicates a movement along a demand curve with the other price and income held constant. In terms of h^1, a change in the quantity demanded means that a variation in x_1 has occurred in response to a modification in p_x (with p_y and m_1 held fixed). The phrase "changes in demand" refers to shifts or movements in the entire demand curve caused by variations in the fixed values of p_y or m_1. The ideas reflected in these phrases are important in the distinction between substitute and complementary goods, and normal and inferior goods described below.

3. Two goods are called substitutes if a rise or fall in the price of one (a movement along its demand curve) results in, respectively, an outward or inward shift in the demand curve of other. As an example, suppose coffee, good y, and tea, good x, are substitutes for consumer 1. Let the current price of coffee be p_y^0. Then a rise in p_y^0 means a reduction in the quantity of coffee demanded along consumer 1's coffee demand curve. Since the price of coffee was originally fixed at p_y^0 behind consumer 1's demand curve for tea, that is, behind the graph of $x_1 = h^1(p_x, p_y^0, m_1^0)$, the increase in the price of coffee causes consumer 1 to substitute tea for coffee regardless of the price of tea, thereby pushing his/her demand curve for tea farther out from the origin.

4. Two goods are called complements if an increase (decrease) in the price of one results in a decrease (increase) in demand, i.e. an inward (outward) shift, in the other. A possible example: coffee and cream. Assuming the consumer uses cream in his/her coffee, a rise in the price of coffee will reduce the quantity of coffee demanded and the need for cream. Since the price of coffee is fixed behind the demand curve for cream, the reduced demand for coffee will decrease the demand for cream at all prices of cream.

5. A good is called normal if a rise (or fall) in the income of the consumer results in an increase (or decrease) in its demand. A possible example: steak. Since income is fixed behind the demand curve for steak, an increase in the consumer's income may induce him/her to demand more steak at all prices.

6. A good is called inferior if a rise (or fall) in the income of the consumer results in a decrease (or increase) in its demand. A possible example: potatoes. Since income is fixed behind the demand curve for potatoes, an increase in the consumer's income may allow him/her to increase consumption of more expensive foods and demand fewer potatoes no matter what their price.

Market demand functions are obtained as the sum of the individual demand functions of all buyers in the market. In the present case, with $x_1 = h^1(p_x, p_y, m_1)$ and $x_2 = h^2(p_x, p_y, m_2)$,

$$x_1 + x_2 = h^1(p_x, p_y, m_1) + h^2(p_x, p_y, m_2). \tag{3.1}$$

Letting $x = x_1 + x_2$ and $D(p_x, p_y, m_1, m_2) = h^1(p_x, p_y, m_1) + h^2(p_x, p_y, m_2)$, the market demand function (3.1) may be written as

$$x = D(p_x, p_y, m_1, m_2).$$

The market demand curve (obtained from D in the same way as the individual demand curve is derived from h^1) is the graph of $x = D(p_x, p_y^0, m_1^0, m_2^0)$ and shifts with changes in p_y^0, m_1^0, or (and) m_2^0. Because of the reversal of axes in graphing individual demand curves, the geometric derivation of the market demand curve from individual demand curves involves a "horizontal" sum as pictured in Figure 3-2 with linear curves:

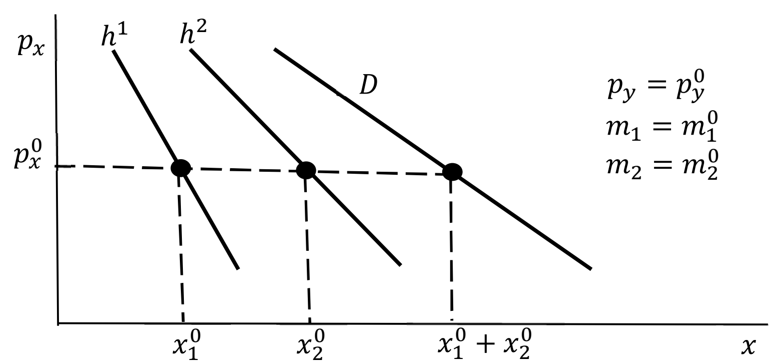


Figure 3-2

The formulation of market and individual supply functions and curves, say in relation to good x and its price p_x, is similar to that of market and individual demand functions and curves except that:

1. Since the sellers are firms in the market under consideration here, individual supply functions and curves emerge from them (Chapter 14) and the dependent variable in those functions and curves, x, represents quantity supplied rather than quantity demanded.

2. Prices other than p_x – but not incomes and not p_y – appear as arguments of supply

functions. Thus a seller's supply function indicates the quantities supplied at each collection of price values. The other prices will be specified in subsequent chapters.

3. Supply curves, obtained by graphing the supply function with p_x varying while holding the other prices fixed, slope upward – the "law of supply."

Consider now the market for good x with equations for market demand and supply functions D and S written more simply as:

$$x = D(p_x) - \text{ market demand and}$$
$$x = S(p_x) - \text{market supply,}$$

where x now represents both quantities demanded and supplied. Other price and income variables, although hidden from view, are still present and held fixed. Let the graphs of these functions be drawn in typical fashion as in Figure 3-3. The intersection point (x^0, p_x^0) is called an

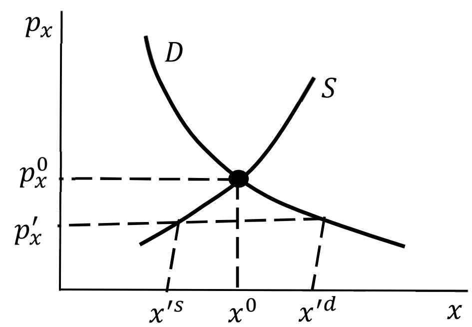


Figure 3-3

equilibrium, with p_x^0 the equilibrium price and x^0 the equilibrium quantity. This terminology is to be understood as follows:

To say something is in equilibrium is to say that it is in a position of rest; all forces acting on it balance each other out and there is no tendency for it to move or change. To illustrate, place a book on a table. The force of gravity is pulling the book down towards the floor. The table is pushing back. These two pressures offset each other and the book on the table is at rest, at equilibrium. In the context of a market, the forces that operate are the competitive forces of demand and supply, that is, the competitive actions taken by buyers to secure their demands and sellers to sell their supplies. Employing these ideas in relation to Figure 3-3, then, it can be seen that (x^0, p_x^0) is the only equilibrium point in that diagram. The demonstration requires 3 steps:

1. If the market price were below p_x^0 at p'_x in Figure 3.3, then the quantity demanded x'^d would be greater than the quantity supplied x'^s. Buyers would not be able to buy the quantity they want and would have to compete with other buyers for the scarce good. They compete by offering to pay a higher price. Since the market price is changing, this cannot be an equilibrium.

2. If the market price were above p_x^0, then the quantity supplied would be greater than the quantity demanded. Sellers would not be able to sell the quantity they want and would

have to compete with other sellers for the scarce buyers. The sellers compete by offering to sell at a lower price. Since the market price is changing, this also cannot be an equilibrium.

3. When the market price is at p_x^0, the quantity demanded equals the quantity supplied, buyers can buy the quantity they want, and sellers can sell the quantity they want. No one has to compete and offer a better price. Nothing is changing so the market is at rest, that is, in equilibrium.

Thus, the competitive forces of demand and supply, that is, buyers offering to pay a higher price when unable to buy the full amount they want and sellers offering to sell at a lower price when unable to sell all that they want, push the market towards and are inoperable at equilibrium.

Note, however, that conceptually it is possible to have equilibrium in a market where demand and supply curves do not intersect. In the right-hand diagram of Figure 3-4, equilibrium

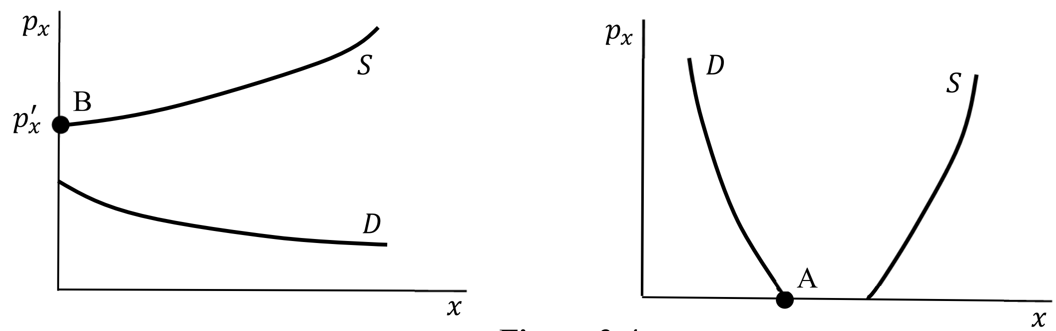


Figure 3-4

occurs at A where $p_x = 0$. As described above, at any positive price, competitive forces (in this case, suppliers lowering their price to attract scarce buyers) drive the market price down to zero. In the left-hand diagram of Figure 3-4, equilibrium occurs at B where $x = 0$. At any positive quantity, the price that buyers are willing to pay is less than that at which sellers are willing to sell. The former price is not high enough to warrant production and sale of that quantity. Sellers lower the price gap by offering to sell at a lower price and reducing the quantity they are willing to sell. In this way, market quantity is pushed down to zero. Only if zero prices and quantities are ruled out of consideration do “equilibrium” and “intersection point” correspond to the same thing.

The diagrams of Figure 3-4 can be used to describe markets that, because of the special circumstances relating to them, do not actually relate to purchases and sales in the real microeconomy. In the right-hand diagram, point A may be thought of as an equilibrium in the market for, say, the air human beings breathe. Air is so plentiful that no one has to buy or pay for it and no one could sell it. “Buyers” obtain all they want free of charge. Point B in the left-hand diagram can be viewed as an equilibrium in the market for electric cars before the time when those cars were actually produced and sold in a real market. Before the real market appeared, the per-unit cost of producing electric cars was so high that at each supply quantity the price at which they could be sold was greater than the price potential buyers were willing to pay. These kinds of markets are of no interest because there is nothing to explain about the production, buying, or selling associated with them. In the present volume, then, zero prices and quantities in markets will not be considered.

Even though it is abstract and cannot exist in reality, the demand-supply model is still employed to understand how goods are bought and sold in the real microeconomy. To be concrete, let x represent gasoline. For a particular week, say, add together all sales of all types of gasoline and the number of gallons sold, and then divide the latter into the former to obtain an average price.[1] Such data is usually interpreted to mean that, in the real microeconomy, a certain quantity of gasoline, call it x^0(the total number of gallons sold), was bought by buyers (and sold by sellers) at the (average) price p_x^0 during the week. The explanation of how this observation came about is obtained by building a demand-supply model. The construction proceeds by first locating the observed point (x^0, p_x^0) in the diagram of Figure 3-5. The next step is to postulate three assumptions:

1. There exists a demand curve D in the market and it passes through the observed point (x^0, p_x^0).

2. There exists a supply curve S in the market and it passes through the same observed point (x^0, p_x^0).

3. The end result in the market of the operation of the competitive forces of demand and supply is equality of the latter, that is, market equilibrium.

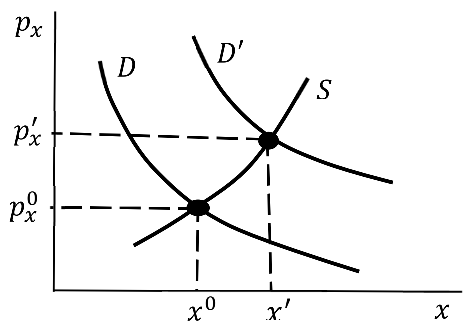


Figure 3-5

The model is now built. Based on its assumptions it can be said that (x^0, p_x^0) came about through or is explained by the interaction of demand and supply forces, that is, through the competition between buyers and sellers described above, that made it possible to think of (x^0, p_x^0) as an equilibrium. If, in a second week, a new point (x', p_x') were observed (Figure 3-5), and if two more assumptions were added:

4. There exists another demand curve D' in the market and it passes through (x', p_x'),

[1] Alternatively, calculate an average price by averaging, perhaps with weights, over all existing prices of all types of gasoline.

5. (x', p_x') lies on the same supply curve S as (x^0, p_x^0),

then (x', p_x') would also be explained as emerging from the competitive interaction of demand and supply. And the movement from (x^0, p_x^0) to (x', p_x') or the rise in price would be explained as resulting from an increase in demand.

It should be emphasized that the above assumptions, along with the idea of the market itself, are abstract mental constructs that do not exist in the reality under consideration. They are figments of the imagination. But as with the parts of the constructed model of the clock in Chapter 1, it does not matter that the assumed demand and supply curves of the model of Figure 3-5 cannot be shown to exist as an accurate representation of what is actually present in the real microeconomy. In spite of that, this demand-supply model still provides an explanation of how the price and quantity of gasoline sold in the real microeconomy came about during the weeks under consideration. And this explanation resonates in the public domain to the extent that it is embraced by most people as **the** explanation of how prices and quantities of goods sold are determined in reality. That is, when someone who is studying a market concludes that an increase in demand for gasoline is the reason for the observed increase in its price, that explanation is generally accepted by the public.

This is the way explanation works in economics. There is no laboratory in which the veracity of this demand-supply model can be tested.[2] Moreover, there is no guarantee that the model provides a correct explanation of what is actually happening. As pointed out in Chapter 1, explanations are not unique. Other models can provide alternative explanations. For example, in periods of relatively rapid price increases, rather than using the model of Figure 3-5 with, say, frequent rightward shifts of the demand curve (due to changing values of p_y and/or individual incomes) against a fixed supply curve, the model whose graph is pictured in Figure 3-6 and

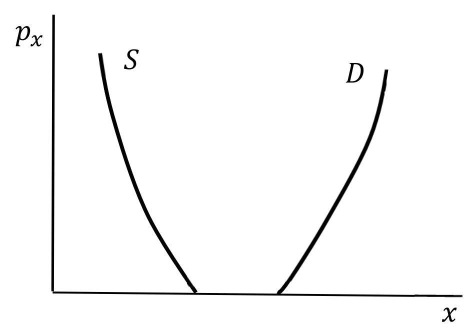


Figure 3-6

which violates the laws of supply and demand, may be thought to provide an alternative and perhaps better understanding of what is going on. In the model of Figure 3-6, buyers are never able to buy all they want to buy and continually offer to pay a higher price. The market price is therefore continually rising. This model, too, cannot be tested to determine its validity.

[2] Certain aspects of models of other economic phenomena can be tested. But, since the economist does not have the same kind of laboratory controls as, say, a physicist or chemist (recall Chapter 1), the results are not nearly as conclusive as in the physical sciences.

In the model of Figure 3-5, the process of equilibrating demand and supply to reach an equilibrium determines which buyers obtain the commodity and which sellers sell it. The buyers willing to pay the equilibrium price obtain the good; the sellers willing to sell at that price sell their product. That is, the market allocates what is bought and sold among buyers and sellers. Anything that interrupts the process of equilibrating demand and supply prevents the market from allocating in this way. Using this model as an explanation of the buying and selling of goods in the real microeconomy, it follows that consequences arise if a government were, for reasons of, say "fairness," to interfere in the in the buying and selling process. There are at least two possibilities:

1. A price ceiling like $\bar{p}_x$ in Figure 3-7 could be set to prevent the market price from rising to what could be the equilibrium price p_x^0. In that case, some kind of non-price rationing scheme will either arise in the purchase-sales process (e.g., first come, first served) or be imposed from outside (e.g., a government sponsored rationing scheme) to allocate the short supply of goods among the too-many demands.

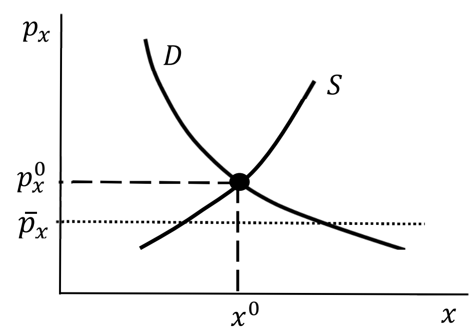


Figure 3-7

2. A price floor, say $\hat{p}_x$ in Figure 3-8, could be set to prevent the market price from falling to what could be the equilibrium price p_x^0. Here, too, some kind of non-price rationing scheme will either arise in the purchase-sales process or be imposed from outside to determine which units of the too-large supply will be bought by buyers.

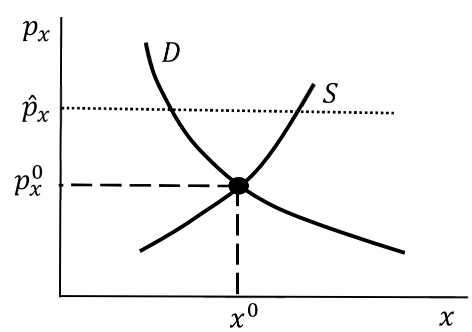


Figure 3-8

An example of the first is rent-control laws; an illustration of the second is minimum wage laws. This is not to pass judgment on whether interference in the buying-selling process is good or bad. The only point is that such interference has consequences that should be recognized.

Chapter 4
Elasticity of Demand

In the context of actual purchases and sales (illustrated in terms of gasoline in Chapter 3) it is often useful to know how one variable might respond to changes in another. For example, before raising the price of steak, a supermarket might want to know how the quantity of steak sales might be affected. The notion of elasticity contains that kind of information. This chapter analyzes the elasticity concept in the framework of the abstract demand-supply model developed earlier. The main focus here is on the elasticity of demand.

Consider the graph of the market demand function $x = D(p_x)$. There are two ways to draw that curve. On the one hand, the dependent variable x can be placed on the horizontal axis and the independent variable p_x on the vertical as described in Chapter 3. This is the usual approach in Economics. The curve drawn in that way is illustrated in Figure 4-1 and will be called the demand curve as it normally is. It. On the other hand, the demand function can be

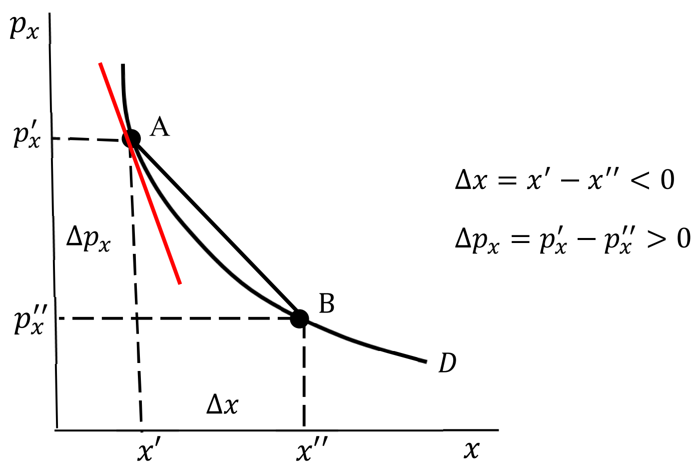


Figure 4-1

graphed in the mathematically standard way by placing the p_x on the horizontal axis and the x on the vertical. This graph will be called the mathematically standard demand curve. Note that the slopes of these two curves are related. In particular, the slope of the demand curve at A in Figure 4-1 (the slope of the red line) is approximated by the slope of the straight-line segment connecting A to B or $^{\Delta p_x}/_{\Delta x}$. But with the mathematically standard demand curve the slope of the straight line between the same two points would appear as $^{\Delta x}/_{\Delta p_x}$. It follows that the slope of the mathematically standard demand curve is the reciprocal of the slope of the demand curve. This applies in the limit as B approaches A. In subsequent discussion, the slope of the mathematically standard demand curve will occasionally be referred to as the reciprocal slope of the demand curve.

The price elasticity of demand is a measure of the responsiveness of quantity to changes in price along that curve – roughly expressed as

$$\frac{\%\text{ change in quantity}}{\%\text{ change in price}}. \tag{4.1}$$

The reciprocal slope of the demand curve ${\Delta x}/{\Delta p_x}$ is also a measure of the responsiveness of quantity changes to price changes. The reason why elasticity is used instead of reciprocal slope will be considered momentarily. There are two ways to make the notion of elasticity precise.

According to the formula of (4.1), the elasticity between points A and B on the demand curve in Figure 4-1, or the arc price elasticity of demand ε_{AB}, is given as the ratio of two percentages:

$$\varepsilon_{AB} = -\frac{\frac{\Delta x}{x}\cdot 100}{\frac{\Delta p_x}{p_x}\cdot 100} = -\frac{\Delta x}{\Delta p_x}\cdot\frac{p_x}{x}, \tag{4.2}$$

where ${\Delta x}/{\Delta p_x}$ is the slope of the straight-line segment connecting the same A to B in the mathematically standard context, that is, the reciprocal of the slope of the straight line between A and B in Figure 4-1. This may be regarded as an approximation to the point price elasticity defined below.

Note that:

1. Since observed demand curves generally slope downward (recall Chapter 3), the minus sign is typically added to make ε_{AB} positive.

2. There is ambiguity in the choice of x and p_x in equation (4.2). It will be convenient here to use $x = x''$ and $p_x = p_x'$ in calculating ε_{AB} so that (4.2) becomes

$$\varepsilon_{AB} = -\frac{\Delta x}{\Delta p_x}\cdot\frac{p_x'}{x''}. \tag{4.3}$$

The point price elasticity of demand, ε_A, at A in Figure 4-1 is obtained by letting Δp_x approach 0 or p_x'' approach p_x' in equation (4.2) so that

$$\varepsilon_A = \lim_{p_x''\to p_x'} -\frac{\Delta x}{\Delta p_x}\cdot\frac{p_x'}{x''} = -D'^{(p_x')}\cdot\frac{p_x'}{x'}, \tag{4.4}$$

where $D'(p_x')$, the first-order derivative at p_x' of $D(p_x)$ with respect to p_x , represents the reciprocal slope of the demand cure at A, or the slope of the red line drawn in Figure 4-1. Unlike arc elasticity, there is no ambiguity in the choice of x and p_x. – the values are those at A in

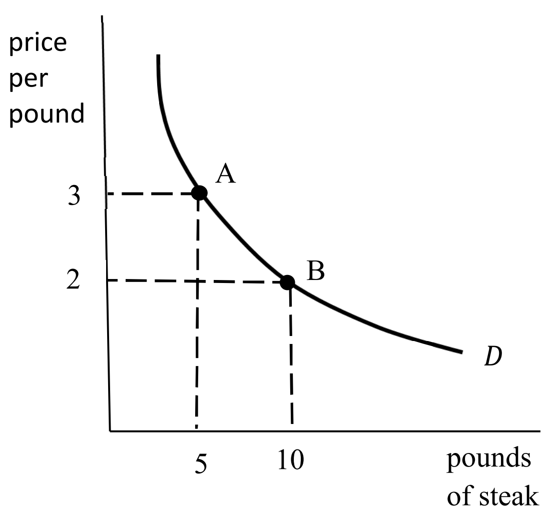
Figure 4-1 Thus, for both point and arc concepts, elasticity is the reciprocal slope, that is the slope of the mathematically standard demand curve adjusted by price over quantity.

Note by way of illustration that the demand curve obtained from the linear demand function $x = \alpha p_x + \beta$, where $\alpha < 0$ and $\beta > 0$ are fixed parameters, is a straight line with constant (negative) reciprocal slope α. But its elasticity varies with movement along the curve. In the point form of (4.4):

$$\varepsilon_A = -\alpha \frac{p_x}{x}.$$

Why use elasticity instead of reciprocal slope as a measure of the responsiveness of quantity changes to price changes? Because slope depends on the units of measurement employed – in the case of the supermarket-steak example referred to above, the change in quantity per unit change in price is different for quantity measured as, say pounds, instead of as ounces. And elasticity, because it is a ratio of percentages, is independent of such units. In either case, the underlying purchasing behavior is identical; it is just being expressed in different ways. Regardless, focusing on elasticity provides a cleaner understanding of that underlying behavior that is unencumbered by the specification of units of measurement.

To illustrate, suppose first that the demand curve for steak in units of pounds appears as drawn in Figure 4-2. Then let steak is measured in ounces, replacing Figure 4-2 by Figure 4-3.



$$\Delta x = 5 - 10 = -5$$

$$\Delta p_x = 3 - 2 = 1$$

Reciprocal slope:

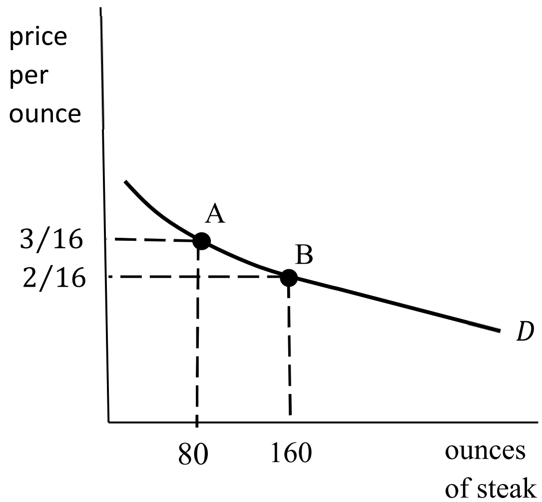
$$\frac{\Delta x}{\Delta p_x} = \frac{-5}{1} = -5$$

Elasticity:

$$\varepsilon_{AB} = -\frac{-5}{1} \cdot \frac{3}{10} = \frac{3}{2}$$

Figure 4-2

The calculation of the reciprocal slopes and arc elasticities between A and B using the equation (4.3) designations for x and p_x appears in the box to the right of each diagram. In Figure 4-2, $p_x' = 3$ and $x'' = 10$; in Figure 4-3, $p_x' = 3/16$ and x'' =160. These demand curves reflect the same buying behavior. So to avoid possible confusion and enhance clarity, the measure of the



$$\Delta x = 80 - 160 = -80$$

$$\Delta p_x = \frac{3}{16} - \frac{2}{16} = \frac{1}{16}$$

Reciprocal slope:

$$\frac{\Delta x}{\Delta p_x} = \frac{-80}{{}^{1}\!/_{16}} = -1280$$

Elasticity:

$$\varepsilon_{AB} = -\frac{-80}{{}^{1}\!/_{16}} \cdot \frac{{}^{3}\!/_{16}}{160} = \frac{1}{2} \cdot \frac{3}{1} = \frac{3}{2}$$

Figure 4-3

response of quantity changes to price changes ought to be the same. In this case, the change in units from pounds to ounces modifies the slope ${}^{\Delta p_x}\!/_{\Delta x}$ of the line segment connecting A to B (not shown in the diagrams) from $-{}^{1}\!/_{5}$ to $-{}^{1}\!/_{1280}$ making the line (and the demand curve) flatter and modifying its reciprocal slope as indicated without changing its elasticity. The latter, then, is clear and informative as it stands and does not have to be qualified by specifying units of measurement. In that sense, it is a better measure of the response of quantity changes to price changes.

Along a demand curve there are two important relationships involving elasticity: The first is that between elasticity and total revenue. Total revenue is defined as

$$TR(x) = xp_x, \tag{4.5}$$

where the point (x, p_x) lies on the demand curve. To understand this relationship in terms of the arc formulation of arc elasticity,[1] use the definition in equation (4.3) (with price and quantity values as pictured in Figure 4-1) to obtain:

$$\varepsilon_{AB} = -\frac{x' - x''}{p'_x - p''_x} \cdot \frac{p'_x}{x''} = \frac{p'_x x'' - p'_x x'}{p'_x x'' - p''_x x''}.$$

Using equation (4.5), this may be rewritten as

[1] That this relationship also holds with respect to point elasticity is not considered here. See Footnote 2 below.

$$\varepsilon_{AB} = \frac{p_x' x'' - TR(x')}{p_x' x'' - TR(x'')}\,, \tag{4.6}$$

and employing (4.6),

$$\begin{array}{lll} \varepsilon_{AB} = 1 & \text{if and only if} & TR(x') = TR(x''). \\ \varepsilon_{AB} < 1 & \text{if and only if} & TR(x') > TR(x''). \\ \varepsilon_{AB} > 1 & \text{if and only if} & TR(x') < TR(x''). \end{array}$$

It follows that:

When $\varepsilon_{AB} = 1$, total revenue is constant when moving from A to B.

When $\varepsilon_{AB} < 1$, price and total revenue move in the same direction along the demand curve as x changes its value from A to B (in Figure 4-1, price falls and total revenue also falls).

When $\varepsilon_{AB} > 1$, price and total revenue move in opposite directions along the demand curve as x changes its value from A to B (in Figure 4-1, price falls and total revenue rises).

In general, these rules apply to all pairs of points on, and moving in either direction along the demand curve. Illustrations are provided in terms of the linear demand curve in Figure 4-5.
.

The second relationship is between elasticity and marginal revenue and is also illustrated in Figure 4-5. Marginal revenue, the derivative of total revenue at x' is defined (in approximate form) as

$$MR(x') = \frac{\Delta TR}{\Delta x} = \frac{TR(x' + \Delta x) - TR(x')}{\Delta x}, \tag{4.7}$$

where $\Delta x = x' - x''$. Its relationship to elasticity at (x', p_x') is given by the formula

$$MR(x') = p_x' \left(1 - \frac{1}{\varepsilon_{AB}}\right),$$

which is derived (also in approximate form) from the definitions of marginal revenue and arc elasticity in Supplemental Note A. [2]

From these relationships and the definition of elasticity, five categories of elasticity can be distinguished. They and their properties are listed in Table 4-1. The table asserts, for example, that where the demand curve is elastic, the arc price elasticity is everywhere finite and greater

---

[2] Although not pursued here, this expression can also be derived in continuous or limiting form, in which case ε_{AB} is replaced by ε_A. In that form it can be used to derive the above relationship between point elasticity and total revenue

than one, the marginal revenue is everywhere positive, and in moving from point to point along the curve, price and total revenue modify in opposite directions. In the perfectly elastic case where $\Delta p_x = 0$, equations (4.3) and (4.4) imply that $\varepsilon_{AB} = \infty$ and $\varepsilon_A = \infty$, respectively. Although the first two columns of the table explicitly refer to the approximate arc concept of elasticity, the entire table can be understood in terms of either arc or point concept.

Table 4-1

Category	ε_{AB}	**Slope of the Demand Curve**	**Marginal Revenue**	**Along the Demand Curve**
perfectly inelastic ($\Delta x = 0$.)	0	∞	---	p_x and $TR(x)$ move in the same direction
inelastic	$0 < \varepsilon_{AB} < 1$	negative	negative	p_x and $TR(x)$ move in the same direction
unitarily elastic	1	negative	0	$TR(x)$ constant
elastic	$1 < \varepsilon_{AB} < \infty$	negative	positive	p_x and $TR(x)$ move in oppo-site directions
perfectly elastic ($\Delta p_x = 0$.)	∞	0	p_x	x and $TR(x)$ move in the same direction

To say that the demand curve is, say elastic, means either that ε_{AB} or ε_A are greater than one everywhere along the curve. Context determines which meaning applies. According to the expression for the slope of the demand curve in Figure 4-1 (in approximate form, $\Delta p_x / \Delta x$), that curve in the perfectly inelastic case is vertical since $\Delta x = 0$, while that in the perfectly elastic case is horizontal with $\Delta p_x = 0$. The former is depicted by the left-hand diagram in Figure 4-4, the latter by the right-hand diagram.

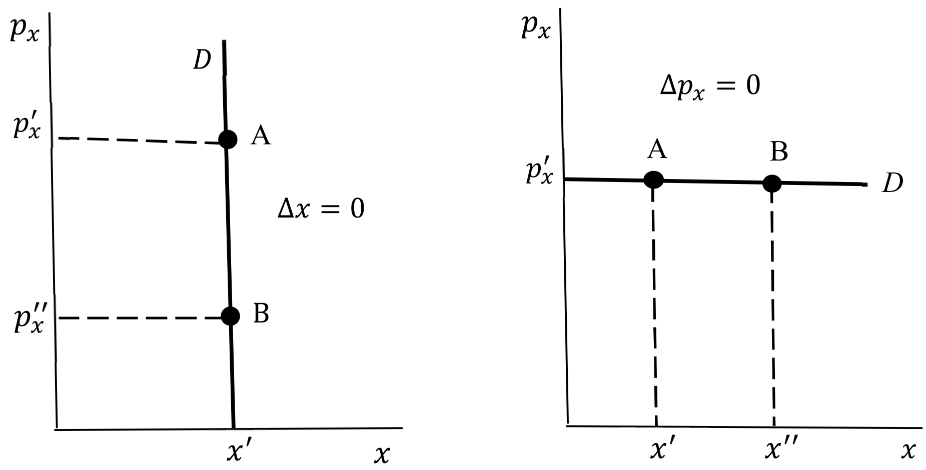


Figure 4-4

The linear demand curve, the geometric properties of its associated total and marginal revenue curves, and the corresponding elasticity ranges exemplify part of Table 4-1 and are illustrated in Figure 4-5. The subscripts on the elasticity symbol ε are dropped for convenience.

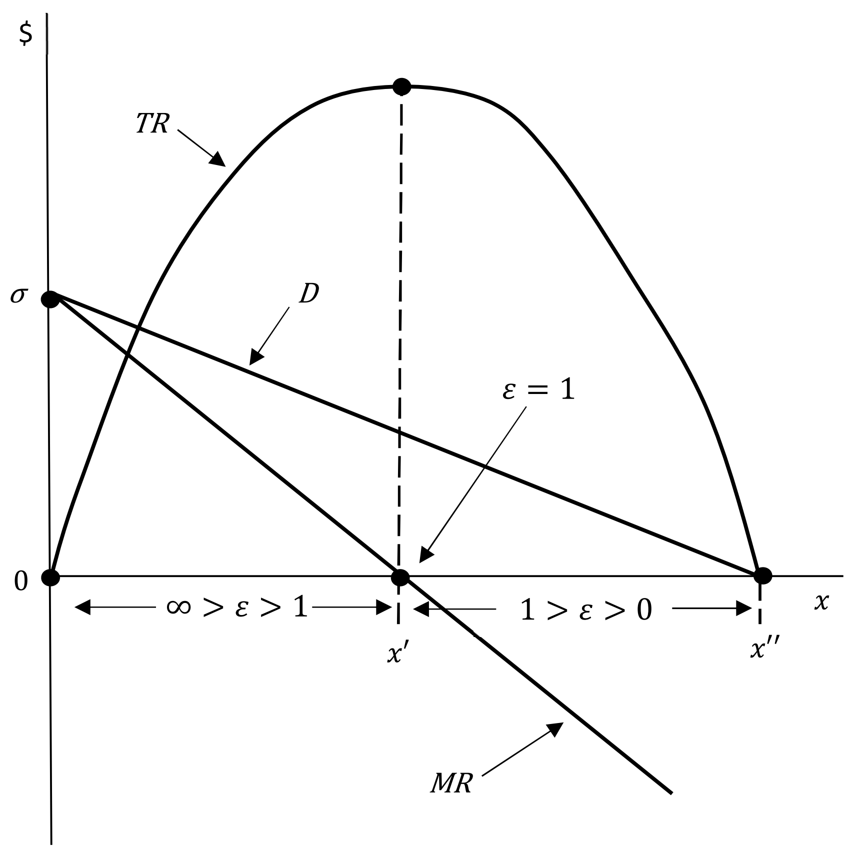


Figure 4-5

With respect to Figure 4-5:

1. Given the demand curve D, the marginal revenue curve *MR* is determined by drawing a straight line from *D*'s vertical intercept at σ through the mid-point x' between the origin and *D*'s horizontal intercept x'' on the x-axis This geometric construction is established in Supplemental Note B.

2. Total revenue along the demand curve is xp_x. Hence the total revenue curve associated with the demand curve, labeled *TR*, must start at the origin since that value of total revenue corresponds to $x = 0$ and $p_x = \sigma$ on the demand curve. Where the demand curve meets the x-axis at x'', that is, where $x = x''$ and $p_x = 0$, total revenue xp_x is also zero. In between, since the value of the marginal revenue at each x is the slope of the total revenue curve at that x, where marginal revenue is positive ($0 < x < x'$), *TR* slopes upward; where marginal revenue is negative ($x' < x < x''$), *TR* slopes downward; and where marginal revenue is zero ($x = x'$), *TR* has a maximum.

3. From the equation $MR = p_x\left(1 - {}^{1}/_{\varepsilon}\right)$, the demand curve is elastic ($\infty > \varepsilon > 1$) where marginal revenue is positive (on the interval $0 < x < x'$); it is inelastic ($1 > \varepsilon > 0$) where marginal revenue is negative (on the interval $x' < x < x''$); and it is unitarily elastic ($\varepsilon = 1$) where marginal revenue is zero (at $x = x'$).

4. Moving along the demand curve in the elastic range ($0 < x < x'$), say by increasing x, price and total revenue move in opposite directions (price falls and total revenue rises). Moving along the demand curve in the inelastic range ($x' < x < x''$), also by increasing x, price and total revenue move in the same direction (price falls and total revenue falls).

There are other notions of elasticity related to the demand function. First, letting p_y, which is held fixed behind the demand curve, now vary with m and p_x held fixed, the arc cross-price elasticity of demand is characterized by

$$\frac{\frac{\Delta x}{x} \cdot 100}{\frac{\Delta p_y}{p_y} \cdot 100}.$$

And second, with m now varying and p_x and p_y held fixed, the arc income elasticity of demand is defined as

$$\frac{\frac{\Delta x}{x} \cdot 100}{\frac{\Delta m}{m} \cdot 100}.$$

These expressions can be rewritten as in equation (4.3) for the arc concept and taken to the limit as in (4.4) for the point concept. The minus sign is left out since the elasticities could normally be either positive or negative.

The arc and point price elasticities of supply are defined similarly to the arc and point price elasticities of demand except that, since supply curves tend to slope upward, no minus sign is included.

Chapter 5
Preferences, Utility, and Indifference Curves

Attention now turns to an explanation of the buying behavior of a single consumer in the abstract world of the Walrasian model set out in Chapter 1. In the two consumer-good context of that world, the consumer's buying behavior is completely described by his/her demand functions for good x and good y. Those functions, recall, constitute the consumer's contribution to the market demand functions of Chapter 3. Using slightly different notation from that of the latter chapter, the consumer's demand functions are denoted here by, respectively, h^x and h^y, where

$$x = h^x(p_x, p_y, m),$$
$$y = h^y(p_x, p_y, m),$$

the superscripts on h now identify commodities, and the earlier subscripts and superscripts representing individuals have been dropped.[1] These functions indicate for each set of values, one for each of p_x, p_y, and m, the quantities of goods x and y the consumer buys or demands. Even though abstract, h^x and h^y are observable in the following sense: Think of the consumer in a supermarket buying goods. For this purchase, p_x, p_y, and m are known since the former are posted for each commodity and the latter has been earned in the factor markets, for example the labor market (observed as hours worked time the wage). It is only necessary to watch what the consumer buys to have an observation of one "point" on his/her demand function. The explanation of these functions will be based on a model employing the assumptions that

1. the individual's tastes or preferences for quantities of goods are fixed and given, and
2. the demand decisions made by the individual are founded on maximization.

The present chapter will discuss the nature of preferences. Chapter 6 will take up first the environment in which the consumer makes buying decisions and then the making of those decisions via maximization.

The consumer is assumed to have preferences among baskets of commodities. These preferences exist only in his/her mind and, unlike buying behavior, but like the inner workings of the clock example of Chapter 1, are not observable. When there are only the two goods x and y, baskets of commodities bought or demanded by the consumer are denoted by (x, y), where x varies over quantities of good x and y quantities of good y. The collection of all possible baskets available,

$$\{(x, y): x \geq 0, y \geq 0\},$$

is called the commodity space.

---

[1] Identifying x as an individual's demand for good x is the third meaning attached to this symbol. In Chapter 3 it was employed to represent both market demand and market supply. And starting in Chapter 9 it will represent the firm's output produced and supplied. Context determines its meaning in any particular usage.

Let (x', y') and (x'', y'') be baskets in the commodity space. Suppose (the assumption is called completeness) that for all such pairs of baskets, the consumer either prefers one to the other or is indifferent between them. Thus it may be that

(x', y') is preferred to (x'', y'') or
(x', y') is indifferent to (x'', y'').

These preferences and indifferences are taken to be transitive: If a basket, call it A, is preferred to another basket B, and basket B is preferred to a third basket C, then basket A is preferred to basket C. And similarly for indifference.

In addition, preferences and indifferences are assumed to be representable by numbers. These numbers, which can be positive, negative, or zero, are called utility numbers and the full representation of preferences and indifferences in terms of them is called the utility function. The utility function is written as

$$\mu = u(x, y),$$

where u assigns a utility number μ to each basket (x, y) in the commodity space. The utility function is related to preferences as follows:

$u(x', y') > u(x'', y'')$ if and only if (x', y') is preferred to (x'', y''),

and

$u(x', y') = u(x'', y'')$ if and only if (x', y') is indifferent to (x'', y'').

Thus, if one basket is preferred to another, the preferred basket has a higher utility value; and if two baskets are indifferent, they have the same utility value.

The only significance of utility numbers μ is that they reflect the ordering of baskets by preference (and indifference) and nothing more. Beyond this the numbers have no meaning. To illustrate, since the utility functions

$$u(x, y) = xy \quad \text{and} \quad u(x, y) = 2xy$$

order baskets (x, y) in the same way, they represent the same preference ordering. For example, suppose basket A consists of (2,3) and basket B contains (4,2). Then with $u(x, y) = xy$, the utility of A is $u(2,3) = 2 \cdot 3 = 6$ and that of basket B is $u(4,2) = 4 \cdot 2 = 8$. Since $u(2,3) < u(4,2)$, B is preferred to A. The same conclusion obtains using $u(x, y) = 2xy$ since, in that case $u(2,3) = 12$ and $u(4,2) = 16$. Utility functions are called ordinal because they only provide information about the ordering of baskets in the commodity space by preference.

Let $(x^0, y^0) > 0$ be a basket in the commodity space with $\mu^0 = \mathrm{u}(x^0, y^0)$. The indifference curve through (x^0, y^0) is

$$\{(x,y) : (x,y) \text{ is indifferent to } (x^0,y^0) \text{ and } (x,y) > 0\},$$

or equivalently,

$$\{(x,y) : u(x,y) = \mathrm{u}(x^0,y^0) = \mu^0 \text{ and } (x,y) > 0\}.$$

An illustration appears in Figure 5-1. (The geometric derivation of an indifference curve

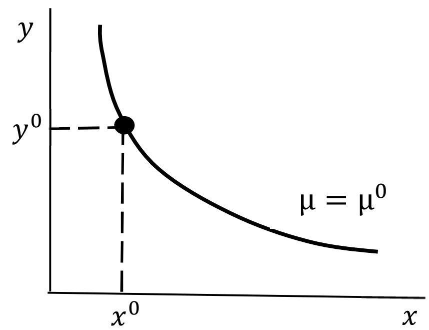


Figure 5-1

from the graph of a utility function may be thought of in terms of Figure 2-4 of Chapter 2. In that case, f repressents the utility function u, and g the indifference function W introduced below.) Each point or basket along the indifference curve in Figure 5-1 has the same utility value μ^0. With μ fixed at μ^0, the relation between x and y as pictured in Figure 5-1 may be thought of as the graph of a function of a single variable $y = W(x)$. For example, suppose $u(x,y) = xy$ and $(x^0,y^0) = (2,6)$. Then $\mu^0 = x^0y^0 = 12$. and the indifference curve through $(2,6)$ contains all baskets such as (6,2, (3,4), and (4,3) whose quantity-roadcut is 12. It is described by the equation $12 = \mathrm{xy}$, or solving for y as a function of x, $y = {}^{12}/_{x}$. More generally, in terms of this utility function, $\mu^0 = \mathrm{xy}$ and

$$y = W(x) = \frac{\mu^0}{x}.$$

There is one indifference curve through each basket $(x,y) > 0$ in the commodity space and the collection of all indifference curves is called the indifference map.

Just as there is a relationship between the utility function and indifference curves, so is there a relationship between the marginals derived from each. The marginal utility with respect to x at $(x^0,y^0) > 0$ in finite incremental or approximate form is

$$MU_x(x^0,y^0) = \frac{\mathrm{u}(x^0 + \Delta x, y^0) - \mathrm{u}(x^0,y^0)}{\Delta x}, \tag{5.1}$$

where Δx can be positive or negative (and in numerical tables of textbooks is often taken to be 1). That with respect to y at $(x^0,y^0) > 0$ is

$$MU_y(x^0, y^0) = \frac{\mathrm{u}(x^0, y^0 + \Delta y) - \mathrm{u}(x^0, y^0)}{\Delta y}.$$

In derivative form (that is, taking the limit as $\Delta x \to 0$),

$$MU_x(x^0, y^0) = \lim_{\Delta x \to 0} \frac{\mathrm{u}(x^0 + \Delta x, y^0) - \mathrm{u}(x^0, y^0)}{\Delta x},$$

and similarly for $MU_y(x^0, y^0)$.

With $\mu^0 = \mathrm{u}(x^0, y^0)$ and derived indifference function $y = W(x)$, where the indifference curve (that is, the graph of W) through (x^0, y^0) appears as pictured in Figure 5-1, the derivative of W or slope of the curve at x^0 in approximate form is

$$\frac{\Delta y}{\Delta x} = \frac{\mathrm{W}(x^0 + \Delta x) - \mathrm{W}(x^0)}{\Delta x},$$

and in limiting derivative form it is

$$W'(x^0) = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{\mathrm{W}(x^0 + \Delta x) - \mathrm{W}(x^0)}{\Delta x}.$$

The relationship between marginal utility values and slopes of the indifference curve can be understood in terms of Figure 5-2:

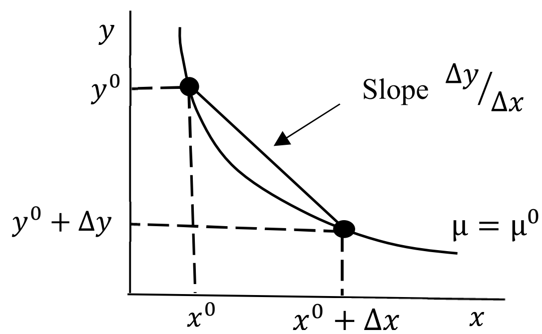


Figure 5-2

Start at (x^0, y^0) on the indifference curve and change basket (x^0, y^0) to $(x^0 + \Delta x, y^0 + \Delta y)$, where $\Delta x > 0$ and $\Delta y < 0$. Then, according to one of the rules of calculus, the change in utility can be approximated by

$$\Delta\mu = \Delta x MU_x(x^0, y^0) + \Delta y MU_y(x^0, y^0). \tag{5.2}$$

But since the two points are on the same indifference curve, $\Delta\mu = 0$. Substituting 0 for $\Delta\mu$ in equation (5.2) and rewriting gives

$$\frac{\Delta y}{\Delta x} = -\frac{MU_x(x^0, y^0)}{MU_y(x^0, y^0)}, \tag{5.3}$$

or, in the limit as $\Delta x \to 0$,

$$W'(x^0) = -\frac{MU_x(x^0, y^0)}{MU_y(x^0, y^0)}. \tag{5.4}$$

Thus the slope of the indifference curve at any basket (in approximate form as in equation (5.2) or in limiting form as in (5.3)) is the negative of the ratio of marginal utilities at that basket. The negative of that slope at (x^0, y^0), or the ratio of marginal utilities without the minus sign, is called the marginal rate of substitution and symbolically written as $MRS(x^0, y^0)$. The marginal rate of substitution is understood as the rate at which the consumer can substitute units of good y for units of good x and remain on the indifference curve.

It is now appropriate to add several properties of the utility function to the list of assumptions made earlier. This is done to ensure that the utility function has characteristics that will permit its maximization. But before doing so, it is necessary to define one more concept.

A basket in the commodity space, call it B, is larger than another basket A, if B contains more of at least one good and no less of the other. For example, all baskets labeled B located in the upper right-hand quadrant above A in Figure 5-3 are larger than A.

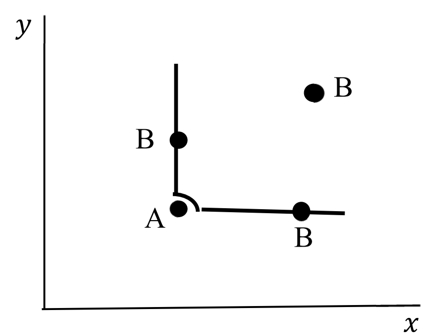


Figure 5-3

The assumptions made thus far in constructing the model to explain consumer buying behavior are now summarized below along with a new assumption that sets out the properties of the utility function alluded to above:

1. The consumer has preferences and indifferences among baskets of commodities such that, for any two baskets either one is preferred to the other or the two are indifferent. (Completeness.)

2. Preferences and indifferences are transitive. (If basket *A* is preferred to basket *B*, and basket *B* is preferred to basket *C*, then basket *A* is preferred to basket *C*. Similarly for indifferences.)

3. Preferences and indifferences are represented by a utility function in the sense that (a) if one basket is preferred to another, then the preferred basket has a higher utility value than the other, and (b) if two baskets are indifferent, then they have the same utility value.

4. The utility function has the following four properties:
   4a. It is continuous and all marginal utilities can be calculated.
   4b. A larger basket of commodities is always preferred to, and therefore has a higher utility value than a smaller one.
   4c. Indifference curves are strictly convex.
   4d. Indifference curves do not touch the co-ordinate axes of the commodity space.

One more assumption will be added in the next chapter. The full set of assumptions appears in Supplemental Note C.

It remains to consider several implications and features of assumptions 4b and 4c. First, in reference to the indifference curves and points A, B, and C in Figure 5-4, since basket A is larger than basket B, assumption 4b implies that the utility of A is greater than that of B. Since B

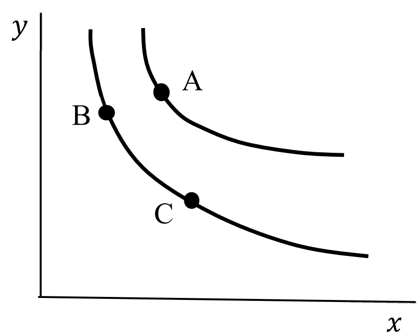


Figure 5-4

and C lie on the same indifference curve, the utility of the two baskets is the same. Therefore, the utility of A is larger than that of C. This is true for all points on the indifference curve containing point A. In other words, indifference curves farther out from the origin are always associated with higher utility.

Second, the marginal utility at (x^0, y^0) with respect to, say x, is, repeating equation (5.1) above,

$$MU_x(x^0, y^0) = \frac{\mathrm{u}(x^0 + \Delta x, y^0) - \mathrm{u}(x^0, y^0)}{\Delta x}. \tag{5.5}$$

On the one hand, if $\Delta x > 0$, then basket $(x^0 + \Delta x, y^0)$ is larger than basket (x^0, y^0) as pictured in Figure 5-5. By assumption 4b, $\mathrm{u}(x^0 + \Delta x, y^0) > \mathrm{u}(x^0, y^0)$. So, using equation (5.5), $MU_x(x^0, y^0)$ is the ratio of two positive numbers and is therefore positive. On the other hand, if

$\Delta x < 0$, then basket $(x^0 + \Delta x, y^0)$ is smaller than basket (x^0, y^0). By assumption 4b, $u(x^0 + \Delta x, y^0) < u(x^0, y^0)$. So, again from (5.5), $MU_x(x^0, y^0)$ is now the ratio of two negative numbers and again is positive. A similar argument applies to $MU_y(x^0, y^0)$. At least in approximate form, then, assumption 4b also guarantees that all marginal utilities are positive for all $(x, y) > 0$ in the commodity space. Although it does not follow by taking the limit of these positive numbers as $\Delta x \to 0$ that $MU_y(x^0, y^0)$ will always be positive,[2] it will be convenient to assume that that conclusion holds for the limiting derivative forms of marginal utilities as well.

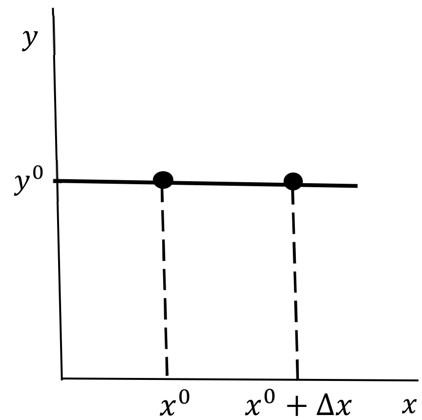


Figure 5-5

Third, since all marginal utilities are positive from the previous implication and since the slope of the indifference curve at any basket is the negative of the ratio of marginal utilities at that basket (equations (5.3) or (5.4)), indifference curves must slope downward everywhere.

Fourth, as suggested in Chapter 2, assumption 4c (strict convexity) will turn out to be a part of the non-derivative second-order maximization condition ensuring that, at baskets where the first-order condition is satisfied, utility will be maximized and not minimized or associated with an inflection point.

Fifth, assumption 4c together with the downward slope of indifference curves means that along any such curve the marginal rate of substitution ($-\Delta y/\Delta x$ or $-W'(x)$) diminishes as x rises or increases as x falls. This implies, referring to Figure 5-6 for example, the less the

---

[2] It is possible for the limit of a sequence of positive numbers to be zero. This is the case, for example, for the limit of the sequence of positive fractions $^1/_2, ^1/_3, ^1/_4, ^1/_5, \ldots$.

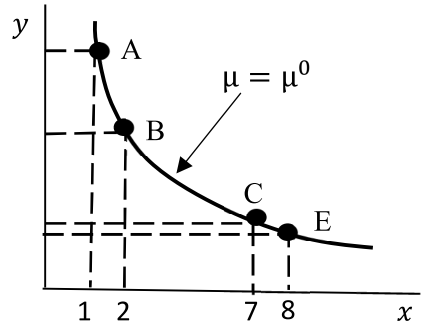


Figure 5-6

individual has of good x when taking one unit away from him/her, that is in moving from B to A rather than from E to C, the more he/she has to be compensated with good y in order to remain on the same indifference curve.

Chapter 6
Budget Constraints, Utility Maximization, and Demand Functions

This chapter continues the construction of a model to explain the consumer demand functions $x = h^x(p_x, p_y, m)$ and $y = h^y(p_x, p_y, m)$ begun in the previous chapter. The determination by the economy's markets of one set of values for the prices and income[1] in those functional equations defines the environment in which the consumer makes buying (and selling) decisions. Once p_x, p_y and m are specified, the baskets of goods in the commodity space that the consumer is able to buy are limited to

$$\{(x, y): xp_x + yp_y \leq m\}.$$

This set is called the consumer's budget set and its outer boundary, where equality prevails and the consumer spends all of his/her income, is given by the equation

$$xp_x + yp_y = m, \tag{6.1}$$

or

$$y = -\frac{p_x}{p_y}x + \frac{m}{p_y}.$$

This outer boundary, that is, equation (6.1), is referred to as the budget line or budget constraint. The graph of the budget constraint is a straight line with slope $-{}^{p_x}/_{p_y}$ and x- and y- intercepts given by $x = {}^{m}/_{p_x}$ and $y = {}^{m}/_{p_y}$ respectively as shown in Figure 6-1.

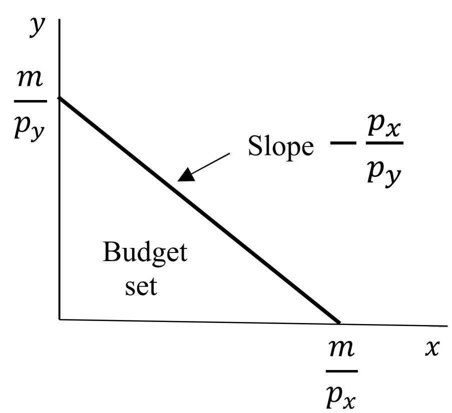


Figure 6-1

Now add one last assumption to those made with respect to the consumer's buying behavior in Chapter 5 (the full list of assumptions appears in supplemental Note C):

[1] As noted in the previous chapter, the consumer's income is determined in the economy's factor markets.

The consumer purchases or demands that basket from his/her budget set that is preferred over all other baskets in the budget set and therefore provides the most utility. In other words, the consumer demands that basket that maximizes his/her utility subject to the budget constraint. [2]

This assumption is sometimes referred to as the "postulate of rationality." Geometrically it implies that the consumer selects that basket (x^0, y^0) from the budget set for prices and income (p_x, p_y, m) that lies on the indifference curve that is as far out as possible from the origin. That basket will maximize utility because, as described in Chapter 5, baskets on indifference curves that are a greater distance from the origin have higher utility values. Observe that the basket (x^0, y^0) that maximizes utility subject to the budget constraint appears at a tangency between the budget line and an indifference curve as pictured in Figure 6-2. That basket is unique because the strictness of the assumption of strict convexity ensures that there can be only one point of tangency between an indifference curve and a budget line.

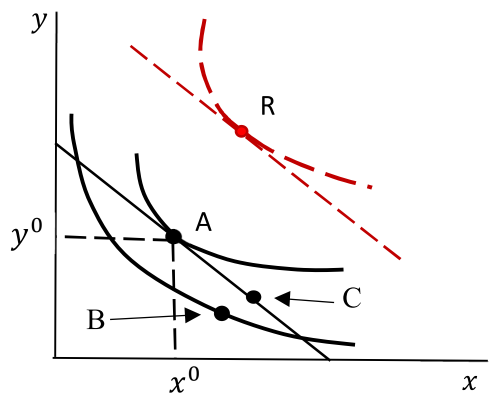


Figure 6-2

Note that by assumption 4b from Chapter 5 (a larger basket of commodities is always preferred to, and therefore has a higher utility value than a smaller one), the utility-maximizing basket must lie on the budget line. If it were in the interior of the budget set, say at B in Figure 6-2, then there would always be a larger basket C on the budget line. Since C is larger than B, assumption 4b implies that C would have a higher utility value and B could not be maximal. Also, by assumption 4d the utility-maximizing basket cannot lie on the x- or y-axes since indifference curves are not permitted to go there.

Pictured geometrically in a three-dimensional graph with the utility function and budget line appearing as shown in Figure 6-3, utility maximization subject to the budget constraint occurs for that utility function and that budget line at basket A. The utility value at the maximizing basket A is μ^0. Of course, this maximization occurs above the budget line and all baskets different from A on the budget line have lower utility than μ^0. To emphasize this, the

[2] The maximization is stated as subject to the budget constraint rather than the budget set because, as will be seen momentarily, under the assumptions that have imposed on the utility function, the utility-maximizing basket will always lie on the budget line.

graph of the utility function only over the budget line and the utility-maximizing basket are redrawn in Figure 6-4 with the coordinates of A identified as (x^0, y^0) and the indifference curve tangent to the budget line at A added. Note that $\mu^0 = u(x^0, y^0)$. In Figure 6-4, all lines and

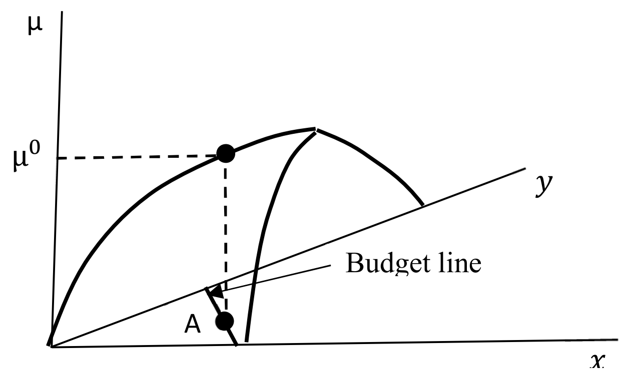


Figure 6-3

curves in the x-y plane are in red; above that plane they are in black. Apart from the axes, the solid black curve in Figure 6-4 is the graph of the utility function of Figure 6-3 over the budget line.

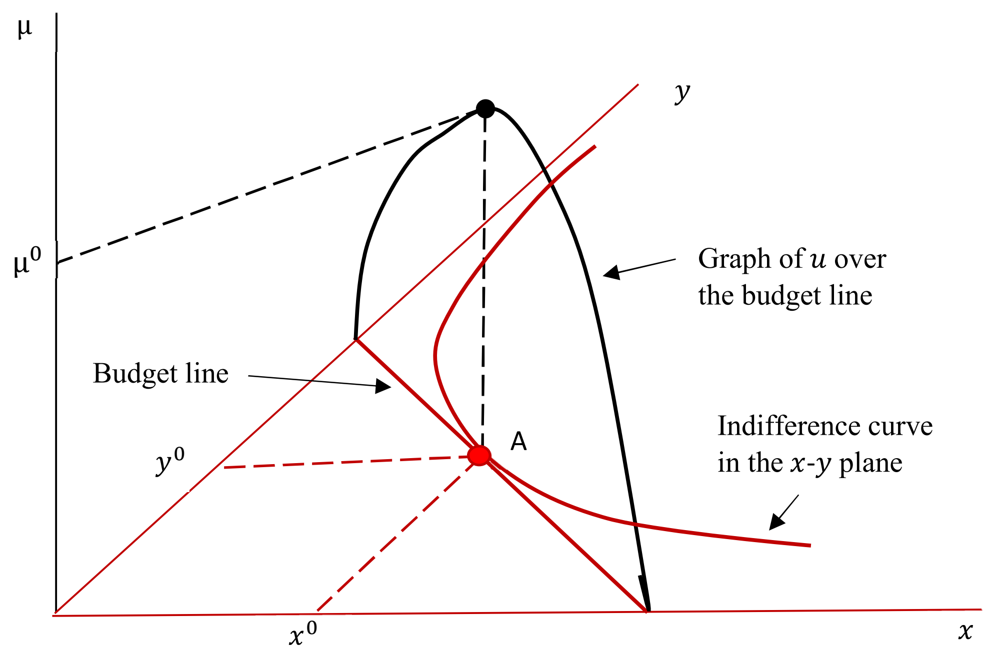


Figure 6-4

A mathematical expression partially describing the basket (x^0, y^0) *at* the tangency in Figures 6-2 and 6-4 is based on the idea that at a tangency between two curves, the slopes of the two curves have to be equal. From the above geometry, the slope of the indifference curve at the utility-maximizing basket (x^0, y^0), namely $-{MU_x(x^0, y^0)}/{MU_y(x^0, y^0)}$ (from equations (5-3) or (5.4) of Chapter 5), is the same as that of the budget line $-{}^{p_x}\!/_{p_y}$ (the coefficient of x in equation (6.1) above). Equating the two,

$$-\frac{p_x}{p_y} = -\frac{MU_x(x^0, y^0)}{MU_y(x^0, y^0)},$$

or

$$\frac{p_x}{p_y} = \frac{MU_x(x^0, y^0)}{MU_y(x^0, y^0)} = MRS_{x,y}(x^0, y^0), \tag{6.2}$$

where, recalling from Chapter 5, the marginal rate of substitution between x and y at the basket (x^0, y^0), or $MRS_{x,y}(x^0, y^0)$, is the ratio of the two marginal utilities. Thus, at (x^0, y^0), the rate at which the consumer substitutes y for x at market prices along the budget line equals rate at which the consumer substitutes y for x along the indifference curve based on his/her preferences.

It should be pointed out that equation (6.2) by itself is not enough to fully describe the tangency at A in Figures 6-2 and 6-3. For example, since the red (dashed) and black straight lines in Figure 6-2 are parallel and therefore have the same slope $-{}^{p_x}\!/_{p_y}$ (of course, only the latter is the specified budget constraint), and since the red (dashed) and black are indifference curves therefore have the same slope $-{}^{MU_x(x,y)}\!/_{MU_y(x,y)}$ at A and R, baskets A and R both satisfy equation (6.2). Because the budget constraint is specified by the prices and income determined in markets and the indifference curve is determined by the maximization process, to fully characterize the basket (x^0, y^0) that maximizes utility subject to the budget constraint for (p_x, p_y, m), it is necessary to combine equation (6.2) with the equation of the budget constraint evaluated at (x^0, y^0), namely $x^0 p_x + y^0 p_y = m$.

Alternatively, rewriting the left-hand equality in equation (6.2), the utility-maximizing basket can also be (partly) described by the equation

$$\frac{MU_x(x^0, y^0)}{p_x} = \frac{MU_y(x^0, y^0)}{p_y}. \tag{6.3}$$

This last equality says that, at the utility-maximizing basket, the marginal utilities per dollar spent on each good must be the same. If that were not the case, say if

$$\frac{MU_x(x^0, y^0)}{p_x} > \frac{MU_y(x^0, y^0)}{p_y},$$

then transferring $1 of expenditure from good y to good x would keep total expenditure constant and the consumer on the same budget line, but would reduce utility by ${}^{MU_y(x^0, y^0)}\!/_{p_y}$ from the reduction in y and, at the same time, increase utility by ${MU_x(x^0,y^0)}/{p_x}$ from the increase in x. The net effect would be to raise utility by

$$\left[{MU_x(x^0,y^0)}/{p_x}\right] - \left[{MU_y(x^0,y^0)}/{p_y}\right].$$

That is, utility would not be maximized at (x^0, y^0).

The above conditions that the marginal of rate substitution equal the price ratio (equation (6.2)) or that the marginal utilities per dollar spent on the two goods be equal (equation (6.3)) together with the budget constraint $x^0p_x + y^0p_y = m$ are first-order conditions that can be derived using calculus. These conditions identify a maximum rather than a minimum or inflection point because enough has been assumed to ensure that the indifference curve tangent to the budget line in Figure 6-2 is associated with the right utility value and has the right shape. That is, assumption 4b ensures that the consumer is on an indifference curve that is as far out from the origin as possible, and assumption 4c guarantees that the indifference curve is strictly convex. If, for example, the indifference curve were strictly concave as drawn in Figure 6-5,[3] then along the budget line, (x^0, y^0) would be associated with minimum rather than maximum

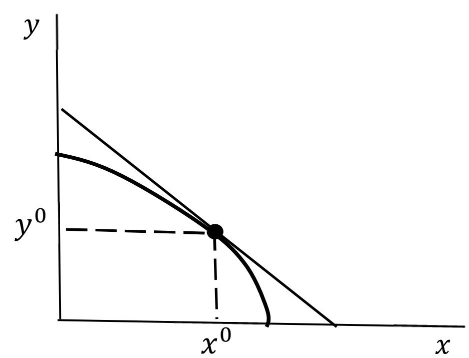


Figure 6-5

utility. In this case, moving away from (x^0, y^0) to any other point on the budget line puts the consumer on an indifference curve farther from the origin and therefore provides higher utility. Thus, assumption 4b and that of strict convexity are playing the role of the second-order conditions guaranteeing that (x^0, y^0) in Figure 6-2 is associated with a maximum and not a minimum.

Having identified and characterized the basket that maximizes utility subject to the budget constraint, utility maximization can now be used to explain consumer buying behavior, that is, consumer demand functions. Thus, for each set of prices and incomes presented by the markets, the consumer demands or buys the basket that maximizes utility subject to his/her budget constraint as determined by those prices and incomes. Mathematically,

[3] Note that in this example, assumption 4d is also violated.

$$\left.\begin{aligned} x^o &= h^x(p_x, p_y, m) \\ y^o &= h^x(p_x, p_y, m) \end{aligned}\right\} \leftrightarrow \ (x^0, y^0) \text{ maximizes } u(x,y) \text{ subject to } xp_x + yp_y = m.$$

Although the notation of Chapter 3 has been slightly altered, this defines the individual consumer demand functions of the demand-supply model of that chapter. In spite of the fact that h^x and h^y and the preferences and utility function behind them as presented here are abstract and do not exist in reality, they can still be used to explain real consumer buying behavior. To be accepted as an explanation of the latter, it is not required that when the consumer actually buys goods he/she determines the utility maximizing basket by calculation using equation (6.2) or (6.3). Rather, it is only necessary to note the model's connections to certain relevant aspects of actual buying situations: First, real consumers do have preferences among the various options they have the opportunity to buy. This appears in the model as preferences and indifferences among baskets of goods that are complete, transitive, and represented by a utility function. Second, consumers' purchases are, in reality, limited by the prices of goods and the income they have to spend. This is incorporated in the model as the budget set and budget constraint. And third, real consumers usually want to obtain the most they can from their income. They generally do not buy what they do not like or desire. This is represented in the model by the constrained maximization of a utility function that is based on preferences. The other elements and assumptions of the model are present only to be able to make individual preferences, the price-income limitation, and the idea of maximization precise so that they can be better understood in analytical terms and their consequences explored. This explanation of real consumer buying behavior and that of actual market observations provided in Chapter 3 are two examples of how explanation works in Economics.

The emergence of demand functions from utility maximization is shown geometrically in terms of price-consumption and demand curves in Figure 6-6. Start with $p_y = p_y^0$ and $m = m^0$

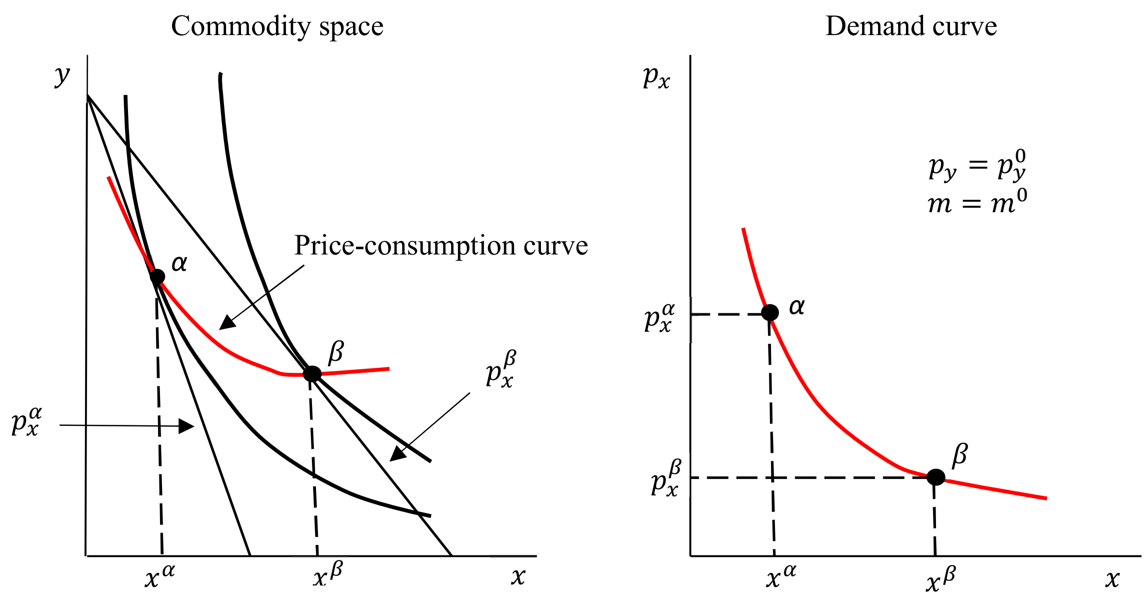


Figure 6-6

that are fixed to obtained a demand curve. Choose a price, say p_x^α in the right-hand diagram, and draw the budget constraint $xp_x^\alpha + yp_y^0 = m^0$ labeled p_x^α and identified with p_x^α, p_y^0 and m^0 in the left-hand diagram. Then utility is maximized subject to this constraint at α. Now lower the price of good x in the right-hand diagram to p_x^β keeping the price of good y and income at their original values p_y^0 and m^0. In the left-hand diagram, the budget constraint rotates out from the origin around its vertical intercept, becomes $xp_x^\beta + yp_y^0 = m^0$. and is labeled with p_x^β. (Since the price of y and income do not change, if $x = 0$ and the consumer spends all of m^0, the same amount of good y, is purchased as before. But if $y = 0$, more of good x is now bought because its price is lower and m^0 has not changed.) With the lower price of x , utility is maximized subject to the p_x^β-budget constraint at β. The demand curve for good x (drawn in red with the price of good y and income fixed, respectively, at p_y^0 and m^0) appears in the right-hand diagram. When the price of good x is p_x^α in that diagram, the quantity of good x demanded is the utility-maximizing quantity x^α in the left-hand diagram; and when the price of good x is p_x^β, the utility-maximizing quantity of good x demanded is x^β. These points on the demand curve are also identified in the right-hand diagram by, respectively, α and β.

Repeating this procedure for all positive price values on the vertical axis in the right-hand diagram generates appropriate tangencies between budget lines and indifference curves in the left-hand diagram and completes the red demand curve in the right-hand diagram. The collection of all tangent points in the left-hand diagram (the red curved line) is called the price-consumption curve. Moving along the demand curve in the right-hand diagram is equivalent to moving along the price-consumption curve in the left-hand diagram.

Chapter 7
Demand Functions and Their Properties

It is useful to begin with a summary of the model explaining consumer demand functions (consumer buying behavior) to this point**:**

To explain $x = h^x(p_x, p_y, m)$ and $y = h^y(p_x, p_y, m)$, assume (from Chapters 5, 6, and Supplemental Note C) the consumer has preferences and indifferences among baskets of commodities that are complete, transitive, and represented by a utility function. Assume also that the utility function has the following properties:

4a. It is continuous and all marginal utilities can be calculated.
4b. A larger basket of commodities is always preferred to, and therefore has a higher utility value than a smaller one.
4c. Indifference curves are strictly convex.
4d. Indifference curves do not touch the co-ordinate axes of the commodity space.

Finally assume the consumer demands (or buys) those baskets that maximize his/her utility subject to the budget constraint defined by each (p_x, p_y, m), the latter being determined in the economy's markets.

Then for any $(p_x, p_y, m) > 0$ in the domain of definition of h^x and h^y, maximization of utility subject to the budget constraint yields a unique basket (x, y) in the budget set such that

$$xp_x + yp_y = m$$

and

$$\frac{MU_x(x,y)}{p_x} = \frac{MU_y(x,y)}{p_y} \qquad \text{or} \qquad \frac{MU_x(x,y)}{MU_y(x,y)} = \frac{p_x}{p_y}.$$

Once the utility maximizing basket (x, y) is obtained from either of these equations (and the budget constraint), the consumer is thought of as demanding or purchasing that (x, y) when confronted with the specified (p_x, p_y, m). That is, using those values for (x, y) and (p_x, p_y, m), the demand functions are given by

$$x = h^x(p_x, p_y, m) \quad \text{and} \quad y = h^y(p_x, p_y, m), \tag{7.1}$$

and are explained as the outcome of maximizing utility subject to the budget constraint at the prices and incomes that consumer faces in the markets.

The properties imposed on h^x and h^y by the assumptions 4a – 4d above will shortly be considered. But there is a matter to deal with before doing so. At the end of the last chapter, it was indicated that moving along the demand curve in (x, p_x)-space is equivalent to moving along the price consumption curve in the commodity space. Another part of the demand function, different from that obtained by holding p_y and m fixed to obtain the demand curve, is related to tangencies in the commodity space that fall along a curve called the income-consumption curve.

To see what is involved, fix $p_x = p_x^0$ and $p_y = p_y^0$ and consider the function $x = h^x(p_x^0, p_y^0, m)$ of the single variable m. The graph of the latter (like the graph of the demand curve, the axes are reversed) is referred to as the Engel curve with respect to x and its geometric derivation parallels that of the demand curve described in the previous chapter. In the right-hand diagram of Figure 7-1, choose two values of m, say m^α and m^β, where $m^\alpha < m^\beta$. The budget lines associated with these values of m and the fixed prices are shown in the left-hand diagram of Figure 7-1. These two lines are parallel since their slopes

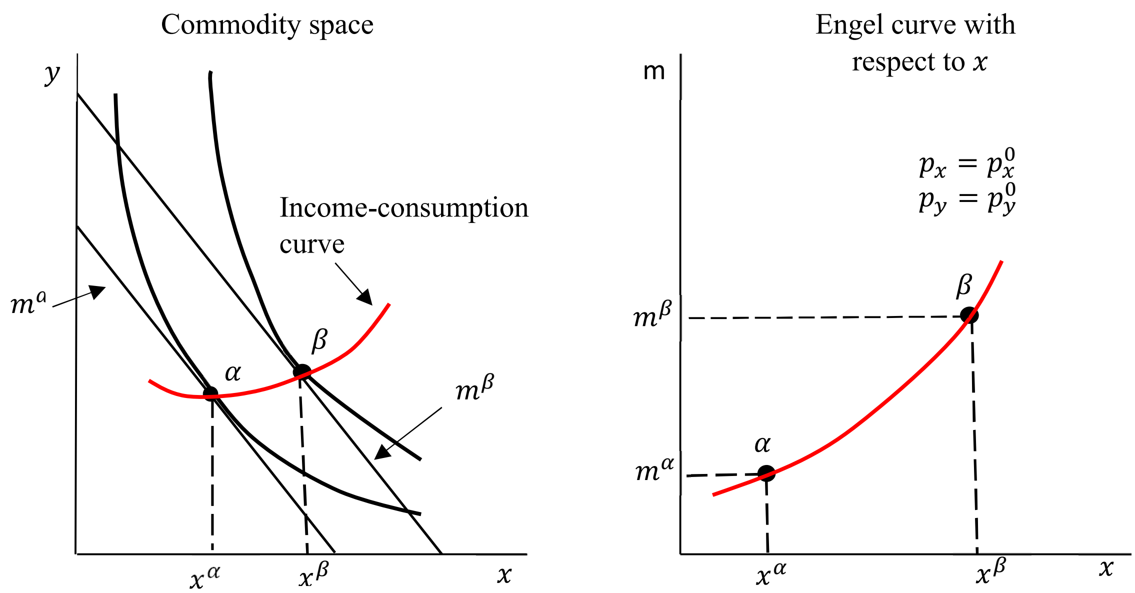


Figure 7-1

$-p_x^0/p_y^0$ are the same, and the one identified with m^β lies farther from the origin since $m^\alpha < m^\beta$. The utility-maximizing baskets occur at the tangencies α and β, and the x-co-ordinates of these baskets, namely, $x^\alpha (= h^x\left(p_x^0, p_y^0, m^\alpha\right))$ and $x^\beta (= h^x\left(p_x^0, p_y^0, m^\beta\right))$, are plotted against, respectively m^α and m^β in the right-hand diagram of Figure 7-1. Repeating this procedure for all values of m generates the income-consumption curve (the red locus of tangencies) on the left and the red Engle curve on the right. As with the price-consumption and demand cures, moving along the Engel curve is equivalent to moving along the income-consumption curve.

Attention now turns to the characteristics of consumer buying behavior. In particular, the assumptions 4a - 4d impose a number of properties on **all** demand functions generated by constrained utility maximization under their supposition. These properties are important because, to the extent that consumer demand functions can be observed as described at the beginning of Chapter 5, if an individual's observed demand functions do not satisfy them, then the model based on 4a - 4d cannot be used to explain that person's demand behavior. A model based on different assumptions would have to be employed. Like the clock example of Chapter 1, if the rotation of the hands of the model did not match the rotation of the hands of the clock whose workings were to be explained (for example, if the rotation were the wrong speed or its direction reversed), then the model would not provide an explanation. A different model would have to be constructed. Four implied characteristics of demand functions h^x and h^y will be considered.

First, in Chapter 6 it was shown that assumption 4b implies that the utility maximizing basket (x, y) lies on the budget line, that is, satisfies the equation $xp_x + yp_y = m$, and that same basket is that which appears to the left of the equal sign corresponding to (p_x, p_y, m) in the demand functions (7.1). Thus, upon substituting those demand functions into the budget-line equation,

$$h^x(p_x, p_y, m)p_x + h^y(p_x, p_y, m)p_y = m. \tag{7.2}$$

This is the statement expressed in terms of demand functions that for all $(p_x, p_y, m) > 0$ in the domain of definition of h^x and h^y, the consumer spends all of his/her income.

Second, since assumption 4d implies that no basket containing zero amount of one or both goods can lie on an indifference curve, and hence that utility maximizing baskets cannot be located on the co-ordinate axes, the consumer must demand or buy something of each good. Mathematically,

$$h^x(p_x, p_y, m) > 0 \quad \text{and} \quad h^y(p_x, p_y, m) > 0,$$

for every $(p_x, p_y, m) > 0$ in the domain of definition of h^x and h^y.

Third, from the way that the budget constraint and utility function have been defined,

$$h^x(\lambda p_x, \lambda p_y, \lambda m) = h^x(p_x, p_y, m)$$

for all numbers $\lambda > 0$ and all $(p_x, p_y, m) > 0$ in the domain of definition of h^x. A similar equation holds for h^y. This is because doubling ($\lambda = 2$), say, both prices and income does not change the budget equation (since the 2s placed in front of p_x, p_y, and m in that equation cancel out), and cannot change the indifference curves (since the utility function and indifference curves do not depend on prices and income – only on commodity baskets). It follows that there can be no change in the tangency between the budget line and indifference curve, no change in the utility maximizing basket and, therefore, no change in the quantities demanded.

To illustrate these ideas, it is easily verified that the demand functions

$$h^x(p_x, p_y, m) = \frac{m}{2p_x} \qquad \text{and} \qquad h^y(p_x, p_y, m) = \frac{m}{2p_y}$$

defined for all $(p_x, p_y, m) > 0$ satisfy all three properties described above.[1] For example, substituting these functions into the left-hand side of equation (7.2) gives

$$\frac{m}{2p_x}p_x + \frac{m}{2p_y}p_y = \frac{m}{2} + \frac{m}{2} = m,$$

establishing the first of the above properties. On the other hand, the linear functions

$$h^x(p_x, p_y, m) = 2p_x + p_y + m \qquad \text{and} \qquad h^y(p_x, p_y, m) = p_x + 2p_y + m, \tag{7.3}$$

also defined for all $(p_x, p_y, m) > 0$, although adhering to the second of the above properties, violate the first and the third. These latter functions, therefore, cannot be derived from constrained utility maximization under the assumptions set out at the beginning of this chapter. Were the functions of (7.3) observed as a description of a consumer's demand behavior, that behavior cannot be explained in terms of the model set out here.

There is one more property of all demand behavior based on constrained utility maximization to consider. But it is worth pausing for a moment to look at the kind of things that can happen when the assumptions that have been made to explain consumer buying behavior are weakened or discarded. Here are four examples, three of which may be thought to allow for more realistic possibilities than those permitted by the model and assumptions set out above. In each example, only those assumptions explicitly indicated have been modified or eliminated.

1. Discard assumption 4b that a larger basket of commodities is always preferred to, and therefore has a higher utility value than a smaller one and allow indifference curves to be partly concave in violation of assumption 4c. Then indifference curves could be closed curves as shown in the left-hand diagram of Figure 7-2. In that diagram, (x^0, y^0) is preferred to all other baskets in the commodity space and has a higher utility value than any other basket. With prices and income defining the red budget line, the consumer would demand or buy the basket (x', y') at the tangency between that line and the lower indifference curve as in the case in which all assumptions 4a – 4d applied. But were the black budget line in force, the utility maximizing basket would fall in the interior of the budget set at (x^0, y^0) and would not lie on the budget line or at a tangency between that line and an indifference curve. In that case the consumer would not spend all of his/her income and equation (7.2) and the property of demand behavior it reflects would not be satisfied. However, the second and third properties would still apply.

2. Keeping assumption 4b, eliminating the assumption that indifference curves are strictly convex, and permitting those curves to touch the co-ordinate axes could result in minimum utility along the budget line at the tangency between an indifference curve and the budget line.

---

[1] These demand functions can be derived from constrained utility maximization under the above assumptions using calculus methods.

(This possibility was described in the previous chapter.) Here the utility maximizing basket would lie on one of the co-ordinate axes as at A in the right-hand diagram of Figure 7-2. This violates the second property that, at the prices and income generating the budget line of the diagram, $h^y(p_x, p_y, m) > 0$. But the first and third properties would still hold.

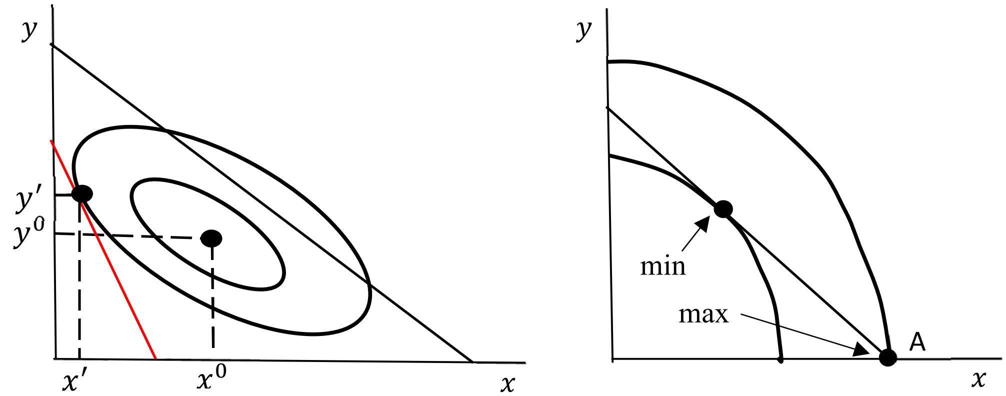


Figure 7-2

3. Dropping just the assumption that indifference curves do not touch the co-ordinate axes of the commodity space would also lead to the possibility that the utility maximizing basket would appear on one of the co-ordinate axes, with or without a tangency. Like the previous example, the consumer does not always have to buy or demand a positive quantity of each good. The left-hand diagram of Figure 7-3 provides an illustration in which, at the constrained utility maximizing basket $(x^0, 0)$, the consumer is spending all income on good x and not buying good y. Again, the second property would not hold and the first and third would still apply. Also, multiple budget lines each associated with different prices and incomes lead to the same utility-maximizing basket. But only one has the same slope at $(x^0, 0)$ as the indifference curve and at that basket is tangent to it.

4. Deleting the strictness of the convexity of indifference curves in assumption 4c (but still requiring convexity) permits linear segments to be present in indifference curves (Chapter 2). This could leave the consumer with multiple baskets that maximize utility subject to the same budget constraint, and no way to choose among them. (The possibility was described in a different mathematical context in Figure 2-11 of Chapter 2.) In the example of the right-hand diagram of Figure 7-3, the straight- line segment connecting A to B on the indifference curve coincides with the budget line drawn through those points. All baskets between A and B maximize utility subject to the budget constraint. Since, in reality, the consumer demands or buys only one basket of commodities, this does not provide a full explanation of the consumer's

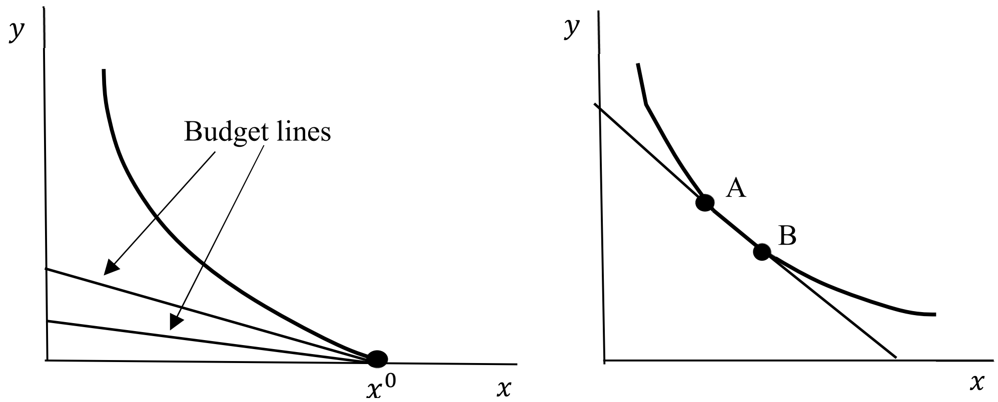


Figure 7-3

buying behavior because it does not indicate which basket the consumer would actually buy.

In all of these cases except the last, demand functions can still be defined in terms of utility maximization subject to the budget constraint. That is, demand or buying behavior can still be explained in terms of constrained utility maximization. But with assumptions different from those of 4a – 4d, the model obtained and the demand functions generated by it will have different characteristics from those of the model developed here. In particular, one or more of the properties of demand functions described earlier need not apply.

Returning to the development of the fourth and final property of demand behavior derived from the assumptions of the model presented here, it should first be pointed out that, although demand curves have almost always been drawn up to now as downward sloping, upward sloping demand curves are still consistent with assumptions 4a – 4d. It all depends on the location of the tangencies between budget lines and indifference curves as prices vary. These locations depend on the characteristics of the utility function not eliminated by assumptions 4a – 4d. To illustrate, fix $p_y = p_y^0$ and $m = m^0$. Using the price value p_x^α in the right-hand diagram of Figure 7-4 along with the fixed p_y^0 and m^0 to draw the budget line labeled p_x^α in the left-hand

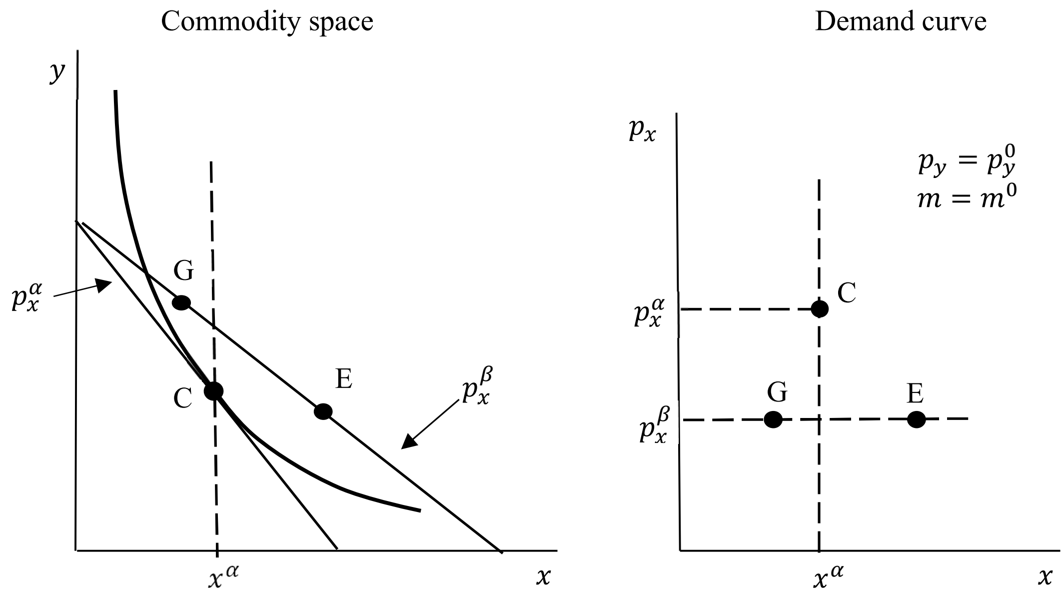


Figure 7-4

diagram (recall the geometric construction of the demand curve at the end of Chapter 6), the tangency with that budget line at C identifies a utility-maximizing quantity demanded of x^α. The latter appears in the right-hand diagram as the x-co-ordinate of C. Lowering the price to p_x^β in the right-hand diagram and keeping p_y^0 and m^0 the same, gives the budget line labeled p_x^β in the left-hand diagram. If the new utility-maximizing tangency along that budget line lies to the right of the vertically dashed line through C, say at E, more will be demanded at the lower price and the demand curve in the right-hand diagram will slope downward from C to E. If the tangency in the left-hand diagram is located to the left of the dashed line at G, less will be demanded and the demand curve will slope upward between G and C in the right-hand diagram. Whether the new tangency lies to the right or left of the dashed vertical line depends on the nature of the consumer's preferences and hence his/her utility function. Thus, a downward sloping demand curve cannot be a property of all demand functions obtained from constrained utility maximization under the assumptions made here.

The fourth property of demand behavior that applies to all demand functions derived from constrained utility maximization in the present model is developed in relation to what are called income and substitution effects. (The notion of substitution effect is entirely different from the concept of substitute goods defined in Chapter 3.) Start as in Figure 7-4 at the utility maximizing basket α in Figure 7-5 where p_x is set at p_x^α and p_y^0 and m^0 are held fixed. Also, as

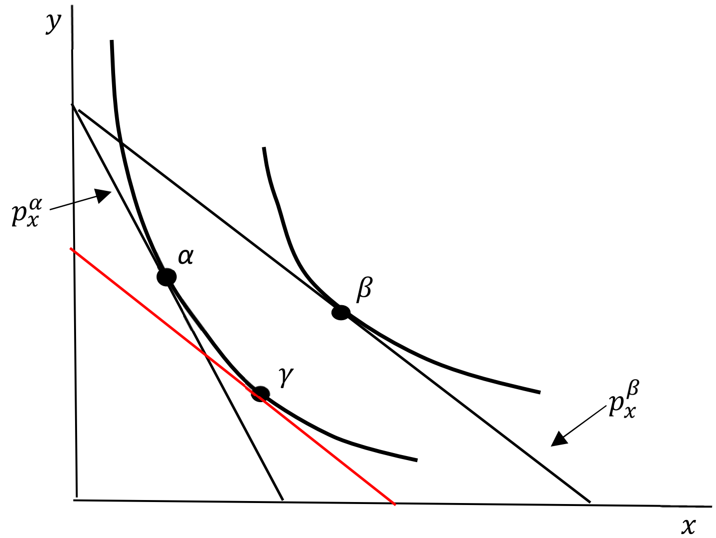


Figure 7-5

in Figure 7-4, let p_x^{α} be lowered to p_x^{β} without changing p_y^0 and m^0. The constrained utility maximizing basket moves to β (recall Figure 6-6 in Chapter 6). Now, with p_y^0 fixed, leave p_x at the lower level p_x^{β} and, in a thought experiment only, take income away from the consumer so that the outer budget line (labeled p_x^{β} in Figure 7-5) moves toward the origin in parallel fashion until it (the red line) is tangent to the indifference curve through α at γ. Thus the passage from α to β caused by the change in the price of good x in Figure 7-5 is split into two parts: The movement from α to γ is called the substitution effect; that from γ to β is referred to as the income effect. In other words:

> The substitution effect is the reallocation of the consumer's demands if the change in the price p_x is "compensated" by a simultaneous adjustment in income (negative in the case of the diagram), forcing him/her after re-maximizing to remain on the original indifference curve at γ, that is, moving the consumer from α to at γ.
>
> The income effect is the change in demands upon restoring the consumer's income, that is, from allowing the consumer to move from γ back to β**.**

Note that the income effect reflects movement along an income-consumption curve; the substitution effect does not coincide with movement along a price-consumption curve. The property of demand in question will be seen to emerge from the substitution effect in the next chapter.

Chapter 8
One More Property of Demand Functions and Consumer Surplus

As noted in the previous chapter, there is one more property that all consumer demand functions derived from constrained utility maximization under the assumptions set out in Chapters 5 and 6 must exhibit. To describe that property requires the use of income and substitution effects. Only the property expressed with respect to the demand curve for good x, that is the graph of $x = h^x(p_x, p_y^o, m^o)$, will be examined. The demand curve for good y has a similar characteristic.

Reminder from Chapter 7:

Income and substitution effects are defined as follows: Start at the utility maximizing basket at α in Figure 8-1 where p_x is set at p_x^α and p_y^0 and m^0 are held fixed. Let p_x^α be lowered

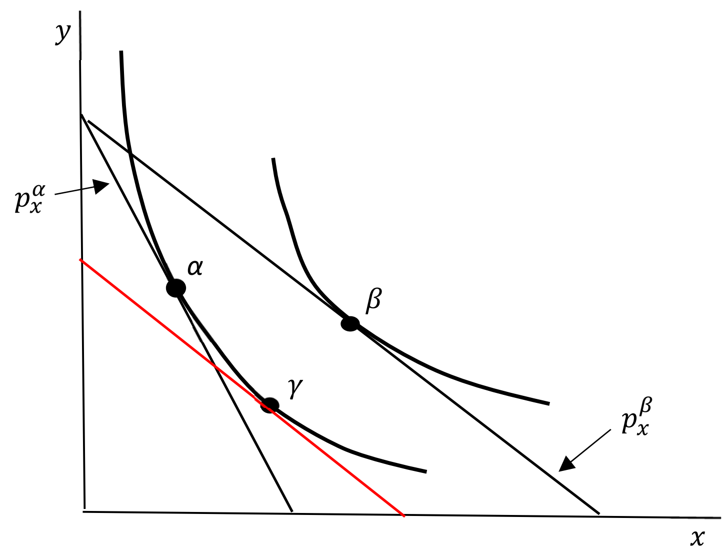


Figure 8-1

to p_x^β without changing p_y^0 and m^0 so that the constrained utility maximizing basket moves to β. With p_x at the lower level p_x^β, take income away from the consumer so that the outer budget line (labeled p_x^β) moves in parallel fashion toward the origin until it (the red line) is tangent to the indifference curve through α at γ. The movement from α to β is broken up into two parts: That from α to γ is the substitution effect; the movement from γ to β is the income effect.

It was also pointed out in Chapter 7 that demand curves could slope upward or downward depending on how the location of the tangencies between indifference curves and budget lines change as p_x varies with p_y^0 and m^0 held fixed. This implied that the slope of the demand curve could not be taken as a property of all demand curves derived from constrained utility maximization under the assumptions that have been made. Recalling the geometric derivation of the demand curve described in Chapter 6 in relation to Figure 6-6, the slope of the demand curve is given by ${}^{\Delta p_x}/_{\Delta x}$. This ratio and its reciprocal ${}^{\Delta x}/_{\Delta p_x}$ have identical signs. It will be convenient in the following discussion, which is concerned with finding a replacement for the sign of the slope of the demand curve that applies universally under the assumptions made, to begin with the reciprocal slope ${}^{\Delta x}/_{\Delta p_x}$. It should be kept in mind that the former is obtained only by reversing the axes when graphing the demand function $x = h^x(p_x, p_y^o, m^o)$ and has the same sign as the latter.

The break-up of the movement from α to β into income and substitution effects caused by the price decline from p_x^α to p_x^β (with p_y^0 and m^0 remaining fixed) in Figure 8-1 can be applied to the reciprocal slope of the demand curve. Relating to the geometry of Figure 8-2 (which is Figure 6-6 redrawn without the price consumption curve and using a different color

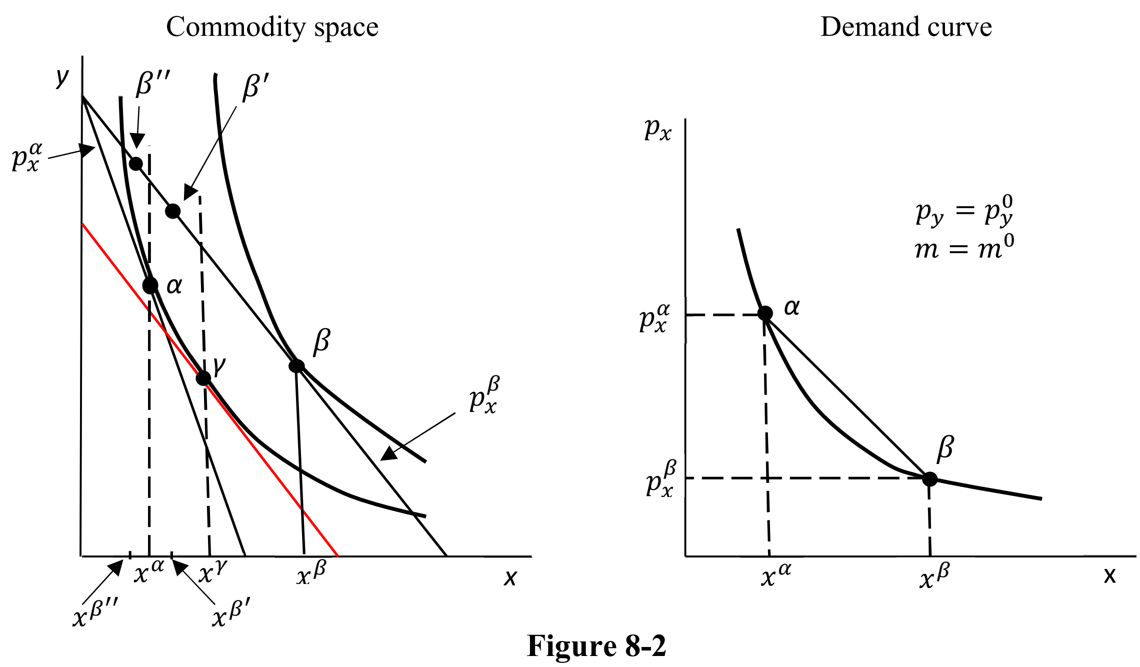


Figure 8-2

scheme), the movement along the demand curve in the right-hand diagram from α to β is coincident with the movement from α to β in the left-hand diagram. The reciprocal of the slope of the demand curve at α in the right-hand diagram is approximated by the reciprocal of the slope of the straight-line segment connecting α to β. Letting x^α, x^β, p_x^α and p_x^β be the values specified in the diagrams of Figure 8-2 and setting $\Delta x = x^{\alpha} - x^{\beta}$ and $\Delta p_x = p_x^{\alpha} - p_x^{\beta}$, the expression for this approximate slope is:

$$\frac{\Delta x}{\Delta p_x} = \frac{x^{\alpha} - x^{\beta}}{\Delta p_x} = \frac{x^{\alpha} - x^{\gamma} + x^{\gamma} - x^{\beta}}{\Delta p_x}$$

or

$$\frac{\Delta x}{\Delta p_x} = \frac{x^{\alpha} - x^{\gamma}}{\Delta p_x} - \frac{x^{\beta} - x^{\gamma}}{\Delta p_x}, \tag{8.1}$$

where x^{γ} is the x-co-ordinate of γ in the left-hand diagram of Figure 8-2. The term $(x^{\alpha} - x^{\gamma})/\Delta p_x$ is called the substitution ratio and $(x^{\beta} - x^{\gamma})/\Delta p_x$ is the income ratio. The former is a measure of the substitution effect $\alpha \to \gamma$; the latter a measure of the income effect $\gamma \to \beta$.

The left-hand diagram in Figure 8-2 suggests possibilities other than β for tangencies after the price of good x falls to p_x^{β}. There could be a tangency at, for example, β' that gives an x value $x^{\beta'}$ between x^{α} and x^{γ}. Alternatively, the tangency could be at β'' producing an x value $x^{\beta''}$ to the left of x^{α}. Each of these possibilities is associated with a different set of preferences, still satisfying all assumptions previously imposed and that, as β moves to β'', result in indifference curves (not shown in the diagram) that become closer together at small x-values and farther apart at large x-values. Note that γ and x^{γ} do not modify as the tangencies change. This is because the substitution effect is determined by the outer budget line and the inner indifference curve which remain fixed as the tangencies in the left-hand diagram of Figure 8-2 vary. Of the three tangencies in that diagram, it is only with the tangency at β'' that, as indicated in the previous chapter, the demand curve will slope upward.

The sign of the reciprocal slope of the demand curve can now be interpreted in relation to the signs of the income and substitution ratios in each of the three cases of Figure 8-2. Recall that the sign of the reciprocal slope is the same as the sign of the slope of the demand curve. That is, if, $\Delta x/\Delta p_x > 0$, then $\Delta p_x/\Delta x > 0$. Using the formulas of equation (8.1) and determining the signs of Δx, Δp_x, $x^{\alpha} - x^{\gamma}$, and $x^{\beta} - x^{\gamma}$ from the left-hand diagram of Figure 8-2, the following table of signs for the three cases pictured in Figure 8-2 is obtained:

Table 8-1

Tangency After the Price Change	Substitution Ratio: $\frac{x^{\alpha} - x^{\gamma}}{\Delta p_x}$	Income Ratio: $\frac{x^{\beta} - x^{\gamma}}{\Delta p_x}$	Reciprocal of the Demand Curve: $\frac{\Delta x}{\Delta p_x}$
β	Negative	Positive	Negative
β'	Negative	Negative	Negative
β''	positive	Negative	Negative

For example, when the tangency is at β, the numerator in the substitution ratio is $x^{\alpha} - x^{\gamma} < 0$ since, according to the left-hand diagram of Figure 8-2, $x^{\alpha} < x^{\gamma}$. Combining this with $\Delta p_x = p_x^{\alpha} - p_x^{\beta} > 0$ because p_x^{α} and p_x^{β} have been selected so that $p_x^{\alpha} > p_x^{\beta}$, yields the result that the substitution ration $x^{\alpha} - x^{\gamma}/_{\Delta p_x}$ is negative. The conclusion that the substitution ratio is negative applies to all three cases in Table 8-1 since, as pointed out above, γ and x^{γ} do not change from case to case. Thus, according to case β'', when the income ratio is sufficiently negative so that upon multiplication by minus one it offsets the negative substitution ratio, according to equation (8.1) the demand curve will slope upward.

In Chapter 7, three properties of demand functions were identified that are implied by constrained utility maximization under the assumptions 4a - 4d of Chapters 5 and 7. Such properties. recall, are essential in determining if that model is a possible explanation of an observed consumer's demand functions. Table 8-1 illustrates a fourth property: Although demand curves may slope downward or upward, the substitution ratio, or what may be thought of as a "part" of the slope of the demand curve drawn without the reversal of axes, is always negative. This is a consequence of the downward slope and strictly convex shape of the indifference curves that emerges from assumptions 4a - 4d. It should be pointed out that, as derived from Figure 8-2, in order to be able to tell if an observed consumer's behavior exhibits negative substitution ratios, it is necessary to know that consumer's indifference curves. These curves are derived from a utility function based on the consumer's preferences. And, as indicated in Chapter 5, these exist only in the consumer's mind and are not observable. As it stands, then, the negativeness of substitution ratios is not really an observable property of demand behavior. However, with more advanced mathematics than that available for use here, this property can be expressed entirely in terms of observable demand functions.[1]

Table 8-1 also provides illustrations of two concepts originally defined in Chapter 3: Goods with positive income ratios (as in case β) are normal goods. Goods with negative income

---

[1] See, for example, Section 2.4 of D.W. Katzner, *An Introduction to the Economic Theory of Market Behavior: Microeconomics from a Walrasian Perspective* (Cheltenham: Elgar, 2006).

ratios (as in cases β' and β'') are inferior goods. In addition, goods with positively sloped demand curves (as in case β'') are Giffen goods.

Before leaving consumer buying behavior, it is appropriate to introduce the concept of consumer surplus which will be employed in later chapters.

Individual or market demand curves are normally interpreted as indicating how much of a good, call it x. buyers will demand at each price. But things could be turned around to think of those curves as describing the price buyers are willing to pay per unit of the good at each quantity. In this interpretation, the demand curve becomes a "willingness-to-pay" curve, p_x becomes the dependent variable which is determined as a function of the independent variable x, and the geometric picture of the demand curve with x on the horizontal axis and p_x on the vertical axis no longer reverses mathematical convention. Consider the same "willingness-to-pay" curve drawn as straight lines in each of the two diagrams of Figure 8-3. Take these lines, each labeled *D,* as market demand curves. In the left-hand diagram, think of the intervals delineated on the x-axis as associated with two units of the good. Then the area of box J is the amount buyers are willing to pay (price times quantity) for the first two units of good x. Similarly, the area of box K is the amount buyers are willing to pay for the second two units. Combining the area of the two boxes gives the amount buyers are willing to pay for the first four units. Now associate the intervals on the x-axis with single units of the good as in the right-hand

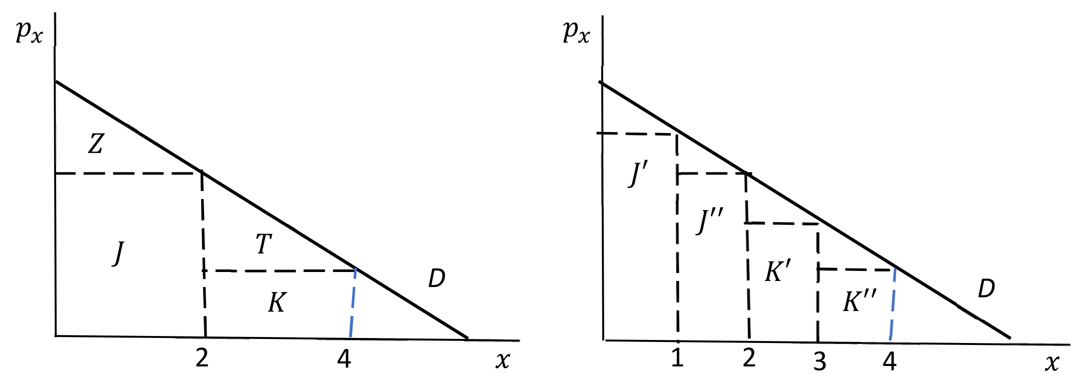


Figure 8-3

diagram of Figure 8-3. Here the boxes J', J'', K', and K'' indicate the amounts buyers are willing to pay for units 1, 2, 3, and 4, respectively. By cutting the number of units associated with an interval in half in the right-hand diagram, parts of the triangular areas Z and T in the left-hand diagram are added to the combined area of boxes J and K, that is, added to the amount buyers are willing to pay for the first 4 units of good x. Thus, the combination of boxes J', J'', K', and K' in the right-hand diagram is a more accurate description of what buyers are willing to pay for those units than the combination of boxes J and K. Making the intervals smaller and smaller, for example, restricting the intervals to be associated with half units, quarter units, and so on, eventually they shrink to single points. When that happens, the entire area under the demand curve up to $x = 4$ becomes the amount buyers are willing to pay for all units up to four units of good x. A similar argument applies to non-linear demand curves. It will be convenient to use the area obtained by restricting intervals to single points as a measure of the amount buyers are willing to pay for all units of the good up to a specified quantity on the x-axis.

Consumer surplus at x^0 is the difference between the amount buyers are willing to pay for all units of the good up to x^0 and the amount they actually have to pay for x^0. To illustrate, consider the market demand curve *D* in Figure 8-4. Suppose the market price is set at p_x^0 with associated quantity x^0. For the increment Δx, buyers are willing to pay approximately the area identified by LMNQRS. But since the market price is p_x^0, they only have to pay area SNQR. Area

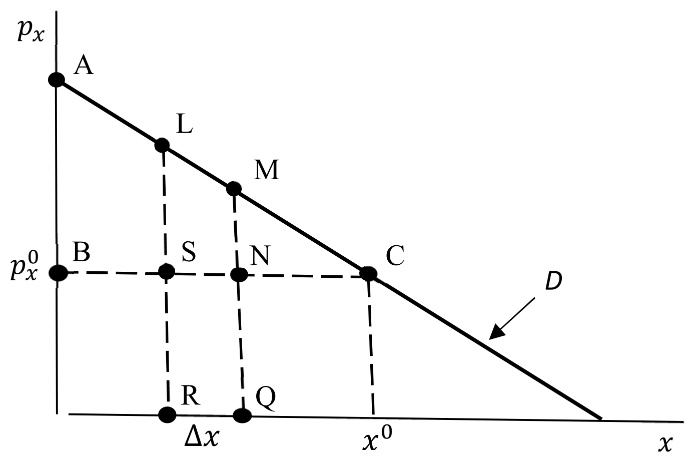


Figure 8-4

LMNS is the consumer surplus for that increment. Shrinking the increments (intervals) to single points, consumer surplus at x^0 is the triangular area ABC. And the greater the consumer surplus, the greater the benefit to consumers. For example, if the price were to fall below p_x^0 in Figure 8-4, area ABC would become larger, the consumer could buy more with the seme income and would therefore be better off.

Chapter 9
Factor Supplies and Production

The next piece to be put in place in constructing the Walrasian model of the operation of the microeconomic economy is an explanation of consumer selling behavior, that is, how consumers decide how much to supply in the factor markets. Remember that in supplying factors, consumers obtain the income with which to buy consumption goods produced by firms. Only the supply of the factor labor and the associated labor market will be considered here. At the individual level, the variables that will be taken to be observable in relation to labor supply are the time worked, the wage, and hence the income secured from work (for example, wage times hours worked). Through the process of abstraction, this leads to an observable labor supply function summarizing all of the consumer's labor-supply behavior and expressing labor-time supplied as dependent on the wage. To explain how this function comes about, the model of consumer buying behavior of the last four chapters will be adapted to the present context.

Let time be represented in terms of hours and consider consumer selling behavior in the labor market during a single day. Divide the use of time into two categories: work and non-work. Call the latter leisure. Then if λ represents hours at leisure and ℓ denotes hours at work or labor hours, $\ell + \lambda = 24$, and, according to this equation, choosing a value for one of λ and ℓ determines the value of the other. In particular,

$$\ell = 24 - \lambda. \tag{9.1}$$

The less leisure time the consumer takes, the more labor time the consumer provides, and the greater his/her income. Thus, each quantity of labor time supplied yields a "basket" containing specific amounts of leisure time and income.

In parallel with the model of consumer buying behavior, the individual may be thought of as choosing among baskets containing quantities of leisure time λ and income m. Once the latter choice is determined, the amount of labor time supplied is obtained from equation (9.1). Assume the individual has complete and transitive preferences and indifferences among baskets (λ, m) in a "commodity" space limited by the number of hours in the day. This commodity space is defined by $\{(\lambda, m): 0 \leq \lambda \leq 24 \text{ and } 0 \leq m\}$ and pictured in Figure 9-1.

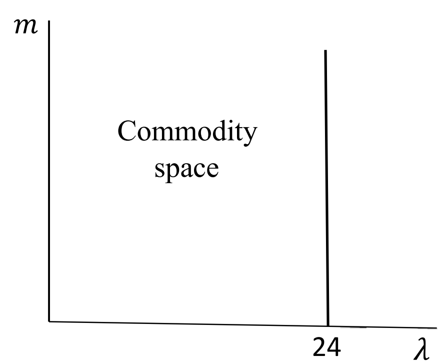


Figure 9-1

Each hour the individual works he/she is paid a wage of p_ℓ per hour as determined in the labor market. Assuming the individual has no source of income other than labor hours sold,

$$m = \ell p_\ell$$

or, substituting for ℓ using equation (9.1),

$$m = (24 - \lambda)p_\ell. \tag{9.2}$$

The last equation may be rewritten as

$$\lambda p_\ell + m \cdot 1 = 24p_\ell. \tag{9.3}$$

This is the "budget line or constraint" in the present context and it looks very much like the budget constraint $xp_x + yp_y = m$ employed in the model explaining consumer buying behavior, that is, equation (6.1) in Chapter 6. In the present case:

i) m and λ function as quantities (x and y).
ii) $24p_\ell$ functions as income (the m in $xp_x + yp_y = m$).
iii) In addition to representing the wage, p_ℓ is also the price of a leisure hour, or what the consumer has to give up in wages in order to have an hour of leisure. Also, the price of a dollar's worth of income is 1. These prices correspond to p_x and p_y respectively. The latter price will be dropped in subsequent budget equations.

Geometrically, the graph of this budget constraint, equation (9.3), is the solid line in Figure 9-2

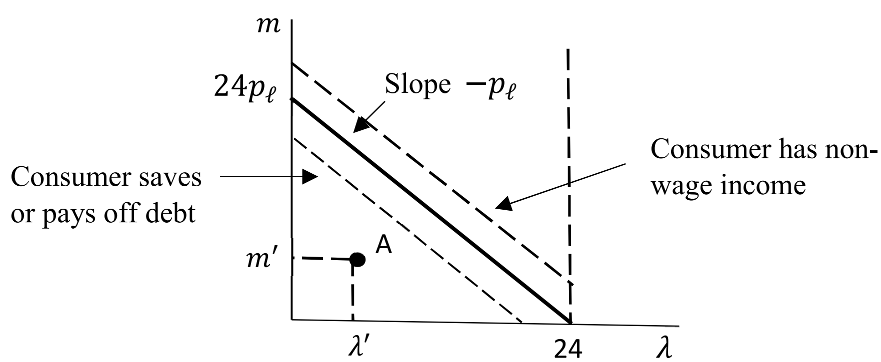


Figure 9-2

having vertical intercept $24p_\ell$, which identifies the income received when the consumer takes no leisure and spends all time working. The horizontal intercept is 24 indicating that the consumer receives no income when spending 24 hours at leisure and no time working) The slope of the line is $-{}^{24p_\ell}/_{24} = -p_\ell$. Were the consumer to have non-wage income, the budget line would shift vertically up by the amount of that income; were he/she to pay off debt or save, the budget line would shift vertically down by the amount paid or saved (recall that income m is intended to reflect the amount of money the consumer has to spend). These shifts are parallel to the original budget line since, in either case, the wage does not change. However, the possibilities of savings, debt, and nonwage income are not considered further in this chapter.

Once p_ℓ is specified, the consumer's income is determined from equation (9.2) by the amount leisure time taken. He/she cannot choose values of m and λ independently of each other. Hence, baskets (λ, m) for which $\lambda p_\ell^0 + m < 24p_\ell$ are not really choice- options for the consumer. However, to keep argument parallel to that of the consumer-buying-behavior model of Chapter 6, baskets below the budget line like A in Figure 9-2 are included in the budget set. But as is indicated momentarily, the outcome of constrained utility maximization places the consumer on the budget line anyway and is therefore the same regardless of whether such baskets are included in the budget set or not.

Thus, given the price of leisure coming out of the labor market, say p_ℓ^0, the consumer is thought of as selecting a basket (λ, m) from the budget set of "possible choices" available or

$$\{(\lambda, m): \lambda p_\ell^0 + m \leq 24p_\ell\}.$$

In addition to completeness and transitivity, assume, as in the buying-behavior model, that the consumer's preferences are represented by a utility function $\mu = u(\lambda, m)$ with the same properties employed earlier. The latter properties are repeated here using the same numbering as in Chapters 5 and 7:

4a. It is continuous and all marginal utilities can be calculated.
4b. A larger basket of commodities is always preferred to, and therefore has a higher utility value than a smaller one.
4c. Indifference curves are strictly convex.
4d. Indifference curves do not touch the co-ordinate axes of the commodity space.

Finally assume the consumer chooses the basket (λ, m) that maximizes his/her utility subject to the budget constraint. (These assumptions, recall, are listed in Supplemental Note C).

In parallel with the earlier explanation of consumer buying behavior, utility is maximized at the tangency α between an indifference curve and the budget line as shown in Figure 9-3. Thus the consumer selects leisure time λ^0, labor supply $24 - \lambda^0$, and income m^0. The first-order equations that describe the tangency at the constrained utility maximizing basket (λ^0, m^0) are the budget constraint (9.3) along with the statement that two slopes are equal. The latter says that the marginal rate of substitution or the ratio of marginal utilities with respect to leisure and income equals the price ratio (remember that the price of income is 1):

$$\frac{MU_\lambda(\lambda^0, m^0)}{MU_m(\lambda^0, m^0)} = p_\ell^0.$$

Note in Figure 9.3 that while the length along the λ-axis from zero to λ^0 represents the utility-maximizing quantity of leisure time demanded, that from λ^0 to 24 is the utility-maximizing quantity of labor time supplied.

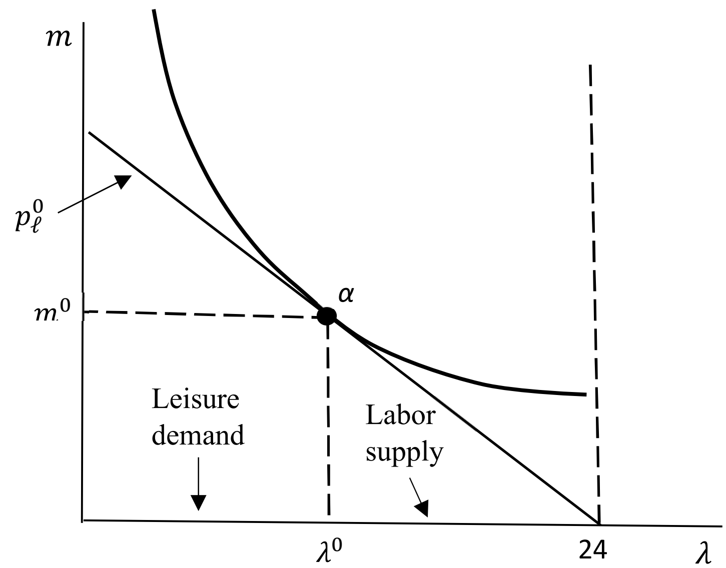


Figure 9-3

Demand functions for leisure and income, h^λ and h^m respectively, are obtained from this maximization in the same way as in the model of consumer buying behavior. That is, to each value of p_ℓ, the functions h^λ and h^m assign the utility maximizing basket quantities λ and m:

$$\lambda = h^\lambda(p_\ell, 24) \quad \text{and} \quad m = h^m(p_\ell, 24).$$

The price of income, namely 1, is implicit in these equations. Note that the number 24 represents the total amount of time the consumer has to allocate between work and leisure. If that time were represented by the symbol *T*, then nothing would change in the above argument except that the 24 would be replaced by *T,* and p_ℓ and λ would be measured in different units. The demand functions would look like

$$\lambda = h^\lambda(p_\ell, T) \quad \text{and} \quad m = h^m(p_\ell, T).$$

Since the supply of labor is found by subtracting the demand for leisure from 24 or *T*, the same subtraction applies to obtain the supply function of labor h^ℓ. That is, using *T* instead of 24 so that equation (9.1) becomes $\ell = T - \lambda$,

$$\ell = h^\ell(p_\ell, T) = T - h^\lambda(p_\ell, T). \tag{9.4}$$

This is the consumer's supply function for labor alluded to at the outset. Of course, T is set by the time period under consideration, p_ℓ is determined by the labor market, and ℓ is the outcome of constrained utility maximization. Real supply behavior is obtained by observing the actual price and corresponding amount of labor time supplied by the consumer in different market situations. Note that the properties of h^λ and h^m are similar and derived in the same way as the properties of the consumer demand function h^x and h^y (Chapters 7 and 8). The properties of h^ℓ (needed to check whether an observed consumer's supply behavior can be explained by this model) are derived from those of h^λ through the use of equation (9.4) and are not pursued here.

The consumer's supply curve of labor is found by fixing $T = T^0$ and graphing $\ell = h^\ell(p_\ell, T^0)$. This curve is derived geometrically in Figure 9-4 in the same way that the demand curve for good x was geometrically constructed in Chapter 6. When the wage rises from p_ℓ^α to p_ℓ^β in the right-hand diagram, the budget line in the left-hand diagram rotates clockwise around T^0 on the horizontal axis: At the higher wage, if the consumer spends all time at leisure and no time at work, his /her income remains at zero; if he/she spends all time at work with no leisure, his/er income is higher than before. Whether the supply curve slopes upward or downward depends on the location of the tangencies before and after the change in p_ℓ. In the figure, the

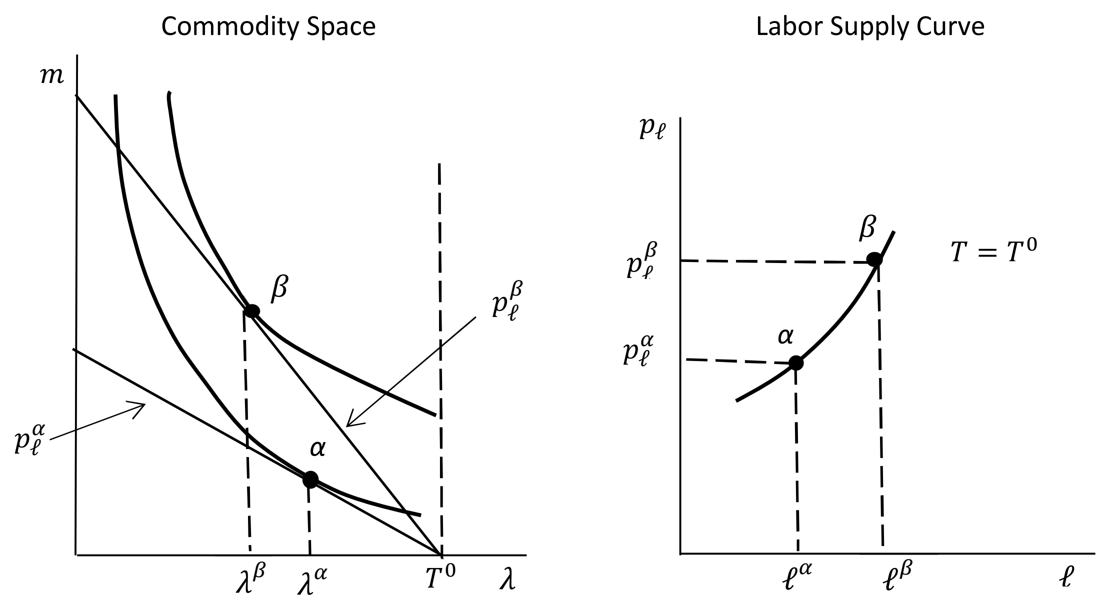


Figure 9-4

supply curve (right-hand diagram) slopes upward because the tangency at β (left-hand diagram) generated by the higher price p_ℓ^β lies to the left of the one at α corresponding to the lower price p_ℓ^α. Thus, recalling that $\ell = T^0 - \lambda$, the consumer demands less leisure at the higher wage p_ℓ^β and supplies more labor.

It is possible for the supply curve to slope upward at low wages and then bend backwards and slope downward at high wages as in Figure 9-5. All possibilities can be explained in terms of income and substitution effects as was done in the model of consumer buying behavior.

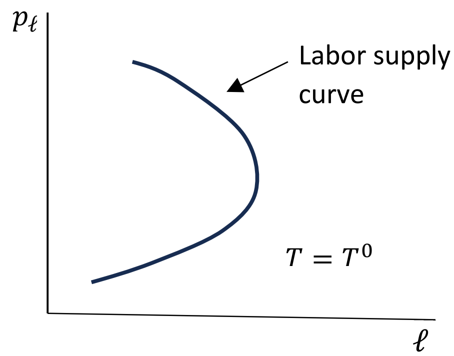


Figure 9-5

To obtain the economy-wide labor supply curve, sum all individual supply curves horizontally, that is, at each price, add the labor time supplied by all consumers to determine the market quantity at that price. This parallels the determination of the market demand curve from the demand curves of all individuals in the market. However, the labor supply curve for any particular industry is not the horizontal sum of the supply cures of all individuals working in that industry. That is because as wages rise along such a horizontally summed curve, persons may move from other industries where the wage is lower to the one under consideration with the now higher wage. Without knowing which individuals will move along with their supply curves, it is not possible to determine new industry quantities supplied. The labor supply curve of an industry depends not only on how its laborers react to wage changes but also on how outsiders react as well.

Attention now turns to an explanation of firm buying and selling behavior in the abstract Walrasian model constructed in this volume. That behavior is summarized in firm input demand and output supply functions. These functions, which, as will be seen, relate input and output prices to input quantities demanded and output quantities supplied, are observable in much the same sense as consumer commodity-demand and labor-supply functions. It turns out that in the model subsequently constructed, other functions can also be considered to be observable. But that observability will not be explored here.

The model explaining the firm's input demand and output supply functions is more complex than that built earlier to explain consumer demand and supply functions. Indeed, it will take seven more chapters to complete. That model includes a description of how the firm decides how much output to produce and how many units of inputs to hire in producing that output. It is assumed that all output produced by the firm will be supplied on the market. At the beginning there will be a very close analogy to the models of consumer buying and selling behavior. Only a firm with two inputs and one output will be considered.

Firm production is based on technology which is defined as the collection of all available information concerning the ways in which inputs can be combined to produce output. Designate quantities of the firm's output by x. Quantities of its inputs are represented by ℓ, labor time, and k. capital units, that is, as indicated in footnote 2 of Chapter 1, units of physical items such as machines and factories. To keep matters simple, all other inputs are ignored. Each input basket (ℓ, k) contains specified amounts of labor and capital. The collection of all possible input baskets, $\{(\ell, k): \ell \geq 0, k \geq 0\}$, is called the input space, the analogue of the commodity space in the consumer buying-selling models. The starting point of the present explanation of firm buying and selling behavior is the firm production function as described by

$$x = f(\ell, k).$$

Here the function f indicates the **maximum** output x that can be obtained for every basket in the input space, given technology.[1] The production function parallels the utility function except that it is derived from technology rather than preferences and is not ordinal. (Output values x are meaningful in the usual way while utility values, recall, have no meaning other than to indicate an individual's ordering of commodity baskets by his/her preferences.) It should be emphasized that the production function is defined for a specific technology. Were that technology to change, the production function would likely modify.

Two time periods in which firm buying and selling behavior can take place will be considered. The long run is a period of time of sufficient duration that the firm can change both inputs. Since in the long run the firm can rid itself of all of its inputs, it is able to go out of business, that is, reduce its output and both labor and capital inputs to zero, if it chooses.

The short run is a period of time of insufficient duration for the firm to change its capital input. It takes time to order a new machine, to receive it, to install it, and integrate it into the production process. And this cannot be done except in a long enough period of time. By contrast, the quantity of labor input can be changed quickly. Apart from whatever training is required, it is only necessary to change an employee's labor hours, hire new employees, or let existing employees go. The short run, then, is defined as a period of time during which only the quantity labor can be varied. Capital necessarily remains fixed. When the firm is not able to change its quantity of capital, it cannot reduce that quantity to zero and cannot, therefore, go out of business. Thus, although the firm can shut down by eliminating all of its labor input and producing no output in the short run, it still cannot go out of business.

Since both inputs can be varied, the production function described above is a long-run production function. The short-run production function is secured from the long-run production function by fixing capital at, say, $\bar{k}$:

$$x = f(\ell, \bar{k}).$$

[1] The word 'maximum' is important in this definition. Since it will be assumed later on that the firm will make input and output decisions on the basis of profit maximization, if the firm is not obtaining maximum output from the inputs it hires, it cannot be maximizing its profit.

Sometimes the short-run production function is abbreviated to $x = f(\ell)$. Here $\bar{k}$ is implicit in the functional symbol f. In either case, $\bar{k}$ can be taken to refer to the size of the firm. A geometric example appears in Figure 9-6 (recall Figure 2-5 in Chapter 2). The graph of the long-run

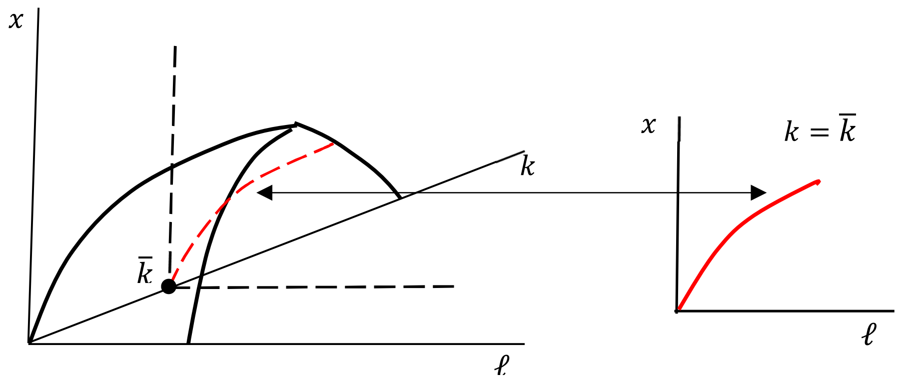


Figure 9-6

production function is drawn with solid curved lines in three dimensions in the left-hand diagram. Also in that diagram, the graph of the short-run production function is that portion of the long-run production function appearing as a red dashed curve in the two-dimensional dashed plane parallel to the $\ell - x$ plane and drawn for capital fixed at $\bar{k}$. The latter plane is reproduced in the right-hand diagram. The graph of the short-run production function with $k = \bar{k}$ is reproduced as the solid red curve in that diagram. With respect to the left-hand diagram, one can think of the horizontally dashed line as the ℓ-axis and the vertically dashed line as the x-axis for the graph of the short-run production function. The axes are labeled this way in the right-hand diagram.

It will be convenient in subsequent chapters to occasionally employ the term ‘production function’ without its long- or short-run designation. Context will determine the time period that applies.

Chapter 10
Production Functions, Isoquants, and Ridge Lines

Consider the long-run production function $x = f(\ell, k)$. The analogue of the indifference curve through (x^0, y^0) from the model of consumer buying behavior is the isoquant through (ℓ^0, k^0). It is defined as the collection of all baskets (ℓ, k) yielding the same output as (ℓ^0, k^0):

$$\{(\ell, k): f(\ell, k) = f(\ell^0, k^0) \text{ and } \ell \geq 0, k \geq 0\},$$

or, with $f(\ell^0, k^0) = x^0$,

$$\{(\ell, k): f(\ell, k) = x^0 \text{ and } \ell \geq 0, k \geq 0\}.$$

Assumptions to be made later will allow for the possibility of graphing isoquants as downward sloping and strictly convex curves like indifference curves.[1] For now, the graphs of isoquants may be thought of in these terms. As with the utility function and indifference curves, the isoquant corresponding to output x^0 can be described as a function of a single variable

$$k = V(\ell)$$

obtained by solving $f(\ell, k) = x^0$ for k as a function of ℓ. For example, if $f(\ell, k) = \ell k$, then the x^0-isoquant is $x^0 = \ell k$ or

$$k = \frac{x^0}{\ell}.$$

The collection of all isoquants is called the isoquant map.

Having only two inputs, the long-run production function with one input held fixed is referred to as a total product function. There are two of them, one with, say, $k = k^0$ and the other with $\ell = \ell^0$:

$$TP^\ell(\ell) = f(\ell, k^0) \quad \text{and} \quad TP^k(k) = f(\ell^0, k), \tag{10.1}$$

defined for all $\ell \geq 0$ and $k \geq 0$ respectively. The one on the left has been referred to as the short-run production function (Chapter 9). Each function generates an average product (per unit of input) function

$$AP^\ell(\ell) = \frac{TP^\ell(\ell)}{\ell} \quad \text{and} \quad AP^k(k) = \frac{TP^k(k)}{k}, \tag{10.2}$$

[1] The assumptions will allow for more complex possibilities as well.

also defined, respectively, for all $\ell > 0$ and $k > 0$, and a marginal product function on the same domain. Marginal product functions correspond to marginal utility functions as can be seen by writing out their definitions and substituting from (10.1). That is, at (ℓ^0, k^0), the marginal product with respect to labor ℓ in approximate form is[2]

$$MP^\ell(\ell^0) = \frac{TP^\ell(\ell^0 + \Delta\ell) - TP^\ell(\ell^0)}{\Delta\ell} = \frac{f(\ell^0 + \Delta\ell, k^0) - f(\ell^0, k^0)}{\Delta\ell},$$

where k^0 is implicit in the symbolism $MP^\ell(\ell^0)$. That with respect to capital k is

$$MP^k(k^0) = \frac{TP^k(k^0 + \Delta k) - TP^k(k^0)}{\Delta k} = \frac{f(\ell^0, k^0 + \Delta k) - f(\ell^0, k^0)}{\Delta k}.$$

Continuing analogously with the model of consumer buying behavior, the slope of the isoquant through (ℓ^0, k^0) at ℓ^0 (the analogue of the slope of the indifference curve through (x^0, y^0) at x^0) is

$$\frac{\Delta k}{\Delta\ell} = \frac{V(\ell^0 + \Delta\ell) - V\ell^0)}{\Delta\ell},$$

and a demonstration parallel to that in the case of the consumer (Chapter 5) shows that

$$\frac{\Delta k}{\Delta\ell} = -\frac{MP^\ell(\ell^0)}{MP^k(k^0)}. \tag{10.3}$$

Thus, where both marginal products are positive, isoquants slope downward. The ratio of marginal products (without the minus sign) or the negative of the slope of the isoquant is called the marginal rate of technical substitution.

For now, only three assumptions, two of which relate to the properties of the long-run production function, are imposed. The remaining assumptions will be introduced later. (The full list of assumptions for the model of firm behavior appears in Supplemental Note D). Assume

1. the firm has a long-run production function $x = f(\ell, k)$ based on a given technology,

and

2. the long-run production function has the following properties:
   - 2a. Zero input produces zero output ($f(0,0) = 0$), and non-negative input produces non-negative output ($f(\ell, k) \geq 0$ for all $(\ell, k) \geq 0$).
   - 2b. It is continuous and all marginal products can be calculated.

---

[2] From here on, the derivative form for marginals obtained, for example, by letting $\Delta\ell$ and Δk in the equations below go to zero will largely be ignored.

Consider now the geometry of the graphs of total, average, and marginal product functions. The shape of the total product curve, which emerges from the graph of the long-run production function as described in Figures 2-5 and 2-6 of Chapter 2, determines the shape of the average and marginal product curves. A common shape for the former[3] appears in Figure 10-1 in which a total product curve passing through points A, C, and B appears in the top diagram.

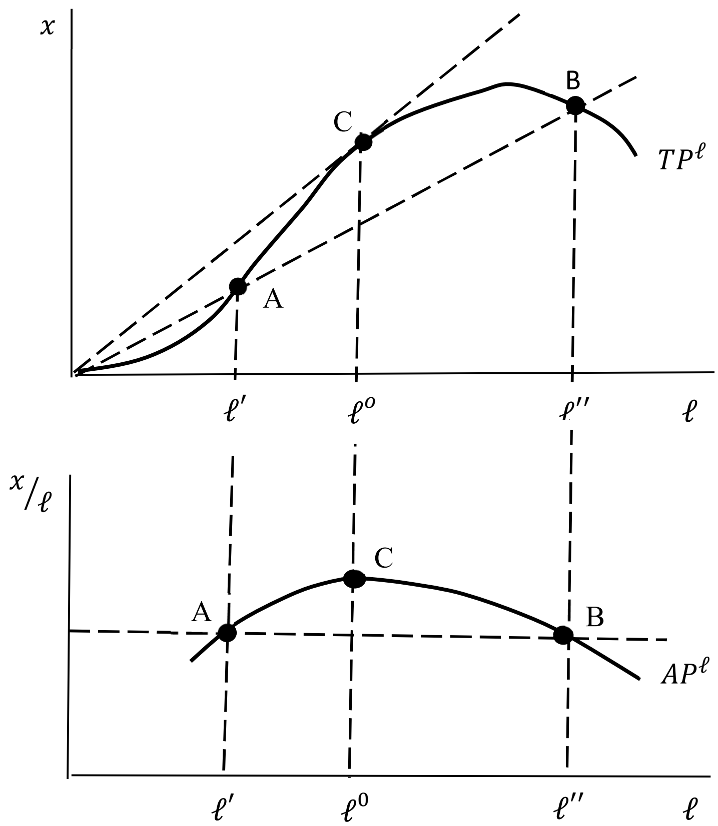


Figure 10-1

Observe that the slope of the straight (dashed) line connecting A and B is the same whether calculated at either point. For example, at A it is the vertical distance from that point to the ℓ-axis at ℓ' or $TP^\ell(\ell')$ divided by the length from the origin to ℓ'. But according to equation (10.2), this is the same as the average product $AP^\ell(\ell')$. Similarly at point B. Since the calculations of the slope of the straight line at A and B yield the same result,

$$AP^\ell(\ell') = AP^\ell(\ell'').$$

This is indicated along the horizontal dashed line through A and B in the graph of the average product curve in the lower diagram of Figure 10-1. As the straight line through A and B in the

[3] The shape of the total product curve in Figure 10-1 is derived from a production function whose graph has a different shape from that of Figure 2-5 of Chapter 2. The latter, recall, is obtained from the strictly concave production function pictured in Figure 2-3.

top diagram becomes steeper and its slope increases, the two values of $AP^{\ell}(\ell)$ rise. When that slope rises to the point at which the line is tangent to the total product curve at C, the average product is at its largest value. This maximum appears at point C for $\ell = \ell^0$ in the lower diagram. Therefore, to determine the value of ℓ at which $AP^{\ell}(\ell)$ has a maximum, locate the point of tangency between the total product curve and a straight line from the origin.

In Figure 10-2, graphs of the total product and average product curves are drawn in the same diagram along with the marginal product curve. The vertical axis, then, although labeled only with x for convenience, has three separate scales – one for each graph – imposed upon it. Remember that the marginal product curve reflects the slopes of the total product curve

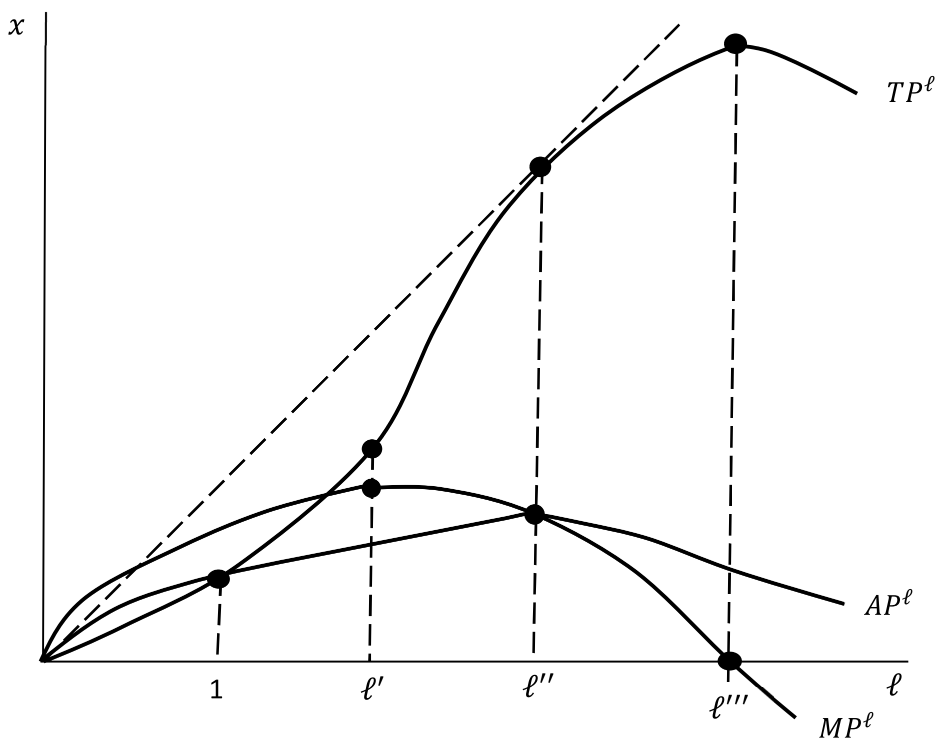


Figure 10-2

as ℓ varies. Thus, in the diagram, as ℓ increases from the origin, the marginal product (slope of the total product curve) increases up to a maximum at $\ell = \ell'$ and declines thereafter. At $\ell = \ell'''$ the marginal product is zero (reflecting the maximum at that point on the total product curve) and crosses the ℓ-axis into negative territory.

Also in Figure 10-2, $AP^{\ell}(\ell)$ has a maximum at $\ell = \ell''$ where the straight line from the origin is tangent to the total product curve. Since equation (10.2) implies that the average product and total product are equal at $\ell = 1$, their curves are drawn to intersect at that value. Furthermore, the marginal product curve intersects the average product curve where the latter has its maximum. This happens for the following reason: Imagine that an average value has been computed from a set of values by totaling the values and dividing by their number. If a new value is added to the set that is larger than the average, the average rises. If it is smaller, the average falls. In the present case, at any ℓ the marginal product $MP^{\ell}(\ell)$ is the extra amount that is being added into the total product $TP^{\ell}(\ell)$ at ℓ. According to Figure 10-2, for an ℓ to the left of ℓ'', the amount $MP^{\ell}(\ell)$ added to the total $TP^{\ell}(\ell)$ is greater than the average $AP^{\ell}(\ell)$. So, since the total product is increased by more than the average, the average must be rising. At an ℓ to the right of ℓ'', the marginal product is less than the average. In that case, when calculating the average from the total, the average must be falling. At $\ell = \ell''$, the marginal and average products are equal. This means there is no change in the average from including the marginal in the total, and the average has a maximum there.

When all total product curves have the shapes as drawn in Figures 10-1 and 10-2, the geometric relationship between these curves and isoquants appears in the three-dimensional diagram of Figure 10-3. In that diagram, everything in the ℓ- k plane (i.e., the input space) is in red; everything Above that plane is in black. The total-product-with-respect-to-labor $TP^{\ell}(\ell)$

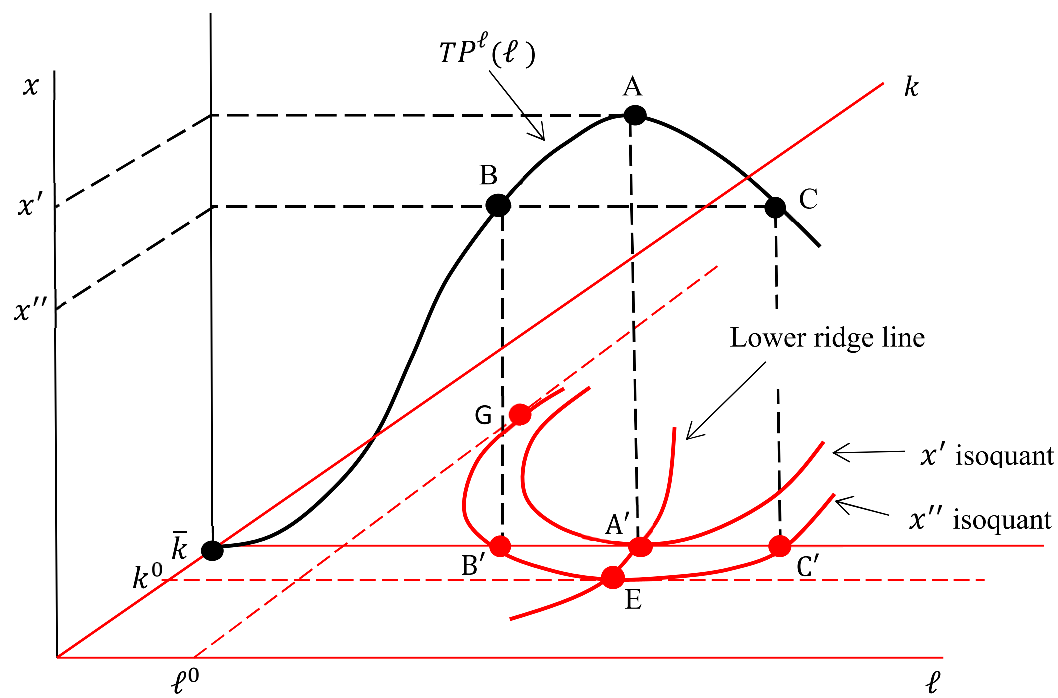


Figure 10-3

curve for the fixed amount of capital $\bar{k}$ (identified on the k axis) is drawn in the plane at $\bar{k}$ that is parallel to the ℓ-x plane. Output x'', which appears at points B and C on the $TP^{\ell}(\ell)$ curve, can be produced with either input basket B′ or C′ located in the input space on the solid red line coming out from $\bar{k}$ and parallel to the ℓ -axis. Since baskets B′ and C′ yield the same output, they lie on the same x'' isoquant. Output x', which is greater than output x'' and appears at point A on the $TP^{\ell}(\ell)$ curve, is produced at only one input basket A′ on the solid red line from $\bar{k}$. Since output is larger with basket A′ than with B′ or C′, it follows that the x' isoquant on which A′ lies is farther out from the origin than the x'' isoquant. The x' isoquant is tangent at A′ to the solid red line parallel to the ℓ-axis from $\bar{k}$ because there is only one basket on that line yielding output x' and all other baskets on that line lie on isoquants associated with smaller outputs.

As ℓ increases from the k-axis along the $\bar{k}$-line in Figure 10-3, output increases up to its maximum over basket A′ (point A on the $TP^{\ell}(\ell\,)$ curve) and declines thereafter. At basket A′, then, the marginal product $MP^{\ell}(\ell\,) = 0$. Were these curves redrawn for, say $k^0 < \bar{k}$, then there would be a basket E at which the x'' isoquant is tangent to the dashed red line emanating from k^0 and parallel to the ℓ-axis. The $TP^{\ell}(\ell\,)$ curve for $k = k^0$ would have a maximum over basket E and its corresponding $MP^{\ell}(\ell\,)$ would be zero there. Putting all such tangency points together yields the lower ridge line which is the collection of all baskets like A′ and E in the red ℓ- k plane that correspond to outputs for which a $TP^{\ell}(\ell\,)$ is maximal and for which the associated $MP^{\ell}(\ell\,) = 0$.

Continuing in Figure 10-3, but with ℓ now fixed at ℓ^0 on the ℓ-axis and with k increasing from that ℓ^0 along the dotted red line parallel to the k-axis, a $TP^k(k)$ curve above the ℓ- k plane plane could be traced out from the graph of the long-run production function. Since that total product curve (not drawn in the diagram) has been assumed to have a shape similar to that of $TP^{\ell}(\ell\,)$, it achieves maximum output and $MP^k(k) = 0$ at a basket, call it G, that, as drawn in the diagram, lies on the x'' isoquant where that isoquant is tangent to the line parallel to the k-axis through ℓ^0. The upper ridge line (also not drawn in the diagram) consists of all such baskets in the input space, that is, all baskets with a $TP^k(k)$ maximal and corresponding $MP^k(k) = 0$.

Taking the red ℓ- k plane and its contents out of Figure 10-3 and reproducing it in two dimensions by itself, yields a picture like that of Figure 10-4. The latter contains the upper

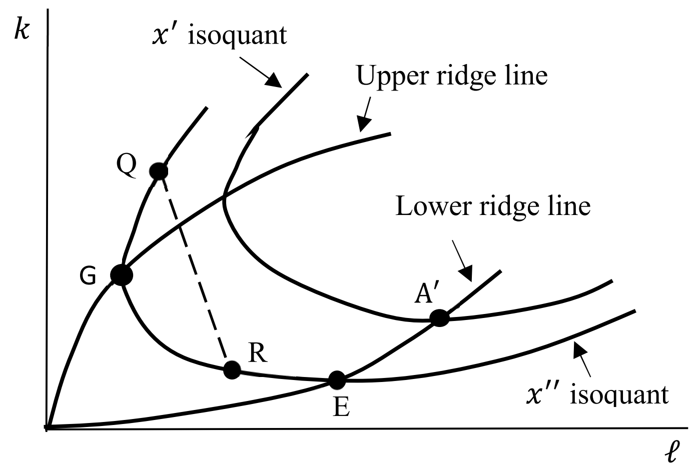


Figure 10-4

ridge line not drawn in Figure 10-3. Also, the straight lines parallel to the axes and tangent to the isoquants at A′, E. and G in Figure 10-3 are not shown in Figure 10-4. Ridge lines may or may not bend enough to intersect. The geometry of Figure 10-3 indicates that throughout the area between the ridge lines in Figure 10-4, both $MP^{\ell}(\ell) > 0$ and $MP^{k}(k) > 0$. This is the only region in the input space where both marginal products are positive. It follows from equation (10.3) that, in that region, isoquants slope downward. The geometry of the diagram also suggests that they are strictly convex there.

Notice that in Figure 10-4, were the firm to consider producing output x'' with input basket Q lying outside the area between the ridge lines (but still on the x'' isoquant), it would see that it could produce the same output using basket R between the ridge lines with less labor and less capital. Since later on (Chapter 13) it will be assumed that the firm operates so as to maximize its profit, it will avoid such wasteful and costly input baskets. For that reason, the area between the ridge lines up to an intersection point if there is one is sometimes referred to as the relevant region of the input space. Thus, to produce any output, the firm will always employ a basket of inputs that lies within the relevant region.

It should also be pointed out that there are long-run production functions that do not have ridge lines. Their total product curves do not exhibit the shapes drawn in previous diagrams and are such that there are no values of ℓ and k for which $MP^{\ell}(\ell) = 0$ and $MP^{k}{}_{k}(k) = 0$. For example, the production function with graph depicted in Figure 10-5 has positive marginal

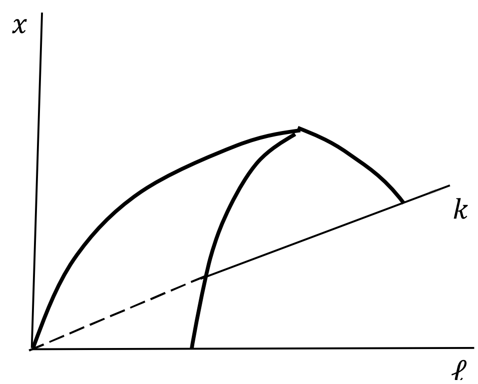


Figure 10-5

products everywhere. Its total product curves for fixed values of k (recall Figure 2-5 from Chapter 2 and Figure 9-5 from Chapter 9) and isoquants (recall Figure 2-4 from Chapter 2) appear, respectively, for $k = k^0$ and $x = x^0$ in Figure 10-6. Here the relevant region of the input space(appearing in the right-hand diagram) is the entire positive or north-east quadrant (excluding the coordinate axes) of the ℓ- k plane. The total product curves for fixed values of ℓ are similar to those for fixed values of k (Figure 2-6 in Chapter 2).

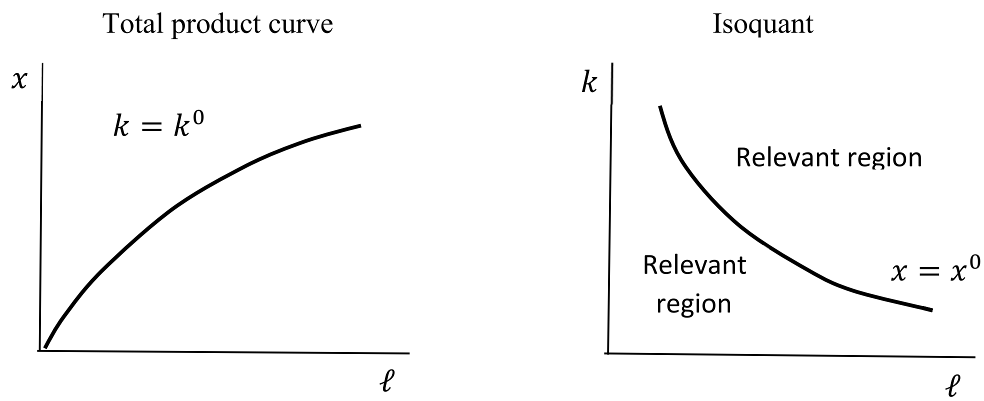


Figure 10-6

In subsequent chapters, isoquants will sometimes be drawn as in the right-hand diagram of Figure 10-6 rather than as in Figures 10-3 and 10-4, thereby ignoring the possibility of ridge lines. This is done to make the drawings simpler and easier to understand. But it should be remembered that the presence of ridge lines could easily be indicated in such diagrams by bending the ends of the isoquants so as to conform to the shapes of those appearing in previous diagrams.

With these ideas in mind, the following assumptions regarding the properties of the long-run production function are added to those stated above:

2c. If ridge lines exist, all marginal products are positive and all isoquants are strictly convex between the ridge lines up to an intersection point if there is one.

2d. If ridge lines do not exist, all marginal products are positive and all isoquants are strictly convex everywhere throughout the input space, and no isoquant touches the co-ordinate axes.

These two assumptions specify properties of the production function for input baskets in the relevant region of the input space. As previously noted, they imply that in that region, isoquants slope downward. Recall that the full list of assumptions for the model of firm behavior may be found in Supplemental Note D.

Chapter 11
Returns to Fixed Factors and Scale, Cost Minimization, Expansion Paths, and Long-Run Cost Functions

Recall that for capital fixed at k^0, the total product function with respect to labor is given by $TP^\ell(\ell) = f(\ell, k^0)$. Now $TP^\ell(\ell)$ is said to have $\left\{\begin{matrix}\text{increasing}\\ \text{constant}\\ \text{decreasing}\end{matrix}\right\}$ (marginal) returns at ℓ^0 to the fixed factor k^0 according as $MP^\ell(\ell)$ is $\left\{\begin{matrix}\text{increasing}\\ \text{constant}\\ \text{decreasing}\end{matrix}\right\}$ at ℓ^0. Two possibilities are illustrated in Figure 11-1. In that diagram, increasing returns are present for $\ell^0 < \bar{\ell}$ and decreasing returns prevail for $\ell^0 > \bar{\ell}$.

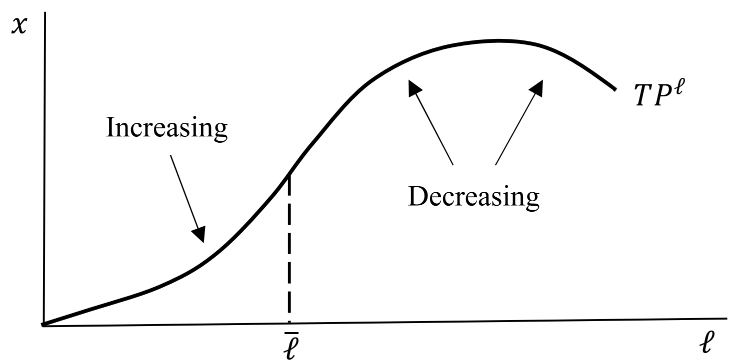


Figure 11-1

The law of diminishing returns is the statement that all total product functions derived from any real-world production function eventually exhibits decreasing returns to the fixed factors as a variable input increases. That this law would seems to hold more universally across space and time than other laws in economics and therefore be more like laws in the physical sciences, is is a consequence of what is called the "crowding effect:" As more and more labor is added to a fixed amount of capital, cooperation may initially increase output at an increasing rate. But eventually, the increased labor will become sufficiently crowded in working with the still fixed amount of capital as to interfere with the efficiency of production, and the increases in output from adding the additional units of labor will start to decline.

While returns to fixed factors involve variations in a single input, returns to scale requires variations in all inputs. With both inputs variable, then, the long-run production function $f(\ell, k)$ exhibits $\left\{\begin{matrix}\text{increasing}\\ \text{constant}\\ \text{decreasing}\end{matrix}\right\}$ returns to scale according as

$$f(\alpha\ell, \alpha k)\begin{Bmatrix} > \\ = \\ < \end{Bmatrix}\alpha f(\ell, k),$$

for all $(\ell, k) > 0$ and $\alpha > 1$. Note that $f(\alpha\ell, \alpha k)$ is the output that results when all inputs are multiplied by α, while $\alpha f(\ell, k)$ is the output that results upon multiplying the original output at (ℓ, k) by α. Thus, under increasing returns to scale, doubling all inputs more than doubles output, and with constant returns to scale, doubling all inputs exactly doubles output. When decreasing returns to scale are present, doubling all inputs less than doubles output. An illustration of a constant returns to scale production function is $f(\ell, k) = \sqrt{\ell k}$. The constant returns to scale property is satisfied since, for example with $\alpha = 2$,

$$f(2\ell, 2k) = \sqrt{2\ell 2k} = \sqrt{4\ell k} = 2\sqrt{\ell k} = 2f(\ell, k).$$

Having dealt with the production function and its properties, the next step in explaining firm buying and selling behavior in the context of the Walrasian model is to use that function along with input prices to examine firm long-run costs. For the time being, the analogy between the explanation of consumer buying behavior and that of firm behavior will continue.

------------------------ -----------------------------------

Let p_ℓ and p_k represent the prices per unit of inputs ℓ and k respectively and let c denote input cost. Then the cost of the input basket (ℓ, k) is

$$c = \ell p_\ell + k p_k. \tag{11.1}$$

Use the symbol π for profit and recall that the price per unit of output x is p_x. Since profit is the difference between sales revenue and input costs, $\pi = xp_x - c$ or

$$\pi = xp_x - (\ell p_\ell + kp_k), \tag{11.2}$$

where (ℓ, k) is the basket of inputs used to produce output x. The assumption that the firm will hire inputs and produce output so as to maximize profit will be introduced later on Chapter 13. But for now, observe that from equation (11.2), if the firm knows the prices of its inputs and output along with the quantity of output it will produce, then profit maximization requires that the firm hire inputs so as to minimize their cost. That is, given its fixed output quantity x, the firm should choose ℓ and k to minimize $\ell p_\ell + kp_k$. This minimization problem needs to be considered first.

The equation $c = \ell p_\ell + kp_k$ is very similar to the consumer budget constraint $m = xp_x + yp_y$ in the model of consumer buying behavior. The graph of the former is called an iso-cost line and its geometry in Figure 11-2 is identical, except for the names of the symbols, to that of $m = xp_x + yp_y$ (Figure 6-1 of Chapter 6). Thus, the slope of the iso-cost line is the negative of the input-price ratio ${}^{p_\ell}/_{p_k}$ while its ℓ- and k- intercepts are given by $\ell = {}^{c}/_{p_\ell}$ and $k = {}^{c}/_{p_k}$ respectively. And just as variations in income with commodity prices fixed shift the consumer's budget constraint in a parallel manner (that is, without changing its slope), so do variations in c

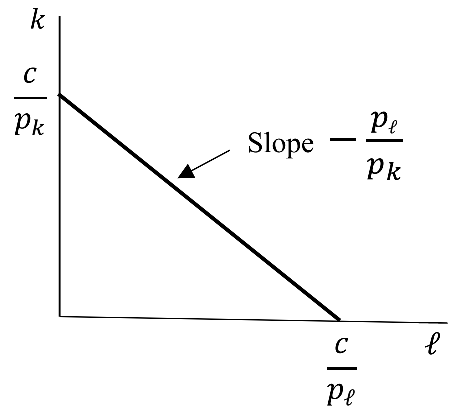


Figure 11-2

with input prices fixed shift the iso-cost line in the same way. For example, as c becomes smaller, the iso-cost line shifts back towards the origin in parallel fashion.

Turning to cost minimization as described above, let output be fixed at x^0. This determines the collection of all input baskets that can be employed produce output x^0 or the x^0 isoquant. Geometrically arguing in an analogous fashion to utility maximization subject to the budget constraint (Chapter 6), the input basket on the x^0 isoquant that minimizes cost of producing x^0 is the one on the lowest possible iso-cost line that intersects or meets the x^0 isoquant. This basket, (ℓ^0, k^0) as shown in Figure 11-3, appears at the tangency between the x^0

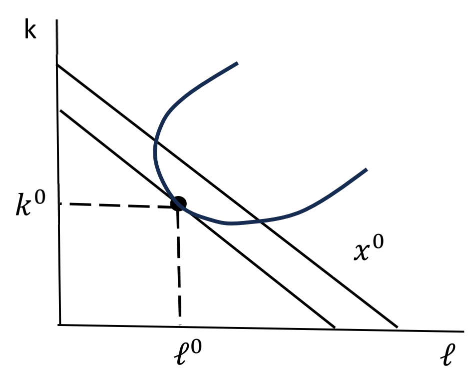


Figure 11-3

isoquant and an iso-cost line. It is called the cost-minimizing basket, the least-cost combination of inputs, or the optimal input combination for output x^0 given p_ℓ and p_k.

Apart from the shape of the isoquant (which, here, has upward sloping parts), the tangency in Figure 11-3 looks exactly like that depicting utility maximization subject to the budget constraint (Figure 6-2 in Chapter 6) except that the names of the variables are different, and the curve and the straight line has different meanings. In the present case, the tangency is said to be the outcome of (input) expenditure or cost minimization subject to an isoquant constraint or a fixed level of output. Thus it is not surprising that, at least between the ridge lines, many of the same assumptions are made here (they were set out in Chapter 10 and are listed in Supplemental Note D) as in the model of consumer buying behavior.

Analogously to the explanation of consumer buying behavior (equations (6.2) and (6.3) in Chapter 6), the statement derived from Figure 11-3 that, at the tangency (ℓ^0, k^0), the slope of the iso-cost line (the negative of the price ratio) equals the slope of the isoquant (the negative of the marginal rate of technical substitution) is part of the first-order condition for minimizing input cost subject to a fixed level of output. Mathematically and cancelling the minus signs, that condition is

$$\frac{MP^{\ell}(\ell^0)}{MP^{k}(k^0)} = \frac{p_\ell}{p_k}, \tag{11.3}$$

or equivalently,

$$\frac{MP^{\ell}(\ell^0)}{p_\ell} = \frac{MP^{k}(k^0)}{p_k}.$$

If the latter equality did not hold, say the marginal product per dollar spent on labor were larger than the marginal product per dollar spent on capital, then transferring a dollar spent on capital to labor would keep costs the same but expand output. But if output can be increased without expanding costs, then the original output could be produced at lower cost. In other words, the cost of producing the original output is not minimized.

As with utility maximization subject to the budget constraint, equation (11.3), although stating that two slopes are equal, is insufficient by itself to fully describe the tangency. To obtain a complete description it is necessary to add the equation of the isoquant constraint evaluated at S(ℓ^0, k^0), that is, $x^0 = f(\ell^0, k^0)$. The latter, together with equation (11.3) form the full set of first-order conditions for cost minimization subject to the fixed level of output x^0.

It has been assumed (Chapter 10) that if ridge lines exist, all marginal products are positive (implying that isoquants are everywhere downward sloping) and all isoquants are strictly convex between the ridge lines up to an intersection point if there is one. When intersecting ridge lines are present, it turns out that isoquants are closed curves and there are two tangencies, and hence two input baskets, such as A and B in Figure 11-4 where the first-order condition (11.3) is satisfied. In that case, the tangency beyond the intersection of the ridge lines at B (that is, outside the relevant region of the input space) occurs where the cost of producing output x^0 is maximized, both marginal products are negative, and the isoquant is strictly concave – not strictly convex. Since, as indicated in Chapter 10, attention is confined to the relevant region of the input space, this possibility is eliminated. In parallel with the explanation of consumer buying behavior given earlier, the assumptions of positivity of marginal products and strict convexity of

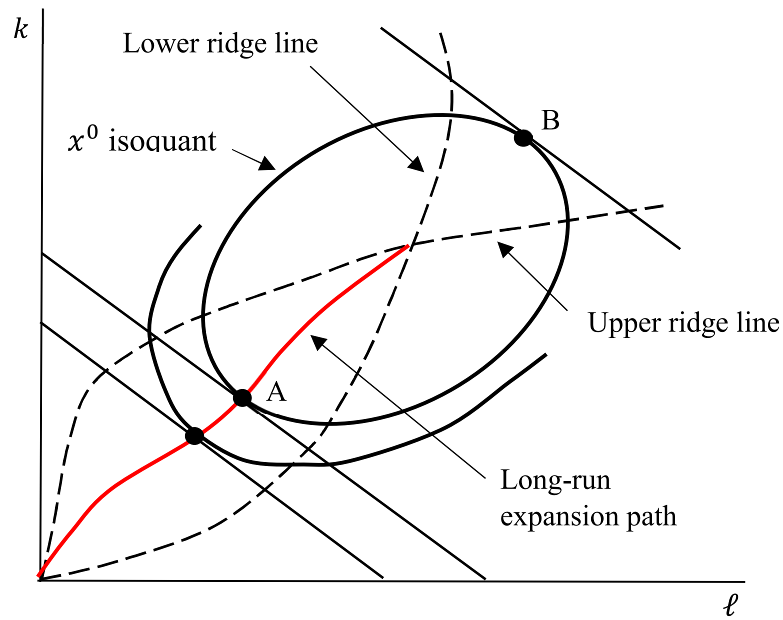


Figure 11-4

isoquants play the role of second-order conditions ensuring that first-order conditions arising at baskets like A lie in the relevant region of the input space are associated with minimum cost.

The firm's expansion path is the curve in the input space that identifies the input baskets the firm employs as it increases its output with constant input prices. Thus, in the long-run, the expansion path, pictured in red in Figure 11-4, is the collection of all cost-minimizing input baskets in the relevant region as output expands with input prices remaining fixed. It is the analogue of the income consumption curve in the model of the consumer's buying behavior. The long-run expansion path necessarily starts at the origin where the firm produces no output and employs no inputs. In the short run, the firm has a fixed amount of capital, say $\bar{k}$, that it is unable to change. So as output expands, say from x'' to x', and the firm moves to a higher isoquant, it must follow the red straight line parallel to the ℓ-axis at $\bar{k}$ as pictured in Figure 11-5. That is the

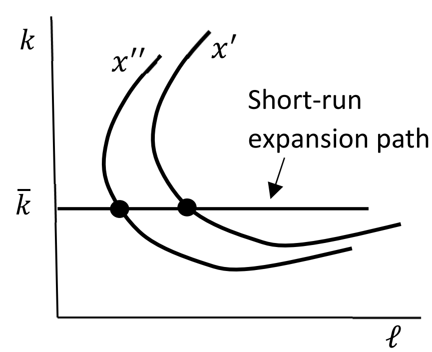


Figure 11-5

firm's short-run expansion path. As input prices vary so as to change the input price ratio, the slope of the iso-cost lines alter as do the locations of the tangencies between them and the isoquants. This causes the long-run expansion path to shift. But changes in the input price ratio have no effect on the short-run expansion path which is determined solely by the fixed amount of capital held by the firm.

------------------------ -----------------------------------

Whereas constrained utility maximization subject to the budget constraint in the model of consumer buying behavior is the basis for defining consumer demand functions, here constrained cost minimization (minimizing cost for a fixed level of output) is used to characterize long-run total cost as a function of output. (Previously cost without the qualifying adjective 'total' has been presented in equation (11.1) as a function of the input basket.) This is accomplished by invoking that part of the assumption of profit maximization that requires the firm to produce each output quantity with the optimal input combination. And it is here that the close analogy between the explanation of consumer buying behavior and the explanation of firm buying and selling behavior ends.

Let p_ℓ and p_k be given. The long-run total cost for producing an output is the cost of the cost-minimizing input basket for producing that output. Using this idea, the long-run total cost function $LRTC(x)$ for producing output quantity x is characterized by

$LRTC(x) = c$ if and only if $c = \ell^0 p_\ell + k^0 p_k$, where (ℓ^0, k^0) is the least cost combination of inputs for producing x given p_ℓ, p_k, and the production function f.

Thus, the long-run total cost function $LRTC(x) = \ell^0 p_\ell + k^0 p_k$, where ℓ^0 and k^0 are determined by a tangency like that of Figure 11-3, is generated by repeated specification of x with p_ℓ and p_k held constant. It is implicit in this definition that, according to the assumptions that have been made, the area outside the relevant region of the input space has been eliminated from consideration. Geometrically, to construct the graph of the long-run total cost function, first choose an output, say x^0. That specification determines the x^0isoquant in the input space pictured in the right-hand diagram of Figure 11-6, where the isoquant is drawn without suggesting the possibility of ridge lines. Since p_ℓ and p_k are fixed, the slopes of all iso-cost lines $c = \ell p_\ell + k p_k$, which are the same as c varies, are also determined. The tangency between the x^0 isoquant and an iso-cost line occurs at (ℓ^0, k^0). Plotting $\ell^0 p_\ell + k^0 p_k$ against x^0 in the left-

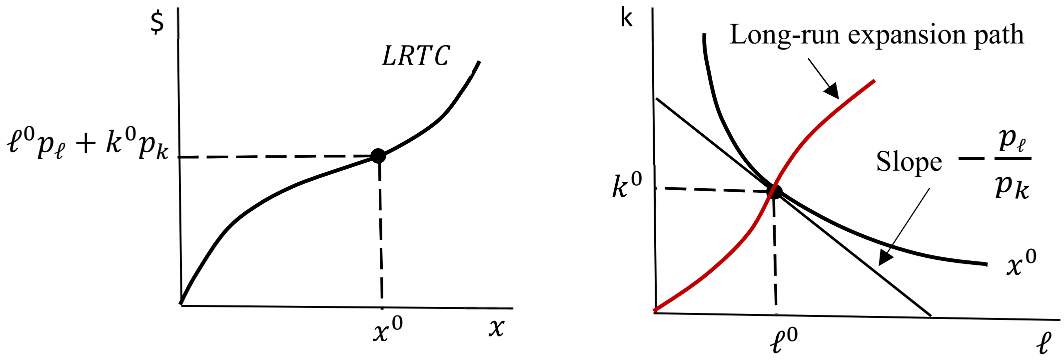


Figure 11-6

hand diagram gives one point on the long run total cost curve. Continuing in this manner for different values of x with p_ℓ and p_k unchanged generates the $LRTC$ curve. That curve starts at the origin because the expansion path in the input space starts at the origin. When the firm uses no inputs, it has no costs.

Note that long-run total costs are defined in terms of input baskets along the expansion path. A movement along one curve implies a corresponding movement along the other. Each change in input prices that alters the input price ratio not only modifies the expansion path (as indicated above), but also changes the long-run total cost function.

Two additional functions are based on the long-run total cost function. The long-run average cost function is given by

$$LRAC(x) = \frac{LRTC(x)}{x}, \tag{11.4}$$

and the long-run marginal cost function in approximate form appears as

$$LRMC(x) = \frac{LRTC(x + \Delta x) - LRTC(x)}{\Delta x}. \tag{11.5}$$

These functions are illustrated geometrically in Figure 11-7 where the shape of the long-run total cost curve in the upper diagram is taken to be that depicted in the left-hand diagram of Figure 11-6. As with total, average, and marginal product curves, the shape of the long-run total cost curve determines the shapes of the long-run average and long-run marginal cost curves. In the lower diagram of Figure 11-7 the long-run average cost curve and the long run marginal cost curve appear together with the separate scales for each recorded on the same vertical axis. Observe that if the curves pictured in Figure 11-7 were turned upside down, they would look quite similar to

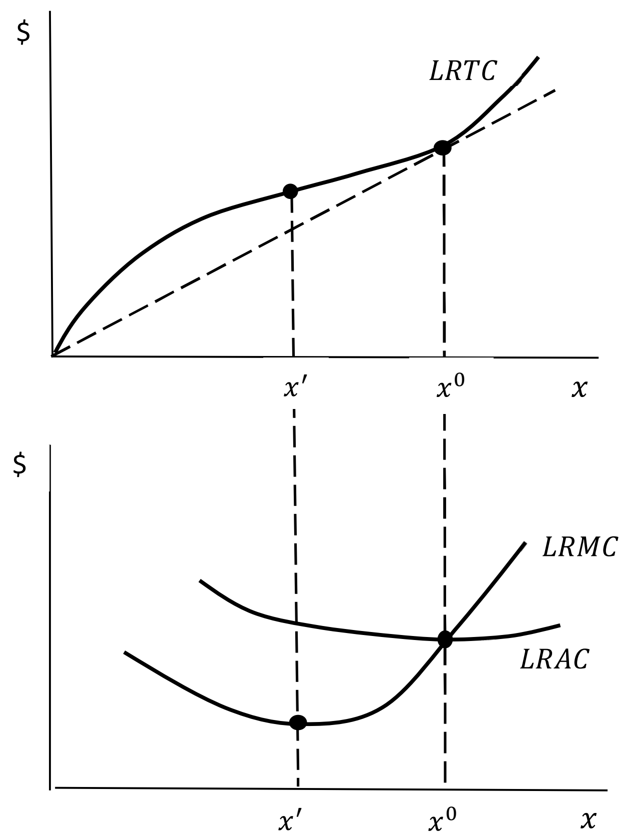


Figure 11-7

the curves drawn in Figure 10-2 of Chapter 10 for total, average, and marginal product curves. As such, the properties of the former parallel those of the latter with appropriate adjustments made to accommodate the "upside-downness" present. Subject to that adjustment, the justifications of the following characteristics of the graphs in Figure 11-7 are the same as those for the total, average, and marginal product case:

1. The graph of $LRMC(x)$ has a minimum at x' where the slope of the graph of $LRTC(x)$ changes from decreasing to increasing.
2. The graph of $LRAC(x)$ has a minimum at x^0 where a straight line from the origin is tangent to the graph of $LRTC(x)$.
3. The graph of $LRMC(x)$ intersects that of $LRAC(x)$ at x^0 where $LRAC(x)$ has its minimum.
4. $LRAC(x) = LRTC(x)$ where $x = 1$ (not shown in the diagram).

Chapter 12
Long- and Short-Run Cost Functions and the Relationship between Them

The long-run average cost curve exhibits $\left\{\begin{matrix}\text{economies}\\ \text{constant econnomies}\\ \text{diseconomies}\end{matrix}\right\}$ of scale where it is $\left\{\begin{matrix}\text{declining}\\ \text{constant}\\ \text{risng}\end{matrix}\right\}$. For example, to the left of $\bar{x}$ in Figure 12-1, economies are present; to the right

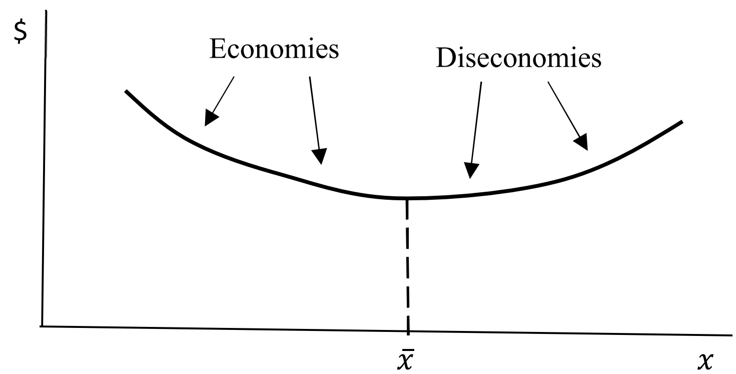


Figure 12-1

diseconomies prevail. Economies and diseconomies of scale in long-run average costs sometimes emerge from returns to scale in the production function.

If the production function has $\left\{\begin{matrix}\text{increasing}\\ \text{constant}\\ \text{decreasing}\end{matrix}\right\}$ returns to scale, and if the expansion path is a straight line from the origin, then the long-run average cost curve necessarily exhibits $\left\{\begin{matrix}\text{economies}\\ \text{constant econnomies}\\ \text{diseconomies}\end{matrix}\right\}$ of scale. Although the reasoning behind this statement will not be pursued here, it may be roughly illustrated by observing that the long-run average cost associated with output x is, from Chapter 11,

$$LRAC(x) = \frac{\ell p_\ell + k p_k}{x}, \tag{12.1}$$

where (ℓ, k) is the least-cost combination of inputs (on the expansion path) for producing x given p_ℓ and p_k. If input quantities were to double, say, the values of ℓ and k, and hence the numerator in equation (12.1), would be multiplied by 2 and the new value of long-run average cost would depend on what happens to output quantity. If the output quantity were to double (constant returns to scale), then $2x$ would appear in the denominator, the 2s would cancel, and there would be no change in the long-run average cost (constant economics of scale):

$$LRAC(2x) = \frac{2\ell p_\ell + 2kp_k}{2x} = \frac{\ell p_\ell + kp_k}{x} = LRAC(x).$$

If the output quantity were to more than double (increasing returns to scale), then a number greater than $2x$ would appear in the denominator and the long-run average cost would fall (economies of scale). And if the output quantity were to less than double (decreasing returns to scale), then a number less than $2x$ would appear in the denominator and the long-run average cost would rise (diseconomies of scale).

Attention now turns to the short run where, recall, capital is fixed at, say, $k = \bar{k}$ regardless of the output produced by the firm. Labor remains variable. This distinction between fixed and variable inputs leads to a separation of production costs into fixed and variable parts. The total fixed cost of producing output quantity x is the cost of the fixed input $\bar{k}$ used to produce that output. The total fixed cost function,

$$TFC(x) = \bar{k}p_k,$$

is therefore constant since the price of capital p_k is fixed by the capital market and $\bar{k}$ does not change with x. As described in Chapter 9, $\bar{k}$ is taken to indicate the size of the firm.

The total variable cost of producing x is the cost of the variable input (labor) used to produce it. The total variable cost function is defined by

$$TVC(x) = \ell^0 p_\ell,$$

where, as indicated in Figure 12-2, ℓ^0 is the only way of producing x along the x isoquant since

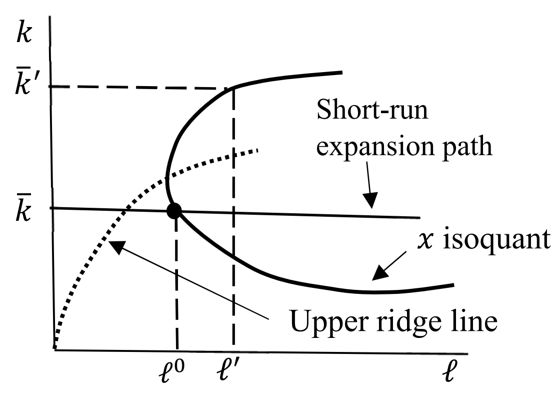


Figure 12-2

k is fixed at $\bar{k}$. Note that $(\ell^0, \bar{k})$ necessarily lies on the firm's short-run expansion path (defined in Chapter 11).[1]

The short-run total cost function for a firm of size $\bar{k}$, symbolized by $SRTC^{\bar{k}}(x)$, is the sum of the total variable and total fixed cost functions:

$$SRTC^{\bar{k}}(x) = TVC(x) + TFC(x). \tag{12.2}$$

Although both TVC and TFC are also for the firm of size $\bar{k}$, it will be convenient to omit the superscript $\bar{k}$ from their symbolisms. In parallel with the long-run, the short-run total costs are derived from input baskets on the short-run expansion path. A movement along the graph of the short-run total cost curve is therefore associated with a movement along the short-run expansion path. However, in this case a change in input prices by itself does not, as noted in Chapter 11, change the short-run expansion path but does alter the short-run total cost curve.

An illustration of the graphs of these functions appears in Figure 12-3. The shape of the

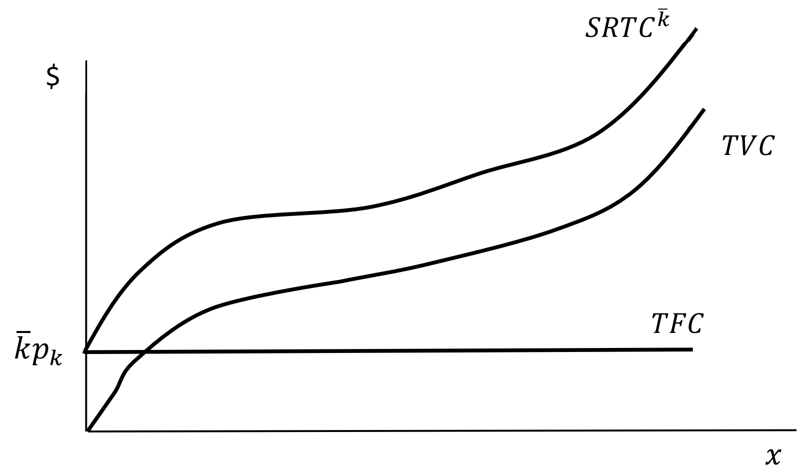


Figure 12-3

TVC curve, which necessarily starts at the origin (since, if the firm produces no output, it does not need labor input and its labor or variable cost is zero), is taken to be that of the long-run total cost curve appearing in upper diagram of Figure 11-7 of Chapter 11. Since total fixed cost does not change with variations in output, the graph of TFC is a straight line parallel to the x-axis. In particular, the firm being unable to change the quantity of capital it employs in the short run, the total fixed cost, and hence total short run cost, remain at $\bar{k}p_k$ even when the firm produces no

[1] If the short-run expansion path intersected the x isoquant outside of the area between the ridge lines as pictured with capital $\bar{k}'$ in Figure 12-2, then by using only, say, $\bar{k}$ of $\bar{k}'$, the firm could produce the same output x with lower labor input (that is, moving from ℓ' to ℓ^0). The firm's fixed cost $TFC(x)$ would remain at $\bar{k}'p_k$ (capital cannot be changed in the short run), but its variable cost $TVC(x)$ would be lowered from $\ell'p_\ell$ to ℓ^0p_ℓ. However, this possibility is not considered further here.

output. Because $SRTC^{\bar{k}}$ is the sum of TVC and TFC (equation (12.2), and because TFC is constant and not changing with x, the graphs of $SSRTC^{\bar{k}}$ and TVC are parallel, having the same slope at each value of x. These latter curves shift with each change in, p_ℓ, p_k, and the production function f. Only $SRTC^{\bar{k}}$ and TFC shift with changes in $\bar{k}$.

Short-run average $SRAC^{\bar{k}}(x)$, average variable $AVC(x)$, and average fixed cost $AFC(x)$ functions are defined in the usual manner with the superscript $\bar{k}$ on the latter two dropped:

$$SRAC^{\bar{k}}(x) = \frac{SRTC^{\bar{k}}(x)}{x},$$

$$AVC(x) = \frac{TVC(x)}{x}$$

$$AFC(x) = \frac{TFC(x)}{x} = \frac{\bar{k}p_k}{x}. \tag{12.3}$$

Dividing equation (12.2) by x gives

$$\frac{SRTC^{\bar{k}}(x)}{x} = \frac{TVC(x)}{x} + \frac{TFC(x)}{x}$$

or

$$SRAC^{\bar{k}}(x) = AVC(x) + AFC(x).$$

In approximate form, short-run marginal cost is

$$SRMC^{\bar{k}}(x) = \frac{SRTC^{\bar{k}}(x + \Delta x) - SRTC^{\bar{k}}(x)}{\Delta x}.$$

In the context of the limiting derivative form for derivatives, differentiating equation (12.2) with respect to x in the normal way,

$$SRMC^{\bar{k}}(x) = \frac{dSRTC^{\bar{k}}(x)}{dx} = \frac{dTVC(x)}{dx} + \frac{dTFC(x)}{dx}.$$

And since $TFC(x)$ is a constant function, its derivative ${}^{dTFC(x)}\!/_{dx} = 0$. Thus, as suggested by the fact that the graphs $SRTC^{\bar{k}}$ and TVC in Figure 12-3 are parallel and have the same slope at each value of x, $SRMC^{\bar{k}}(x)$ could be calculated by differentiating either $SRTC^{\bar{k}}(x)$ or $TVC(x)$.

From equation (12.3), the graph of $AFC(x)$ has the shape of a rectangular hyperbola as shown in Figure 12-4. When $x = 1$, applying (12.3) gives $AFC(1) = \bar{k}p_k$. In addition, since $SRAC^{\bar{k}}(x) = AVC(x) + AFC(x)$, as x becomes large the downward slope of the graph in Figure

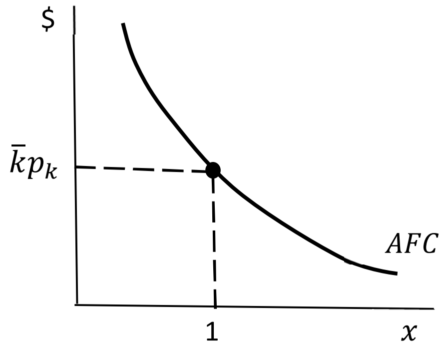


Figure 12-4

12-4 implies that the difference $SRAC^{\bar{k}}(x) - AVC(x) = AFC(x)$ becomes small. This is sometimes referred to as the "spreading of overhead effect."

A geometric example of short-run total, average and marginal cost curves based on the curves in Figure 12-3 appears in Figure 12-5. (The TFC and AFC curves are not drawn.) The presence of two scales on the vertical axis in the lower diagram and the relationships among

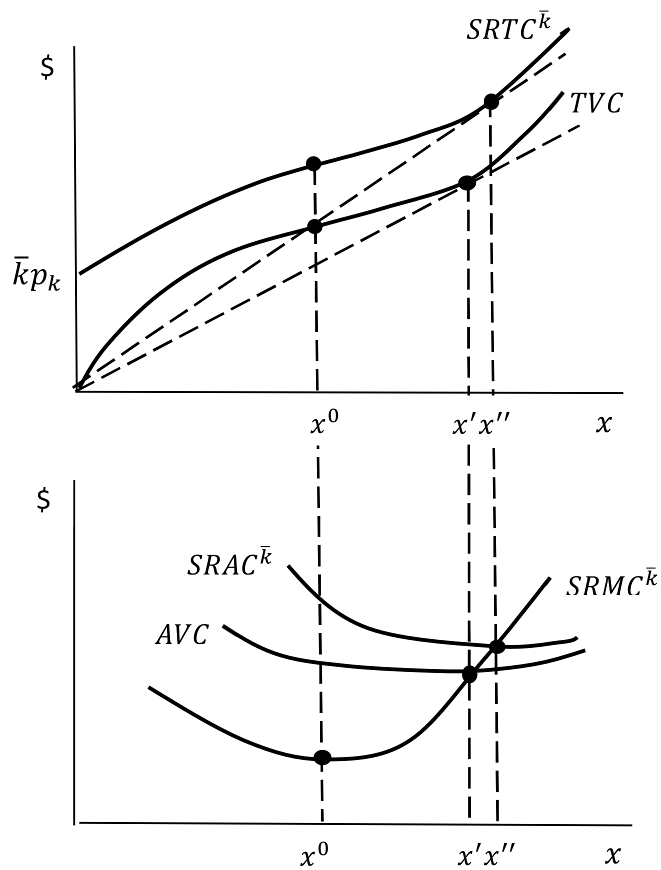


Figure 12-5

these curves are the same as in the long run case (Figure 11-7 of Chapter 11) with the following additions:

1. The fact that the $SRTC^{\bar{k}}$ and TVC curves are parallel has two geometric implications: First, there is only one $SRMC^{\bar{k}}$ curve which has a minimum at x^0. And second, because the $SRTC^{\bar{k}}$ curve is farther from the x-axis than the TVC curve, the tangency on the former with a straight line from the origin necessarily lies above and to the right of that on the latter.

2. From the second implication of (1), the minimum point on the $SRAC^{\bar{k}}$ curve (over x'') occurs above and to the right of that on the AVC curve (over x').

3. The $SRMC^{\bar{k}}$ curve goes through the minimum points of both $SRAC^{\bar{k}}$ and AVC curves.

4. The $SRAC^{\bar{k}}$ and AVC curves become closer together as x increases due to the spreading of overhead effect.

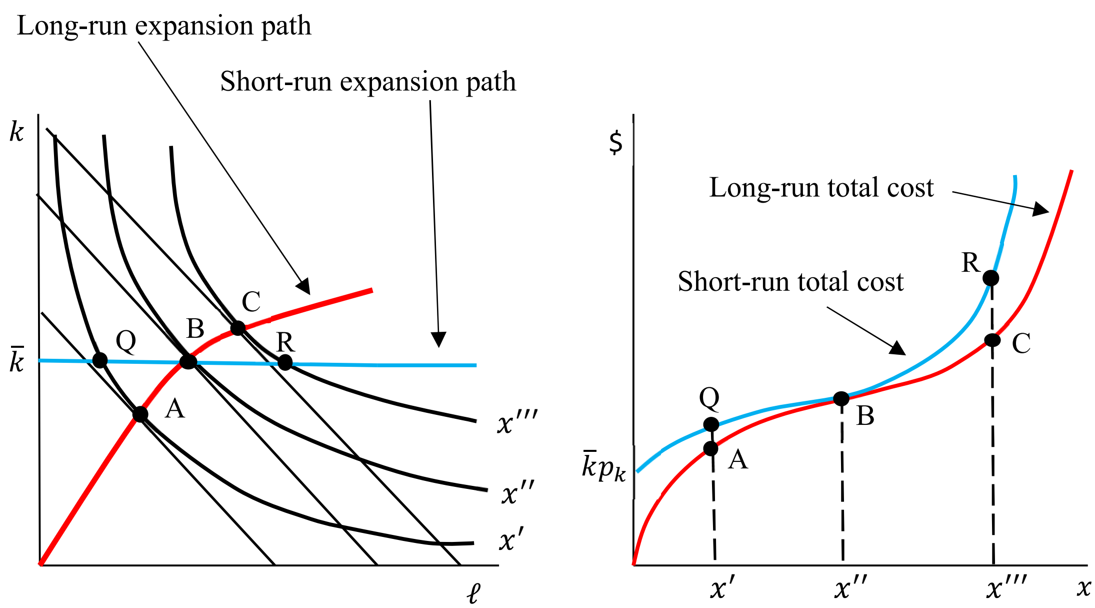


Figure 12-6

The difference between the short-run and long-run total cost curves derives from the difference between the expansion paths on which each is based. As output increases in the short run, k, recall, is fixed at $\bar{k}$ and the firm moves along its short-run expansion path $k = \bar{k}$. Increasing output in the long run causes both inputs to rise as the firm moves from one cost-minimizing tangency to another along its long-run expansion path. Since the long-run expansion reflects cost minimization and the short-run expansion path does not (because it is based on a fixed quantity of capital), short-run total costs are never smaller than long-run total costs and, in most cases, larger. The following geometry (Figure 12-6) and its accompanying argument illustrate these ideas.

In the input space of the left-hand diagram three isoquants are drawn ignoring the possibility of ridge lines. Each is denoted by the outputs produced along them, namely, $x',\ x''$, and x'''. Each isoquant is tangent to an iso-cost line, whose slope is the negative of the fixed input price ratio p_ℓ/p_k. The tangencies occur at capital and labor input combinations A, B, and C respectively. The red curve on which A, B, and C lie is the long-run expansion path. In the short run, with capital or firm size fixed at $\bar{k}$, outputs $x',\ x'',$ and x''' are produced with respective capital and labor input combinations Q, B, and R. The blue straight line on which Q, B, and R appear is the short-run expansion path.

The long-run cost of producing the three outputs is, in each case, the cost of the least-cost basket of capital and labor inputs for producing those outputs, that is, the cost of baskets A, B, and C in the left-hand diagram, respectively. These costs are associated with their outputs and identified in the right-hand diagram at points A, B, and C on the red long-run total cost curve in the same way as shown in Figure 11-6 of Chapter 11. Consider now the long-run and short-run costs of producing output x'. Since, in the left-hand diagram, basket A is cost-minimizing and basket Q is not, the short-run total cost of producing output x' with basket Q has to be larger than the long-run total cost of producing that output with basket A. Hence, in the right-hand diagram, the short-run total cost of producing output x' lies at point Q on the blue short-run total cost curve which is necessarily above the red long-run total cost of producing that output at point A. The same argument (that C in the left-hand diagram is cost-minimizing while R is not) implies that the total cost of producing output x''' on the short-run total cost curve at point R has to be above the long-run total cost of producing that output at point C on the long-run curve. The basket of inputs B in the left-hand diagram is used to produce output x'' in both long and short runs. It follows that the long-run and short-run total costs of producing that output are identical, and that the long-run and short-run total cost curves are tangent at point B in the right-hand diagram. Therefore, except where the two curves are tangent at B, the short-run total cost curve always lies above the long-run total cost curve. (Note that, in the right-hand diagram, when output falls to zero, so does long-run total cost since the firm is able to reduce both ℓ and k to zero. But in the short run with capital fixed at $\bar{k}$, the firm still has to cover its fixed cost even if it is not producing anything. So its short-run cost remains at $\bar{k}p_k$ when output falls to zero.)

According to the above geometry and argument, with input prices fixed, each value of $\bar{k}$ determines a different short-run total cost curve tangent to the long-run total cost curve. The short-run total cost curve tangent to the long-run total cost curve at B in the right-hand diagram of Figure 12-6 is reproduced in the top diagram of Figure 12-7 and labeled $SRTC^{\bar{k}}$. The short-run total cost curve tangent to the long-run total cost curve at A does not appear in the right-hand diagram of Figure 12-6. But it is drawn as a second blue curve tangent to the red log-run total cost curve at A in the top part of Figure 12-7 and is identified by the symbol $SRTC^{\bar{k}'}$. The red long-run total cost curve is labeled $LRTC$. The translation of curves in the top diagram of Figure 12-7 into blue short-run average cost curves, $SRAC^{\bar{k}}$ and $SRAC^{\bar{k}'}$, and red long-run average cost

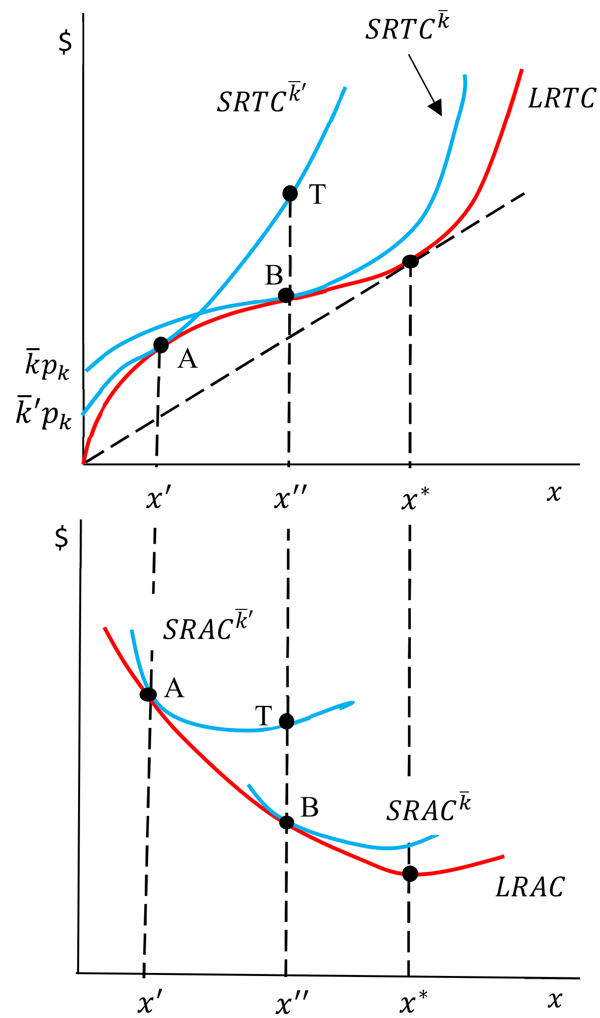


Figure 12-7

curve $LRAC$ appears in the bottom diagram of the figure. For output x'', for example, the short-run and long-run averages are equal (the two curves are tangent) because the short- and long-run totals are equal and, to obtain the averages, both totals are divided by the same output x''. A little to the left or right of x'', the short-run averages are greater than the long-run averages because the short-run totals are larger than the long-run totals. Note that the minimum points of the short-run curves are not tangent to the long-run curve which has its own minimum at x^*. Also, although not explicitly identified in Figure 12-7, the minimum points of the short-run average cost curves in the lower diagram occur at the same outputs at which there are tangencies not appearing in the diagram between the corresponding short-run total cost curves and straight lines from the origin in the upper diagram.

In addition to its earlier characterization as the cost of the cost-minimizing basket for producing output x, the long-run total cost $LRTC(x)$ at x can also be seen, in reference to the upper diagram of Figure 12-7, as the minimum short-run total cost of producing x over all firm sizes $\bar{k}$. In that diagram, the short-run total cost of producing x'' with firm size $\bar{k}'$ occurs at point

T. Other firm sizes will give different total costs at points above and below T. But the long-run total cost of producing output x'', namely $LRTC(x'')$, occurs at point B, which also appears on the short-run total cost curve associated with firm size $\bar{k}$, and corresponds to the lowest total cost of producing x'' over all of these possibilities. (Recall that the short-run cost of producing any output cannot be lower than the long-run cost of producing that output since the latter is cost-minimizing over all relevant input baskets.) The same statement characterizing the long-run total cost of an output as the minimum of all short-run costs for producing that output can be made with respect to long-run average costs in reference to the lower diagram of Figure 12-7.

To add marginal cost curves into the picture, the short-run total and average cost curves tangent to the long-run total and average cost curves at A in both diagrams of Figure 12-7 are redrawn in Figure 12-8 below. Recall that the marginal costs at values of x represent the slopes

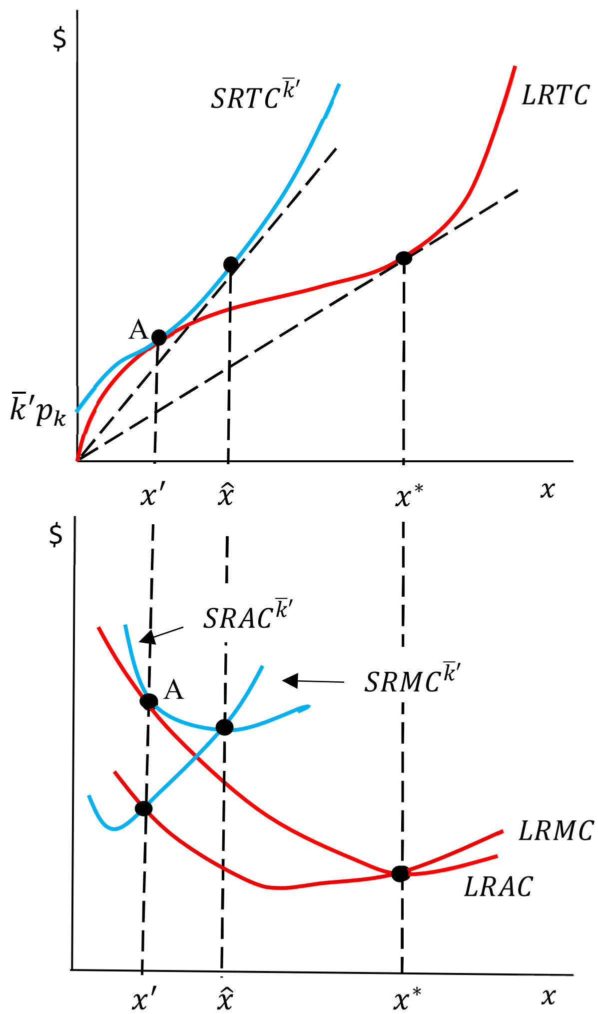


Figure 12-8

of the total cost curves at those values. Thus, referring to the top diagram of Figure 12-8, at x' it is necessary that L$RMC(x') = SRMC(x')$ because the short- and long-run total cost curves are tangent at A and therefore have the same slopes. This implies that the blue short-run and red long-run marginal cost curves in the lower diagram of Figure 12-8 intersect when $x = x'$. For $x < x'$, the slope of the short-run total cost curve in the upper diagram is flatter than that of the long-run curve. This means that L$RMC(x) > SRMC(x)$ for those values of x and, in the lower diagram, the long-run marginal cost curve lies above the short-run marginal cost curve. Similarly, when $x > x'$, the slope of the short-run total cost curve is steeper than that of the long-run curve. Hence L$RMC(x) < SRMC(x)$ and the long-run marginal cost curve lies below the short-run marginal cost curve. Observe that, in the lower diagram, the blue short-run average cost curve is tangent to the red long-run average cost curve at the same value x' at which the two marginal cost curves intersect. This is because $SRTC^{\bar{k}'}\left(x^{'}\right) = LRTC(x^{'})$ in the upper diagram forcing $SRAC^{\bar{k}'}\left(x^{'}\right) = LRAC(x^{'})$ as described above, and the two total cost curves are tangent at x' ensuring $SRMC^{\bar{k}'}\left(x^{'}\right) = LRMC(x^{'})$. Also in the lower diagram, the two marginal cost curves pass through the minimum points of their respective average cost curves as defined by the tangencies in the upper diagram between the total cost curves and straight (dashed) lines emanating from the origin. And, although not explicitly identified in Figure 12-8, the minimum points of the short- and long-run marginal cost curves in the lower diagram occur at the same levels of output at which the slopes of the respective total cost curves change from decreasing to increasing in the upper diagram as described for Figure 11-7 in Chapter 11.

Chapter 13
Profit Maximization

Having dealt with the perfectly competitive firm's costs in the last two chapters, attention now turns to its revenue and profit. Begin by adding the following assumptions to the model developed thus far to explain its buying and selling behavior:

3. Long-run and short-run total cost curves appear as previously drawn so that average and marginal cost curves can be determined and have the shapes attributed to them.[1]

4. The firm hires (demands) inputs and produces and sells (supplies) output so as to maximize its profit.

All assumptions required for the present explanation of firm behavior have now been provided. The earlier assumptions were introduced in Chapter 10. The complete list appears in Supplemental Note D.

Consider the firm's revenue receipts first. Those receipts depend on the demand curve it faces. Now in general, the demand curve facing the firm is conceptually different from the market and individual consumer demand curves defined in earlier chapters. The latter, recall, indicates the amount that all buyers or a single buyer demand at each price. But except in the case of monopoly to be discussed in Chapter 22, the firm does not face the market demand curve or that of a single buyer. Rather it faces a demand curve that shows the quantity of output that it can sell at each price. (Thus, summing the demand curves facing each firm in a market does **not** generally yield the market demand curve.) Under perfectly competitive conditions, the firm is so small and there are so many of them that it has no impact on price regardless of how it might change the quantity of output it brings to the market. Thus, the firm is able to sell any amount it can produce and the demand curve it faces is a flat line parallel to the x-axis at the level of the

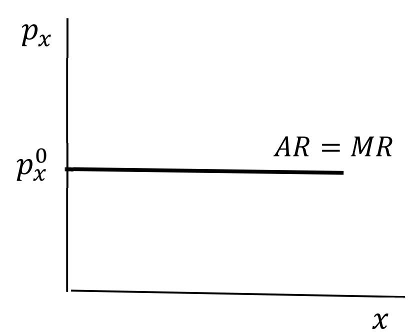


Figure 13-1

---

[1] This assumption will occasionally be relaxed in later chapters to allow for argumentative and diagrammatic simplicity. Note that the long-run cost curves of constant returns to scale production functions do not satisfy this assumption since, as indicated at the beginning of Chapter 12, such production functions generate constant economies of scale.

market price. This demand curve is always known to the perfectly competitive firm. An illustration appears in Figure 13-1 where p_x^0 is the market price.

Along the demand curve facing the firm, the firm's total revenue is given by the function

$$TR(x) = xp_x \tag{13.1}$$

where p_x is determined in the output market. Because the firm is unable to influence or change the market price, it takes the price of its output as fixed. It follows that the graph of its total revenue function (13.1) is a straight line emanating from the origin with slope p_x as pictured in Figure 13-2

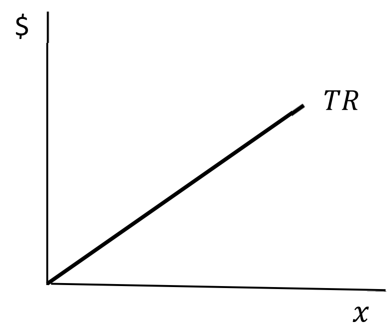


Figure 13-2

The firm's average revenue function is

$$AR(x) = \frac{TR(x)}{x}$$

or, using equation (13.1),

$$AR(x) = p_x. \tag{13.2}$$

In this expression x is the independent variable and p_x the dependent variable. Since average revenue, like total revenue, is defined along the demand curve facing the firm, each value of x satisfying equation (13.2) has to be the quantity of the firm's output it can sell at the price value p_x. Therefore, solving $AR(x) = p_x$ for x as a function of p_x yields the demand curve facing the firm. And, as it appears in Figure 13-1, the graph of the demand curve facing the firm is the same as the average revenue curve of equation (13.2) except that, as with the geometry of the market and individual consumer demand curves, the axes have been reversed and x has become the dependent variable although located on the horizontal axis and p_x the independent variable on the vertical.

Since p_x is constant, the marginal revenue function, $MR(x)$, thought of in approximate form as the additional revenue obtained per unit of additional output, can also be obtained equivalently in derivative form as

$$MR(x) = \frac{dTR(x)}{dx} \tag{13.3}$$

or, upon differentiating (13.1) with respect to x,

$$MR(x) = p_x. \tag{13.4}$$

Combining equations (13.2) and (13.4), it is clear that in this case $MR(x) = AR(x)$ at each x and, with $p_x = p_x^0$ say, the graphs of each are the same straight line parallel to the x-axis at the level of p_x^0. This is illustrated in Figure 13-1.

Turning to the firm's profit, recall that π was defined as a function of x, ℓ, and k in equation (11.2) of Chapter 11 as $\pi = xp_x - (\ell p_\ell + kp_k)$, where (ℓ, k) is the basket of inputs used to produce output x. Since the firm operates to maximize its profit, in the long run it will always produce x with the cost-minimizing combination of inputs. Thus $p_\ell + kp_k$ can be replaced by the expression for the cost of that basket, $LRTC(x)$, developed in Chapter 11. Using equation (13.1) to substitute $TR(x)$ for xp_x, profit π can now be written as a function of x alone or

$$\pi(x) = TR(x) - LRTC(x). \tag{13.5}$$

In this expression the cost of producing each output has already been minimized. To maximize profit, then, it only remains to find the value of x that does the maximizing. The first step is to identify the first-order condition that characterizes the critical value x^0. That is secured by equating the derivative of (13.5) to zero:

$$\frac{d\pi(x^0)}{dx} = \frac{dTR(x^0)}{dx} - \frac{dLRTC(x^0)}{dx} = 0. \tag{13.6}$$

Using equation (11.5) of Chapter 11 and (13.3) above, the equation obtained from the right-hand equality in (13.6) may be rewritten as $MR(x^0) = LRMC(x^0)$ or, from equation (13.4),

$$LRMC(x^0) = p_x. \tag{13.7}$$

When x^0 is associated with maximum profit, equation (13.7) can be derived in words: If $MR(x) > LRMC(x)$ at some x, then producing an extra unit of output would expand revenue by more than its cost. Hence profit would increase and could not have been maximized at x. If $MR(x) < LRMC(x)$, then the last unit produced costs more than the revenue obtained from its sale. Profit could then be increased by not producing it. Once again, profit could not have been maximized at x. A parallel argument applies when x^0 identifies a minimum.

For the short run, set $k = \bar{k}$ and replace $LRTC$ and $LRMC$ in the above symbolism and equations by $SRTC^{\bar{k}}$ and $SRMC^{\bar{k}}$ from Chapter 12. The fact that that $SRTC^{\bar{k}}(x) = TVC(x) + TFC(x)$ from equation (12.1) in Chapter 12 does not introduce further modification of equations (13.5) – (13.7).

Second-order conditions relating to maximum or minimum profit can be equivalently expressed either in terms of strict concavity and convexity of the profit function, or in terms of the second order derivative of that function. With respect to the former the maximum occurs in the region where the profit function is strictly concave; the minimum appears where it is strictly convex. With respect to derivatives and in the long run (similar statements apply to the short run by replacing long-run functions with short-run functions as described above), the second-order condition ensuring a maximum obtained from the second-order derivative of equation (13.5) is

$$\frac{d^2\pi(x^0)}{dx^2} = \frac{d^2TR(x^0)}{dx^2} - \frac{d^2LRTC(x^0)}{dx^2} < 0. \tag{13.8}$$

Equation (13.8) is an alternative way of stating that the profit function is strictly concave around the point of maximum profit x^0. Since, using equation (13.4),

$$\frac{d^2TR(x^0)}{dx^2} = \frac{dMR(x^0)}{dx} = \frac{dp_x}{dx} = 0$$

because p_x is determined by the market and constant from the point of view of the firm, and since, from equation (11.5) of Chapter 11,

$$\frac{d^2LRTC(x^0)}{dx^2} = \frac{dLRMC(x^0)}{dx},$$

the second-order condition (13.8) implies that

$$\frac{dLRMC(x^0)}{dx} > 0. \tag{13.9}$$

To ensure that x^0 represents a minimum, the inequality of (13.9) should be reversed.

The geometry of long- and short-run profit maximization appears in Figures 13-3 and 13-4 respectively. Cost curves are drawn in red. Observe that in each case, there are two values of x at which marginal cost equals price, that is, at which first-order derivative conditions are satisfied. One corresponds to maximum profit, the other to minimum profit. The signs of $dLRMC(x^0)/dx$ in the above discussion of second-order conditions mean that, around the maximizing value x^0, the long-run marginal cost curve must slope upward and hence that, in this region, the long-run total cost function is strictly convex. At the minimizing value, the sign in (13.9) is reversed and, around that value, the long-run marginal cost curve slopes downward and the long-run total cost function is strictly concave.

There are several things to notice in Figure 13-3. First, the vertical difference between the TR and $LRTC$ curves is the same as the vertical height (positive or negative) of the π curve at each x in accordance with equation (13.5). Second, at $\tilde{x}'$ and $\tilde{x}$ profit is zero so that for both x values $LRTC(x) = TR(x)$. Using $TR(x) = x\bar{p}_x$ to eliminate $TR(x)$ from $LRTC(x) = TR(x)$ and dividing the result by x gives, from equation (11.4) of Chapter 11 and equation (13.2) above,

$LRAC(x) = \bar{p}_x$. This is represented in the lower part of the diagram by the intersections of the $LRAC$ curve and the MR $(= AR)$ line a botht $\tilde{x}$ and $\tilde{x}'$. Third, π is maximized and $LRMC(x) = \bar{p}_x$ at x^0; and π is minimized and $LRMC(x) = \bar{p}_x$ at $x^{0'}$. At these two values of x the $LRMC$ curve and the MR line intersect. Fourth, for reasons discussed in Chapter 12, the $LRAC$ curve has a minimum at $\hat{x}$ and the $LRMC$ has a minimum at x^*. (When drawing Figure 13-3 to account for these characteristics, consider using the following procedure: After lining up the two pairs of coordinate axes, draw, in the upper part of the diagram $LRTC$ first, followed by TR and then π. Next, in the lower part of the diagram draw in the following order the horizontal line at $\bar{p}_x$, $LRMC$, and $LRAC$.)

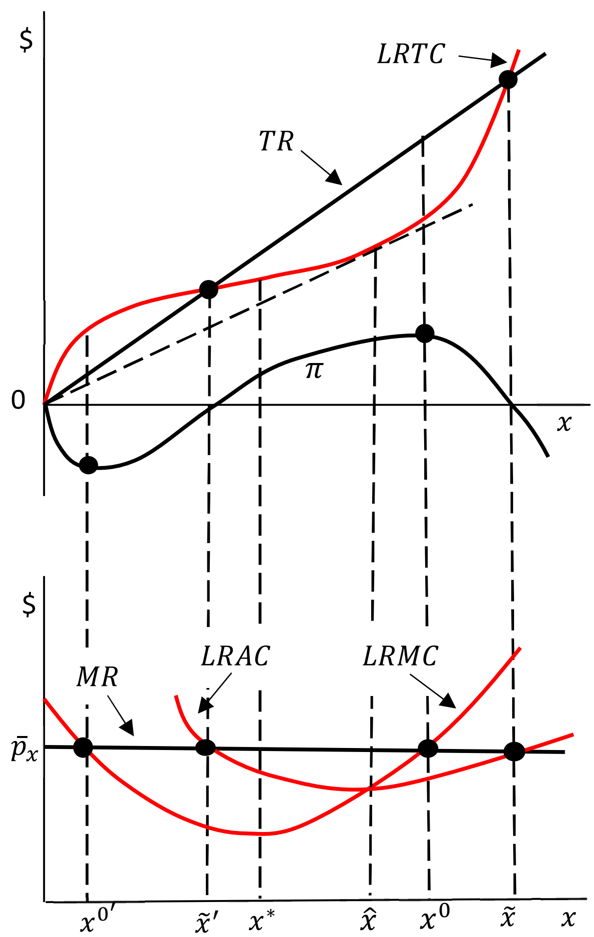


Figure 13-3 (long run)

Focusing now on the short run with $k = \bar{k}$, where $SRTC^{\bar{k}}$ and $SRMC^{\bar{k}}$replaces $LRTC$ and $LRMC$ in equations (13.5) – (13.7), and $SRTC^{\bar{k}}(x) = TVC(x) + TFC(x)$. The geometric parallel to Figure 13-3 is shown in Figure 13-4. With respect to Figure 13-4, the properties of the cost curves carry over from those described in Chapter 12 as follows: Observe first that $SRTC^{\bar{k}}$ (in blue) and TVC (in gold) are parallel curves. They differ by the amount of the fixed cost $\bar{k}p_k$ at each x (Figure 12-3 of Chapter 12). Second, the tangency between the line from the origin and

$SRTC^{\bar{k}}$ at $\hat{x}$ lies to the right of that between the line from the origin and TVC at $\hat{x}'$ (Figure 12-5 of Chapter 12). Hence the minimum of $SRAC^{\bar{k}}$ (also at $\hat{x}$) lies to the right of the minimum of AVC (at $\hat{x}'$). Third, the $SRAC^{\bar{k}}$ (in blue) and AVC (in gold) curves become closer together as x increases because their difference is the average fixed cost ($AFC = \bar{k}p_k/x$), and that becomes smaller as x increases (Figures 12-4 and 12-5 of Chapter 12). Fourth, the minimum of the $SRMC^{\bar{k}}$ curve occurs at $x^*$where the slope of the $SRTC^{\bar{k}}$ curve changes from decreasing to increasing.

There are also parallels with Figure 13-3. First, the vertical difference between the TR and $SRTC^{\bar{k}}$ curves is the same as the vertical height (positive or negative) of π at each x. Second, as in the long-run case described above (Figure 13-3), $SRTC^{\bar{k}} = TR$, $\pi = 0$, and $SRAC^{\bar{k}} = \bar{p}_x$ at $\tilde{x}$ and $\tilde{x}'$. And third, π is maximized and $SRMC^{\bar{k}} = \bar{p}_x$ at x^0; and π is

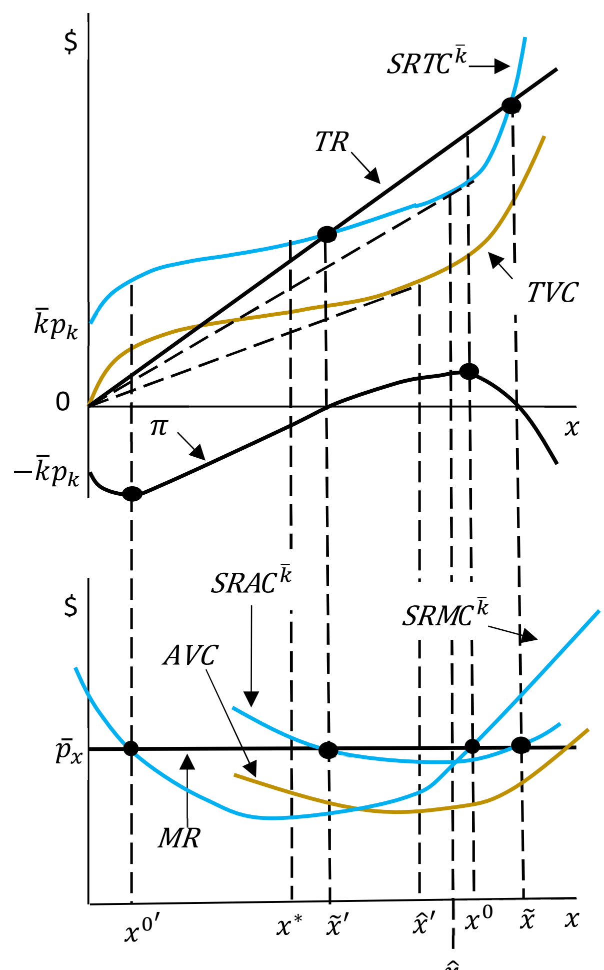


Figure 13-4 (short run)

minimized and $SRMC^{\bar{k}} = \bar{p}_x$ at $x^{0'}$. (One way to draw Figure 13-4 and account for these characteristics is, after lining up the two pairs of coordinate axes, to draw the parallel $SRTC^{\bar{k}}$ and

TVC curves first. Then, in the following order, draw TR, π, the horizontal line at $\bar{p}_x$, $SRMC^{\bar{k}}$, $SRAC^{\bar{k}}$, and AVC.)

In both the long and short runs, once the profit-maximizing value of x is determined, the basket of inputs that corresponds to that output (and, as will be seen in subsequent chapters, also maximizes firm profit when expressed as a function of inputs) is found from the intersection of the isoquant corresponding to the profit-maximizing output x^0 and the appropriate expansion path – (ℓ^0, k^0) in the long run and ℓ^0 in the short run – as shown in Figure 13-5.

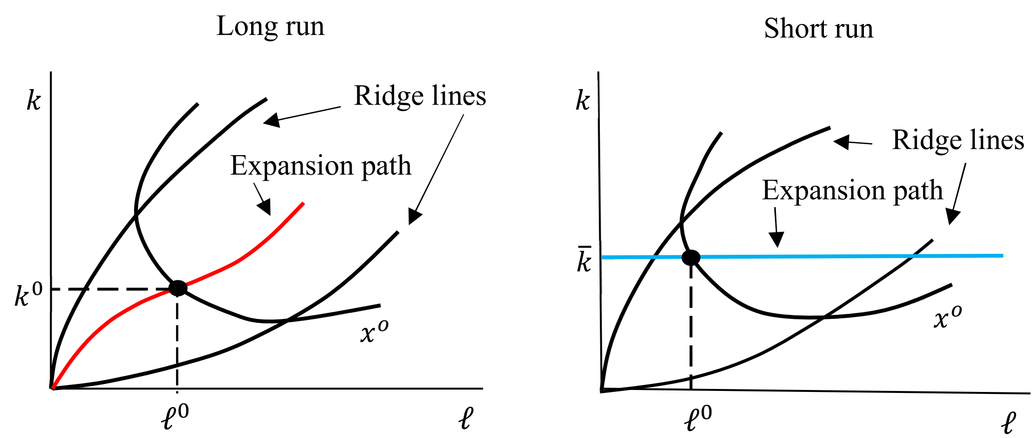


Figure 13-5

The entire argument may be summarized as follows: Under the assumptions of the model set out from Chapters 9 up to this point (and listed in Supplemental Note D), the explanation of firm buying and selling behavior developed here views the firm as faced with a fixed technology and fixed input and output prices. In that context, it also sees the firm as pursuing the procedure set out below to arrive at the profit-maximizing output quantity to produce and the profit-maximizing input quantities to employ (the steps of this procedure are reproduced in Supplemental Note E):

1. If ridge lines exist in the input space, eliminate the regions outside of the area between them and beyond any intersection point if there is one. This reduces the input space to its relevant region. (Recall from Chapter 10 that if ridge lies do not exist, the relevant region is the entire input space excluding the co-ordinate axes.)

2. Using input price information and cost minimization (long run) or fixed capital information $\bar{k}$ (short run), calculate the appropriate expansion path in the relevant region and confine attention to it.

3. Using the production function, expansion path, and input price information, calculate all cost functions and curves expressing cost as a function of output.

4. Using output price information, calculate all revenue functions and curves.

5. Using cost and revenue information, calculate the profit function and the profit-maximizing output x^0.

6. From the intersection of the isoquant relating to the profit-maximizing output and the long- or short-run expansion path, calculate the profit-maximizing input quantities.

The output and input quantities obtained from this procedure are those that will be, respectively, sold and bought by the firm in response to output and input prices dictated by the markets. This is the basis for the specification of the firm's output supply and input demand functions discussed in the nest chapter.

As was the case with the model of consumer buying and selling behavior in Chapters 5 and 6, to be accepted as an explanation of real firm behavior, it is not necessary that firms actually follow this exact procedure with function satisfying all of the assumptions made above. It is only required that enough of the model's ideas, such as fixed and variable cost, profit maximization, and cost minimization resonate, as they often do, in a reasonable way with what is present in reality. The rest of the structure (including the assumptions) only provides a framework within which firm behavior can be analyzed in precise terms and pursued in appropriate directions.

Changes in the profit-maximizing quantities of this model come about through variations in the elements that are fixed in the procedure described above. For example, were technology to modify, the firm's production function could change. The latter modification works its way through the six steps described above, thereby altering the output and input quantities obtained from the maximization of profit. The manner in which price variations affect these quantities are examined in greater detail below.

First, any change in a single input price, say p_ℓ (with technology, p_k, and p_x held fixed), works its way throughout this model in the following steps:

<u>Long run:</u>

1. Since p_ℓ changes with $p_k$held fixed, the input price ratio ${}^{p_\ell}/_{p_k}$changes and hence so does the slope of the iso-cost lines.
2. This changes the locations of the tangencies between iso-cost lines and isoquants.
3. This changes the long-run expansion path.
4. This changes the long-run cost curves – in particular, the long-run marginal cost curve.
5. This changes the intersection of the long-run marginal cost curve and the marginal revenue line, and hence the profit-maximizing output.
6. This changes the isoquant relating to the profit-maximizing output.
7. Since there is a new long-run expansion path (Step 3) and a new isoquant relating to the profit-maximizing output (Step 6), there is a new intersection between them, and hence a change in the profit-maximizing quantities of labor and capital inputs.

Short run with $k = \bar{k}$:

1. There is no change in the short-run expansion path because the quantity of capital input is fixed.
2. The short-run total cost curve still changes because p_ℓ changes. That curve pivots around the fixed-cost point on the vertical (\$) axis (both $\bar{k}$ and p_k remain fixed – only p_ℓ varies) changing its slope at each value of x as shown in Figure 13-6. In that diagram the $SRTC^{\bar{k}}$ curve farther from the x-axis is associated with a higher price of labor.

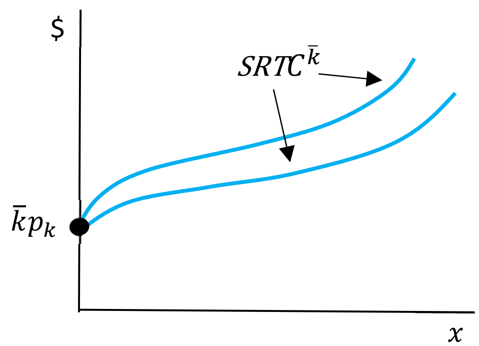


Figure 13-6

3. This change in the slopes of the short-run total cost curve changes the short-run marginal cost curve.
4. This changes the intersection of the short-run marginal cost curve and the marginal revenue line, and hence the profit-maximizing output.
5. This changes the isoquant relating to the profit-maximizing output.
6. This changes the intersection of the isoquant relating to the profit-maximizing output and the short-run expansion path, and hence changes the profit-maximizing quantity of labor input.

In the long run, a change in the price of capital (with no alteration in technology or the other prices) works its way through the profit-maximizing process in a manner similar to that of variations in the price of labor. In the short run, however, since capital is fixed, such a change only modifies the firm's fixed costs. It does not affect its marginal cost. The firm's profit-maximizing output and input quantities therefore remain the same, although profit itself is necessarily altered.

In addition, any change in the output price p_x (with technology, p_ℓ, and p_k held fixed) also has an impact on the profit-maximizing quantities of the model. But that impact is less extensive than that for labor and long-run capital input price changes.

Long run:

1. Since there are no changes in technology and input prices, there are no changes in

isoquant-iso-cost tangencies, the long-run expansion path, or the long-run cost curves.

2. But the change in p_x still changes the revenue curves – in particular, the marginal revenue curve.
3. This changes the intersection of the long-run marginal cost curve and the marginal revenue line, and hence the profit-maximizing output.
4. This changes the isoquant relating to the profit-maximizing output.
5. This changes the intersection of the long-run expansion path and the isoquant relating to the profit- maximizing output, and hence changes the profit-maximizing quantities of labor and capital inputs.

<u>Short run with $k = \bar{k}$:</u>

1. There is no change in the short-run expansion path because the quantity of capital input is fixed.
2. Since there are no changes in technology and input prices, there are no changes in the short-run cost curves.
3. But the change in p_x still changes the revenue curves – in particular, the marginal revenue curve.
4. This changes the intersection of the short-run marginal cost curve and the marginal revenue line, and hence the profit-maximizing output.
5. This changes the isoquant relating to the profit-maximizing output.
6. This changes the intersection of the isoquant relating to the profit-maximizing output and the short-run expansion path, and hence changes the profit-maximizing quantity of labor input.

Chapter 14
Output Supply and Input Demand Functions, Short-Run Output Supply Curves, Short-Run Equilibrium, Taxation, and Producer Surplus

According to the model of firm buying (demanding) and selling (supplying) behavior developed in the last five chapters, as output and input prices change, new output and input quantities are determined through profit maximization. This process of variation generates output supply and input demand functions. In the long run the input demand functions are

$$\begin{Bmatrix} x = g^x(p_x, p_\ell, p_k) \\ \ell = g^\ell(p_x, p_\ell, p_k) \\ k = g^k(p_x, p_\ell, p_k) \end{Bmatrix} \text{ if and only if } \begin{Bmatrix} x, \ell, k \text{ are long run profi } maxi- \\ mizing \text{ values given } p_x, p_\ell, p_k. \end{Bmatrix}, \tag{14.1}$$

for all $p_x, p_\ell,$ and p_k. As will be seen in Chapter 15, the long-run output supply function $x = g^x(p_x, p_\ell, p_k)$ requires some modification if all long-run specifications are to be fully met. For the short run there is an output supply function but only one input demand function, namely that for labor, for which the same symbols g^x and g^ℓ are used:

$$\begin{Bmatrix} x = g^x(p_x, p_\ell, p_k) \\ \ell = g^\ell(p_x, p_\ell, p_k) \end{Bmatrix} \text{ if and only if } \begin{Bmatrix} x, \ell \text{ are short run profi } maxi- \\ mizing \text{ values given } p_x, p_\ell, p_k. \end{Bmatrix}, \tag{14.2}$$

also for all $p_x, p_\ell,$ and p_k. There is no short-run demand function for capital since that input is fixed over the short-run time period. The models that generate these long- and short-run functions provide explanations of long- and short-run firm buying and selling behavior.

The output supply and input demand functions of the firm are observable in the same sense as the demand and supply functions of the consumer: At any moment, watching what the firm buys and sells at the prices dictated by the economy's markets provides one observed "point" on these functions. The properties of these functions as derived from the assumptions made are important for the same reason that the properties of the consumer functions are important: If a firm's observable behavior expressed in terms of such functions does not have the same properties as those of functions (14.1) or (14.2), then the model does not explain that firm's behavior. However, an extensive discussion of these properties will not be undertaken here. Only three examples will be provided. That for the short-run functions g^x is taken up next. A property of the long-run output supply function (that accounts for the modification referred to above) and one for the short-run labor demand function g^ℓ are deferred to the next chapter.

Consider, then, the short run with $k = \bar{k}$. To obtain the supply curve for the firm's output from g^x in equation (14.2), fix the two input prices $p_\ell = p_\ell^0$ and $p_k = p_k^0$, and plot the graph of $x = g^x(p_x, p_\ell^0, p_k^0)$ with axes reversed. It will now be shown that this short-run supply curve is the upward sloping portion of the firm's short-run marginal cost curve above minimum average variable cost. Since the argument will be presented geometrically in reference to short-run cost curves, it will be useful first to describe how total costs, total revenue, and profit for a profit-maximizing output x^0 can be read off of the diagram containing average and marginal costs as pictured in Figure 14-1.

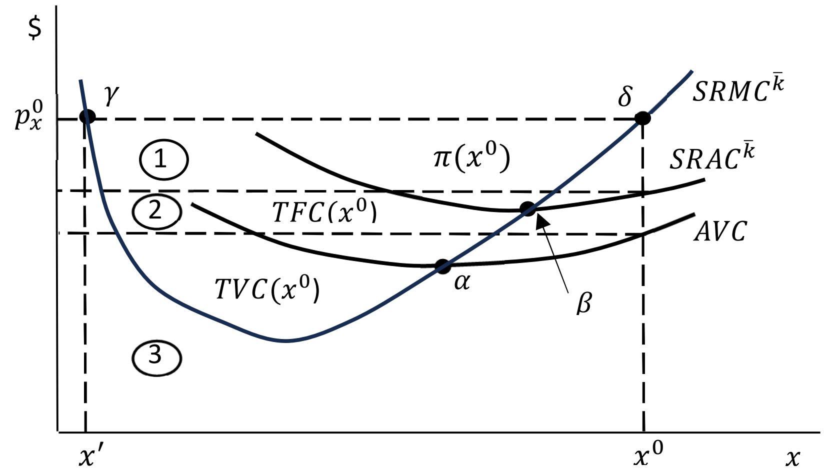


Figure 14-1

In Figure 14-1, profit is maximized at output x^0 where $SRMC(x^0) = p_x^0$. Total revenue at x^0 is $p_x^0 x^0$, as represented by the area of the combined rectangles 1 + 2 + 3. Short-run total cost is short-run average cost at x^0 times x^0 or the area of rectangles 2 + 3. Profit at x^0 is the difference between total revenue at x^0 (rectangles 1 + 2 + 3) and short- run total cost at x^0 (rectangles 2 + 3) or the area of rectangle 1. Total variable cost at x^0 is average variable cost at x^0 times x^0 or the area of rectangle 3. Total fixed cost at x^0 is the difference between short-run total cost at x^0 (rectangles 2 + 3) and total variable cost at x^0 (rectangle 3) or the area of rectangle 2.

The first thing to establish in showing that the short-run output supply curve is the firm's short-run marginal cost curve above minimum average variable cost is the relationship between the firm's output supply and short-run marginal cost in each profit-maximizing position. Still referring to Figure 14-1, for an output price such as p_x^0 whose associated horizontal line intersects the upward sloping portion of the SRM$C^{\bar{k}}$ at δ, the firm maximizes its profit at x^0, where x^0 is the value that satisfies the equation $SRMC^{\bar{k}}(x) = p_x^0$. According to the assumption that the firm produces and supplies the profit-maximizing output (listed as number 4 at the beginning of Chapter 13 and in Supplemental Note D), the firm will supply output x^0. More generally, for any such price the firm's output supply will be that for which

$$SRMC^{\bar{k}}(x) = p_x.$$

Output x' for which the intersection of the horizontal line associated with p_x^0 occurs in the negatively sloped portion of the short-run marginal cost curve at γ in Figure 14-1 is not, as shown in Chapter 13, profit maximizing. Hence, only points on the upward sloping portion of the SRM$C^{\bar{k}}$ curve, that is, ignoring a cut-off described below, have the same coordinates as points on the graph of the supply function $x = g^x(p_x, p_\ell^0, p_k^0)$. Therefore, the two curves are identical for profit-maximizing outputs. It follows that the short-run supply curve is upward sloping, and this is the observable property of the firm's output supply function referred to above.

However, it turns out that, as previously indicated, the supply curve is the upward sloping portion of the SRM$C^{\bar{k}}$ curve only down to the latter's intersection with the average variable cost curve at the point of minimum average variable cost α in Figure 14-1. Outputs associated with intersections below α will not be supplied by the firm. To see why, observe that as p_x falls from p_x^0 in Figure 14-1, all of the rectangles become smaller. When the horizontal line associated with p_x^0 reaches the level of β, the total revenue and short-run total cost rectangles become equal as in Figure 14-2, and $\pi(x^0) = 0$. At this point, referred to as the break-even point, the firm is said to

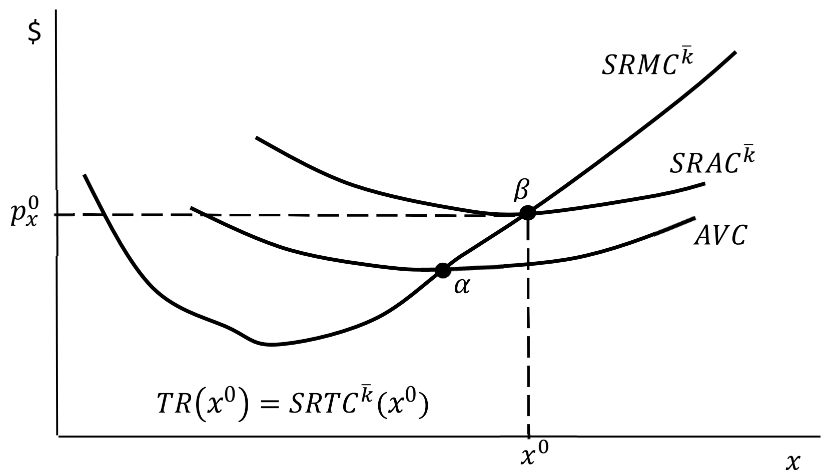


Figure 14-2

be earning a normal profit.

It is worth pausing for a moment to consider the notion of normal profit in some detail. Normal profit may be thought of as the minimum return necessary for the owners of the firm to keep their money invested in the firm. This may be understood as follows: The owners of the firm have invested their funds by buying the capital $\bar{k}$ that the firm uses in production and supplying it to the firm for an expected return of $p_k^0\bar{k}$, where p_k^0 is the given market price of capital. Assume the owners also provide the entrepreneurial input to the firm which results in the firm's profit π. That profit is the reward to the owners for their entrepreneurial contribution. Thus, the total expected return on the owners' investment in the firm is

$$p_k^0\bar{k} + \pi(x).$$

Were the owners to remove their money from the firm for investment elsewhere (of course, they could only do this in the long run), they would have to sell $\bar{k}$ on the market and would receive $p_k^0\bar{k}$ for it. As long as $\pi(x) > 0$, their return is greater than $p_k^0\bar{k}$, and assuming positive profits do not exist as investment opportunities in other industries, the owners are doing better by keeping their money where it is than by selling $\bar{k}$ on the market and putting their funds elsewhere. Thus when $\pi(x) > 0$, there is there is no incentive for long-run considerations of investment removal. When $\pi(x) < 0$, since the variable cost (that is, the cost of the labor input) has to be paid before the fixed cost, the firm is unable to pay for is capital at the market rate. In that case, the owners can do better by selling $\bar{k}$ on the market and long run considerations of removal come into play. Assume the owners keep their investment in the firm when $\pi(x) = 0$. Then $p_k^0\bar{k}$ is the minimum return necessary for the owners to keep their investment in the firm, that is, the normal profit. In the circumstance in which $\pi(x) > 0$, the firm is said to be earning abnormal, excess, or economic profit. When $\pi(x) < 0$, the firm is said to be incurring a loss.

Returning to the short-run output supply curve, when p_x^0 falls to the level of β in Figure 14-2 the owners will want to keep the firm operating and not let it shut down. As described above, when operating at zero profit, the owners still receive a payment of $p_k^0\bar{k}$. Were the firm to shut down, all revenue would vanish, and the owners of the firm would not obtain any return on their investment.

But even when p_x^0 falls below the break-even point β as in Figure 14-3 and the firm is suffering losses, it will still not shut down. (The firm cannot go out of business because this is the short run and the owners of the firm do not have enough time to sell $\bar{k}$ on the market.) Now in the short run the firm still has to pay its variable cost first and then cover as much of the fixed cost of its capital as it can. At p_x^0 in Figure 14-3, the profit-maximizing output is x^0.The sum of $TVC(x^0)$, rectangle 3, plus $TFC(x^0)$, rectangles 1 + 2, or $SRTC^{\bar{k}}(x^0)$, is larger than $TR(x^0)$,

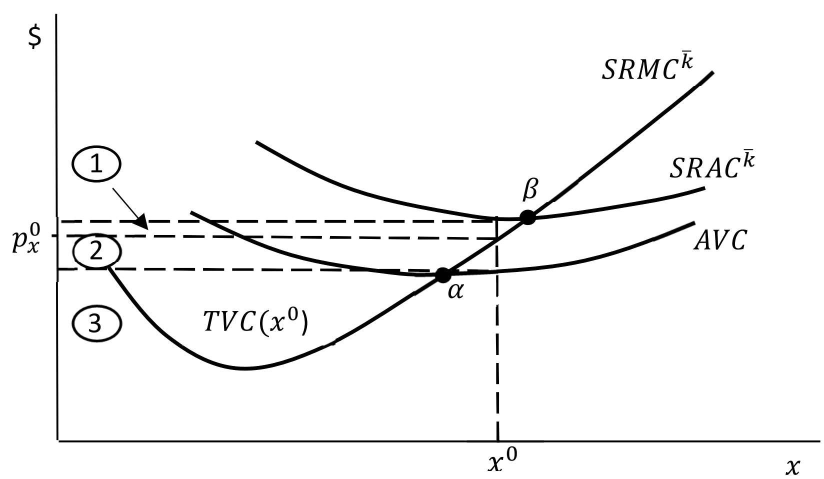


Figure 14-3

rectangles 2 + 3. The firm's loss is given by rectangle 1. This loss is absorbed by the owners of the firm because the firm can pay them only rectangle 2 of its total fixed cost. Were the firm to get rid of its labor and shut down so that $x = 0$, then at that output $TR(0) = TVC(0) = 0$, the firm's loss would be all of its fixed cost, $\pi(0) = -p_k^0\bar{k}$ (rectangles 1 + 2), and the owners of the firm would receive no return on their investment. Thus, as long as $TR(x^0) > TVC(x^0)$, the owners will recover at least part of the fixed cost and, for that reason, insist that the firm continue production. Point α in the diagram is known as the shut-down point. Were p_x^0 to fall below that level, the firm could not even recover in revenue all of its variable cost. It would therefore lose more than just its fixed cost. It could reduce its loss to the amount of the fixed cost by shutting down and ceasing production, thereby eliminating all variable cost. It follows, then, that the firm would not supply any output on the market when the horizontal line associated with p_x^0 intersects the SRM$C^{\bar{k}}$ curve below α. This completes the argument establishing that the firm's short-run supply curve is its short-run marginal cost curve above minimum average variable cost (α in Figure 14-3).

Recall that equilibrium is a position of rest. All forces present balance each other out and there is no tendency for anything to change. As was seen in Chapter 3, a market is in equilibrium when supply equals demand. Likewise, given input and output prices, if a firm is hiring inputs and selling output so as to maximize its profit, there is no incentive for any change. The firm is "at rest" and may be said to be in equilibrium. Interpret the perfect information assumption of perfect competition to imply that all firms have access to the same technology and therefore have the same production functions, cost functions, and cost curves. Then, with all firms having identical cost curves, the relationship between the typical or representative firm and the market or industry at short-run simultaneous equilibrium, where every firm in the market is maximizing its profit and supply equals demand in the market as a whole, appears in Figure 14-4. Here, x_f denotes firm output and x_m represents market quantities. The symbol $\Sigma\ SRMC^{\bar{k}}$ indicates that

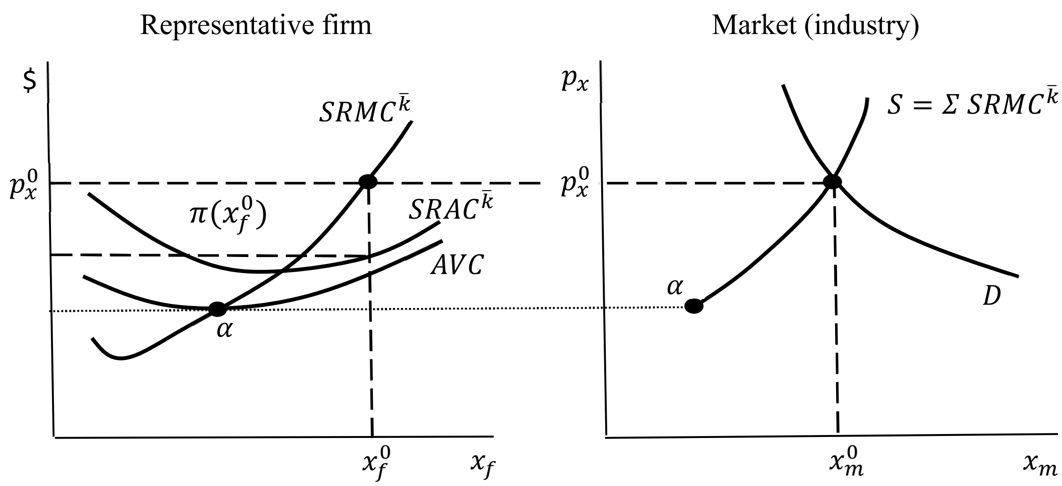


Figure 14-4

the supply curve S is the (horizontal) sum of the supply curves of all individual firms in the market, each starting at the minimum-average-variable-cost point α. Profit maximization for the representative firm occurs at x_f^0, market equilibrium is at price p_x^0 and quantity x_m^0 where supply equals demand, and

$$x_m^0 = \sum x_f^0,$$

where the sum is taken over all firms in the industry. Abnormal firm profit $\pi(x_f^0) > 0$.

Consider a market and its firms in short-run equilibrium. Suppose the government were to impose a tax of t cents per unit of output sold. This is called a specific tax. Firms would collect the tax and send the revenue it generates to the government. The tax would raise the short-run marginal costs of the firm, that is, to the additional cost of producing and selling each unit of output, the firm adds the tax that has to be sent to the government. In terms of Figure 14-4, the $SRMC^{\bar{k}}$ curve would shift up by t. The market supply curve S would also shift up by the amount of the tax.[1] The new equilibrium is shown in Figure 14-5 which uses only the short-run

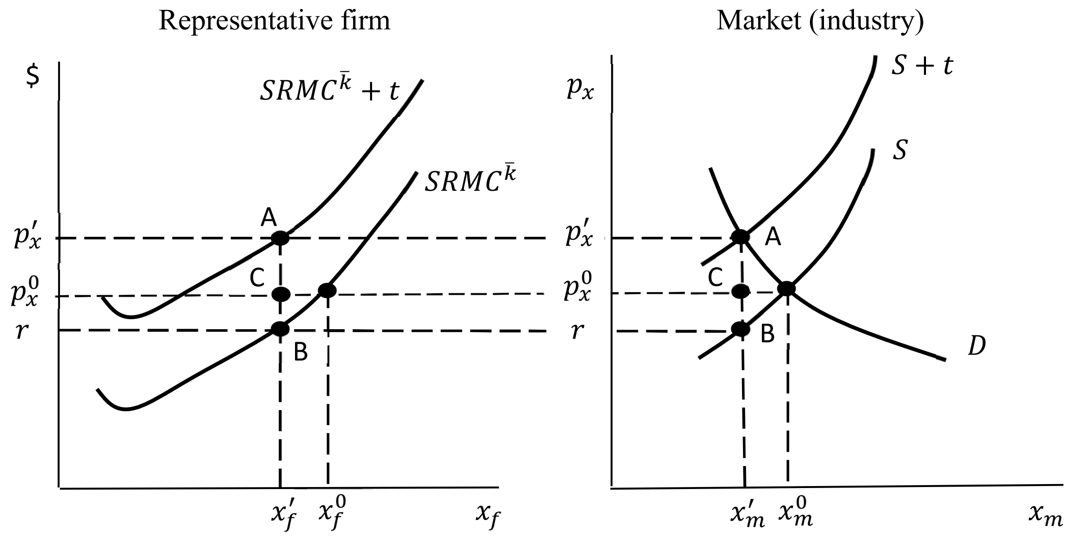


Figure 14-5

marginal cost and market supply and demand curves from Figure 14-4. The minimum-average-variable-cost points are not shown. In both diagrams of Figure 14-5, p_x^0 is the market equilibrium price before the tax is imposed; the after-tax equilibrium price is p_x'. The before- and after-tax market equilibrium quantities are, respectively, x_m^0 and x_m'; the before- and after-tax profit=maximizing firm quantities are x_f^0 and x_f'. And the vertical distance between points A and B, and between the before- and after-tax curves everywhere, is t. Thus

$$t = p_x' - r.$$

where r is the revenue per unit of output that the firm obtains after sending the tax to the government.

What is called the "incidence" of the tax refers to who pays it. That is understood as follows: The increase in the market equilibrium price $p_x' - p_x^0$, or the vertical distance between A

[1] This is because the market supply curve is obtained by summing the quantities supplied at each price by all firms in the market. Since, for every firm quantity supplied, the price that calls forth that quantity is increased by the amount of the tax, the same must be true for the sum of those quantities.

and C in Figure 14-5, is considered to be what the consumers or buyers pay. But that, of course, is less than the amount of the tax t. The remainder $t - (p'_x - p^0_x)$ or the distance between C and B in Figure 14-5 is said to be paid by the firms. The sum of these distances yields the amount of the tax:

$$p'_x - p^0_x + \; t - (p'_x - p^0_x) = t.$$

Producer surplus is the difference between the market price of a good and a unit's marginal cost summed over all units produced. In Figure 14-6, producer surplus for the firm in the short run is illustrated in terms of the firm producing output x^0 with, to keep things simple, the short-run marginal cost curve extended to the vertical axis. Producer surplus for the unit of output labeled 1 is the length of the straight line connecting A and B. As with consumer surplus, taking each point on the x-axis as a unit and summing over all units yields the area between the horizontal line at p^0_x and the short-run marginal cost curve SRM$C^{\bar{k}}$.

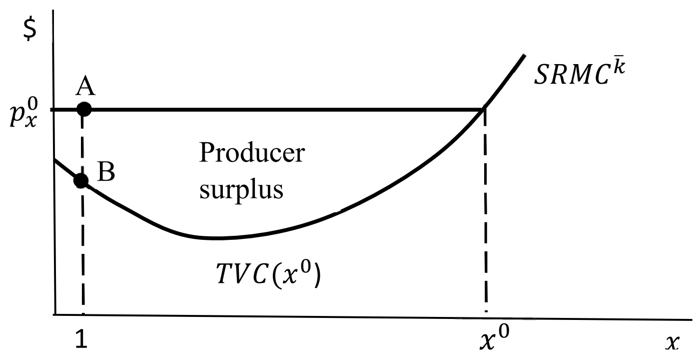


Figure 14-6

It can be shown with more advanced mathematics that, in the short run, $TVC(x^0)$ equals the area in Figure 14-6 under the short-run marginal cost curve up to x^0. The area representing producer surplus combined with that representing $TVC(x^0)$ is the same as the area representing total revenue $p^0_x x^0$. Using that and the fact that $\pi(x^0) = TR(x^0) - TVC(x^0) - TFC(x^0)$, producer surplus is the sum of fixed cost plus profit, that is,

$$TR(x^0) - TVC(x^0) = TFC(x^0) + \pi(x^0). \tag{14.3}$$

Producer surplus in the long run is even simpler. Start with the short-run marginal cost curve in Figure 14-6 which depends on the capital the firm is using in the long run. Change that to the long-run marginal cost curve. Since there is no fixed cost in the long run, $TVC(x^0)$ becomes $LRTC(x^0)$ and equation (14.3) reduces producer surplus to

$$TR(x^0) - LRTC(x^0) = \pi(x^0).$$

Since the market supply curve is the horizontal sum of all short-run firm marginal cost curves, the vertical length from any quantity on the x-axis to the market supply curve is the short run marginal cost of producing that quantity. Hence producer surplus at market equilibrium is the area between the market supply curve and the horizontal line connecting the demand-supply curve intersection with the vertical axis. The diagram corresponding to Figure 14-6 for short-run producer surplus at the market level appears in Figure 14-7 with the market supply curve, extended to the vertical axis as with the short-run marginal cost curve in Figure 14-6.

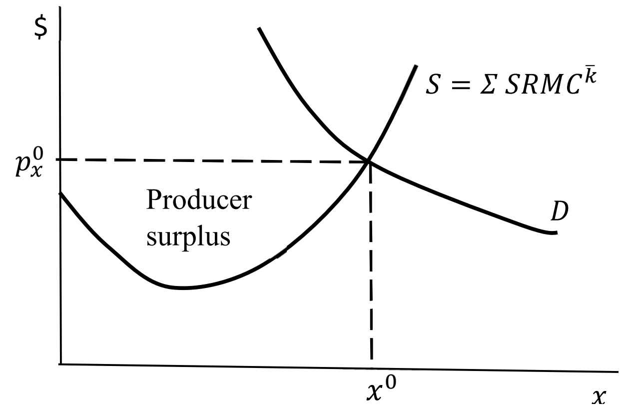


Figure 14-7

Producer surplus benefits consumers in the same way that consumer surplus does because producer surplus accrues to the owners of the firm who are, themselves, consumers. Consider a short-run equilibrium (x^0, p_x^0) in a market with, for simplicity, a linear demand curve D and a linear supply curve S both of which extend to the vertical axis. The combination of consumer surplus (Chapter 8) and producer surplus (Figure 14-7), which represents the total benefit to consumers at that equilibrium, is the area between the demand and the supply curve to the left of the equilibrium point A as shown in Figure 14.8. (Recall that consumer surplus is the difference between what consumers are willing to pay and what they actually have to pay for all units of the good up to x^0.)

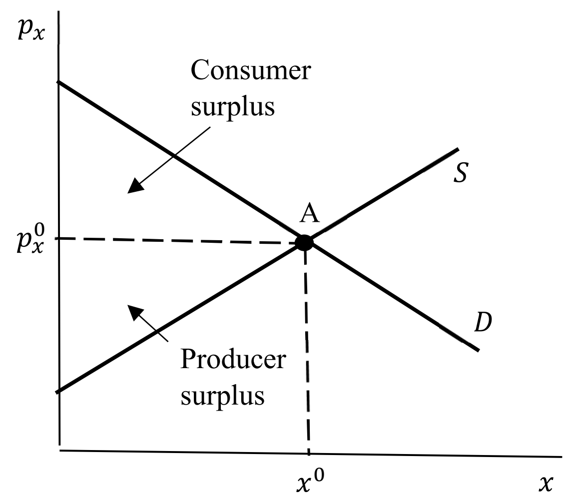


Figure 14.8

Chapter 15
Short-Run Input Demand Functions and Curves, and Long-Run Equilibrium

Fix $p_x = p_x^0$ and $p_k = p_k^0$ and consider the short-run input demand function for labor

$$\ell = g^\ell(p_x^0, p_\ell, p_k^0),$$

with $k = \bar{k}$. The graph of this function (with axes reversed) is the firm's short-run labor input demand curve. In discussing this function and describing its graph, it will be convenient to drop the naughts on p_x^0 and p_k^0.

Recall that input demand along with output supply is determined by profit maximization. With prices p_x, p_ℓ, and p_k specified, the profit function was originally written as

$$\pi = xp_x - \ell p_\ell - kp_k,$$

where (ℓ, k) is the basket of inputs used to produce output x. In previous chapters, using the intersection of isoquants with the short-run expansion path permitted $\ell p_\ell + kp_k$ to be expressed as a function of x in the form of $SRTC^{\bar{k}}(x)$. This also allowed π to be written as a function of x and profit could then be maximized with respect to that variable. The intersection of the isoquant for the profit-maximizing output and the short-run expansion path determined the quantity of labor input that went along with maximum profit. And this quantity of labor was demanded by the firm. But there is another way to obtain the profit-maximizing quantity of labor input. Substituting the total product function $x = TP^\ell(\ell) = f(\ell, \bar{k})$ for x in the above equation, profit can also be written as a function of ℓ:

$$\pi(\ell) = TP^\ell(\ell)p_x - \ell p_\ell - \bar{k}p_k,$$

where k has been fixed at $\bar{k}$ as required in the short run. Now setting the derivative of $\pi(\ell)$ with respect to ℓ at zero to obtain a critical value of ℓ gives

$$\frac{d\pi(\ell)}{d\ell} = \frac{dTP^\ell(\ell)}{d\ell}p_x - p_\ell = 0, \tag{15.1}$$

or, since ${dTP^\ell(\ell)}/{d\ell} = MP^\ell(\ell)$,

$$p_x MP^\ell(\ell) = p_\ell. \tag{15.2}$$

Of course, equation (15.2) is a first-order condition that could identify, at different values of ℓ associated with minimum as well as maximum profit. As in the case of expressing profit as a function of output, the sign of the second-order derivative of $\pi(\ell)$ at a value of ℓ satisfying (15.2) determines whether that ℓ corresponds to a maximum or minimum. Using (15.1), if for some $\ell = \ell^0$

$$\frac{d^2\pi(\ell^0)}{d\ell^2} = \frac{dMP^\ell(\ell^0)}{d\ell} p_x < 0, \tag{15.3}$$

then at ℓ^0 profit is maximized. If the sign is reversed, then ℓ^0 is associated with minimum profit.

The expression $p_x MP^\ell(\ell)$ has two names. One is the value of the marginal product function written $VMP^\ell(\ell)$; the other is the marginal revenue product function written $MRP^\ell(\ell)$.[1] In either case, $p_x MP^\ell(\ell)$ reflects the additional revenue accruing to the firm upon selling the output produced by hiring an additional unit of labor input, or from hiring the last unit of that input. Focusing on the case in which inequality (15.3) holds, equation (15.2) asserts that if the firm is to maximize its profit, it should hire labor up to the point at which the firm pays the last unit hired what that unit is worth in terms of revenue receipts to the firm. Using the former nomenclature so that $VMP^\ell(\ell) = p_x MP^\ell(\ell)$, equation (15.2) becomes

$$VMP^\ell(\ell) = p_\ell. \tag{15.4}$$

If, for example, $VMP^\ell(\ell) > p_\ell$, then the additional cost of hiring one more unit of labor (p_ℓ) is less than the additional revenue obtained by hiring it. Profit would be increased upon hiring that unit. If $VMP^\ell(\ell) < p_\ell$, then the last unit of labor hired costs more than the revenue it produces, and profit would be increased by letting it go. In either case, profit is not maximized.

Profit maximization with respect to ℓ and that with respect to x are two different ways of obtaining the same result. It can be shown (see Supplemental Note F) that:

1. If the firm hires labor up to the point at which $VMP^\ell(\ell) = p_\ell$, then the firm is also, at the same time, producing an output for which SRM$C^{\bar{k}}(x) = p_x$.
2. If the firm produces output up to the point at which SRM$C^{\bar{k}}(x) = p_x$, then the firm is also, at the same time, hiring labor such that $VMP^\ell(\ell) = p_\ell$.

Thus, having one form of profit maximization automatically guarantees having the other.

Turning to the geometry, since p_x is a fixed number and $VMP^\ell(\ell) = p_x MP^\ell(\ell)$, the general shape of the graph of $VMP^\ell(\ell)$ is similar to that of $MP^\ell(\ell)$. The shape of the latter is taken here to be the same as that portion of the MP^ℓ curve pictured in Figure 10-2 of Chapter 10 above the ℓ-axis. Thus, the graph of $VMP^\ell(\ell)$ is that of $MP^\ell(\ell)$ stretched upward if $p_x > 1$ as in Figure 15-1, or compressed downward if $p_x < 1$. The negative slope to the right of ℓ^* reflects

---

In general, there is a conceptual difference between the value of the marginal product and the marginal revenue product. Definitionally, $VMP^\ell(\ell) = p_x MP^\ell(\ell)$ on the one hand and $MRP^\ell(\ell) = MR(x)MP^\ell(\ell)$ on the other. In a perfectly competitive environment, $MR(x) = p_x$ from equation (13.4) of Chapter 13 so that the two concepts amount to the same thing. As will be seen in Chapter 22, under imperfectly competitive conditions such as monopoly $MR(x) < p_x$. In that context, the appropriate concept is $MRP^\ell(\ell)$ and $VMP^\ell(\ell)$ no longer represents the additional revenue accruing to the firm upon selling the output produced by hiring an additional unit of labor input.

diminishing returns to the fixed factor $\bar{k}$, and is consistent with the law of diminishing returns (Chapter 11).

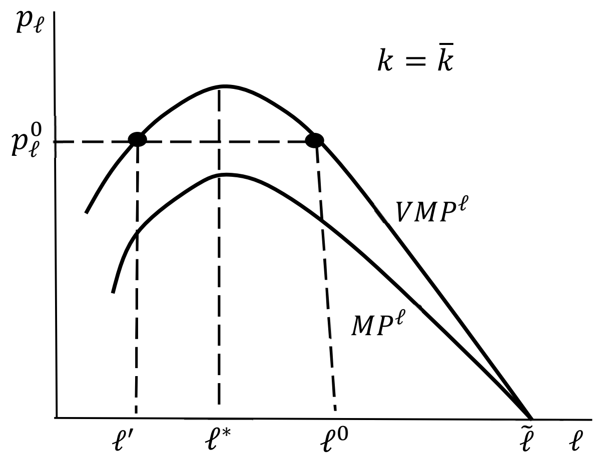


Figure 15-1

Recall that the marginal product curve in Figure 10-2 of Chapter 10 crosses the horizontal axis and becomes negative. At the point of crossing $MP^\ell(\ell) = 0$ and the basket $(\ell, \bar{k})$ lies on the lower ridge line (Figure 10-3). Negative values of MP^ℓ occur outside the relevant region of the input space and can be ignored. Observe in Figure 15-1 that $MP^\ell(\tilde{\ell}\,) = 0$. Since $VMP^\ell(\ell) = p_x MP^\ell(\ell)$, this implies $VMP^\ell(\tilde{\ell}) = 0$. Thus the $VMP^\ell(\ell)$ function values are nonnegative and the VMP^ℓ and MP^ℓ curves meet on the ℓ-axis at $\tilde{\ell}$.

As with the relationship between the firm's SRMC$^{\bar{k}}$ curve and its output supply curve in the lower part of Figure 13-4 of Chapter 13, at a price of p_ℓ^0 in Figure 15-1, the firm will maximize its profit with a labor input at which the horizontal line at p_ℓ^0 intersects the VMP_ℓ curve, thereby satisfying equation (15.4). But also, in parallel with Figure 13-4, there are two values of ℓ in Figure 15-1 that satisfy (15.4), namely ℓ' and ℓ^0. The latter is associated with maximum profit because the VMP^ℓ curve is downward sloping at ℓ^0 and, as in inequality (15.3),

$$\frac{dMP^\ell(\ell^0)}{d\ell} p_x = \frac{dVMP^\ell(\ell^0)}{d\ell} < 0; \tag{15.5}$$

the former corresponds to minimum profit because the VMP^ℓ curve is upward sloping at ℓ' and the sign in (15.5) is reversed. Since ℓ^0 is profit-maximizing, and since it has been assumed in Chapter 13 that the firm will only demand labor input that maximizes its profit, ℓ^0 will be demanded by the firm at p_ℓ^0. Thus the coordinates of points on the downward sloping portion of the VMP^ℓ curve are the same as those on the firm's input demand curve, that is, on the graph of $\ell = g^\ell(p_x^0, p_\ell, p_k^0)$ with the naughts on p_x and p_k emphasizing fixed price values restored. With one caveat, the firm's input demand curve is therefore the downward sloping portion of the firm's VMP^ℓ curve above the ℓ-axis as shown as the red curve in the right-hand diagram of Figure 15-2. The caveat is an appropriate cutoff that can arise from the cutoff of the firm's output supply curve at the minimum of its average variable cost curve.

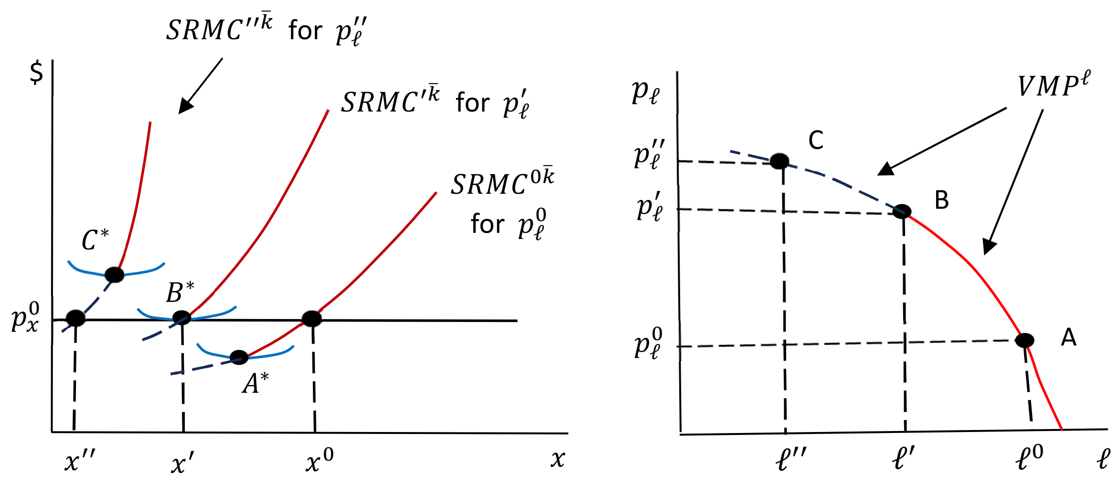


Figure 15-2

To see how that cutoff comes into play, consider Figure 15-2. Suppose the firm, with capital fixed at $\bar{k}$, is at point A on its VMP^ℓ curve employing ℓ^0 units of labor at price p^0_ℓ (right-hand diagram) and producing and selling the profit-maximizing output x^0 at price p^0_x where $SRMC^{0\bar{k}}(x^0) = p^0_x$ (left-hand diagram). In the left-hand diagram, the firm's short-run output supply curve (in red) is its $SRMC^{0\bar{k}}$ curve (in red and black dashed) cut off at A^* where it intersects the minimum point on the firm's average variable cost curve (in blue). As the price of labor rises along the VMP^ℓ curve, the firm's costs at each level of output increase, and its marginal and average variable cost curves shift upward, forcing the minimum point on the latter to rise. But its output price p^0_x, which is determined in the output market, and $\bar{k}$ do not change. At p'_ℓ the short-run marginal-cost-output-supply curve has risen to $SRMC'^{\bar{k}}$and the cutoff point on the latter at B^* lies on the line parallel to the x-axis at p^0_x. The cutoff point B^* is the shut-down point for $SRMC'^{\bar{k}}$ although, as described in Chapter14, the firm continues to operate and produce output x' with labor input ℓ'. Any further increase in p_ℓ, say to p''_ℓ as shown at C in the right-hand diagram, will raise the marginal-cost-output-supply curve still higher to $SRMC''^{\bar{k}}$, pushing the cutoff point to C^* above the horizontal line at p^0_x. When this happens, the short-run marginal cost curve will intersect the p^0_x-line below the minimum point of the also higher average variable costs curve. Total revenue has become less than total variable cost at the profit-maximizing output x'' and input ℓ'', and the firm will shut down (Chapter 14). Thus, quantities of labor associated with points on the curve to the left of B in the right-hand diagram, although falling along its VMP^ℓ curve, are not demanded by the firm. Only the portion of the curve that is red in that diagram constitutes the firm's labor demand curve.

The fact that the input demand curve slopes downward is the property, referred to early in Chapter14, of the short-run input demand function $\ell = g^\ell(p_x, p_\ell, p_k)$ that is implied by the assumptions of the model of firm buying and selling behavior. If the observed input demand function of a firm does not satisfy this property, then the present model does not provide an explanation of that firm's buying behavior.

Return to expressing profit as a function of x and consider the long run. Recall from Chapter 13 that, given output and input prices p_x^0, p_ℓ^0, and p_k^0, long-run profit maximization determines long-run output x^o where the horizontal line at p_x^0 intersects the $LRMC$ curve, and determines long-run inputs ℓ^0 and k^0 where the x^o isoquant intersects the long run expansion path. This geometry is pictured in Figure 15-3. Now the firm does not base its actual production

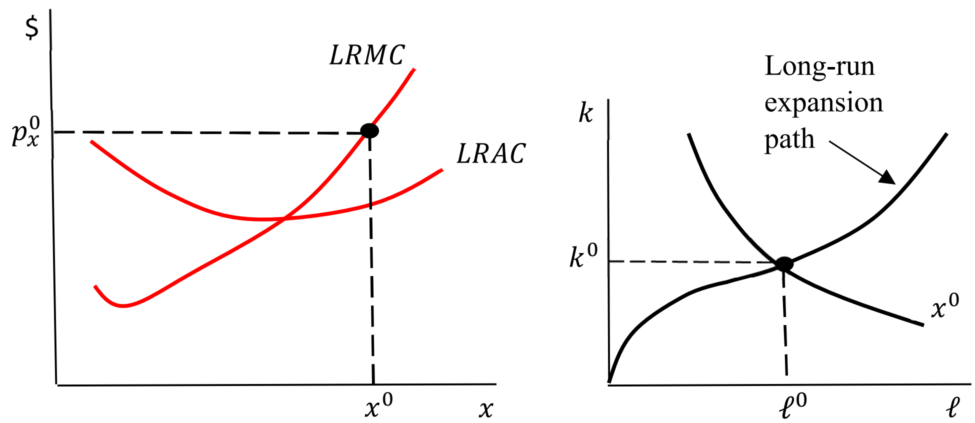


Figure 15-3

on its long-run cost curves. Rather, the k^0 in the right-hand diagram of Figure 15-3 indicates the quantity of capital or the firm size needed for long run profit maximization. The firm first has to create a production facility of size k^0 and then hire labor to produce its output with that facility. But once it does so, it is now operating in the short run with fixed capital k^0, the k^0 becomes the fixed capital $\bar{k}$ of the short-run analysis developed in Chapter 12, and the firm's supply curve is its short-run marginal cost curve for the built facility above minimum average variable cost (Chapter 14).

The cost curves for this aspect of the firm's operation were described in Chapter 12 where the blue short-run average cost curve for the firm of size $\bar{k}$ in the lower part of Figure 12-8 is tangent there at A to the red long-run average cost curve at an output x' smaller than that at minimum long-run average cost. In that diagram, the two average cost curves (blue and red) slope downward at x' and the short- and long-run marginal cost curves intersect directly below the tangency. A slightly different version is presented in Figure 15-4 where firm output is larger that that associated with minimum long-run average cost (and $\bar{k}$ is the same as k^0 in Figure 14-3). The tangency at A between (blue) short- and (red) long-run average cost curves occurs in the upward sloping portion of those curves at output x^0, and the short- and long-run marginal cost curves intersect above the tangency at the same output. The justification for the relationship between the short- and long-run cost curves in Figure 15-4 is identical to that for the relationshoip between the short- and long-run cost curves in Figure 12-8.

Think of Figure 15-4 as depicting the representative firm in a market or industry. Continuing the interpretation of the perfect information assumption of Chapter 14 that all firms are identical in a perfectly competitive market, every firm in that industry is earning, in addition to normal profit $p_k\bar{k}$, an abnormal profit of $\pi(x^0) > 0$ at x^0. Assume all firms outside of this industry have only normal profit. That is, with p_k fixed by the capital market and therefore the same for all investors, providing capital k for an outside firm will yield only $p_k k$ in return. Because all markets are perfectly competitive and because there is perfect information in

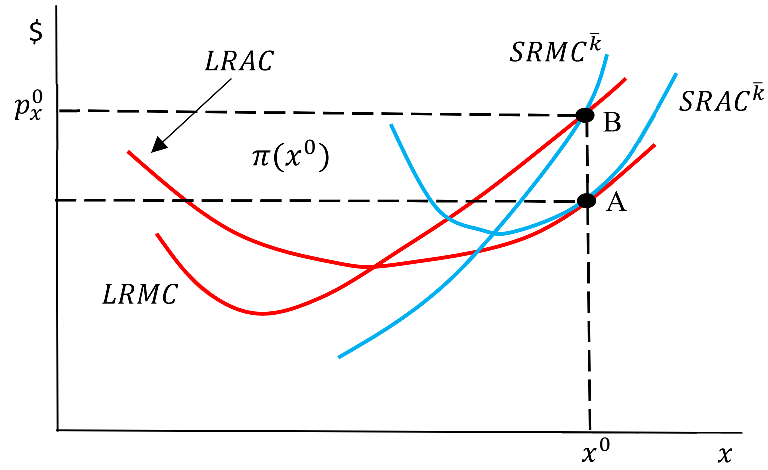


Figure 15-4

perfectly competitive markets, outside investors will be aware of the abnormal profit in the firms of Figure 15-4 and consider the industry an investment opportunity. Were they to enter the market by providing the capital for a new firm, they would earn more than the normal profit $p_k k$ they would receive by investing in outside firms. And, in the long run, the free-entry-into-and-exit-from-the-market property of perfect competition allows them to do just that.

Of course, the entry of one firm, because it is such a small part of the market, will have no impact. But as more new firms enter, there will be more short-run output supply curves to add into the market supply curve, eventually causing the market output supply curve to shift out. With the downward-sloping market demand curve remaining fixed, the market price will fall and firm abnormal profit will shrink. This will continue until abnormal profit no longer exists anywhere in the industry, that is, until $\pi(x) = 0$ for all firms. (The perfect information assumption ensures that all firms, including new entries, have access to the same technology and resources, and therefore have the same fixed capital values, production functions, cost functions and curves, and hence the same supply curves and values for minimum average variable cost.) It is clear from Figure 15-4 that all profit will disappear when the price of output p_x falls to the level of minimum long-run average cost p_x^0, that is, where

$$p_x^0 = \min_x LRAC(x).$$

Similarly, if the market price fell below p_x^0 (perhaps outside investors made mistakes and too many firms entered the market to stop the price from falling below p_x^0), all firms would be suffering losses. In that case, some investors will remove their investment from the industry, that is, sell the capital they had provided in the capital market, thereby causing enough firms to go out of business to force the market price to rise back to p_x^0 and profit to return to normal.

This has an important implication for the firm long-run output supply function $x = g^x(p_x, p_\ell, p_k)$. If the entry and exit of firms driving the market price of output to the minimum point on the long-run average cost curve is to be taken into account, then the firm's output supply can no longer be thought of as the response to a price that is the outcome of the interaction of demand and supply in the market. Rather, the supply of the firm's output along with the output price p_x itself is determined by that which sets the minimum point of the long-run average cost curve, namely input prices and the production function. The firm is still maximizing long- and short-run profit because the long- and short-run marginal cost curve still passes through the minimum points on the tangent long- and short-run average cost curves so that long-run and short-run marginal costs still equal output price. Since, in this case, the output price has lost its independent variable status and does not figure into the determination of the long-run output supply, that price should not appear as an argument in the long-run output supply function. That function should more properly be written as $x = g^x(p_\ell, p_k)$. And this, in turn, implies that when account is taken of free entry and exit, there is no long-run firm, and hence no long-run market output supply curve showing quantities supplied at each price.[2] As described earlier, firm and market supply curves are based on short-run marginal cost curves. This revision of g^x is the modification of the long-run output supply function referred to earlier in Chapter 14.

Long-run equilibrium in relation to the market and its firms, then, which requires all activity to be at rest with no tendency to change, has three parts: First, firms must be producing where short- and long-run marginal cost equals market price so that short- and long-run profits are maximized. Second, market demand has to equal short-run market supply so that competitive market forces are inactive. And third, market price must be at the level of minimum long-run average cost so that profits are exactly normal and no entry or exit of firms occurs. An illustration of this long-run equilibrium in terms of the representative firm and industry is shown in Figure 15-5. As in Figure 14-4, which depicts a short-run equilibrium situation, x_f denotes firm output quantities and x_m represents market quantities. Figure 15-5 continues the color coding of short- and long-run cost and supply curves of Figure 15-4. Remember that the short-

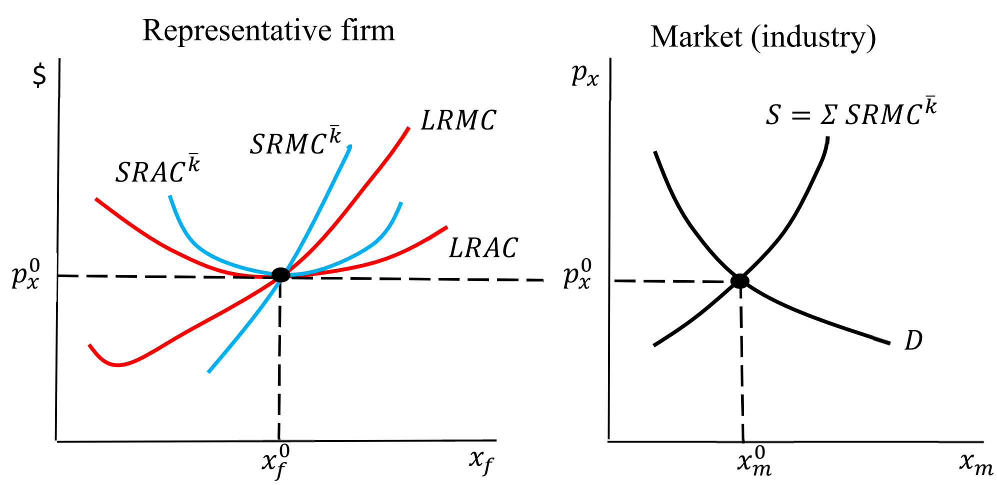


Figure 15-5

[2] There is still, however a conceptually different long-run industry (market) output supply curve that will be considered in the next chapter.

run marginal cost curve in the left-hand diagram of Figure 15-5 is the output supply curve of the representative firm. Of course, the market supply curve in the right-hand diagram is the horizontal sum of the short-run marginal cost curves of all firms in the industry. It is denoted by $S = \Sigma\, SRMC^{\bar{k}}$ in Figure 15-5 and is cut off at the representative firm's minimum average variable cost, although the latter curve is not shown in the left-hand diagram.

It should also be noted that long-run equilibrium involves two implicit minimizations. First, firms must be somewhere on their long-run average cost curves. And to do so, the input costs of producing each level of output has to be minimized (Chapter 11). Second, since free entry into and exit from the industry forces abnormal profit to vanish, the output produced at equilibrium has to minimize long-run average cost.

Based on this discussion, it is clear that the determination of the (equilibrium) price in an output market depends on the time frame under consideration:

1. In the short run, market price is determined by the interaction of the competitive forces of supply and demand as described in Chapter 3.

2. In the long run, the competitive forces of demand and supply still operate with respect to demand and short-run supply curves. But in the end, due to the entry and exit possibilities of firms, market price is determined by the costs of production, that is, by input prices in conjunction with technology or the production function, as they determine minimum long-run average cost.

Chapter 16
Long-Run Output Supply and Input Demand, and Economic Rent

At the end of Chapter 15 it was indicated that there is no long-run supply curve for the individual firm and no long-run market supply curve in a perfectly competitive industry. But there is a sense in which a long-run industry supply curve, that is, a supply curve for the market as a whole or all firms in the industry combined, can be characterized. Toward that end, recall that under the interpretation of perfect competition employed in this volume, all firms have the same cost curves (Chapter 14). Recall also that the long-run equilibrium market price of output is located at the level of minimum firm long-run average cost (Chapter 15).

Now consider an industry and its firms in an initial long-run equilibrium at price p_x^0 and quantities x_m^0 and x_f^0as pictured in Figure 15-5 and reproduced as part of Figure 16-1. In the left-

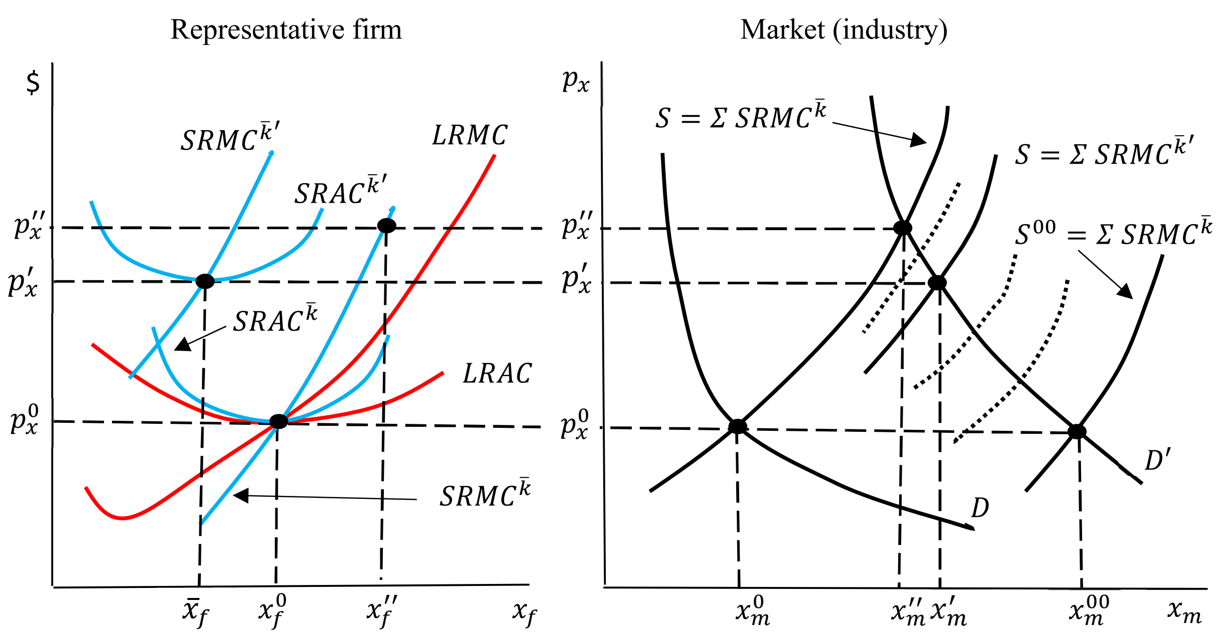


Figure 16-1

hand diagram of Figure 16-1, the profit of the representative firm at this equilibrium is normal, that is, $\pi(x_f^0) = 0$. Suppose market demand increases from D to D' as indicated in the right-hand diagram. Then in the short run, the market price rises to p_x'', and since there can be no change in firm capital input, there is no change in existing firms' cost curves. In particular, $LRAG, LRMC, SRAC^{\bar{k}}$, and $SRAC^{\bar{k}}$ in the left-hand diagram remain fixed. But firm output rises to x_f'', market quantity rises to x_m'' in the right-hand diagram, profit increases, and abnormal profit $\pi(x_f'') > 0$ is now present. As discussed in Chapter 15, in the long run new firms enter, the short-run market supply curve eventually shifts to the right, and the market equilibrium price begins to fall, lowering abnormal profit. The assumption that all firms are identical means that entering firms have the same cost functions and curves as the existing firms. The new long-run equilibrium after all changes work themselves out and normal profit is restored in all firms depends on what happens in the input markets. There are three possibilities:

1. The **constant cost industry.** In this case, the industry is such a small part of the input markets that the increase in input demands stemming from the entry of new firms has no effect on input prices. In parallel with the short-run, $LRAG, LRMC, SRAC^{\bar{k}}$, and $SRAC^{\bar{k}}$ remain fixed throughout the entire long-run period, and the minimum points on firms' long-run average cost curves do not change. New firms identical to existing firms enter and the market price falls until the long run equilibrium price p_x^0 is re-established. Of course there are now more firms producing the industry's output. So although the short-run supply curves of the individual firms have remained the same, the short-run market supply curve has moved from S to S^{00}, and the long-run equilibrium output x_m^0 has increased to x_m^{00}.

2. The **increasing cost industry**. Here input prices rise due to the increase in demand for inputs from both existing and entering firms. Costs rise, and long-run and short-run average and marginal cost curves shift upward for all firms. [1] This squeezes abnormal profit between the falling output price as new firms enter and the rising costs as input prices increase. Long-run equilibrium reestablished where existing and new firms employ a quantity of capital $\bar{k}'$ different from $\bar{k}$ that maximizes long-run profits and for which normal profit is restored at a price p_x' equal to the minimum on the new long-run average cost curve. That minimum, indeed the entire curve, $SRAC^{\bar{k}'}$ in the left-hand diagram of Figure 16-12,has shifted upwards (the long-run average cost curve does not appear in Figure 16-1 – only the $SRAC^{\bar{k}'}$ curve which would be tangent to the new long-run average cost curve at its minimum point is shown). Abnormal profits have been wiped out, the market price p_x' is higher than p_x^0, and the market quantity x_m' is lower than x_m^{00}. Note that even though the output of the representative firm has fallen at the new long-run equilibrium in the example of Figure 16-1, the long-run market equilibrium quantity has become larger (it has increased from x_m^0 to x_m') due to the increase in market demand and the entry of new firms.

3. The **decreasing cost industry** (not shown in Figure 16-1). In this case, input prices fall perhaps because the increased demand for, say capital, from the entry of new firms allows for the introduction of cost-saving technology in the capital-producing industry. With the lower input prices, the long- and corresponding short-run cost curves and their minimum points fall, the long- run market equilibrium price is lower than the p_x^0, and the long-run market equilibrium quantity is greater than the x_m^{00} of the right-hand diagram of Figure 16-1.

[1] It is implicit here that when the price of capital changes, both new and existing firms are assumed to be subject to the new price. For existing firms, then, even though they are using the same capital as before, their capital costs have modified accordingly.

With these ideas in mind, were all possible increases in market demand between D and D' for the constant- and increasing-cost industries allowed in the right-hand part of Figure 16-1, and were the decreasing-cost industry case added, curves formed by the long-run market equilibrium points in each case would be generated. The curves for each of the three cases are illustrated in green in Figure 16-2. They end at D' where the short-run market supply curves S', S'', and S''' in, respectively, the increasing-, constant-, and decreasing-cost industry cases intersect D'. For example, in the constant cost case the collection of all long-run equilibrium points would lie along a straight line parallel to the x_m-axis at the level of the long-run equilibrium price p_x^0 between D and D' because the minimum point of firm long-run average costs curves do not change as new firms enter. The green lines are often referred to as long-run industry supply curves. But they are not supply curves in the usual sense of the term. Typically

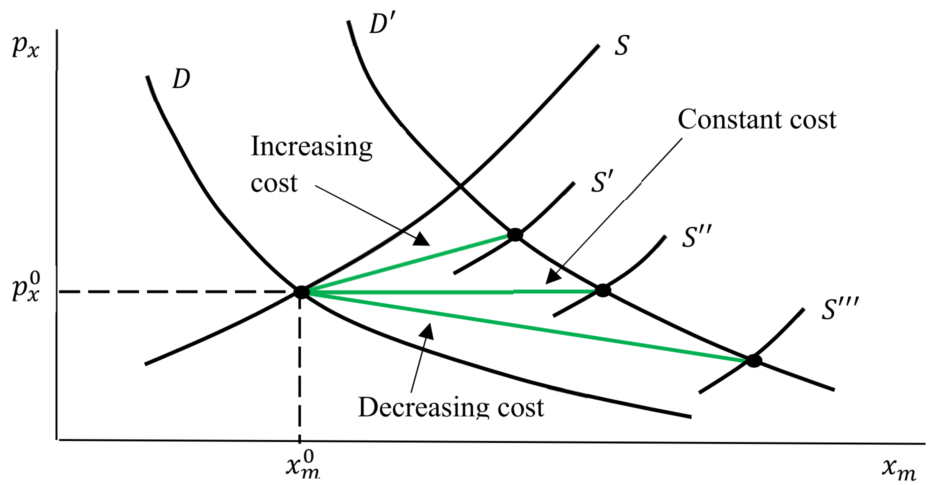


Figure 16-2

market supply curves show the market quantities supplied at various prices as dictated by the market. But in the present long-run circumstance, the output market cannot specify a collection of output prices to which quantities supplied can be associated. Rather, it is for each pair of input prices that a long-run output price and firm-profit-maximizing quantity supplied are determined, namely, those relating to minimum firm long-run average cost. Summing the latter gives the market quantity supplied at long-run equilibrium. Thus the long-run-supply-quantity response at the market level is to both input prices – not to the output price. The long-run industry supply curves described here do indicate long-run industry supply, but only in terms of long-run equilibrium values.

Even though there is no firm long-run output supply curve, the firm does have long-run demand curves for its inputs. Only that for labor will be considered here. In the short-run, recall, the input demand curve for labor is the firm's value-of-marginal-product-with-respect-to-labor curve (identified by the symbol VMP^{ℓ}) for a fixed production facility represented by a specific value for capital k (Chapter 15). Recall also that, in the long run once that value of k is determined by profit maximization, the firm builds the production facility associated with that k-value and operates with respect to the short-run curves identified with it. Thus the profit-maximizing long-run quantity of labor demanded emerges using the VMP^{ℓ} curve associated with that k-value. Were the profit-maximizing value of k to change diring the long-run period, the firm would build a new facility and its VMP^{ℓ} curve would adjust accordingly. The long-run quantity of labor demanded, then, is always identified with a short-run VMP^{ℓ} or short-run labor demand curve. But the derivation of the long-run labor demand curve is complicated by the fact that, in the long-run context, both labor and capital are variable.

Before explaining why, it is worth pointing out that there is a partial parallel here between consumer demands for goods and long-run firm demands for inputs. Although there is no long-run-short-run distinction in the context of the consumer, both consumers and firms are making decisions to demand baskets containing two variable elements. For the consumer, as the price of one good, say good x, falls with the price of the other good, y, and the consumer's income fixed, the movement from the initial basket demanded to that after the decline in price was broken up into income and substitution effects (Chapters 7 and 8). Not only is there a change in the quantity of x demanded, but there is also, according to Figures 7-5 or 8-2, a change in the quantity of y demanded. Since the price of y has remained fixed, this means there has been a shift in the demand curve for that good due to the change in the price of x, leading to a different quantity demanded for y at the same price. With respect to the firm, the decomposition of the change in the quantity of labor demanded due to a reduction in its price is split into three parts and, although the price of capital remains fixed, the quantity of capital demanded alters as did that of y for the consumer.

Suppose, then, that a firm is producing at long-run maximum profit and the price of labor p_{ℓ} falls with p_k and p_x remaining fixed. The first two effects are analogous to the substitution and income effects of the model of consumer buying behavior of Chapter 7. In Figure 16-3, let

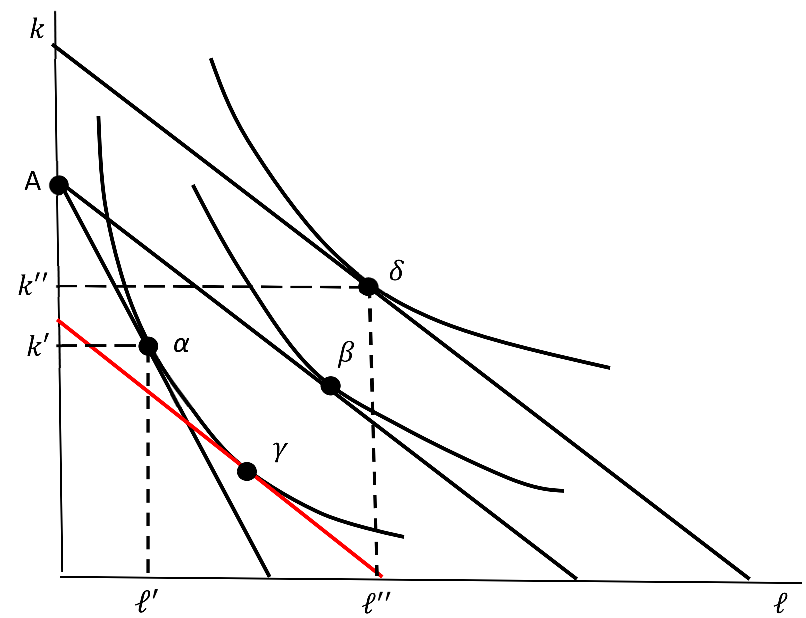


Figure 16-3

the isoquant through α be that for a profit-maximizing output and the coordinates of α, that is (ℓ', k'), represent the cost-minimizing input basket for that output given prices p_x, p_ℓ, and p_k. Then (ℓ', k') lies on the iso-cost line through point A on the k-axis. Suppose p_ℓ falls with the other prices held fixed and suppose the full cost of input baskets, like income in the case of the consumer, remains constant. (In the present context, this supposition is unrealistic and will be with the analogy to the consumer model, a red straight line is drawn parallel to the rotated iso-cost line and tangent to the original isoquant at γ. In this context the movement from α to γ is called the factor substitution effect, and that from γ to β the output effect.

Now remove the supposition that costs have remained constant and remember that the price of labor has fallen while the prices of capital and output have not changed. Then in the new profit-maximizing position, costs will be minimized at a tangency between an iso-cost line (whose slope is the same as that for the iso-cost line through β) and the isoquant corresponding to the new profit-maximizing output. Without further assumptions, it is not possible to tell whether that tangency will lie farther out from the origin than the iso-cost line through β or closer to it. For convenience, it is drawn in the diagram at δ. The movement from β to δ is referred to as the profit effect.

The combination of these effects, factor substitution, output, and profit, show, as pictured in Figure 16-3, that a fall in the price of labor will usually be accompanied by a change in the quantities of both labor and capital. In the example of the diagram, capital increases from k' to k''. This means that the firm changes the size of its production facility in order to maintain long-run profit maximization. And since the short run labor demand curve (its VMP^ℓ curve cut off as pictured in Figure 15-2 of Chapter 15) depends on that production facility as represented in the value of k, if the firm is on one short-run demand curve in its initial long-run profit-maximizing position, after the change in p_ℓ, even with no change in p_k, it will, in its new long-run profit-maximizing position, be on a different short-run demand curve. Thus, in the long run, the short-run demand curve for labor shifts with changes in the price of labor. Using this fact, the long-run labor demand curve can be described as follows: Let input prices, including the price of labor p'_ℓ, be specified. Suppose, as shown in Figure 16-4, the initial profit-maximizing position is at α where the firm is hiring ℓ' units of labor. Its capital is k' and the associated short-run labor

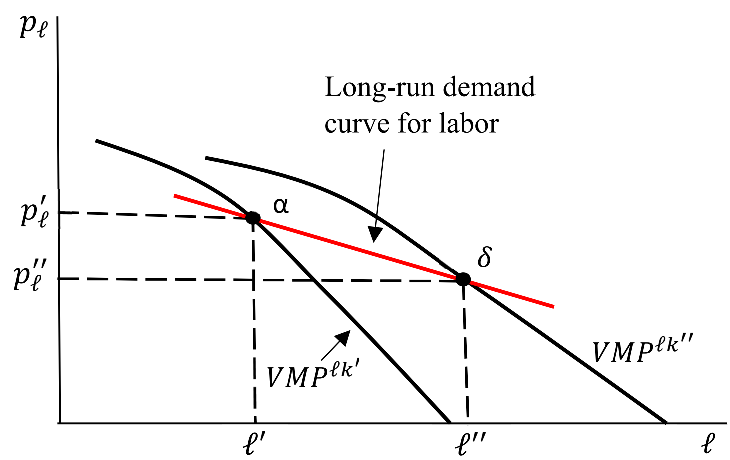


Figure 16-4

demand curve is that labeled $VMP^{\ell k'}$. Then (ℓ', p'_ℓ) is one point on its long run labor demand curve drawn in red in Figure 16-4. Now suppose labor's price falls to p''_ℓ. Capital changes to k'' and the new short-run labor demand curve, labeled $VMP^{\ell k''}$, could either shift back towards the origin or farther away from it. The above picture assumes the latter. The long-run labor demand quantity is now ℓ'' where $VMP^{\ell k''}(\ell) = p''_\ell$ at δ along the new short-run demand curve for labor. Thus (ℓ'', p''_ℓ) is a second point on the long run demand curve for labor. In this way, continually changing the price of labor will generate the full long-run demand curve for labor (the red curve in Figure 16-4). In general, each point of the latter also lies on a different VMP^ℓ curve associated with a different value of capital.

In a factor market, economic rent, or, for short, just rent, is the amount firms have to pay for a unit of the input less the minimum amount necessary to bring it onto the market to sell (supply) summed over all units up to the market equilibrium quantity. In the labor market depicted in Figure 16-5, rent for the unit at ℓ' is the vertical length between α and β.

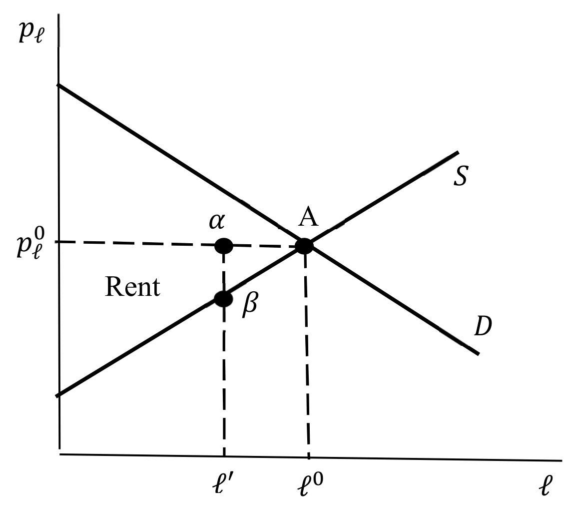


Figure 16-5

Aggregating over all units up to ℓ^0, rent, which is a payment to the sellers of the input, is the triangular area between the supply curve and the horizontal line connecting p_ℓ^0 to the equilibrium point A (for convenience, the supply curve has been drawn as a straight line extended to the vertical axis). Note that if labor were replaced by a produced good like paper in a diagram like Figure 16-5, then what would have been called rent in the context of the labor market, becomes producer surplus received by the selling firms (recall the last part of Chapter 14). Thus, in the labor market, say, rent can be thought of as a kind of surplus for the individuals supplying labor in that $\alpha - \beta$ is the difference between what the individuals obtain by selling the ℓ' unit of labor and the price (at the level of β) they require to supply that unit on the market. In Figure 16-5, rent is only a portion of the full cost of the input $p_\ell^0 \ell^0$ to buyers.

If the supply of the input were fixed and did not vary with is price as, for example, land is sometimes thought to be, then the market supply curve would be a vertical straight line at the fixed quantity of land available and the entire amount paid by buyers $p_\ell^0 \ell^0$ would be rent. This is pictured in Figure 16-6 where ℓ is now taken to represent quantities of land instead of labor and

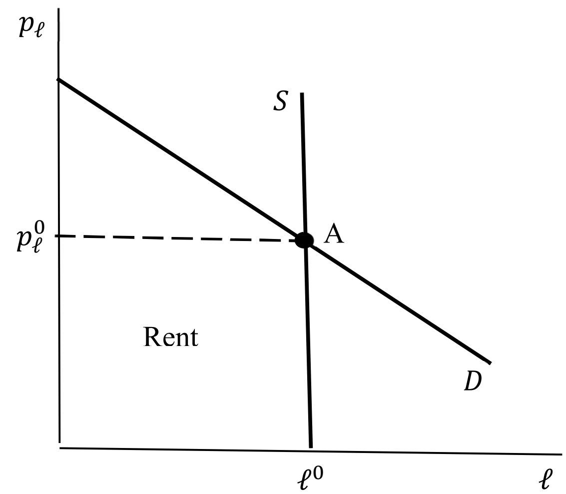


Figure 16-6

ℓ^0 is the fixed supply of land.

Chapter 17
The Perfectly Competitive Microeconomy and General Equilibrium

A. It is now appropriate to take stock of what has been accomplished thus far. The starting point is the problem of explaining how the real microeconomy operates. That world is so complex that it is impossible to describe it without simplification or abstraction. The abstraction employed here appears in the circular flow diagram of Figure 17.1 (reproduced from Figure 1-1)

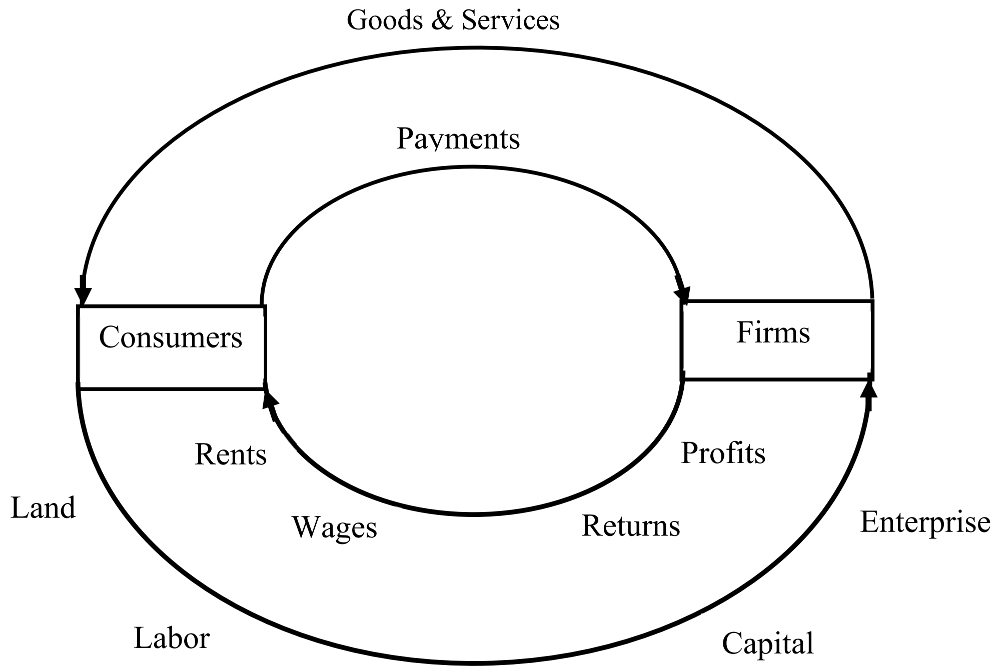


Figure 17-1

in which goods flow from firms to consumers and back to firms, and payments flow from consumers to firms and back to consumers. All goods and payments flow through markets of which there are the 3 kinds listed in Table 17-1 (reproduced from Table 1-1). This abstraction

Table 17-1

Market	Buyers	Sellers
Final Commodities	Consumers	Firms
Intermediate commodities	Firms	Firms
Factors	Firms	Consumers

is understood in terms of a model built to elucidate its inner workings. The model is then used to explain that which can be observed in the real microeconomy. What is considered to be observable and what serves as the focus of explanation is

1. the occurrence of market prices and individual, firm, and market quantities of goods throughout the microeconomy,
2. individual and firm buying and selling behavior as summarized in demand and supply functions.

The model that has been built assumes perfect competition everywhere and consists of several sub-models which are constructed as follows:

B. To explain how a market operates to determine an observed market pair (p_x, x), a model is built by assuming that in that market there exists both a demand and a supply curve passing through (p_x, x)). It is also assumed that the market functions by working through the competitive forces that equilibrate demand and supply. (For example, if buyers cannot buy all they want to buy of a good at its current price, they offer to pay a higher price.) Then the occurrence of observation (p_x, x) is explained as an "equilibrium" resulting from the resolution of the demand-supply competitive forces.

C. To explain consumer buying behavior as summarized in observable consumer demand functions, a model of that behavior is constructed from the following assumptions:

1. The consumer has preferences and indifferences among baskets of commodities such that, for any two baskets, he/she can say if one is preferred to the other or if the two are indifferent. (Completeness.)

2. Preferences and indifferences are transitive. (If basket *A* is preferred to basket *B*, and basket *B* is preferred to basket *C*, then basket *A* is preferred to basket *C*. Similarly for indifferences.)

3. Preferences and indifferences are represented by a utility function in the sense that (a) if one basket is preferred to another, then the preferred basket has a higher utility value than the other, and (b) if two baskets are thought to be indifferent, then they have the same utility value.

4. The utility function has the following 4 properties:
    - 4a. It is continuous and all marginal utilities can be calculated.
    - 4b. A larger basket of commodities is always preferred to, and therefore has a higher utility value than a smaller one.
    - 4c. All indifference curves are strictly convex.
    - 4d. Indifference curve do not touch the co-ordinate axes of the commodity space.

5. The consumer purchases or demands that basket from his/her budget set that provides

the most utility, that is, that maximizes consumer's utility subject to the budget constraint.

Under these assumptions, for each $(p_x, p_y, m) > 0$ the utility-maximizing basket (x, y) can be characterized as occurring at a tangency between a budget line and indifference curve. That basket is then identified as that which is demanded by the consumer at the specified (p_x, p_y, m). This explains the consumer's demand functions (buying behavior) as $x = h^x(p_x, p_y, m)$ and $y = h^y(p_x, p_y, m)$. The functions h^x and h^y have certain properties that are implied by the assumptions listed above.

D. To explain observable consumer selling behavior – only the supply of labor is considered – a model is built employing the same assumptions as in **C** above except that the basket (λ, m) is used in place of (x, y), where m is income from time worked and λ represents non-work or leisure time. With the total time available for all activities symbolized by T, time worked $\ell = T - \lambda$. The price of a unit of leisure time p_ℓ is the wage and the price of a unit of income (one dollar) is \$1. Utility maximization subject to the budget constraint now gives the demand function for leisure time. Subtracting that function from the total time available yields the supply function for labor time, $\ell = h^\ell(p_\ell, T)$, thereby explaining the consumer's supply-of-labor behavior. This function too has specific properties implied by the assumptions made.

E. Given output and input prices (p_x, p_ℓ, p_k). to explain the firm's observable selling and buying behavior as summarized in, respectively, its output supply and input demand functions, a long- or short-run model of the firm is constructed from the following assumptions:

1. The firm has a long-run production function $x = f(\ell, k)$ based on a given technology.

2. The long-run production function has the following properties:
   - 2a. Zero input produces zero output ($f(0,0) = 0$), and nonnegative input produces nonnegative output ($f(\ell, k) \geq 0$ for all $(\ell, k) \geq 0$).
   - 2b. f is continuous and all marginal products can be calculated.
   - 2c. If ridge lines exist, all marginal products are positive and all isoquants are strictly convex between the ridge lines up to an intersection point if there is one.
   - 2d. If ridge lines do not exist, all marginal products are positive, all isoquants are strictly convex everywhere, and no isoquant touches the co-ordinate axes.

3. Long-run and short-run total cost curves appear as drawn in previous chapters so that average and marginal cost curves can be determined and have the shapes attributed to them.

4. The firm hires or demands inputs and produces and sells or supplies output so as to maximize its profit.

Using the following procedure, the long- or short-run profit maximizing output x and basket of inputs (ℓ, k) are identified:

1. If ridge lines exist in the input space, eliminate the regions outside of the area between them and beyond the intersection point if there is one.

2. Using input price information (long run) or fixed capital information (short-run), calculate the appropriate expansion path and confine attention to it.

3. Using the production function, input price, and expansion path information, calculate all cost functions and curves (long- or short-run) expressing cost as a function of output.

4. Using output price information, calculate all revenue functions and curves.

5. Using cost and revenue information, calculate the profit function and the profit-maximizing output x.

6. From the intersection of the isoquant relating to the profit-maximizing output and the appropriate expansion path, calculate the profit-maximizing input quantities – (ℓ, k) in the long run and ℓ in the short run

The profit-maximizing output x and input basket (ℓ, k) define the output and input quantities that the firm will, respectively, sell and buy at the given prices (p_x, p_ℓ, p_k). This explains the firm's long- and short-run output supply input demand functions (behavior). The former are:

$$x = g^x(p_x, p_\ell, p_k), \quad \ell = g^\ell(p_x, p_\ell, p_k) \quad \text{and} \quad k = g^k(p_x, p_\ell, p_k).$$

The short-run output supply and labor demand functions (behavior) are, respectively,

$$x = g^x(p_x, p_\ell, p_k) \quad \text{and} \quad \ell = g^\ell(p_x, p_\ell, p_k),$$

where the symbols g^x and g^ℓ are used in both long- and short-run contexts and k is fixed only in the short run. The functions g^x, g^x, and g^ℓ have certain properties that are implied by the assumptions made. In particular, the long-run output supply function has to account for the long-run, zero-profit requirement implied by the perfect-competition assumption of free entry into and exit from the market.

F. Each of these sub-models by itself describes a part of the abstract microeconomy in partial equilibrium. (The notion of partial equilibrium was introduced in Chapter 1.) That is, an isolated market is in partial equilibrium (or at rest with no tendency to change) when market supply equals market demand (and all other variables throughout the remaining microeconomy are held fixed). Similarly, an isolated consumer is in partial equilibrium when purchasing quantities of goods or selling quantities of factors that maximize his/her utility subject to the budget constraint. And an isolated firm is in partial equilibrium when hiring inputs and producing and selling output so as to maximize its profit.

But partial equilibrium concepts are not enough to explain the operation of the entire microeconomy. The model to be used to explain that economy, or what has in Chapter 1 been called the general-equilibrium or Walrasian model, consists of the combination of all of these sub-models operating in conjunction with each other. That is, in the short run, all markets are operating so as to equate supply with demand, all consumers make buying and selling decisions based on constrained utility maximization, all firms make hiring, production, and selling decisions based on profit maximization, and all of this is happening simultaneously. When this occurs, the model is said to be in general equilibrium. To be more precise (the following definition is reproduced in Supplemental Note G),

the entire model of the micro-economy is at (short-run) general equilibrium at
- (i) quantities of final goods bought and factors sold by each consumer,
- (ii) quantities of inputs bought and outputs produced and sold by each firm, and
- (iii) market quantities and prices of each good,[1]

provided that
- (a) each consumer is buying final goods and selling factors so as to maximize utility subject to his/her budget constraint,
- (b) each firm is hiring inputs and producing and selling outputs so as to maximize its profit,
- (c) supply equals demand in every market, and in the long run only,

.

As in the case of an isolated market, observations of prices and individual and market quantities across the entire real microeconomy at each moment of time are explained as the outcome of the workings of this model, that is, as a general equilibrium. In other words, the prices and quantities that are observed when looking at the real microeconomy are understood as the result of constrained utility maximization by all consumers, profit maximization by all firms, and the interaction of the competitive forces that make supply equal to demand in all of the economy's markets.

For the long-run general equilibrium, a fourth proviso has to be added to eliminate possible movements of capital across markets:

(d) all profits in all firms are zero.

This is necessary because with free entry and exit into markets, the presence of abnormal profits or losses lead to the expansion or contraction of market supplies that result in changes in market prices and quantities.

The next four chapters will consider what are sometimes referred to as the Walrasian model's "welfare properties" of general equilibrium. This will permit an evaluation, from the perspective of society as a whole, of the general equilibrium outcome achieved in the Walrasian model. Since that model is taken as an explanation of each complete set of observations of the

[1] The incomes for each individual appearing in consumer demand functions are obtained from summing the appropriate price-times-quantity values arising in the factor markets.

real microeconomy, that is, since those observations are interpreted as a general equilibrium, these welfare properties and that evaluation may be thought of as applying to them. Thus, accepting the Walrasian model as the explanation of the prices and quantities that are observed in the real microeconomy, what is achieved in that economy can be evaluated in terms of the Walrasian model. The criterion of evaluation to be employed here will be referred to as Pareto optimality (after the economist who is erroneously thought to have developed the idea[2]) or efficiency and its three manifestations will be introduced in the next two chapters.

Generally, by the phrase "welfare properties" is meant how well off society as a whole is. The latter is determined by how well off the individuals of that society are. And that, in turn, depends on the baskets of commodities those individuals have to consume. Such baskets come into being from the economic activity of the Walrasaian model in which the output produced by firms is distributed among consumers through the economy's markets. An individual's well-off-ness in a particular distribution will be measured by utility that person receives from the basket of commodities he/she obtains in the distribution. However, this kind of evaluation has to allow for the possibility that some consumers may be better off with one distribution while others are better off with another and, in such cases, which of the two distributions is better for society as a whole still has to be determined.

Care also has to be taken when using utility as a measure of how well off individuals are. Recall that utility is an ordinal measure in that utility numbers have limited meaning. Nevertheless, each of the many possible utility functions (e.g., $u(x, y)$ or $2u(x, y)$) that represent a particular consumer's preferences and indifferences can serve as a measure of that person's well-off-ness. And it will not matter which utility function is employed as long as the function, once selected, is not changed throughout the discussion. In addition, the possibility of interpersonal comparisons of utility raises another issue. For example, if consumer 1 gets 10 units of utility from an apple and consumer 2 gets 20 units, since utility is a highly personal thing, this does not necessarily mean that consumer 2 is better off with the apple than consumer 1. This limitation has to be respected when evaluating the well-off-ness obtained from society's perspective of different distributions.

In subsequent discussion, it will be convenient to employ a version of the long-run model summarized above in which the zero-profit requirement for long-run equilibrium is ignored and all long-run cost functions and curves, although still derived from constrained cost minimization, are re-interpreted as short-run cost functions and curves. That is, both capital and labor can be bought and sold but new firms cannot enter and old firms cannot leave industries. The number of firms in each market (industry) remains fixed. In addition, it will also be assumed that all marginal cost curves slope upward everywhere and decline to zero as output decreases.[3] With these modifications in force, equilibrium in the full Walrasian model remains a position in which all consumers, firms, and markets are at rest, and the definition of general equilibrium given above applies without condition (d).

---

[2] Jaffè, W., "Pareto Translated: A Review Article," *Journal of Economic Literature* 10 (1972), pp. 1190-1201.

[3] This last assumption is consistent with the requirement that productions functions be strictly concave and therefore do not have ridge lines – a possibility that has been noted in Chapter 10 but is not explored in this book.

To keep matters simple, the only case to be considered is that in which the quantities of both labor and capital supplied by each individual are fixed and do not vary with changes in market prices. This means that the choice between income and leisure described in Chapter 9 no longer determines labor supply and can be ignored. Since market supplies are the sums of individual supplies, it follows that the market supplies in the labor and capital markets are also fixed. These fixed quantities will be denoted by ℓ^* and k^* respectively. For example, everyone might work eight hours per day regardless of the wage paid. Then the market quantity of labor-hours supplied during a day is eight hours times the number of people working and remains fixed at that level regardless of the market price (wage) of labor. The market input supply curves, then, are straight vertical lines at the levels of the quantities of the factors supplied, for example, ℓ^* in Figure 17-2. Of course, the market prices of factors still vary as market demand curves shift. Thus, only output quantity variations among consumers, among firms, and between consumers and firms in the upper part of the circular flow diagram of Figure 17-1 including the consumer and firm boxes are considered. In addition, quantity variations of the fixed input supplies hired by firms are also considered, but not market level variations in those supplies between consumers and firms that are pictured in the lower part of Figure 17-1.

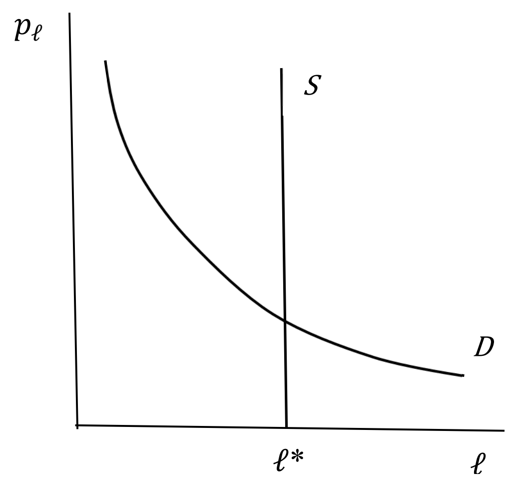


Figure 17-2

Moreover, the simplifying assumptions earlier chapters are continued here with certain amplifications. In particular, the following are worth noting: There are only two persons, two consumer goods, two firms, and two factors. Each consumer buys both goods and supplies (in fixed amounts) both factors. There is only one firm in each industry producing and suppling its unique product using both factors. Thus, there are two output markets each with two buyers and one seller, and there are two factor markets each with two buyers and two sellers. Finally, the assumption that all markets are perfectly competitive is also continued. This last assumption is understood in this context to mean that each consumer and firm, in spite of their small numbers and relatively large size, takes prices as fixed as determined by the markets. Buying and selling behavior is therefore exactly what it would be if the original perfectly competitive requirements of a "large" number of "small" buyers and sellers in each market were present.

Chapter 18
Consumption Edgeworth Box and Consumption Pareto Optimality

To begin a discussion of the welfare properties of general equilibrium in the simplified model described at the end of the previous chapter, imagine the economy's firms produce a specified amount x^0 and y^0 of, respectively, the two goods x and y. These are distributed through the economy's markets to the two consumers labeled person 1 and person 2. Let

x_1 and y_1 represent the quantities of the two goods going to person 1, and
x_2 and y_2 represent the quantities of the two goods going to person 2.

Then since, as described in the circular flow diagrams of Chapters 1 and 17, everything produced by firms is bought by consumers, $x_1 + x_2 = x^0$ and $y_1 + y_2 = y^0$, and (x_1, y_1, x_2, y_2) represents a distribution of x^0 and y^0 among persons 1 and 2. Denote the utility functions of the two individuals by

$$\mu_1 = u^1(x_1, y_1) \quad \text{and} \quad \mu_2 = u^2(x_2, y_2),$$

and assume they have all of the properties attributed to utility functions earlier (Chapter 5, or Supplemental Note C). Even though μ_1 and μ_2 are ordinal in nature, as long as their values remain the same for each basket (x_1, y_1) or (x_2, y_2) throughout the present discussion and are not, say, doubled in the middle of it, they may still serve as indicators of individual well-off-ness. For example, the more a person has of goods x and y, the better off that person is.

There are many ways to allocate or distribute x^0 and y^0 between persons 1 and 2. Several possibilities are illustrated in Figure 18-1 which contains the commodity space for person 1 on

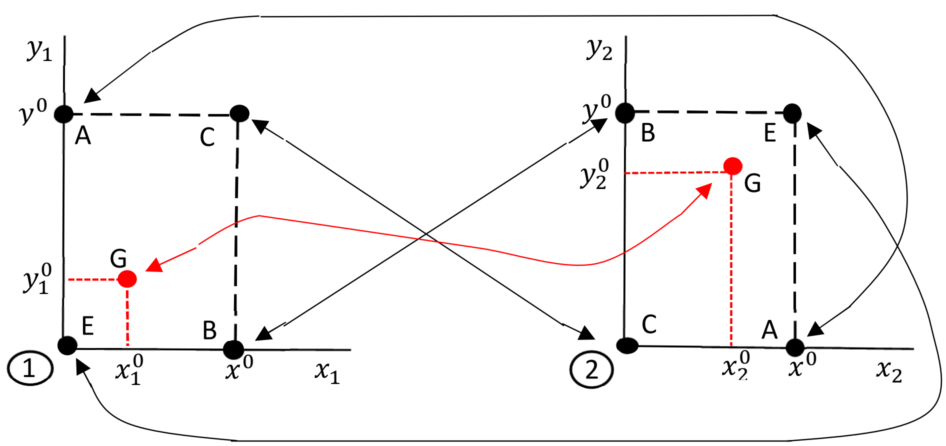


Figure 18-1

the left and that for person 2 on the right (indicated by encircled numbers at their respective origins). Remembering that $x_1 + x_2 = x^0$ and $y_1 + y_2 = y^0$, the distribution

$(x_1, y_1, x_2, y_2) = (0, y^0, x^0, 0)$ is identified by the letter "A" in both commodity spaces. In this distribution person 1 has all of good y that has been produced and no x, and person 2 has all x that has been produced and no y,

$(x_1, y_1, x_2, y_2) = (x^0, 0, 0, y^0)$ is identified by the letter "B" in both commodity spaces. In this distribution person 1 has all of good x that has been produced and no y, and person 2 has all y that has been produced and no x,

$(x_1, y_1, x_2, y_2) = (x^0, y^0, 0, 0)$ is identified by the letter "C" in both commodity spaces. In this distribution person 1 has all that has been produced of both goods while person 2 has nothing,

$(x_1, y_1, x_2, y_2) = (0, 0, x^0, y^0)$ is identified by the letter "E" in both commodity spaces. In this distribution person 2 has all that has been produced of both goods while person 1 has nothing,

$(x_1, y_1, x_2, y_2) = (x_1^0, y_1^0, x_2^0, y_2^0)$ is identified by the letter "G" in both commodity spaces. In this distribution person 1 has (x_1^0, y_1^0) and person 2 has (x_2^0, y_2^0), where $x_1^0 + x_2^0 = x^0$ and $y_1^0 + y_2^0 = y^0$.

In each distribution the paired baskets of commodities of that distribution are connected by a line. Similar identification of distributions can be made for all other pairs of baskets in the boxes outlined by the solid and dashed lines in the two commodity spaces.

However, this is a rather awkward way of identifying distributions. When there are only two goods and two persons, the geometry can be simplified considerably. First rotate the right-hand commodity space of person 2 180° as in Figure 18-2 maintaining the lines connecting

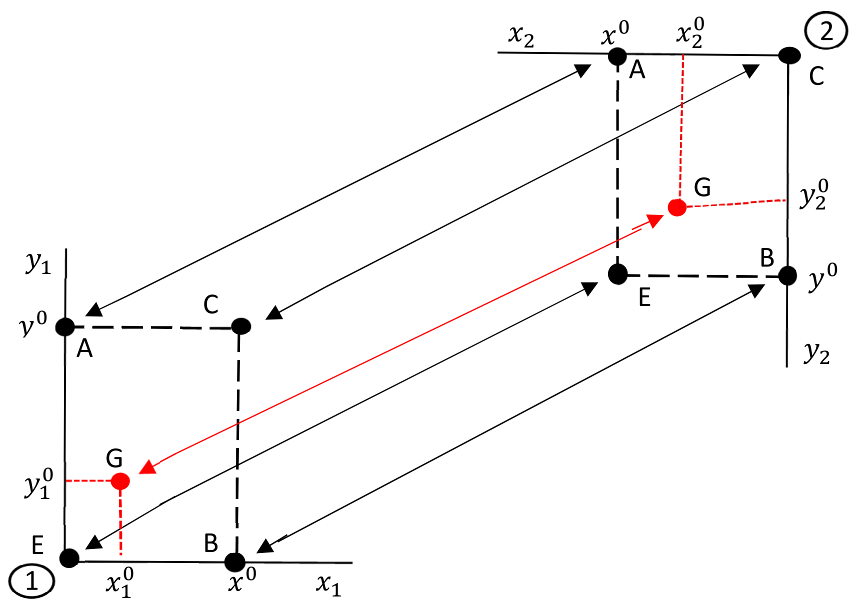


Figure 18-2

paired baskets. Then place it on top of the left-hand commodity space of person 1 as in Figure 18-3 (drawn to a slightly larger scale) so that the baskets identified with the letters A, B, C, E,

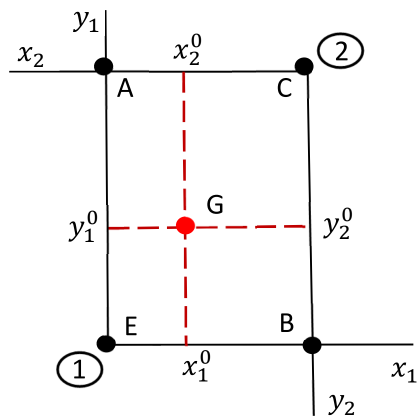


Figure 18-3

and G in one commodity space coincide with their counterparts in the other. The resulting picture is called the consumption Edgeworth box.[1] Observe that each point in the box has four coordinates – two read off the coordinate system of person 1 and two coming from that of person 2. These coordinate values specify a unique allocation or distribution $(x_1^0, y_1^0, x_2^0, y_2^0)$ of x^0 and y^0 among the two persons such that $x_1^0 + x_2^0 = x^0$ and $y_1^0 + y_2^0 = y^0$. The length of the box is determined by quantity x^0, its height by quantity y^0. This is illustrated in Figure 18-4 for a box with different values of the economy's outputs x and y from those of Figure 18-3.

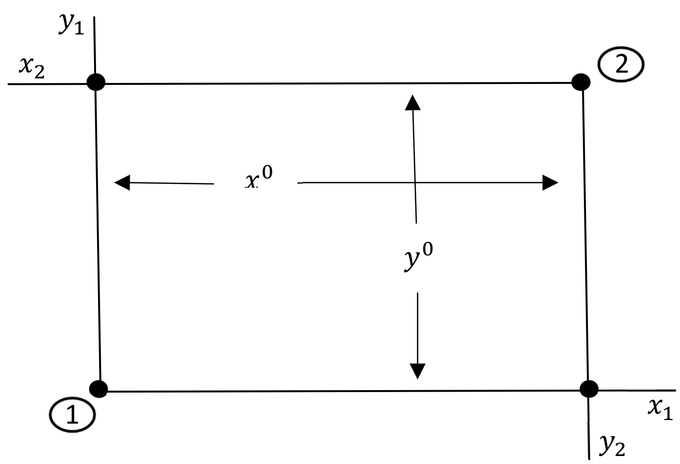


Figure 18-4

Individuals evaluate distributions only with respect to the baskets of commodities they receive in them. The evaluations are expressed in terms of the values assigned to those baskets by their utility functions. Indifference curves in the consumption Edgeworth box reflecting the

---

[1] According to Jaffè cited in footnote 2 of Chapter 17, Edgeworth is the economist who initially introduced the idea of what is called Pareto optimality (mentioned in Chapter 17 and to be defined shortly). Pareto was the first to draw what is referred to as the Edgeworth box.

evaluations are determined by the individuals' utility functions. Those for person 1 are pictured in Figure 18-5 as usual as downward-sloping, strictly convex curves. However, the indifference

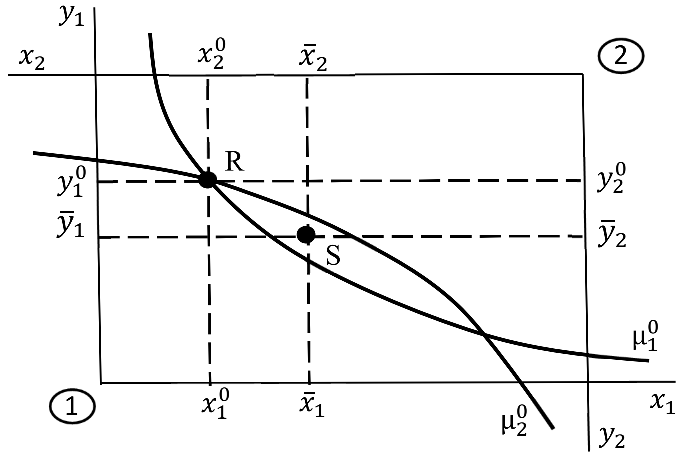


Figure 18-5

curves for person 2 have to be drawn upside down to conform to that person's upside-down commodity space. With respect to that space, person 2's indifference curves are still downward sloping and strictly convex. But in the context of the consumption Edgeworth box diagram, they appear as downward sloping and strictly concave.

Suppose the distribution of goods x and y between persons 1 and 2 occurs at point R in the consumption Edgeworth box of Figure 18-5. The indifference curves of the two individuals through that distribution, (x_1^0, y_1^0) for person 1 and (x_2^0, y_2^0) for person 2, are labeled μ_1^0 and μ_2^0 respectively. In this situation, if person 1 were to trade $y_1^0 - \bar{y}_1$ to person 2 for $\bar{x}_1 - x_1^0$ or, equivalently, if person 2 were to trade $x_2^0 - \bar{x}_2$ to person I for $\bar{y}_2 - y_2^0$ where

$$y_1^0 - \bar{y}_1 = \bar{y}_2 - y_2^0 \quad \text{and} \quad \bar{x}_1 - x_1^0 = x_2^0 - \bar{x}_2,$$

that is, if the distribution moved from point R to point S, then both individuals would be on a higher indifference curve and therefore better off. In fact, any trade that moves the distribution into the area (including its boundary) between the two indifference curves that contains the point S will make at least one person better off without making the other worse off. This illustrates the gains from trade.

But if the indifference curves are tangent at a distribution $(x_1^0, y_1^0, x_2^0, y_2^0)$, labeled Z in Figure 18-6, there is no trade from Z that can make one person better off (higher utility) without making the other worse off (lower utility). Compared to distribution Z, distributions α and γ make both persons worse off, while distributions β and δ make one person better off and the other worse off. Distribution $(x_1^0, y_1^0, x_2^0, y_2^0)$ in Figure 18-6 is referred to as efficient or Pareto optimal in consumption. In general, a distribution is Pareto optimal in consumption if it is impossible through redistribution of its quantities of goods among consumers to make one

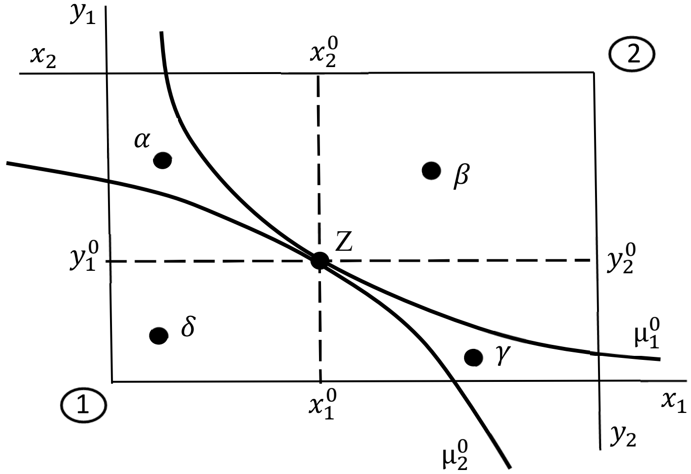


Figure 18-6

person better off without making the other worse off. As shown in Figure 18-6, consumption Pareto optimal distributions are identified in the interior of the consumption Edgeworth box by a tangency between two indifference curves, one for each person. Thus, at such a distribution, the slopes of the two curves, that is, the negatives of their marginal rates of substitution or ratios of marginal utilities (equation (5.3) of Chapter 5), must be equal. That is, eliminating the minus signs,

$$\frac{MU_x^1(x_1^0, y_1^0)}{MU_y^1(x_1^0, y_1^0)} = \frac{MU_x^2(x_2^0, y_2^0)}{MU_y^2(x_2^0, y_2^0)}. \tag{18.1}$$

Consumption Pareto optimal distributions can also be characterized as the outcome of maximizing one person's utility with the other person's utility no lower than a particular value. For example, in Figure 18-6, let the minimum utility for person 2 be that along the indifference curve labeled μ_2^0. Think of that indifference curve as a constraint analogous to the consumer's budget constraint in the explanation of consumer buying behavior of Chapter 6 with every distribution to its left as a possibility providing higher utility to person 2. Maximizing person 1's utility over these possibilities leads to the tangency at Z which, of course, is a consumption Pareto optimal distribution. (Alternatively, maximizing person 2's utility subject to person 1's utility constraint would lead to the same result.) The second-order condition ensuring a maximum is derived from the assumptions made on the utility function in the same way that the second-order condition for utility maximization subject to the budget constraint is obtained in Chapter 6. The specific argument leading to the second-order condition is not pursued here.

Regardless of whether consumption Pareto optimality is formulated in terms of maximization or not, the utility value of one person, in the previous maximization example μ_2^0, has to be specified to identify the indifference curve needed to locate the tangency. Like equation (6.2) of Chapter 6 with regard to utility maximization subject to the budget constraint, equation (18.1) by itself does not fully characterize the distribution at the tangency in Figure 18-6. A full description, that is the complete set of first-order conditions, has to include the equation of the indifference curve constraint evaluated at the tangency point, say, to be consistent with the previous discussion,

$$\mu_2^0 = u^2(x_2^0, y_2^0). \tag{18.2}$$

To complete the description, it is also necessary to add in the equations that confine the distribution $(x_1^0, y_1^0, x_2^0, y_2^0)$ to the consumption Edgeworth box, namely,

$$x_1^0 + x_2^0 = x^0 \quad \text{and} \quad y_1^0 + y_2^0 = y^0, \tag{18.3}$$

where x^0 and y^0 are the outputs produced by the economy's firms. These equations also assert that all produced output of goods x and y is distributed in distribution $(x_1^0, y_1^0, x_2^0, y_2^0)$,

The collection of all Pareto optimal in consumption distributions (tangencies between two indifference curves) in the consumption Edgeworth box is called the consumption contract curve. An illustration with three pairs of tangent indifference curves is pictured in Figure 18-7.

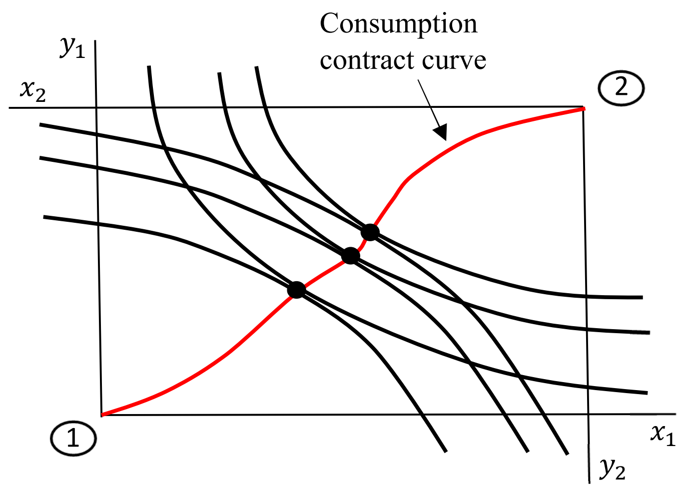


Figure 18-7

To keep the diagram simple, the individual indifference curves are not labeled. The consumption contract curve (in red) necessarily connects and includes the two origins of the box. For example, if the distribution is at the origin of person 2's commodity space, then person 1 receives all of the economy's produced output. And it is not possible to make person 2 better off (higher utility) by taking a small amount of x and/or y away from person 1 and giving it to person 2 without making person 1 worse off (lower utility). Thus, that distribution is Preto optimal and lies on the contract curve. There is no tangency between two indifference curves here because person 2 has no indifference curve through the origin of his/her commodity space.[2] But the idea of Pareto optimality in consumption still applies. It also follows from previous discussion that whenever the two persons are placed at a distribution of x^0 and y^0 that lies off of the consumption contract curve, it is in both their interests to trade until the distribution of goods among them arrives on the contract curve.

[2] Recall that it has been assumed that indifference curves do not touch the co-ordinate axes of the commodity space. In particular, the co-ordinate axes cannot appear as indifference curves.

Chapter 19
Production Edgeworth Box, Production Pareto Optimality, Production Possibilities, and General Pareto Optimality

Now redraw the consumption Edgeworth box of Chapter 18 by replacing

1. total outputs produced x^0 and y^0 by total fixed factors supplied ℓ^* and k^*.
2. consumers 1 and 2 by firms producing goods x and y,
3. commodity spaces by input spaces with labor and capital inputs (ℓ_x, k_x) for the firm producing x and (ℓ_y, k_y) for the firm producing y, and
4. utility functions $\mu_1 = u^1(x_1, y_1)$ and $\mu_2 = u^2(x_2, y_2)$ and their indifference curves by production functions, now written $f^x(\ell_x, k_x)$ for the firm producing x and $f^y(\ell_y, k_y)$ for the firm producing y, and their isoquants,

as shown in Figure 19-1. This diagram is known as the production Edgeworth box. Its

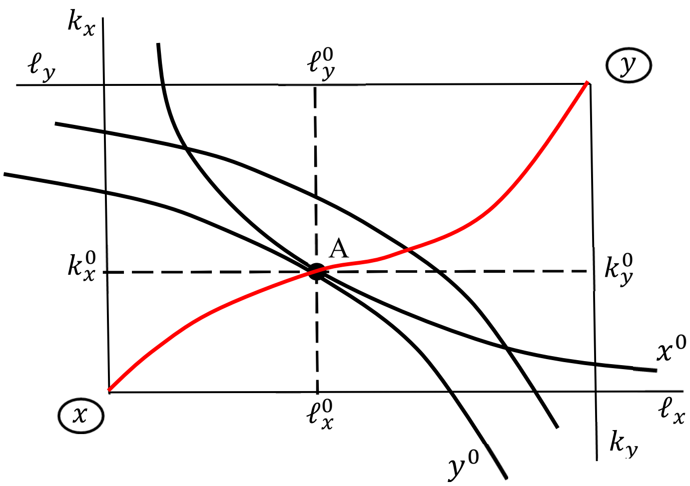


Figure 19-1

characteristics are analogous to those of the consumption Edgeworth box described in Chapter 18:

1. The dimensions of the box are determined by the fixed supplies of factors ℓ^* and k^* coming from consumers.

2. Each point in the box represents a distribution $(\ell_x, k_x, \ell_y, k_y)$ of the fixed factor supplies among the two firms, where $\ell_x + \ell_y = \ell^*$ and $k_x + k_y = k^*$.

3. Point A in the box, where two isoquants are tangent, is efficient or Pareto optimal in production, that is, by redistributing the fixed factor supplies, it is not possible to increase the output of one firm without lowering that of the other. That distribution can

also be described in terms of maximizing the output of one firm subject to the fixed output or isoquant constraint of the other.

4. The curve in red connecting the two origins and made up of the tangencies of isoquants is the collection of all production Pareto optimal distributions in the box. It is called the production contract curve.

5. The equation describing the equality of slopes at the tangency distribution $(\ell_x^0, k_x^0, \ell_y^0, k_y^0)$ is, eliminating the minus signs (from equation (10.3) of Chapter 10),

$$\frac{MP_\ell^x(\ell_x^0)}{MP_k^x(k_x^0)} = \frac{MP_\ell^y(\ell_y^0)}{MP_k^y(k_y^0)}, \tag{19.1}$$

where, as with the production functions, the superscripts x and y have been added to distinguish between the marginal products of the two firms.

6. The equations that fully describe the distribution $(\ell_x^0, k_x^0, \ell_y^0, k_y^0)$ at the tangency are (19.1) along with the equation, say for the firm producing x, of the isoquant constraint passing through that distribution and those that confine $(\ell_x^0, k_x^0, \ell_y^0, k_y^0)$ to the production Edgeworth box:

$$x^0 = f^x(\ell_x^0, k_x^0), \tag{19.2}$$

$$\ell_x^0 + \ell_y^0 = \ell^* \quad \text{and} \quad k_x^0 + k_y^0 = k^*. \tag{19.3}$$

Thus, in parallel to the case of consumers where marginal rates of substitution are equal at consumption Pareto optimal distributions (equation (18.1) of Chapter 18), with respect to firms, marginal rates of technical substitution are equal at distributions of inputs that are Pareto optimal in production.

In Figure 19-1, traveling the production contract curve from the origin of the firm producing x to that of the firm producing y, output x rises while output y falls. Plotting those outputs against each other yields what is called the production possibility or transformation curve. An illustration is provided in Figure 19-2. Starting at point α on the production contract curve in the production Edgeworth box on the left and proceeding through points β, γ, δ, and ε, parallels moving along the transformation curve from α to ε on the right. The outputs along the isoquants at points β, γ, and δ in the production Edgeworth box on the left are the coordinates of, respectively, points β, γ, and δ on the transformation curve on the right. The isoquant at α for the firm producing y and that at ε for the firm producing x are not drawn in the left-hand diagram. That is because, as described for the consumption contract curve at the end of Chapter 18, there can be no tangency at those distributions since there is no isoquant for the firm producing x at α and for the firm producing y at ε.[1] The numerical outputs at α and ε are also

[1] In parallel with the case of the consumer (footnote 2 of Chapter 18), this follows from the assumption that isoquants do not touch the co-ordinate axes of the input space.

not indicated in either diagram, although, at α, the output of the firm producing x is zero as is that for the firm producing y at ε. And at α, the output of the firm producing y is that obtained if all of the economy's resources were devoted only to the production of y. A similar property holds at ε for the output of the firm producing x. Note that if either ℓ^0 or k^0 were to change, the dimensions of the production Edgeworth box would alter, and the production contract and

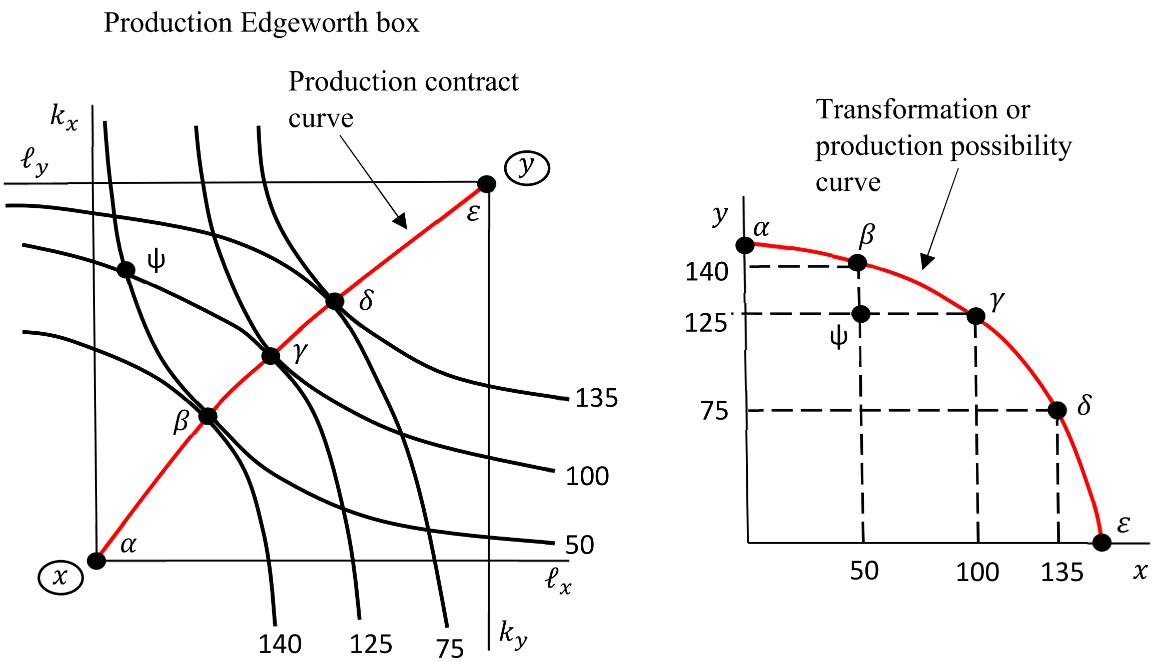


Figure 19-2

transformation curves would alter. The transformation curve is often drawn as a strictly concave curve like that in Figure 19-2.[2] The function whose graph is the transformation curve is written

$$y = T(x) \tag{19.4}$$

for all outputs x that can be produced by firm x given the quantities of labor and capital available.

It has been implicitly suggested above that there is a parallel between the consumption Edgeworth box (Chapter 18) where each tangency can be expressed in terms of the maximization of one person's utility subject to the condition that the other person's utility is no lower than a specified value, and the production Edgeworth box where tangencies can be viewed as the outcome of maximizing one firm's output where the other firm's output is no lower than a specified amount. The parallel first-order conditions here are equations (19.1)-(19.3). Since tangencies along the production contact curve are translated into points along the transformation

[2] Although not demonstrated here, this is entirely consistent with production functions that are themselves strictly concave. Indeed, strict concavity of the latter implies strict concavity of the transformation function. The assumption of strictly concave production functions yielding a strictly concave transformation function is maintained here and in Chapter 20.

curve, this maximization may be restated in terms of the latter as follows: Each point (x, y) on the transformation curve $y = T(x)$ may be thought of as indicating the maximum amount of y that can be obtained from the remaining quantities of the fixed supplies of labor and capital after output x is produced.

All points beneath the transformation curve such as ψ in Figure 19-2 are attainable but inefficient. They correspond to distributions in the production Edgeworth box that are off the production contract curve and for which the isoquants passing through them intersect but are not tangent. This is illustrated in Figure 19-2 where the output coordinates at ψ in the right-hand diagram ($x = 50$ and $y = 125$) are associated with the isoquants corresponding to those outputs that intersect at ψ in the left-hand diagram. Thus, it is always possible at ψ through redistribution of inputs to increase the output of one firm without lowering that of the other. In addition, the factor supplies are insufficient to allow the economy to achieve points beyond the transformation curve.

The transformation curve also shows how the economy converts units of good x into units of good y (or vice versa) when all production takes place as efficiently as possible. This conversion is characterized by the slope (which is negative) of the transformation curve. That slope indicates the rate at which the economy gives up good x in exchange for good y at each output combination (x, y). The negative of this slope is called the marginal rate of transformation and can be shown to be (the demonstration is not provided)

$$\frac{MC^x(x)}{MC^y(y)},$$

where $MC^x(x)$ is the marginal cost at output x for the firm producing x with the short run (SR) and long run (LR) symbolisms dropped.[3] The same letters are discarded in $MC^y(y)$ for the firm producing good y.

There is a more general form of Pareto optimality that encompasses both Pareto optimality in consumption and Pareto optimality production. Its name is just Pareto optimality without the phrase 'in consumption' or 'in production' attached to it and its definition is as follows (it also appears in Supplemental Note G):

A its distribution $(x_1^0, y_1^0, x_2^0, y_2^0)$ of (x^0, y^0) among consumers is Pareto optimal or efficient (in general) provided that:

A. (x^0, y^0) lies on the transformation curve, that is, satisfies equation (19.4),
B. there is no other distribution of (x^0, y^0) **and no distribution of any other pair of outputs on the transformation curve** at which one person is better off (higher utility) without the other person being worse off (lower utility).

[3] Recall that, as described in Chapter 17, the model of Chapters 18-21 combines features of both the long- and short-run models of earlier chapters.

According to previous argument, to say that (x^0, y^0) lies on the transformation curve means that (x^0, y^0) is associated with a distribution of the economy's fixed factor supplies at which the isoquants of the two firms are tangent. That is, Part A of the definition of Pareto optimality implies Pareto optimality in production. And ignoring the words in bold typescript, Part B of that definition asserts that the distribution $(x_1^0, y_1^0, x_2^0, y_2^0)$ of (x^0, y^0) is also Pareto optimal in consumption relative to the consumption Edgeworth box determined by (x^0, y^0). The words in bold characterize the additional content of Pareto optimality that does not appear in Pareto optimality in consumption and production combined. That additional content states that it is not possible to move from $(x_1^0, y_1^0, x_2^0, y_2^0)$ to any distribution of any other pair of outputs on the transformation curve without lowering the utility of at least one person. Alternatively put, the difference between Pareto optimality in consumption and Pareto optimality is that to determine a Pareto optimal distribution, the former compares the distribution in question to distributions from the single consumption Edgeworth box generated by (x^0, y^0) while the latter compares that distribution to distributions from all Edgeworth boxes generated by all points on the transformation curve. Thus, although the general form of Pareto optimality includes both Pareto optimality in consumption and production, the combination of the latter two by themselves do not fully constitute the former.

It can be shown (this is also not demonstrated here) that at a Pareto optimal distribution $(x_1^0, y_1^0, x_2^0, y_2^0)$, the slope of the transformation curve at (x^0, y^0) must be the same as that of each person's indifference curve at (x_1^0, y_1^0) for person 1 and (x_2^0, y_2^0) for person 2. Mathematically (and dropping the minus signs),

$$\frac{MC^x(x^0)}{MC^y(y^0)} = \frac{MU_x^1(x_1^0, y_1^0)}{MU_y^1(x_1^0, y_1^0)} \quad \text{and} \quad \frac{MC^x(x^0)}{MC^y(y^0)} = \frac{MU_x^2(x_2^0, y_2^0)}{MU_y^2(x_2^0, y_2^0)}, \tag{19.5}$$

where $x_1^0 + x_2^0 = x^0$ and $y_1^0 + y_2^0 = y^0$. Thus, at a Pareto optimal distribution, the marginal rate of transformation equals both marginal rates of substitution. In other words, the rate at which the microeconomy can transform the production of good x into the production of good y is the same as the rate at which consumers can substitute x for y without changing their level of utility. But as long as outputs x^0 and y^0 are split up between the two persons, this cannot mean a tangency between their indifference curves and the transformation curve.[4] For example, according to equation (19.5) with minus signs restored, the slope of the transformation curve

$$-\frac{MC^x(x^0)}{MC^y(y^0)} \tag{19.6}$$

occurs at (x^0, y^0), while that of person 1's indifference curve

$$-\frac{MU_x^1(x_1^0, y_1^0)}{MU_y^1(x_1^0, y_1^0)} \tag{19.7}$$

[4] Since the general form of Pareto optimality includes Pareto optimality in consumption, there is still a tangency between the two person's indifference curves at distribution $(x_1^0, y_1^0, x_2^0, y_2^0)$ in the (x^0, y^0) consumption Edgeworth box as described in Chapter 18.

occurs at (x_1^0, y_1^0). Even though the ratio of (19.6) equals that of (19.7), only if person 1 receives all of both outputs so that

$$(x^0, y^0) = (x_1^0, y_1^0)$$

will a tangency between that person's indifference curve and the transformation curve exist.[5]

It is instructive to see geometrically how each Pareto optimal distribution of (x^0, y^0) is associated with a production Pareto optimal distribution $(\ell_x^0, k_x^0, \ell_y^0, k_y^0)$ in the (ℓ^0, k^0) production Edgeworth box and a consumption Pareto optimal distribution $(x_1^0, y_1^0, x_2^0, y_2^0)$ in the (x^0, y^0) consumption Edgeworth box. Thus, in Figure 19-3, the requirement of Part A of the definition of Pareto optimality that (x^0, y^0) be on the transformation curve (shown in the middle third of the diagram on the left) means that the isoquants associated with the outputs x^0 and y^0 are tangent at some distribution $(\ell_x^0, k_x^0, \ell_y^0, k_y^0)$ in the production Edgeworth box whose length and height are determined, respectively, by the fixed supplies of factors ℓ^0 and k^0, and where $\ell_x^0 + \ell_x^0 = \ell^0$ and $k_x^0 + k_y^0 = k^0$. This appears in the lower two-thirds of Figure 19-3. As previously suggested, even though the distribution $(\ell_x^0, k_x^0, \ell_y^0, k_y^0)$ does not explicitly appear in the above definition of Pareto optimality, it is implicit in Part A of that definition.

Part B of the (general) definition of Pareto optimality states (in part) that there is a consumption Edgeworth box whose length is output x^0 and height is output y^0 and a distribution $(x_1^0, y_1^0, x_2^0, y_2^0)$ of (x^0, y^0) (where $x_1^0 + x_2^0 = x^0$ and $y_1^0 + y_2^0 = y^0$) in that box at which person 1's indifference curve through (x_1^0, y_1^0) is tangent to person 2's indifference curve through (x_2^0, y_2^0). This relationship appears in the upper two-thirds of Figure 19-3 where the indifference curves are labeled with associated utility values μ_1^0 for person 1 and μ_2^0 for person 2.

The following is a complete statement of the equations that characterize Pareto Optimality at output basket (x^0, y^0) with distribution $(x_1^0, y_1^0, x_2^0, y_2^0)$. The statement combines the equality-of-slope equations (18.1) from Chapter 18 and the equality-of-slope equations (19.1) and (19.5) above, equation (19.2) of the isoquant constraint passing through and evaluated at the distribution at the tangency in the production Edgeworth box, the two equations (18.3) of Chapter 18 and the two of (19.3) that ensure that the distributions $(x_1^0, y_1^0, x_2^0, y_2^0)$ and $(\ell_x^0, k_x^0, \ell_y^0, k_y^0)$ exhaust the outputs produced and factor supplied and therefore lie in the respective consumption and production Edgeworth boxes, and the equation of the transformation curve (19.4). It is also necessary to include an equation of one of the indifference curves passing through the distribution at the tangency in the consumption Edgeworth box. In this regard, rather than thinking of maximizing person 1's utility with person 2's utility no lower than μ_2^0 as in Chapter 18, let person 2's utility be maximized subject to person 1's utility being no lower than μ_1^0. The relevant indifference-curve-constraint equation becomes $\mu_1^0 = u^1(x_1^0, y_1^0)$ and is substituted for Chapter 18's equation (18.2) for person 2. (Here μ_1^0 instead of μ_2^0 has to be independently specified in order to fully describe the tangency between two indifference curves

[5] Of course, a full description of that tangency still requires, in addition to the equation representing the equality of two slopes, the equation of one of the two tangent curves.

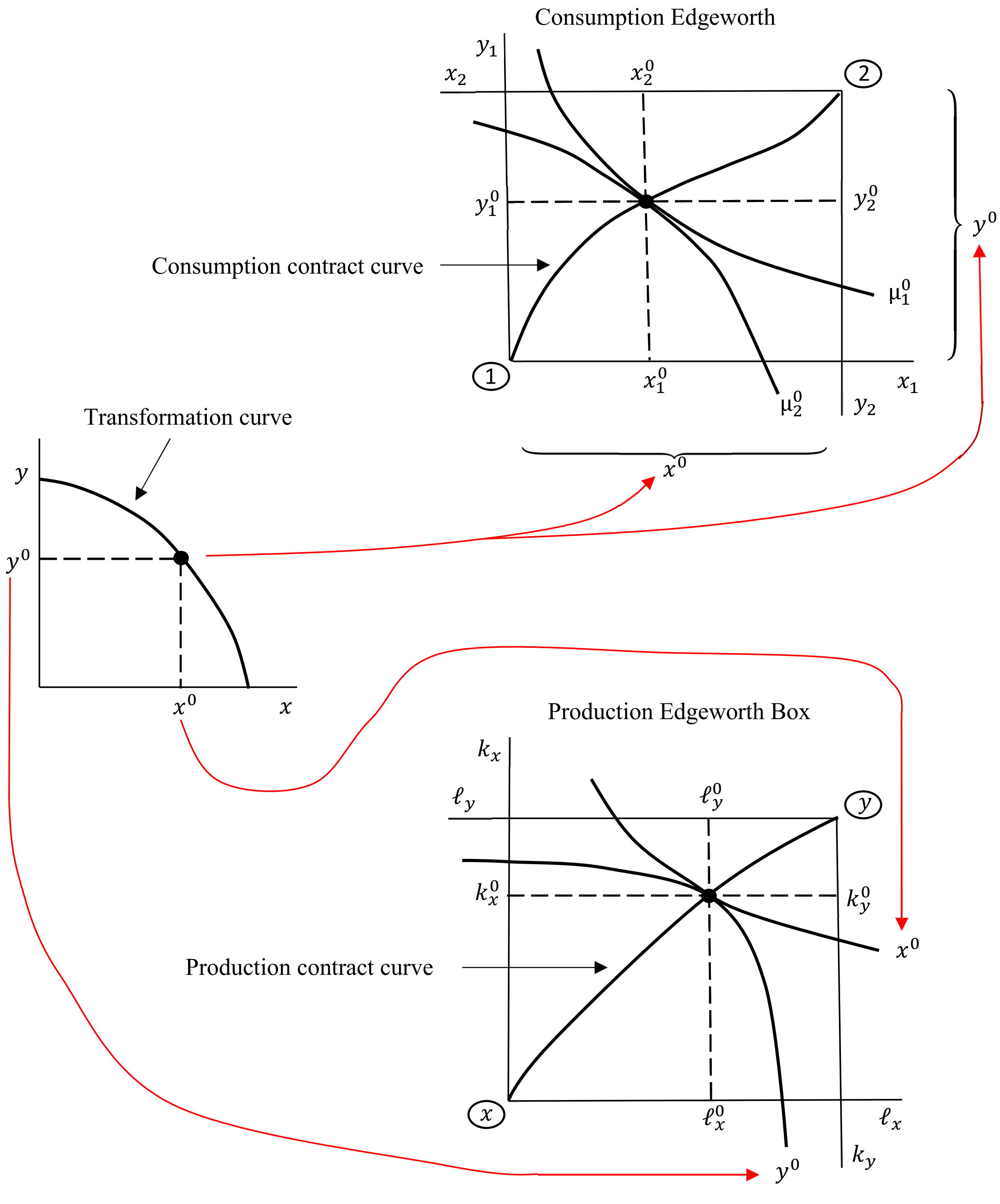

Consumption Edgeworth
y_1
x_2
x_2^0
2
y_1^0
y_2^0
y^0
Consumption contract curve
μ_1^0
1
x_1^0
x_1
μ_2^0
y_2
Transformation curve
x^0
y
y^0
x^0
x
Production Edgeworth Box
k_x
ℓ_y
ℓ_y^0
y
k_x^0
k_y^0
x^0
Production contract curve
x
ℓ_x^0
ℓ_x
y^0
k_y

Figure 19-3

in the consumption Edgeworth box.[6]) Also, capital will now be reintroduced explicitly as an argument in all marginal product functions, a practice that is consistent with the definition of total and marginal product functions at the beginning of Chapter 10. Note that the second equation of (19.5) is included implicitly below since it is implied by the combining the first and last equation of the list.[7]

$$\frac{MU_x^1(x_1^0,y_1^0)}{MU_y^1(x_1^0,y_1^0)}=\frac{MU_x^2(x_2^0,y_2^0)}{MU_y^2(x_2^0,y_2^0)},$$

$$u^1(x_1^0,y_1^0) = \mu_1^0 \text{ (the fixed utility value for person 1)},$$

$$x_1^0 + x_2^0 = x^0 \text{ (the output produced by firm } x),$$

$$y_1^0 + y_2^0 = y^0 \text{ (the output produced by firm } y),$$

$$\frac{MP_\ell^x(\ell_x^0,k_x^0)}{MP_k^x(\ell_x^0,k_x^0)}=\frac{MP_\ell^y\left(\ell_y^0,k_y^0\right)}{MP_k^y\left(\ell_y^0,k_y^0\right)},$$

$$x^0 = f^x(\ell_x^0,k_x^0),$$

$$\ell_x^0 + \ell_y^0 = \ell^* \text{ (the fixed supply of } \ell),$$

$$k_x^0 + k_y^0 = k^* \text{ (the fixed supply of } k),$$

$$y^0 = T(x^0),$$

$$\frac{MC^x(x^0)}{MC^y(y^0)}=\frac{MU_x^1(x_1^0,y_1^0)}{MU_y^1(x_1^0,y_1^0)}.$$

These equations suggest the earlier comment that the general form of Pareto optimality includes both Pareto optimality in consumption (fully described by the first four equations in the above list) and Pareto optimality in production (fully described by the second four equations). But Pareto optimality in consumption and production together do not, by themselves, imply Pareto optimality in general because the equation on the last line in the above list is still missing. It is also worth pointing out that the first nine equations in the list provide a complete mathematical representation of Figure 19-3: The first four equations relate to the consumption Edgeworth box; the second four relate to the production Edgeworth box and the ninth to the transformation curve.

---

[6] The parallel value of x^0 which identifies the isoquant on the production Edgeworth box does not have to be independently specified. It is determined in conjunction with the other equations of the system. See footnote 7 below and footnote 3 in Chapter 21.

[7] Given the fixed values of μ_1^0, ℓ^*, and k^*, this is a system of 10 simultaneous equations in the 10 variables $x^0, y^0, x_1^0, y_1^0, x_2^0, y_2^0$, $\ell_x^0, k_x^0, \ell_y^0$, and k_y^0. If specific forms of the marginal utility, marginal product, and marginal cost functions are introduced, it may be possible to solve the system to determine the values of the 10 variables.

Chapter 20
Nonwastefulness, Unbiasedness, and Utility Possibilities

The notion of Pareto optimality provides a criterion that may be used to evaluate general equilibrium outcomes. That evaluation is based on what are called the first and second fundamental theorems of welfare economics. These are considered here in the two-person, two-final good, two-input context of the last three chapters having fixed factor supplies ℓ^* and k^*. The first theorem is the statement that

General equilibrium under perfect competition is Pareto optimal.

Recall from the end of Chapter 19 that the equations that characterize Pareto Optimality at output basket (x^0, y^0) with distributions $(x_1^0, y_1^0, x_2^0, y_2^0)$ and $(\ell_x^0, k_x^0, \ell_y^0, k_y^0)$ are:

$$\frac{MU_x^1(x_1^0, y_1^0)}{MU_y^1(x_1^0, y_1^0)} = \frac{MU_x^2(x_2^0, y_2^0)}{MU_y^2(x_2^0, y_2^0)}, \tag{20.1}$$

$$\mu_1^0 = u^1(x_1^0, y_1^0), \tag{20.2}$$

$$x_1^0 + x_2^0 = x^0, \tag{20.3}$$

$$y_1^0 + y_2^0 = y^0, \tag{20.4}$$

$$\frac{MP_\ell^x(\ell_x^0, k_x^0)}{MP_k^x(\ell_x^0, k_x^0)} = \frac{MP_\ell^y(\ell_y^0, k_y^0)}{MP_k^y(\ell_y^0, k_y^0)}, \tag{20.5}$$

$$x^0 = f^x(\ell_x^0, k_x^0), \tag{29.6}$$

$$\ell_x^0 + \ell_y^0 = \ell^*, \tag{20.7}$$

$$k_x^0 + k_y^0 = k^*, \tag{20.8}$$

$$y^0 = T(x^0), \tag{20.9}$$

$$\frac{MC^x(x^0)}{MC^y(y^0)} = \frac{MU_x^1(x_1^0, y_1^0)}{MU_y^1(x_1^0, y_1^0)}, \tag{20.10}$$

where μ_1^0. ℓ^*, and k^* are specified independently of these equations. One way to establish the first fundamental theorem is to show that each of these equations holds at every general equilibrium.

Towards that end, let

(1) quantities of outputs consumed by individuals $(x_1^0, y_1^0, x_2^0, y_2^0)$ and market

quantities of outputs produced by firms (x^0, y^0) at prices p_x^0 and p_y^0, and

(2) quantities inputs employed by firms $(\ell_x^0, k_x^0, \ell_y^0, k_y^0)$ and market quantities of those inputs supplied by consumers (ℓ^*, k^*) at prices p_ℓ^0 and p_k^0

be a general equilibrium as described in parts (i) - (iii) of the definition of that concept in Chapter 17. Using the characteristics (a) – (c) required of all general equilibria according to the Chapter 17 definition, it will now be demonstrated that at these quantity and price values, all equations (20.1) – (20.10) are satisfied.

Observe first that characteristic (a) of the definition of general equilibrium requires that at (x_1^0, y_1^0) and (x_2^0, y_2^0), consumers 1 and 2 are, respectively, buying final goods that maximize their utility subject to their budget constraint. Applying the equality-of-slopes equation that partly describes that maximization (equation (6.2) of Chapter 6) to both consumers results in

$$\frac{MU_x^1(x_1^0, y_1^0)}{MU_y^1(x_1^0, y_1^0)} = \frac{p_x^0}{p_y^0} \text{ and } \frac{MU_x^2(x_2^0, y_2^0)}{MU_y^2(x_2^0, y_2^0)} = \frac{p_x^0}{p_y^0}. \tag{20.11}$$

Because both consumers face the same market prices, it follows that

$$\frac{MU_x^1(x_1^0, y_1^0)}{MU_y^1(x_1^0, y_1^0)} = \frac{MU_x^2(x_2^0, y_2^0)}{MU_y^2(x_2^0, y_2^0)},$$

thereby establishing equation (20.1). Since consumer 1 obtains the basket (x_1^0, y_1^0) in the specified general equilibrium, he/she also attains the utility value along the indifference curve on which that basket lies. Denote that value by μ_1^0. Thus

$$\mu_1^0 = u^1(x_1^0, y_1^0),$$

securing equation (20.2).

Next, according to characteristic (b) of the definition of general equilibrium, x^0 and y^0 have to be profit-maximizing outputs. Since profit maximization implies that x^0 and y^0 are produced with their respective cost-minimizing input baskets (ℓ_x^0, k_x^0) and (ℓ_y^0, k_y^0), the equality-of-slopes equations which partly describe this minimization (from equation (11.3) in Chapter 11) are

$$\frac{MP_\ell^x(\ell_x^0, k_x^0)}{MP_k^x(\ell_x^0, k_x^0)} = \frac{p_\ell^0}{p_k^0} \text{ and } \frac{MP_\ell^y\left(\ell_y^0, k_y^0\right)}{MP_k^y\left(\ell_y^0, k_y^0\right)} = \frac{p_\ell^0}{p_k^0},$$

so that

$$\frac{MP_\ell^x(\ell_x^0, k_x^0)}{MP_k^x(\ell_x^0, k_x^0)} = \frac{MP_\ell^y\left(\ell_y^0, k_y^0\right)}{MP_k^y\left(\ell_y^0, k_y^0\right)}$$

establishing equation (20.5). Because the firm producing good x is producing its output x^0 with the input basket (ℓ_x^0, k_x^0), that input basket lies on its x^0-isoquant or

$$f^x(\ell_x^0, k_x^0) = x^0,$$

which is equation (20.6). And since both firms' isoquants are tangent at the input baskets they employ (from equations (20.5) and (20.6) just established), their outputs (x^0, y^0) are located on the transformation curve (Chapter 19). This yields equation (20.9):

$$y^0 = T(x^0).$$

Also, profit maximization further means that $MC^x(x^0) = p_x$ and $MC^y(y^0) = p_y$ at, respectively, x^0 and y^0 (see, for example, equation (13.3) in Chapter 13).[1] Dividing the second equation into the first gives

$$\frac{MC^x(x^0)}{MC^y(y^0)} = \frac{p_x^0}{p_y^0}, \tag{20.12}$$

and combining equations (20.11) and (20.12) results in equation (20.10)):

$$\frac{MC^x(x^0)}{MC^y(y^0)} = \frac{MU_x^1(x_1^0, y_1^0)}{MU_y^1(x_1^0, y_1^0)}.$$

To obtain equations (20.3) and (20.4), use characteristic (c) of the definition of general equilibrium stating that, at such an equilibrium, supply equals demand in all markets. Since, at the general equilibrium specified, x_1^0 and x_2^0 are the quantities demanded of good x by the two consumers, the market quantity demanded is $x_1^0 + x_2^0$. Moreover, at this general equilibrium the profit-maximizing output which is the same as the quantity supplied by the firm producing x is x^0. Since equality of the two is required,

$$x_1^0 + x_2^0 = x^0.$$

Similarly, in the market for good y:

$$y_1^0 + y_2^0 = y^0.$$

Finally, at the specified general equilibrium, summing the profit-maximizing quantities of inputs ℓ and k demanded by the two firms gives, respectively, the market demands $\ell_x^0 + \ell_y^0$ and $k_x^0 + k_y^0$. And the market quantities supplied by the two consumers combined are the fixed quantities ℓ^* and k^*. Equilibrium in the factor markets now yields equations (20.7) and (20.8):

[1] As in Chapter 19, the short-run (SR) and long-run (LR) designations in the cost-function symbolism of earlier chapters is dropped.

$$\ell_x^0 + \ell_y^0 = \ell^*$$

and

$$k_x^0 + k_y^0 = k^*.$$

Therefore, all of the equations required for Pareto optimality are satisfied so general equilibrium is Pareto optimal.

Use the phrase "perfectly competitive price system" to refer to the mechanism of the competitive forces of supply and demand by which equilibrium prices and hence equilibrium quantities and their distribution are established in the Walrasian model. Implicit in this mechanism are the profit maximization by firms and constrained utility maximization by consumers that generate supply and demand.[2] If general equilibrium were not Pareto optimal, then at any equilibrium distribution it would be possible through changing quantities of outputs produced, or redistributing inputs among firms or final goods (outputs) among consumers, to make at least one person better off in terms of utility without making the other worse off. In this sense there would be waste. Thus it may be said that the first fundamental theorem of welfare economics asserts that the perfectly competitive price system is **nonwasteful**.

As previously described in Chapter 17, the perfectly competitive price system is taken as an explanation of how the real microeconomy operates, and each complete set of observations of prices and quantities in that economy is understood as a general equilibrium. It follows that the workings of the real microeconomy, or what is represented in the Walrasian model by the perfectly competitive price system, can be said to have this nonwastefulness property. That is, each observed distribution of commodities among consumers is Pareto optimal and it is not possible through redistribution to make one person better off without making another worse off.

Note that if at least one of the two firms, say the firm producing x, were operating in an imperfectly competitive output market (such markets will be considered in due course), then the general equilibrium achieved in that case would not necessarily be Pareto optimal. For example, if the firm producing good x were a monopolist, then, as will be seen in Chapter 22, $MC^x(x^0)$ would be less than the output price p_x^0 . In that case, the above argument establishing equation (20.12) need not hold and the ratio of marginal costs would not necessarily equal the output-price ratio. The latter is required to obtain equations (20.1) and (20-10). Without them, Pareto optimality cannot be established, and the first fundamental theorem would not hold.

The second fundamental theorem of welfare economics states:

---

[2] Apart from the description of the competitive forces of demand and supply in Chapter 3 and the implicit profit and constrained utility maximization behind them, no further characterization of this mechanism will be given here.

For each Pareto optimal distribution of outputs and associated input quantities, there are prices at which that distribution and its output and input quantities are part of a perfectly competitive general equilibrium.

This theorem will not be demonstrated here. But the theorem indicates that all Pareto optimal distributions, even those in which one person receives only very small quantities of goods, can arise as general equilibria. Thus, the perfectly competitive price system need not lead to an equal or near equal distribution of goods among consumers.

The fact that all distributions on a variety of different consumption contract curves (each corresponding to a different pair of outputs on the transformation curve) are possible general equilibria means that the perfectly competitive price system is not biased away from any particular distribution along any curve. It is in this sense that the second fundamental theorem asserts that the perfectly competitive price system is **unbiased.** Applied to the real microeconomy, the inner workings of that economy do not push distributions away from any outcome that might be observed.

Thus, as long as what happens in the real microeconomy is interpreted in terms of the Walrasian model that has been built to understand it, the nonwastedfulness and unbiasedness properties of general equilibrium say something about the world in which real microeconomic activity takes place. The former asserts that it is not possible by changing inputs, outputs, or their distribution to make one person better off without making someone else worse off. And from the latter it is understood that unequal distributions of goods across consumers are not only possible but also have the same status in terms of Pareto optimality or efficiency as more equal distributions.

Knowing that general equilibrium under perfect competition is Pareto optimal does not indicate which equilibrium is best from society's point of view. In terms of the consumption Edgeworth box, distributions on the consumption contract curve such as B In Figure 20-1 can be

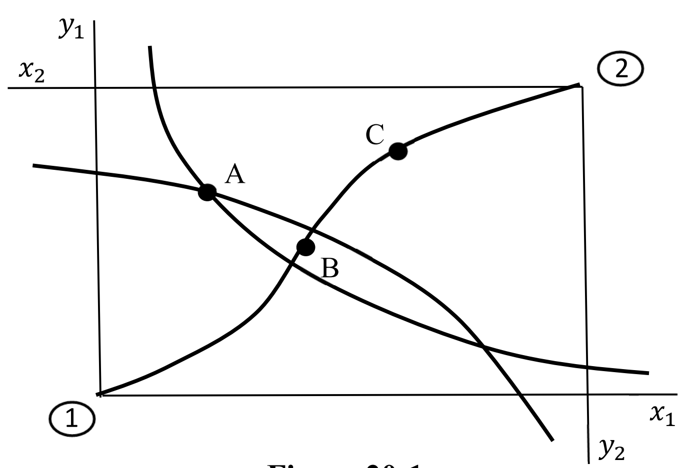


Figure 20-1

said to always be better than distributions off of the contract curve such as A because both individuals have higher utility at B than at A. But without introducing additional elements into the Walrasian model, there is no way of determining which of two distributions on the contract curve, such as B and C, is better since, in moving from one to the other, one person's utility necessarily rises while that of the other falls. To see that this statement also applies to distributions that are Pareto optimal in the general sense, consider the following:

Recall that the utility functions for persons 1 and 2 are written respectively as $\mu_1 = u^1(x_1, y_1)$ and $\mu_2 = u^2(x_2, y_2)$. These functions are ordinal and there are many of them that represent the same underlying preferences and indifferences.[3] It will simplify matters to specify particular utility functions for the two individuals with the property that

$$u^1(0,0) = 0 \quad \text{and} \quad u^2(0,0) = 0. \tag{20.13}$$

In what follows, these functions remain fixed and are not permitted to change.

Now choose a basket of outputs or point (x', y') on the transformation curve and construct the consumption Edgeworth box associated with it. In the same way that the transformation curve was obtained from the production contract curve in Chapter 19, plot the two persons' utility values corresponding to tangent indifference curves that arise when moving along the consumption contract curve from the origin of person 1 to that of person 2. In parallel with the transformation curve, the result is a diagram containing a quadrant with co-ordinates μ_1 and μ_2 and a curve in that quadrant called a utility possibility curve. It is illustrated in Figure 20-2. Note that the utility possibility curve slopes downward because as one person's utility rises

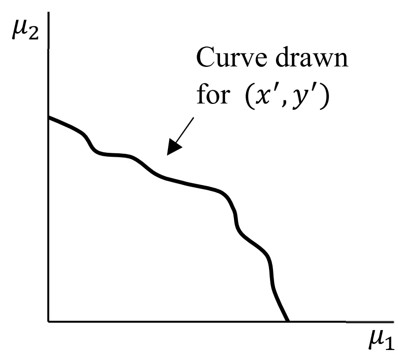


Figure 20-2

along the contract curve, that of the other must fall. But unlike the strictly concave transformation curve in Figure 19-2, the utility possibility curve in Figure 20-2 is drawn as a wavy line because its exact shape is generally taken to be unknown.[4] It should be emphasized

---

[3] Recall that, for example, $\alpha u^1(x_1, y_1)$ represents the same preferences and indifferences for each $\alpha > 0$.

[4] It was pointed out in footnote 2 of Chapter19 that strict concavity of production functions implies strict concavity of the transformation function. But unlike productions functions which are fixed by technology and as noted above, there are many utility functions satisfying (20.13) that represent the same underlying preferences and indifferences of the two individuals. Some of these could be strictly concave. Due to the variety of conceivable utility functions, the parallel assumption of strict concavity from footnote 2 of Chapter 19 is not imposed on utility functions and the utility possibility curve.

that, with fixed specifications of utility functions, each utility possibility curve is identified with a single point (x', y') on the transformation curve and no other.

Now, keeping the specified utility functions the same, choose a second point (x'', y'') on the transformation curve and construct the utility possibility curve for that point. Do this for all points on the transformation curve. This provides a web of utility possibility curves. The outer boundary of all of these curves is called the grand utility possibility curve. It is pictured as the red line in Figure 20-3 along with two ordinary utility possibility curves, each associated with its

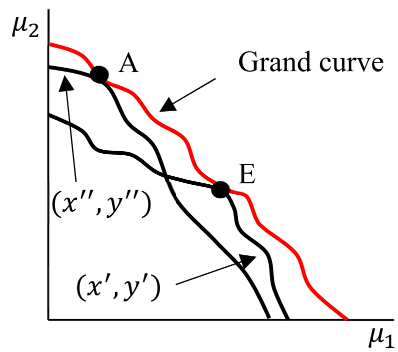


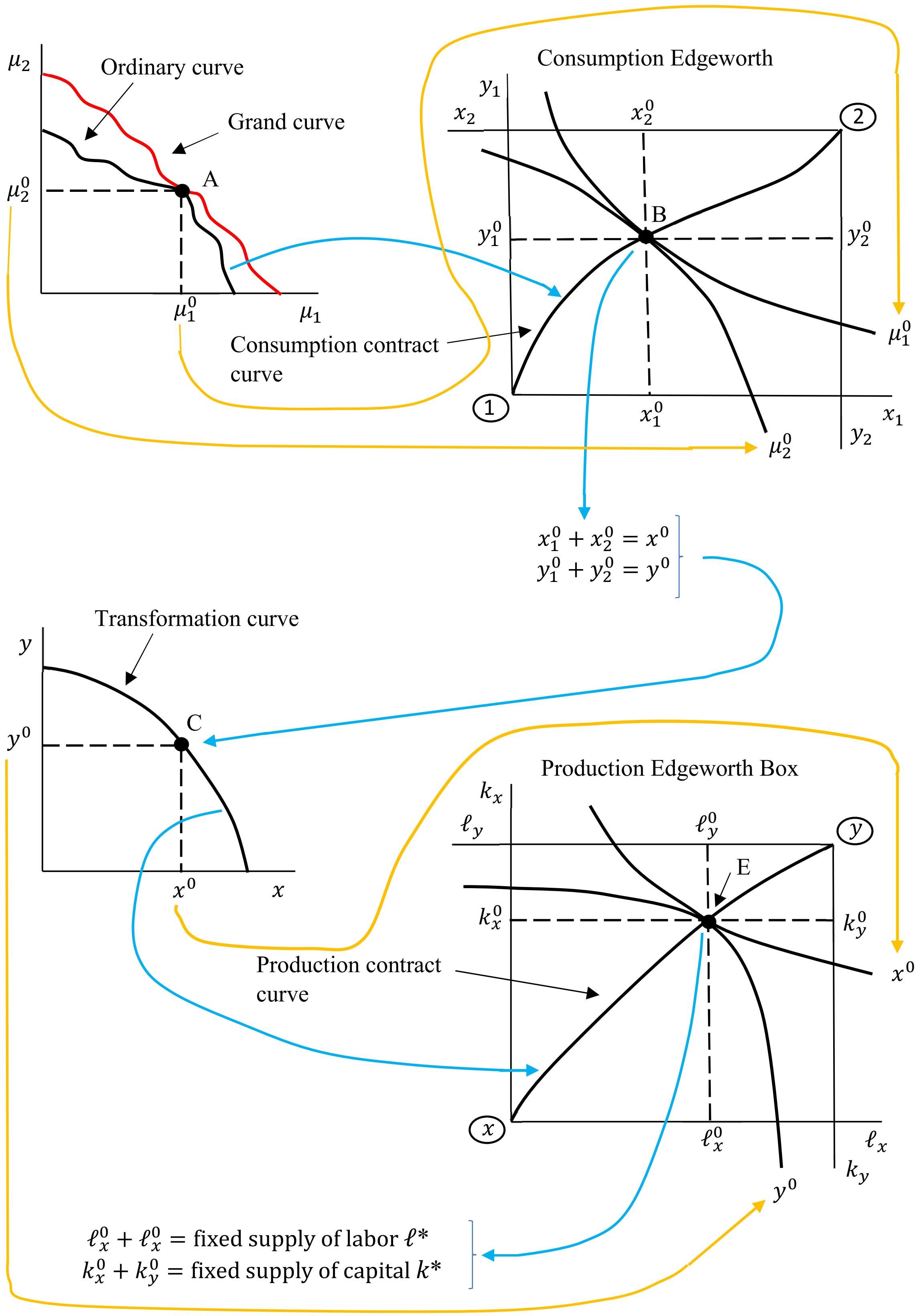
Figure 20-3

own unique point (x', y') or (x'', y'') on the transformation curve. Observe that each point such as A or E on the grand curve also lies on an ordinary utility possibility curve and, in turn, is associated with

(i) a unique point on the transformation curve and hence a unique production-Pareto-optimal distribution of inputs among firms in the corresponding production Edgeworth box, and

(ii) a unique consumption Edgeworth box and a unique consumption-Pareto-optimal distribution in that box of the outputs that have been identified on the transformation curve among consumers.

To see how these associations work out geometrically, consider Figure 20-4. Start at point A with coordinates (μ_1^0, μ_2^0) on the grand utility possibility curve in the upper-left diagram. That point also lies on an ordinary utility possibility curve. The ordinary curve, in turn, corresponds to the consumption contract curve in the consumption Edgeworth box of the upper-right diagram. And the coordinates (μ_1^0, μ_2^0) identify two tangent indifference curves, one, for consumer 1, along which the utility value is μ_1^0 and the other, for consumer 2, along which the utility value is μ_2^0. The point of tangency of the two curves, B, lies on the consumption contract curve. The coordinates of that point determine the distribution $(x_1^0, y_1^0, x_2^0, y_2^0)$ of quantities of commodities among the two consumers. The sums $x_1^0 + x_2^0$ and $y_1^0 + y_2^0$ of the individuals' quantities of the two goods in the distribution, namely, x^0 and y^0, define the dimensions of the consumption Edgeworth box and comprise the coordinates (x^0, y^0) of the associated point C on the economy's transformation curve pictured in the center-left diagram.



μ_2
Ordinary curve
Grand curve
μ_2^0
A
μ_1^0
μ_1
Consumption Edgeworth
y_1
x_2
x_2^0
2
B
y_1^0
y_2^0
Consumption contract curve
μ_1^0
1
x_1^0
x_1
y_2
μ_2^0
$x_1^0 + x_2^0 = x^0$
$y_1^0 + y_2^0 = y^0$
Transformation curve
y
C
y^0
x^0
x
Production Edgeworth Box
k_x
ℓ_y
ℓ_y^0
y
E
k_x^0
k_y^0
x^0
Production contract curve
x
ℓ_x^0
ℓ_x
k_y
y^0
$\ell_x^0 + \ell_x^0$ = fixed supply of labor ℓ^*
$k_x^0 + k_y^0$ = fixed supply of capital k^*

Figure 20-4

The transformation curve, in turn, corresponds to the production contract curve in the production Edgeworth box in the lower-right diagram of Figure20-4 (recall Figures 19-2 and 19-3 from Chapter 19). And the coordinates of the point (x^0, y^0) on the transformation curve identify two isoquants tangent at point E on the contract curve, one for the firm producing good x along which the output value is x^0, and the other for the firm producing good y,\ along which the output value is y^0. The coordinates of the point E determine a distribution $(\ell_x^0, k_x^0, \ell_y^0, k_y^0)$ of the fixed factor supplies (the total amounts of input quantities available from consumers) among the two firms. Note that the sums $\ell_x^0 + \ell_x^0$ and $k_x^0 + k_y^0$ of the firms' quantities of the two inputs in the distribution, namely, $\ell*$ and $k*$ are precisely the market fixed factor supplies and define the dimensions of the production Edgeworth box.

Now consider points such as A and E on the grand curve of Figure 20-5. Moving from one to the other requires transferring to a different ordinary curve and relocating to a distribution among consumers of a different basket of outputs along the transformation curve. Were A and E on the same ordinary curve (a possibility not pictured in Figure 20-5), the outputs on the transformation curve would not change but there would still be a rearranging of their distribution. In either case, as can be inferred from Figure 20-5, such a move cannot increase one person's utility without lowering that of the other. Clearly, points on the grand curve must be Pareto optimal. In \addition, starting at a point off of the grand curve that lies on an ordinary

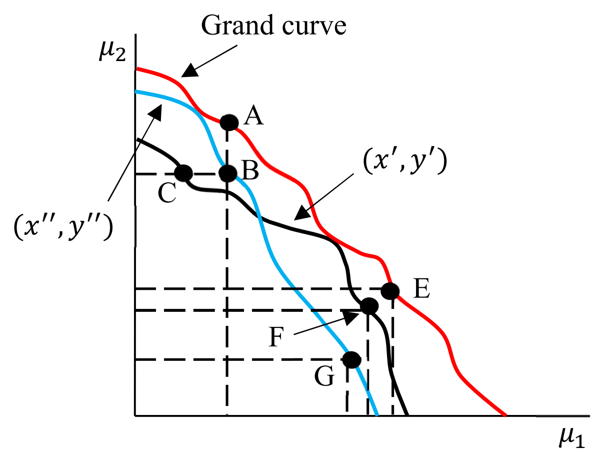


Figure 20-5

curve, it is always possible to move to a point on another ordinary curve (or on the grand curve) and increase the utility of one person without lowering that of the other or raise the utility of both persons simultaneously. In the example of Figure 20-5, moving from C to B increases person 1's utility keeping person 2's the same, while moving from B to A makes person 2 better off without hurting person 1. And moving from G to F to E increases the utility of both persons in each case. Clearly, outputs and their distributions corresponding to at points C, B, G, and F are not Pareto optimal,

Thus, although points off the grand curve but still on an ordinary curve are associated with distributions that are Pareto optimal in consumption, only points on the grand curve can be Pareto optimal in the general sense. And since moving from one point on the grand curve to another always increases one person’s utility at the expense of the other’s, it is not possible without adding something more to the Walrasian model to determine which points on the grand curve along with their associated distributions of outputs are better from society’s perspective. This is so even when comparing points in the middle of the grand curve with points at the ends of the curve where one individual’s utility value is zero and that person has zero quantities of both goods. Hence the same conclusion derived from the consumption contract curve in the consumption Edgeworth box of Figure 20-1 applies in relation to the general form of Pareto optimality to the entire microeconomy. One way to think about “adding something more” is considered in the next chapter.

Chapter 21
Welfare Maximization and Market Failure

What is observed when looking at the real microeconomy are prices and quantities (which includes distributions of produced goods among and incomes of consumers). To explain where they came from, a Walrasian model has been built and the prices and quantities observed are interpreted as those of a general equilibrium in that model. The model indicates that there are many possible general equilibria and each one along with its distribution is Pareto optimal or efficient.[1] The question is: among the possible general equilibria is there, from the broad perspective of society as a whole, a general equilibrium that is better than, that is, preferred to the one that has been observed. Of course, in moving to such an equilibrium and its accompanying distribution, at least one person would have to be made worse off while at least one other would become better off. Thus, the determination that one general equilibrium is preferred to another requires the making of value judgments – judgments based on cultural and moral commitments about whether it is, from society's point of view, beneficial and proper to take something away from one person in order that someone else could have more. Once such a determination is made, it would become desirable to make the changes needed to move to the better or more preferred general equilibrium. And if among all possibilities, a best or most preferred general equilibrium could be found, the goal would be to attain it.

In the model developed here, one way to incorporate the relevant value judgments and express society's preferences among distributions is with a "welfare function." A welfare function is like an individual's ordinal utility function but for society as a whole. It expresses society's preferences among the various distributions that emerge from economic activity. Since each distribution contains baskets of goods for every individual, and since utility functions assign utility values to the baskets in the distributions individuals obtain, the welfare function can be defined in terms of those utility amounts,[2] one amount for each individual. Although it is not usually possible in practice to specify a real society's welfare function, that function is often implicit in its behavior. For example, in the US, it has been decided that society's welfare is increased by taking money away from those who are able to consume significantly more and giving it in some form to those who wind up consuming significantly less. This can be seen in the country's income taxation system in which persons with higher income pay more tax, and various assistance programs for low-income persons such as housing choice vouchers and food stamps.

---

[1] Up to this point the concepts of general equilibrium and Pareto optimality have been considered only in the context of a world containing two persons, two firms, two produced goods, and two inputs. But these ideas can also be applied more generally to situations in which there are many persons, many firms, many produced goods, and many inputs. The statement of the definition of general equilibrium (Supplementary Note 6) applies in this more general context without modification. The statements of the definitions of Pareto optimality in consumption (Chapter 18) and production (Chapter 19) also carry over without change. That for Pareto optimal (in general) requires relatively minor modification. (However, the systems equations that characterize these notions of Pareto optimality become more complex.) The following discussion may be understood in terms of the more general versions of these notions unless otherwise stated.

[2] As in Chapter 20, the utility functions for individuals have to be fixed for a welfare function to be specified.

Return to the two-good, two-person world of previous chapters and continue with the assumptions from Chapter 20 that $u^1(0,0) = 0$ and $u^2(0,0) = 0$. In formal terms, let society's welfare function W be expressed as

$$\omega = W(\mu_1, \mu_2),$$

where ω varies over ordinal welfare values, and μ_1 and μ_2 vary over utility values for persons 1 and 2 respectively. This function is assumed to have the same properties earlier imposed on individual utility functions (Supplemental Note C), and the indifference curves obtained from W, called welfare-indifference curves, are derived in the same way and have the same properties as the indifference curves required of individual utility functions (Chapter 5). The function W can be used to determine the produced output quantities and their distribution among consumers that maximizes society's welfare subject to a constraint in a similar fashion to that of the determination of individual consumer demand quantities through utility maximization subject to the budget constraint.

The possibilities for baskets (μ_1, μ_2) that are achievable by consumers given the economy's resources consist of those that lie on the utility possibility curves. Each such basket is associated with a distribution of outputs that could be produced by the firms in the economy (Chapters 19, 20). The grand curve, that is, the outer boundary of all of the ordinary curves, imposes a limit on the achievable (μ_1, μ_2) baskets and plays the same role here as the consumer's budget constraint in the model of consumer buying behavior. There are not enough factor supplies to obtain a basket of utilities beyond that curve. And in parallel with the consumer model, society's welfare will be maximized at the basket (μ_1^0, μ_2^0) on the grand utility possibility curve that also lies on the welfare indifference curve that is farthest out from the origin. This appears at tangency A in Figure 21-1.[3] From the second fundamental theorem of welfare

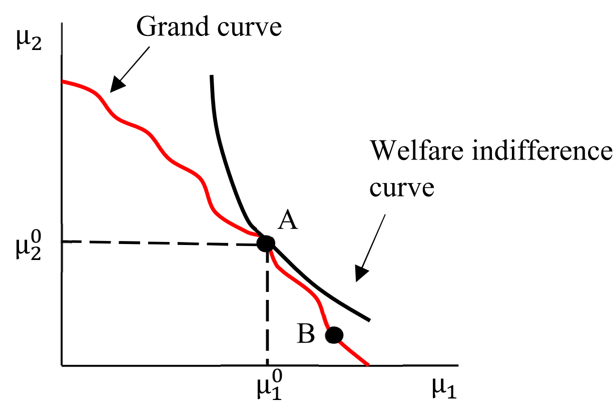


Figure 21-1

economics (Chapter 20), the outputs and their distribution associated with (μ_1^0, μ_2^0) and

[3] Like the mathematical statement of other tangencies in this volume, the equations that fully characterize the tangency in Figure 21-1 are the statement that the two slopes are equal at A together with the equation associated with the grand-utility -possibility-curve constraint evaluated at the tangency. Adding the former of these equations to the system of equations in Chapters 19 and 20 (e.g., 20.1-20.10) that are associated with Pareto optimality, would now determine the value of μ_1^0. That is, μ_1^0 no longer has to be specified independently. And the solution values of the system not only are identified with Pareto optimality but are also those that maximize welfare.

appropriate market prices are part of a perfectly competitive general equilibrium. Those welfare-maximizing values (that is, μ_1^0, μ_2^0 and their associated general equilibrium price, output, input, and distribution values) may or may not be the ones that are actually observed in the real microeconomy. In other words, even though the prices, quantities, and distributions that are actually observed are interpreted as a general equilibrium and, from the first fundamental theorem, Pareto optimal, it does not follow that the utility values from that distribution, although necessarily lying on the grand utility possibility curve, will always be located at a tangency between a welfare indifference curve and the grand curve. That is, maximum welfare is not guaranteed in the real microeconomy. For example, the distribution that is observed could provide a utility basket such as B in Figure 21-1 instead of at the welfare-maximizing basket A.

It is also the case, as has been suggested in Chapter 20, that general equilibrium under perfect competition, even if it maximizes social welfare, does not necessarily lead to an equal or near-equal distribution of goods among people. In the two-person situation, distributions associated with points near either end of the grand curve where one individual has very small quantities of both goods are potential outcomes.

If knowledge of the welfare function were available, then government policies could be devised to ensure that society would achieve maximum welfare. But since it is generally not possible to know what the welfare functions is,[4] judgments to undertake policies that might improve social welfare like food-stamp or housing-voucher programs are generally made one at a time and largely independently of each other.

It should be kept in mind that the relationship between general equilibrium and Pareto optimality as defined by the first and second fundamental theorems of welfare economics applies only to the situation in which all of the economy's markets are perfectly competitive. And it is well known that the characteristics of many (if not most) markets in the real microeconomy do not approximate perfectly competitive conditions very well. So it is worth considering the kinds of adjustments that might be introduced into the Walrasian model in order to make it, with respect to market characteristics, more consistent with microeconomic reality.

Towards that end, a market is said to fail when there is a "malfunction" in it that prevents the perfectly competitive outcome that is, the price and quantity that would arise if the market were perfectly competitive, from emerging. Introducing the possibility of market failure into the Walraasian model, then, will bring that model closer to what is actually experienced in the real microeconomy. But. as has been suggested in Chapter 20, the first fundamental theorem of welfare economics implies that when a real market fails, Pareto optimality cannot be guaranteed in the real microeconomy, and it is usually possible by changing outputs and/or their distribution among consumers to make one person better off (in terms of utility) without making at least one other worse off.

There are four kinds of market failures that will be considered. Although they will be examined more closely in the following chapters, each is briefly summarized here.

---

[4] For example, the voting paradox described in Chapter 25 suggests that the welfare function cannot be determined by majority-rule, pair-wise voting.

1. **Imperfectly competitive market structures.** The types of imperfect competition to be considered here arise when there are no longer a large number of small sellers or a standardized product in the market under consideration. There are several forms that such imperfectly competitive markets can take on. And in each circumstance, firms are able to raise the price of their output without the demand for that output falling to zero. This will imply, in the case of, for example, the firm producing and selling good x, that short- or long-run marginal cost is less than the market price at the profit maximizing output. Assuming the firm producing y remains a perfect competitor, the argument establishing the first fundamental theorem of welfare economics, which relies on the equation $MC^x(x^0) = p_x$, breaks down, consumer marginal rates of substitution at the resulting distribution no longer equal the marginal rate of transformation at the economy's produced outputs, and the outcome cannot be consistent with general efficiency or Pareto optimality.

2. **Externalities.** An externality is a cost or benefit arising from a transaction that is imposed on individuals or firms not involved in that transaction. For example, the production and sale of electricity that creates air-borne pollutants that fall as acid rain imposes costs on the individuals living in the acid-rain area. Similarly, the production and sale of output by an upstream firm that dumps production pollutants into the water raises the costs and hence output price of the down-stream firm that needs clean water to manufacture its output. And spending money to fix up one's yard may make the neighborhood properties more valuable or the neighborhood itself more beautiful thereby benefiting neighbors. Since these costs or benefits lie outside of the market in which the transactions take place, they are not included in, respectively, the marginal cost or market price[5] of the transaction. Once again, inefficiency or absence of Pareto optimality arises.

3. **Public goods.** Generally, a public good is one that cannot be produced for one person without making it available to everyone free of charge. Street lighting, TV and radio broadcasting, and national defense are examples. Since individuals do not have to pay directly for public goods, there is no demand curve indicating the market quantity demanded of such a good for each possible price. Firms cannot, therefore, respond to market demands by equating price to marginal cost. Although governments usually step in to arrange for the production of a public good, since its demand is unknown, the first fundamental theorem does not apply and the Pareto optimal output quantity cannot usually be determined, let alone achieved.

4. **Imperfect information.** When the perfect information requirement of perfect competition is violated and buyers or sellers do not have full information on all characteristics (including prices) of products, non-maximizing transactions can occur. For example, consumer 1, say, may think that buying (x_1^0, y_1^0) will be utility-maximizing. But after purchase, because he/she did not know everything about the products being bought, that is found to be false. So for that purchase, it turns out that $MU_x^1(x_1^0, y_1^0) \big/ MU_y^1(x_1^0, y_1^0) \neq p_x / p_y$ and that consumer's marginal rate of substitution actually achieved cannot equal person two's marginal rate of substitution and the marginal rate of transformation. The latter conditions are necessary for the

[5] It will be seen in Chapter 25 that the market price of a good reflects the benefits per unit of that good to the individuals buying it.

argument establishing the first fundamental theorem to apply (equations (20.1) and (20.9) of Chapter 20). Thus, for a reason different from those in the first three forms of market failure described above, the argument establishing that general equilibrium is Pareto optimal no longer holds up.

In all of these instances of market failure the Pareto optimality or efficiency of perfect competition breaks down. That fact establishes the case for government intervention in the failed market to mitigate the failure. The purpose of the intervention would be to move the microeconomy from a position off the grand utility possibility curve to, or closer to a non-wasteful Pareto optimal situation on the grand curve. By doing so the government might increase the utility of at least one person without lowering that of any other. But this intervention only reduces or eliminates one particular kind of waste. It does not guarantee that the intervention will result in welfare maximization or even making it to the grand curve. In the context of Figure 21-1, it might push the utilities of the two persons from where they are closer to a point like B rather than to maximum welfare at A.

The first form of market failure to be taken up here is imperfectly competitive market structures. It is necessary to begin by characterizing the types of imperfect competition that will be considered in the next three chapters. To do so, start with the definition of a perfectly competitive market given in Chapter 1. With respect to the properties of perfect competition, maintain the assumptions of a large number of small buyers and of perfect information. But permit the possibilities of a small number of large sellers (including a single seller), non-standardized or differentiated products across sellers, and limited entry into and exit from the market. Consider only output markets in which the buyers are consumers and the sellers are firms. The different types of market structures to be considered in subsequent chapters appear in Table 21-1. Thus, for example, keeping all of the characteristics of perfect competition intact

Table 21-1

	Number and Size of Sellers		
Type of Product	**Large Number of Small Sellers**	**Small Number of Large Sellers**	**One Seller**
Standardized (Homogeneous)	Perfect Competition	Perfect Oligopoly	Monopoly
Differentiated (Heterogeneous)	Monopolistic Competition	Monopolistic Oligopoly	
	Free Entry and Exit	Limited Entry of Sellers	Entry of Sellers Is Blocked

except that of standardized products and requiring the product of each firm to be slightly different from that of all other firms in the market results in monopolistic competition. And additionally changing the number and size of sellers to be, respectively, small and large, and switching to limited entry leads to monopolistic oligopoly.

In Table 21-1, and apart from perfect competition, free entry and exit is maintained only for monopolistic competition. There cannot be free entry of firms for monopoly. For once abnormal profit leads to the entry of a second firm, the monopoly would no longer be a monopoly. Entry has to be limited for both forms of oligopoly for a similar reason. And it is implicit in the case of monopoly that the product being produced has no close substitutes available for purchase in outside markets.

In all cases, perfect competition will be assumed in all input markets. This will mean that cost functions and cost curves will be constructed as in earlier chapters without modification. Thus, the only changes that occur in previous discussions of the firm appear in the calculation of revenue functions and curves. This, of course, will imply alterations in profit-maximizing positions.

Chapter 22
Monopoly

Monopoly, the first form of market failure to be considered, is a market situation with a large number of small buyers (it will be assumed that they are consumers) in which there is one firm producing a product for which there are no close substitutes and in which entry into the market by other producers is blocked.

As noted at the end of Chapter 21, because it is assumed that the monopolist buys its inputs in perfectly competitive markets, its cost functions and curves are identical to that of the perfectly competitive firm. The only difference between explanations of the behavior of the two types of firms is in terms of the demand curves they face and their revenues. Recall from Chapter 13 that a firm producing under perfectly competitive conditions, no matter what it does, can have no impact on the market price of its output. That price is dictated by the market, and the firm can sell all it wants at that price, call it p_x^0. The demand curve facing the firm is its average revenue as curve pictured in Figure 22-1. Total revenue along the curve as x varies is $TR(x) = xp_x^0$,

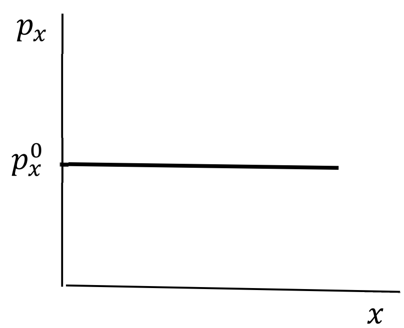


Figure 22-1

Where p_x^0 is constant. The graph of $TR(x)$ is an upward sloping straight line emanating from the origin. The firm's marginal revenue curve is the same as the demand curve it faces since $MR(x) = AR(x) = p_x^0$.

Unlike the perfectly competitive firm, the monopolist, as will be seen, sets the market

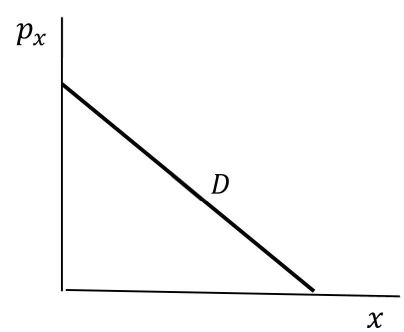


Figure 22-2

price. Also by contrast, the demand curve facing the monopolist, the only seller in the market, is the market demand curve.[1] An example of that curve, labeled as D, is shown in Figure 22-2. (For purposes of simplicity here and in subsequent diagrams appearing in the next two chapters, the monopolist's demand curve along with market demand curves and demand curves facing firms under other forms of imperfect competition, will be drawn as downward sloping straight lines.) As the monopolist moves along this demand curve, price is not constant, that is, $TR(x) = xp_x$, where p_x changes as x varies. To sell more, the monopolist has to lower its price. Suppose, then, that the monopolist, while selling a specified output quantity at a particular price, decides to increase the quantity of output it sells by, say, one unit. Then it must sell that extra unit at a lower price. But the increase in revenue obtained from selling the extra unit is offset by a loss incurred because, according to the demand curve it faces, it must also lower the price of all other units sold at that price. This suggests the following formula for the monopolist's marginal revenue:

$MR(x) =$ the lower price p_x less the loss in revenue on the other units it could have been selling at the original price.

It follows that the monopolist's marginal revenue is always less than the selling price. This conclusion also emerges in the geometric argument below.

Before pursuing the implications of these divergences from the explanation of the behavior of the perfectly competitive firm, since the linear demand curve will be used repeatedly in this and subsequent chapters, it is worth recalling its characteristics in relation to the total and marginal revenue curves and the various elasticity values that arise along it. These originally appeared in Chapter 4. That discussion is restated below along with Figure 4-5 which is reproduced here as Figure 22-3. With respect to the latter:

1. Given the demand curve D, the marginal revenue curve *MR* is determined by drawing a straight line from *D*'s vertical intercept at σ through the mid-point x' between the origin and *D*'s horizontal intercept x'' on the x-axis (Supplemental Note B).

2. Since $TR(x) = xp_x$ along the demand curve, the total revenue curve *TR* must start at the origin where $x = 0$ along the demand curve and meet the x-axis at x'' where $p_x = 0$ also along the demand curve. In between, where marginal revenue is positive ($0 < x < x'$), the *TR* curve slopes upward; where marginal revenue is negative ($x' < x < x''$), it slopes downward; and where marginal revenue is zero ($x = x'$), *TR* has a maximum.

3. From the equation $MR = p_x\left(1 - {}^{1}\!/_{\varepsilon}\right)$, the demand curve is elastic ($\infty > \varepsilon > 1$) where marginal revenue is positive (on the interval $0 < x < x'$); it is inelastic ($1 > \varepsilon > 0$) where marginal revenue is negative (on the interval $x' < x < x''$); and it is unitarily elastic ($\varepsilon = 1$) where marginal revenue is zero (at $x = x'$).

[1] In the case of monopoly, the demand curve facing the firm, that is, the quantity the firm can sell at each price, is the same as the quantity buyers will buy at each price. As in Chapter 13, this curve is known to the firm.

4. Moving along the demand curve in the elastic range ($0 < x < x'$) by increasing x, price and total revenue move in opposite directions (price falls and total revenue rises). Moving along the demand curve in the inelastic range ($x' < x < x''$), also by increasing x, price and total revenue move in the same direction (price falls and total revenue falls).

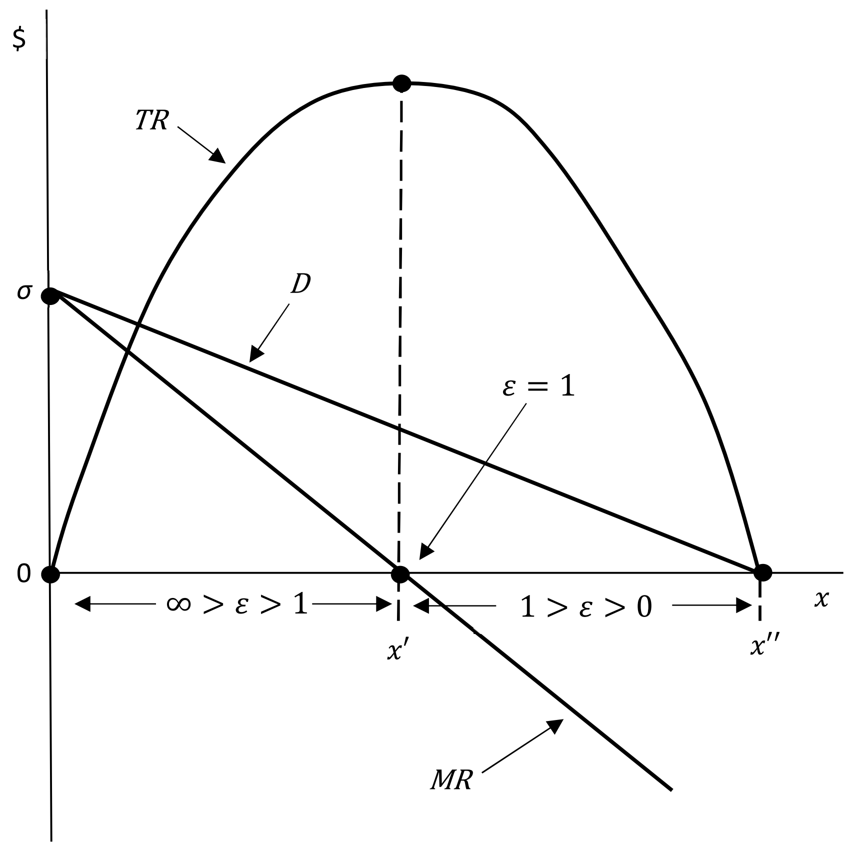


Figure 22-3

Observe that, as noted above and implied by the equation $MR = p_x\left(1 - {}^{1}/_{\varepsilon}\right)$, at each output x in Figure 22-3,

$$MR(x) < p_x. \tag{22.1}$$

Turning to the model explaining the monopolist's buying and selling behavior, impose the same assumptions on the monopolist as were required of the perfectly competitive firm (Supplemental Note D) and determine its cost functions in the same way. The short run profit of the monopolist as a function of output when $k = \bar{k}$ is given by

$$\pi(x) = TR(x) - SRTC^{\bar{k}}(x), \tag{22.2}$$

where $SRTC^{\bar{k}}(x)$ is calculated as it was in Chapter 12, and $TR(x)$ is obtained as described above. Profit maximization requires finding the output $\bar{x}$ that equates the derivative of (22.2) to zero, that is, that satisfies the first-order condition

$$MR(\bar{x}) = SRMC^{\bar{k}}(\bar{x}). \tag{22.3}$$

The second-order condition ensuring that $\bar{x}$ is actually associated with maximum profit and not a minimum or inflection point is more complex than that for the perfectly competitive firm and appears in Supplemental Note H.[2] Satisfaction of that condition, of course, ensures that the graph of $\pi(x)$ in equation (22.2) is strictly concave around the output $\bar{x}$ (see Figure 22-4 below). But, as shown in Supplemental Note H, it is not necessary that the marginal cost curve always slope upward at the profit-maximizing value $\bar{x}$ as under perfect competition. The mathematics of long-run profit maximization (that is, first- and second-order derivative conditions) is similar. Thus replace $SRTC^{\bar{k}}(x)$ and $SRMC^{\bar{k}}(x)$ by, respectively, $LRTC(x)$ and $LRMC(x)$ to obtain, for example,

$$MR(\bar{x}) = LRMC(\bar{x}).$$

The geometry of short-run profit maximization appears in Figure 22-4 in which the shape of the total revenue curve is that of Figure 22-3 (in that it starts at the origin, rises to a maximum, and then declines), the cost curves are the same as those for the perfectly competitive firm in Figure 13-4 of Chapter 13, and the total variable cost and average variable cost curves are not drawn. The differences between Figure 13-4 (profit maximization for the perfectly competitive firm) and Figure 22-4 arise only due to the different revenue curves. Note that:

1. $TR(x)$ has a maximum and $MR(x) = 0$ at x'.
2. $\pi(x)$ is maximized and $SRMC^{\bar{k}}(x) = MR(x)$ at $\bar{x}$.

(When drawing this diagram so as to include the appropriate characteristics of the curves involved, the following procedure could be employed: After lining up the two pairs of coordinate axes, draw the $SRTC^{\bar{k}}$ and TR curves first. Then, in the following order, draw π, MR [making sure that the MR line crosses the horizontal axis at x' where the TR curve has a maximum], the demand curve D [so that the distance from the point where the demand curve meets the horizontal axis is twice that from the origin to where marginal revenue is zero], $SRAC^{\bar{k}}$ [identifying that curve's minimum point at $\hat{x}$ and intersecting it with the demand curve at $\tilde{x}'$ and $\tilde{x}''$ where profit is zero]. Finally, draw $SRMC^{\bar{k}}$ so that it intersects MR at the profit-maximizing output $\bar{x}$ and $SRAC^{\bar{k}}$ at the latter's minimum point at $\hat{x}$.)

---

[2] In subsequent discussion, the possibility that $\bar{x}$ could be associated with minimum profit is ignored. Indeed, in the geometry of Figure 22-4 to be presented shortly, the total revenue and short-run total cost curves are drawn in such a way that a minimum satisfying equation (22.3) is not present.

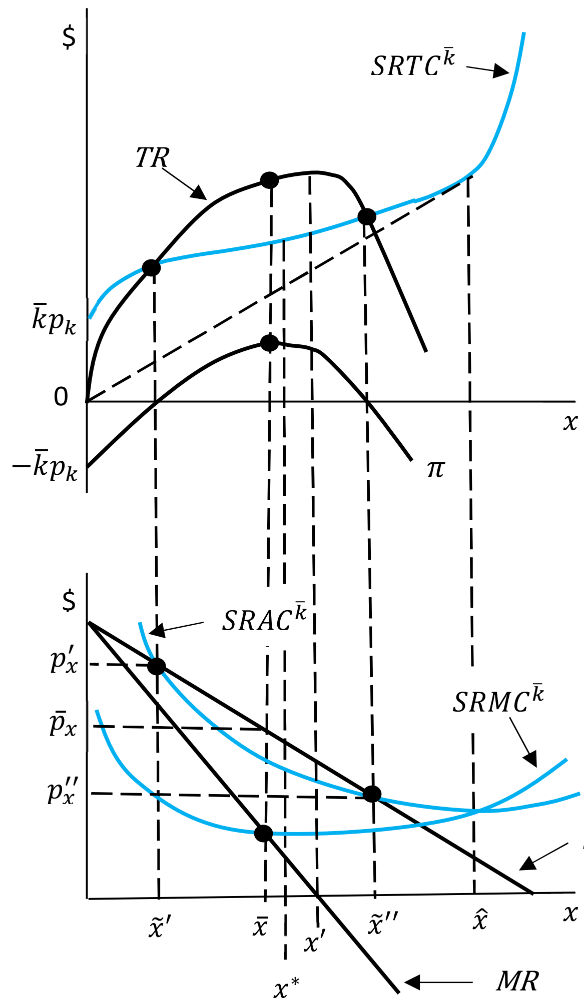


Figure 22-4

The following characteristics of the monopoly can be thought of in reference to Figure 22-4:

1. The profit-maximizing output $\bar{x}$, which requires $MR(\bar{x}) = SRMC^{\bar{k}}(\bar{x})$, can never be revenue maximizing since maximum revenue must occur at an x for which $MR(x) = 0$ and, for all values of x including the profit maximizing $\bar{x}$, $SRMC^{\bar{k}}(x) > 0$.

2. The monopolist will never produce in the inelastic portion of the demand curve *D* since marginal revenue is negative there and short run marginal cost is always positive.

3. The monopolist can lose money, that is earn a negative profit. This will happen if the short run total and average cost curves lie entirely above the total revenue and demand (average revenue) curves, respectively.

The output price set by the monopolist is that price at which, according to the demand curve it faces (D in Figure 22-4), all of (and no more than) its profit-maximizing output $\bar{x}$ will be sold. This is illustrated in Figure 22-24 and repeated in Figure 22-5 at A where $\bar{x}$ denotes the profit-maximizing output and the monopoly price is set at $\bar{p}_x$. Were the monopolist to charge more than $\bar{p}_x$, it could not, according to the demand curve it faces, sell all of the profit-maximizing output and profit would not be maximized. Were it to charge less for $\bar{x}$, its revenue would be less that it should be and, maximum profit would not be achieved. Notice in Figure 22-5 that $\bar{p}_x > SRMC^{\bar{k}}(\bar{x})$ at $\bar{x}$. The same inequality is obtained mathematically by combining (22.1) and (22.3), Thus, under monopoly the argument establishing the first fundamental theorem of welfare economics (Chapter 20) breaks down and inefficiency or an absence of Pareto optimality arises. That is, accounting for the remainder of the microeconomy's markets at whatever distribution is in force, consumer marginal rates of substitution would not equal the marginal rate of transformation at the economy's outputs. Even if all other markets in the microeconomy were perfectly competitive, it would still be possible by rearranging inputs, outputs, or their distribution to make one person better off (higher utility) without making anyone else worse off (lower utility).

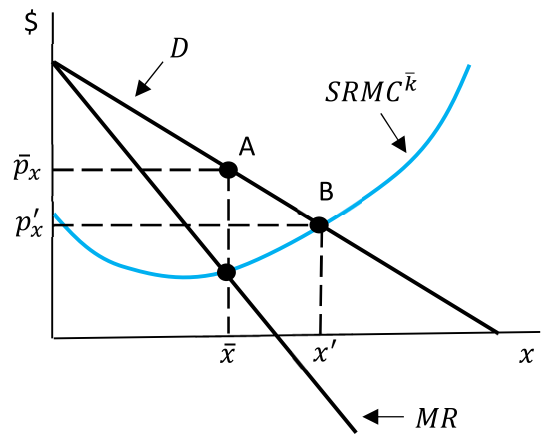


Figure 22-5

However, it should be pouted out that this inefficiency could be offset by economies of scale (Chapter 12). That is, due to its size and the nature of its cost curves, the monopolist could be using mass-production techniques that would not be available were it a smaller perfectly competitive firm. These techniques would enable it to produce its output at a much lower cost and therefore sell that output at a much lower price than would be the case under perfect competition. That, in turn, would increase consumer surplus (Chapter 8), thereby benefiting consumers and making up, to some extent, for the loss of Pareto optimality.

It has already been indicated in Chapter 21 that the presence of inefficiencies provides a justification for government intervention aimed at their reduction or elimination. The idea would be to lower the discrepancy between marginal cost and price in order to move closer to the perfectly competitive ideal. With respect to Figure22-5, that ideal occurs at B where $p'_x = SRMC^{\bar{k}}(x')$. In spite of the downward sloping demand curve, as long as the market for good y remains perfectly competitive, equation 20.12 of Chapter 20 holds, the first fundamental theorem of welfare economics applies, and Pareto optimality is restored. To achieve a movement towards

B, government intervention might mean breaking up the one large firm into several smaller ones as was done with AT&T when it had a monopoly in the United States telephone service industry,[3] or regulating the price that may be set as is currently done for monopolists who charge for delivering electricity to homes and businesses.

Returning to the analytics of monopoly, it should also be noted that there is no supply curve for the monopolist as there is for the perfectly competitive firm. This is because the monopolist does not determine the quantity it will supply for each price dictated by the market. Rather, the monopolist takes profit-maximizing action against the entire market demand curve (not against specific individual prices) to set both the output it will supply and the per-unit price it will charge for that output. This results in an equilibrium since the market demand curve indicates the quantities consumers will buy at each price. Once the price is set by the monopolist, consumers act on the quantity they demand at that price and the market is at rest. It follows that the competitive forces of demand and supply described in Chapter 3 no longer set the market price at the intersection of demand and supply curves in a market with a monopoly as they do under perfectly competitive conditions. A similar conclusion applies in the long run to be considered momentarily.

Since the labor market has been assumed to be perfectly competitive, the labor input the monopolist hires to produce the profit-maximizing output is determined the same way as that for the perfectly competitive firm – at the intersection of the isoquant associated with the profit-maximizing output $\bar{x}$ and the short run expansion path. In Figure 22-6, which is similar to the right-hand diagram in Figure 13-5 in Chapter 13, the monopolist hires ℓ^0 units of labor.

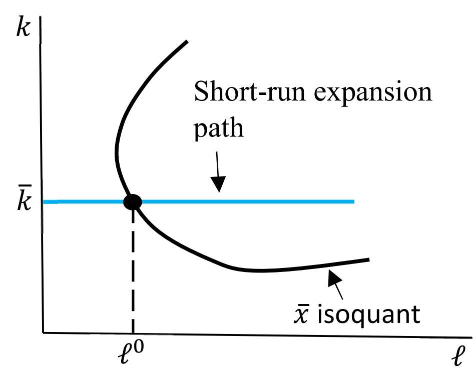


Figure 22-6

Like the short run, the long run analysis of monopoly also parallels that of the perfectly competitive firm (Chapter 13). In this case, the monopolist selects not only a profit-maximizing output and an output price, but also a profit-maximizing firm size k^0 with which to produce that output. Thus, as previously suggested, replace $SRTC^{\bar{k}}(x)$ with $LRTC(x)$ in equation (22.2) and adjust equations and graphs accordingly. Geometrically, profit maximization appears at $\bar{x}$ in Figure 22-7 and the output price is set at $\bar{p}_x$. The profit-maximizing quantities of labor and

[3] This break up resulted in lower telephone service costs, and a greater variety of telephone services and telephones available for consumers.

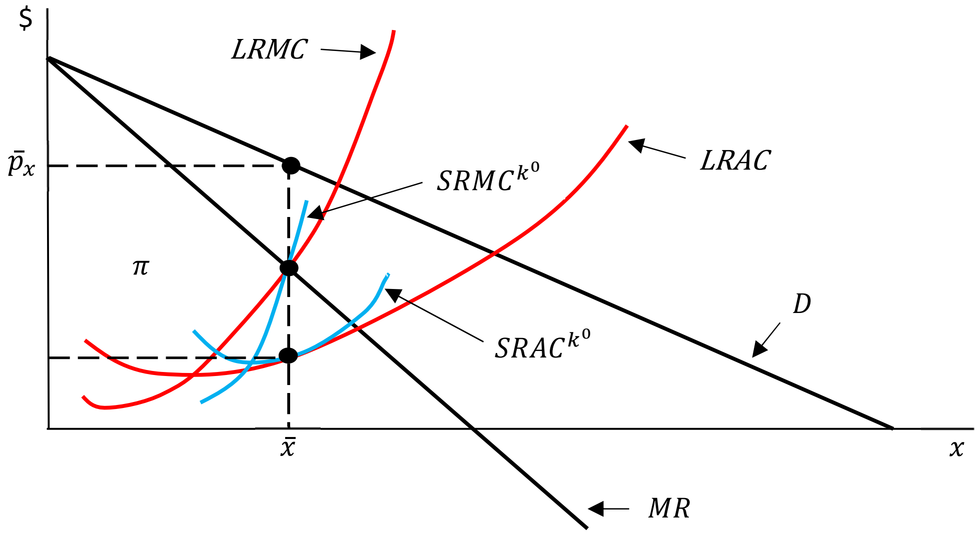


Figure 22-7

capital to be employed are determined, as in the perfectly competitive case, at the intersection of the long-run expansion path and the $\bar{x}$ isoquant. Thus in Figure 22-8, which is similar to the left-hand diagram in Figure 13-5 in Chapter 13, the firm hires (ℓ^0, k^0), and the short-run cost curves of the firm of size k^0 appear in blue in Figure 22-7. With respect to Figure 22-7, the long- and

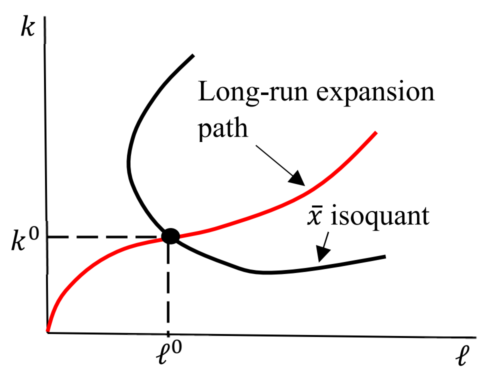


Figure 22-8

short-run average cost curves are tangent and the long- and short-run marginal cost curves intersect at $\bar{x}$ (Chapter 12). Of course, abnormal profit (the box labeled π in Figure 22-7) is not eaten away by the entry of new firms as under perfect competition. In the case of monopoly all entry is blocked.

Comparisons of the market outcomes for monopoly as opposed to those for perfect competition are not straightforward and require numerous assumptions. One such comparison is as follows:

Consider two alternative organizations of an industry – one as perfectly competitive with a large number of identical small firms; the other as a monopoly with all of the small firms combined into one. Assume input prices do not change when passing from one circumstance to the other. Assume also that the production function is the same for all firms and the monopoly. The relevant part of the production function for the small firms is associated with inputs relatively close to the origin of the input space; those for the monopoly are much farther out. Assume that, for fixed capital, the production function generates short run cost curves shaped like those drawn in Figure 22-4. Assume that, when both inputs vary, the production function has constant returns to scale so that long run average costs have constant economies of scale and are therefore constant as x varies (Chapter 12). Call the constant δ so that $LRAC(x) = \delta$. Then $LRTC(x) = x\delta$ and hence $LRMC(x) = \delta$. That is, the long-run average cost and the long-run marginal cost curves are the same. With the long-run average cost curve a straight line at δ parallel to the x-axis, each short-run average cost curve generated by a different firm size (Chapter12) is tangent at its minimum point to the long-run average cost curve. Geometrically, the relationship between the long-run and short-run average cost curves appears in Figure 22-9 where the two curves are drawn in red and blue respectively. (The short-run marginal cost curves pass through the minimum of their respective short-run average cost curves and are not drawn.)

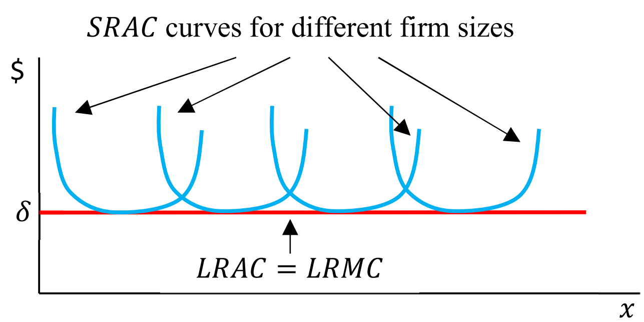


Figure 22-9

Since the production function is the same for both the identical small firms and the monopoly, the long-run marginal cost curve is also the same for both. The identical small perfectly competitive firms will have short-run average costs curves relatively close to the vertical axis based on the small amount of capital they employ. With its much larger amount of capital, the short-run average cost curve for the monopolist will be farther to the right. This is shown in Figure 22-10 where the same red line labeled $LRAC = LRMC$ applies to both the representative perfectly competitive firm and the monopoly. This line was identified with δ in Figure 22-9.

When there is one firm in the industry, long-run profit is maximized for the monopolist at x_M in Figure 22-10 where marginal revenue based on the market demand curve equals long-run marginal cost. The monopolist sets the market price at p_x^M. The long-run profit maximizing value of capital, call it k'', determines the short-run average cost curve along which the firm produces. It is shown in Figure 22-10 as tangent at A to the long-run average cost curve and labeled $SRAC^{k''}$. Were its associated short-run marginal cost curve drawn, it would intersect the short-run average cost curve at A.

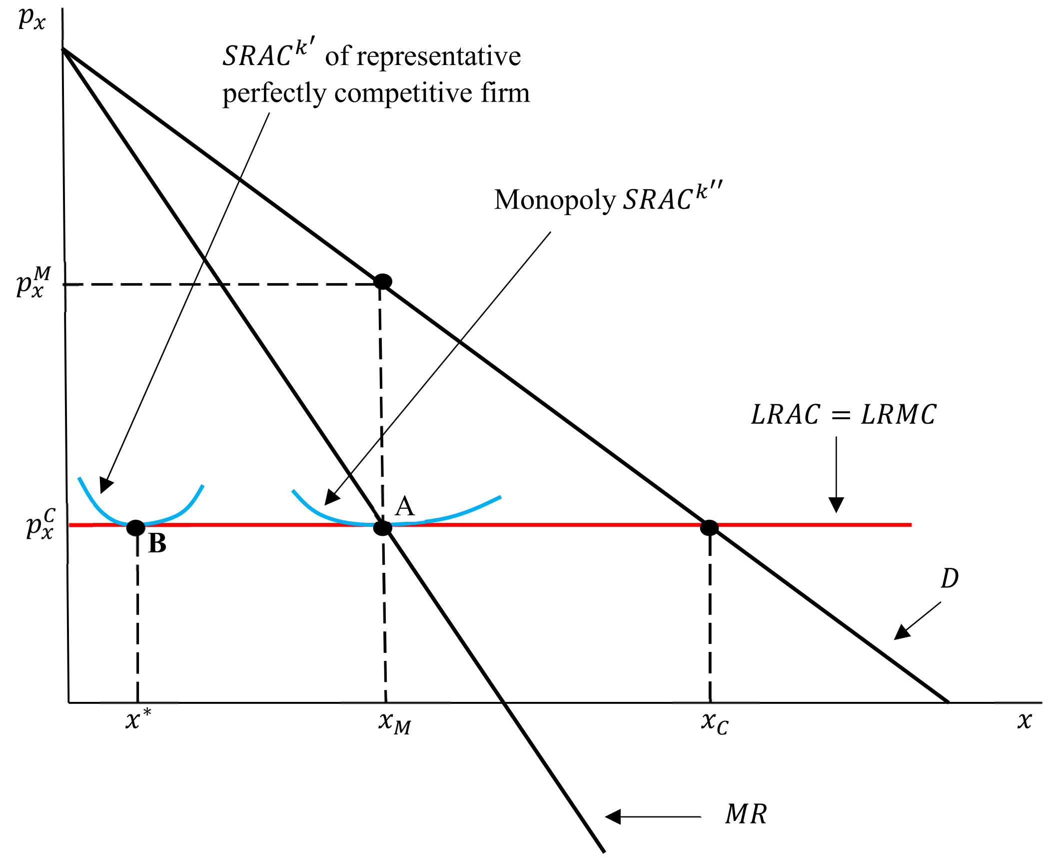


Figure 22-10

Now consider the industry as organized perfectly competitively. At long-run equilibrium under perfect competition, price equals long-run marginal cost at the level of minimum of long-run average cost (Chapter 13 and 15). Since, in Figure 22-10 the long-run average cost curve is a horizontal line at p_x^C, every point on it is a minimum, and since the long-run marginal and average cost curves are identical, all conditions for long-run equilibrium (market supply equals market demand and, for the representative firm, $p_x^C = LRMC(x^*)$ and $\pi(x^*) = 0$) are met at the intersection of the red $LRMC$ line and the demand curve, that is, at (x_C, p_x^C). Assuming the number of identical small firms in the industry (market) is known, dividing that number into x_C gives the output of each perfectly competitive firm. That output, that is, the output of the representative firm, is taken to be x^* in Figure 22-10. The appropriate short-run average cost curve for producing this output is $SRAC^{k'}$, the one whose associated firm size is $k'$and is tangent to the long-run average cost curve at point B. Again, the relevant short-run marginal cost curve which would pass through B, is not shown.

Under the highly restrictive assumptions imposed here, in the long run market price is higher and market (output) quantity lower under monopoly. As can be seen in Figure 22-10, profit is positive under monopoly and vanishes as it should under perfect competition.

Chapter 23
The Social Cost of Monopoly and Monopolistic Competition

The social cost of monopoly due to inefficiencies (the absence of Pareto optimality) can be measured in terms of the loss of consumer surplus (Chapter 8) net of any gain in producer surplus (Chapter 14) when the outcome under monopoly is compared to that arising were the monopolist to charge the perfectly competitive price. That price would be based on its short-run marginal cost curve or what would be the supply curve if the monopolist were a perfectly competitive firm. Since those surpluses are seen as benefits to the consumer, the greater the loss of surplus, the greater the social cost. In Figure 32-1, (x_C, p_x^C) is the perfectly competitive

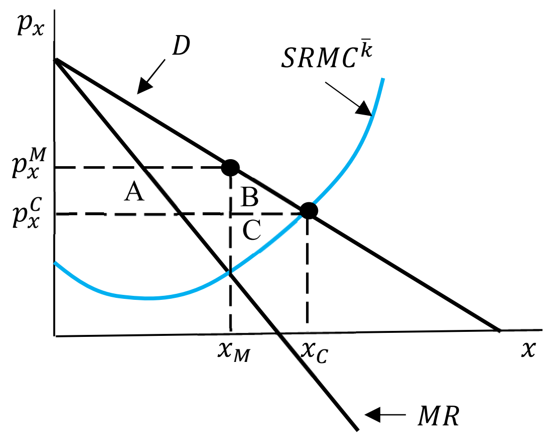


Figure 23-1

outcome where, as described in Chapter 22, the monopolist, behaving like a perfect competitor, equates it's short run marginal cost[1] to the market price p_x^C. The monopoly outcome, where $SRMC^{\bar{k}}(x) = MR(x)$ occurs at (x_M, p_x^M). In moving to the monopoly position from that in which the monopolist charges the perfectly competitive price, the loss in consumer surplus is the enclosed area $A + B$. The gain in producer surplus accounting for the loss of area C is $A - C$. Combining the two, the (net) social cost of monopoly resulting from the monopoly form of market failure, also referred to as the deadweight loss, is

$$(A + B) - (A - C) = B + C.$$

Suppose the government were to point to the market failure arising from the presence of inefficiency in terms of Pareto optimality as the justification for a proposal to require the monopoly to move its position from (x_M, p_x^M) towards (x_C, p_x^C) . It could suggest, for example, regulating the monopoly by forcing it to charge a lower price. But the monopolist might respond by using its abnormal profit (which, as has been seen in Chapter 14, is part of producer surplus) for political or lobbying activity to protect that profit rather than returning it to the owners of the firm. In that case, consumers would benefit less because the funds owners (who are consumers)

[1] As in Figure 14-6, the marginal cost curve is extended to the vertical axis for geometric simplicity.

receive would be smaller thereby making the social cost of monopoly greater. This is called rent-seeking behavior. To the extent that it is successful, it leads to what may be referred to as government failure to reduce the effects of the monopoly- induced market failure.

Attention now turns to the second form of market failure to be considered here, namely, monopolistic competition. This is a situation in which all of the properties of perfect competition are maintained except that each firm is now producing a "unique" product that is at least slightly different from, and closely substitutable for all others in the market. Because its product is unique, it will be seen that, like the monopolist, the monopolistically competitive firm is able to set the price of its output. Here, the requirement that there are a large number of small sellers means that the firm faces a demand curve determined by its market share. That share is assumed to be fixed by market characteristics beyond the control of the firm.[2] As in the case of monopoly, the demand curve facing the firm is taken to be a downward sloping straight line known by the firm.

Market power is the ability of the firm to raise the price it charges for its output and still maintain some or most of its customers. Generally, the source of market power is the availability or lack thereof of substitute commodities to which buyers can switch as the firm raises its price. Under perfect competition, where all firms in a market sell identical products, the firm has no market power and the demand curve it faces is a straight line parallel to the output axis at the level of the market price. Were such a firm to raise its price above the market price, it would lose all of its sales. In the case of monopoly, where there is only one firm selling a product for which there is no close substitutes, the demand curve the firm faces is the market demand curve. The firm's market power is maximal in the sense that the sales it loses upon raising its price arise only because some customers no longer want to buy the product at the higher price. The market power of the monopolistically competitive firm is in between that of the perfectly competitive firm and the monopolist. That is because if the firm were to raise its price, (a) some customers may no longer want to but any of the market's products as in the case of monopoly and (b) the availability of close substitute commodities makes it possible for some buyers, who still want to buy the product, to switch to other firms' products. As the monopolistically competitive firm raises its price, then, it will lose more sales than if it were a monopolist. The buyers who continue to buy from the monopolistically competitive firm as it raises its price are considered to be loyal to that firm's "brand."

Since elasticity indicates the responsiveness of quantity changes to price changes at each point on the demand curve (Chapter 4), it may be taken to be a measure of the firm's market power at those points. The lower the elasticity, the lower the percentage response to the same price increase and the greater the market power. Thus, the elasticity along the demand curve facing the monopolistically competitive firm would lie between that of the perfectly competitive firm (infinite elasticity) and the monopolist (the elasticity of the market demand curve).

---

[2] The possibility that the firm might be able to increase its market share or the demand for its output through advertising is not considered.

The firm operating under monopolistically competitive conditions may be thought of as a "weak" monopolist in the sense that no other firm sells the exact same product, although there are other products that can substitute for it. The model built to explain the monopolistically competitive firm's behavior, then, is like that constructed for the monopolist except for the particularization of the demand curve facing the firm. The mathematics for the maximization of its profit is identical. For the short run with $k = \bar{k}$,

$$\pi(x) = TR(x) - SRTC^{\bar{k}}(x),$$

and profit is maximized at that value x^0 for which the first-order condition ${d\pi(x)}/{dx} = 0$ is satisfied or

$$MR(x^0) = SRMC^{\bar{k}}(x^0),$$

with second-order conditions appropriately satisfied as described in Chapter 22. Geometrically, diagrams similar to Figures 22-4, 22-5 and 22-6 of Chapter 22 depicting short-run profit maximization for the monopolist apply, with the demand curve facing the monopolistically competitive firm replacing that of the monopolist. The latter two diagrams as modified for present use are reproduced here. In Figure 23-2, the profit-maximizing output is x^0 and the price

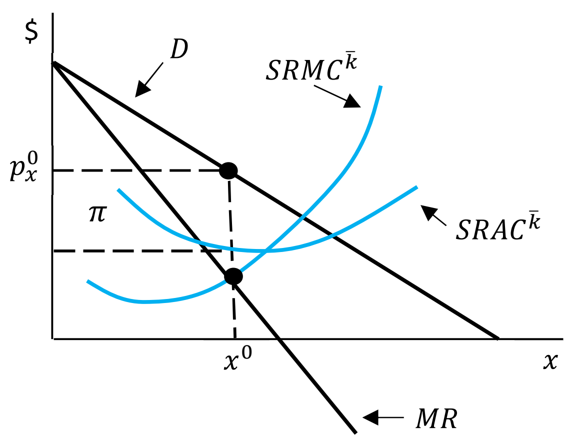


Figure 23-2

set by the monopolistically competitive firm is p_x^0. As with monopoly, $p_x^o > SRMC^{\bar{k}}(x^0)$ and inefficiency is present. Observe that, as drawn, the representative firm has abnormal profit since $\pi(x^0) > 0$. The firm's profit- maximizing quantity of labor input is ℓ^0 in Figure 23-3. Figure 23-3 is also similar to the right-hand diagram of Figure 13-5 in Chapter 13 for the perfectly competitive firm.

In the long run, the above mathematics is the same with $LRTC(x)$ and $LRMC(x)$ replacing $SRTC^{\bar{k}}(x)$.and $SRMC^{\bar{k}}(x)$ But here, if abnormal profit were present (as in Figure 23-2), outsiders would see and take advantage of an investment opportunity and cause new firms, each producing new and different substitute products (with the same cost and revenue functions and curves as the existing firms), to enter the market. This would result in a decline in market share of the previously existing firms, and a shift of the demand curves facing them and

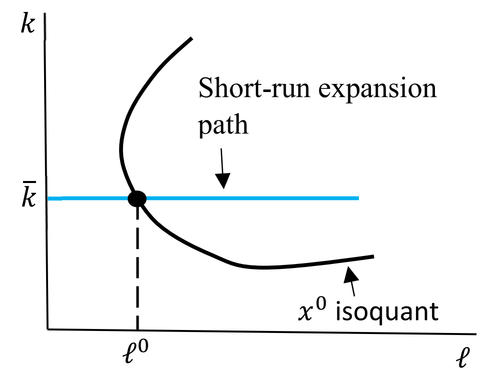


Figure 23-3

associated marginal revenue curves towards the origin. Abnormal profits would fall. Entry would continue until all abnormal profits of all firms old and new in the market disappeared. The firms and industry would then be at long-run equilibrium. At this point, the geometry of the representative firm would appear as shown in Figures 23-4 and 23-5. In Figure 23-4, the long-

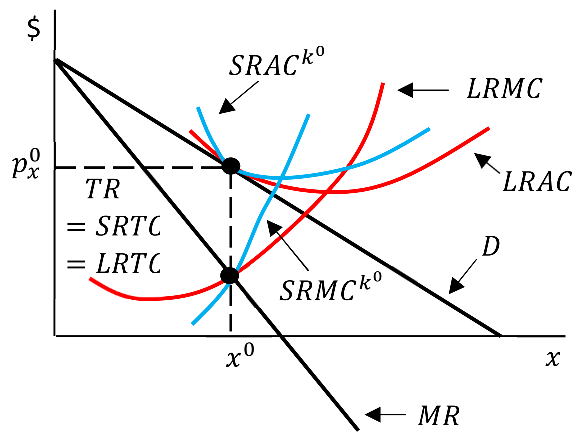


Figure 23-4

and short-run average cost curves are tangent to each other and to the demand curve facing the firm at the profit-maximizing output x^0, and the long-run and short-run marginal cost curves intersect at that same level of output (Chapter 12). The tangencies indicate that $\pi(x^0) = 0$, that is, that the area representing total revenue in Figure 23-4 is identical to that representing short- and long-run total cost. Thus, abnormal profit is no longer present. In Figure 23-5, which is similar to Figure 22-8 (Chapter 22) and the left-hand diagram in Figure 13-5 (Chapter 13), the firm employs its profit-maximizing quantity of capital k^0 as determined by the intersection of the isoquant for the profit-maximizing output and the long-run expansion path, and this, in turn,

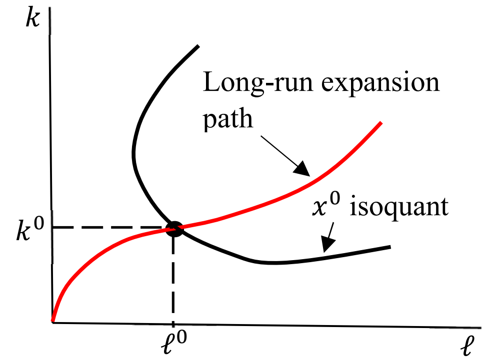


Figure 23-5

determines the firm's size and the$SRAC^{k^0}$ and $SRMC^{k^0}$ curves in Figure 23-4.

The long-term position of a representative firm under monopolistic competition (shown in Figure 23-4) as compared to that of the same firm under long-run perfectly competitive conditions (from Figure 15-5 of Chapter 15) is pictured in Figure 23-6. This comparison assumes

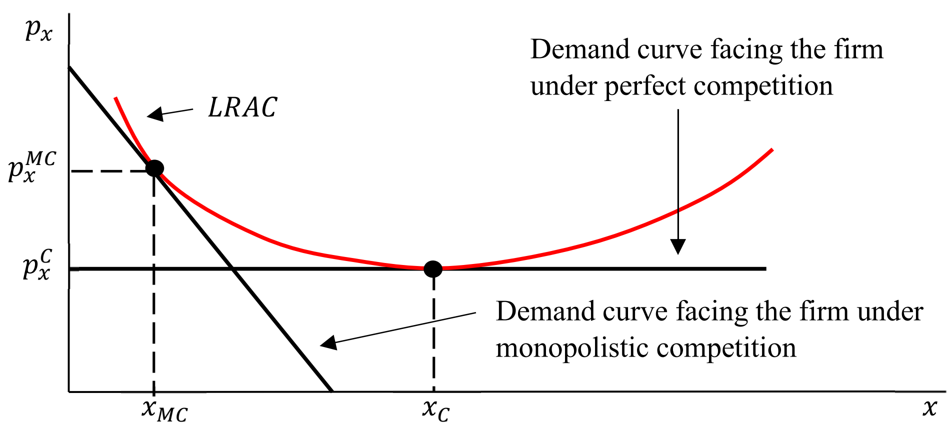


Figure 23-6

that, in spite of the differences in the products produced, all monopolistically competitive firms have the same cost curves as they would if they operated under perfectly competitive conditions. That is, the long-run average cost curves are the same for both types of firms. Short-run curves are not shown. When the industry is monopolistically competitive, the representative firm's long-run output-price position is located at a tangency between the downward sloping demand curve it faces and its long-run average cost curve. When the industry is organized perfectly competitively, the representative firm produces at minimum long-run average cost. Thus the representative firm charges a higher price p_x^{MC} and produces a smaller output x_{MC} under monopolistic competition than under perfect competition, where its output price is p_x^C and its output is x_C.

It has been previously noted that since output price is greater than marginal cost in a monopolistically competitive industry, inefficiency (a lack of Pareto optimality) is present. That is, by changing output quantities, changing input quantities, or distributing them differently across the microeconomy, it would be possible to increase at least one person's utility without lowering that of any other person. But there is a benefit offsetting the loss of Pareto optimality under monopolistic competition, namely, that in a monopolistically competitive market there is a large number of different product varieties among which consumers are able to choose. Rather than providing only one manifestation of a product for all buyers as in the perfectly competitive or monopoly case, the monopolistically competitive market like, for example, the market for breakfast cereals,[3] provides a huge variety of product manifestations so as to satisfy a large number of different tastes or preferences.

Precise comparisons of the long run price and quantity outcomes of the same industry organized under perfect competition, monopolistic competition, or monopoly are complicated and will not be pursued here. Suffice it to say that because of the differences in market power in the different competitive situations, and because, in the long run, the monopolist generally retains abnormal profit while the perfectly and monopolistically competitive firms do not, one might expect that, as pictured along a market demand curve in Figure 23-7, the price of the product to be highest were the industry organized perfectly competitively (at p_x^C). The quantity

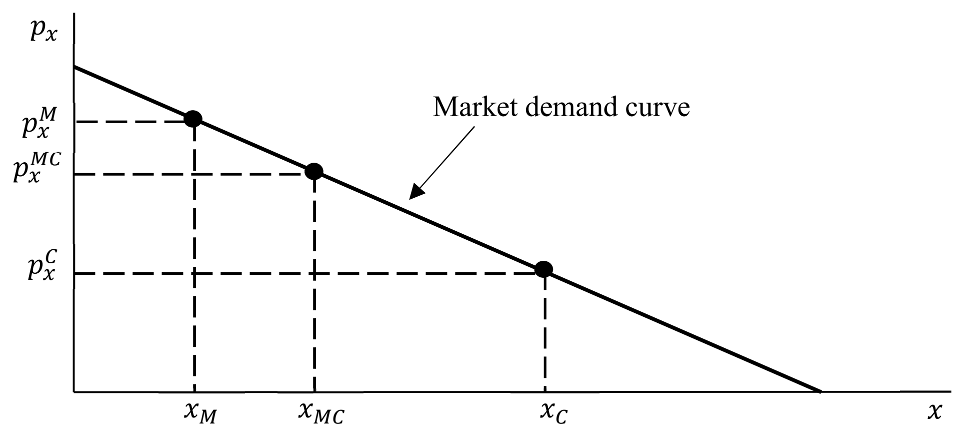


Figure 23-7

produced would be lowest under monopoly (at x_M) and highest in the perfectly competitive situation (at x_C). And the average price and full market quantity over all firms under a monopolistically competitive organization would be in between at (x_{MC}, p_x^{MC}). But it should be noted that recognizing differences in production and cost functions across the different competitive organizations and within the monopolistically competitive industry, and differences in demand curves facing the monopolistically competitive firms could easily confound this expectation.

[3] Although there are very large firms producing breakfast cereals, this market still has characteristics of a monopolistically competitive industry.

Chapter 24
Oligopoly

Turning to the third form of market failure, a market or industry having a large number of small buyers facing a small number of large sellers with limited seller entry into the market is an oligopoly. Products could either be standardized (perfect oligopoly) or differentiated (monopolistic oligopoly). Like monopoly and monopolistic competition, in an oligopolistic market the competitive forces of supply and demand do not set the market price at the intersection of demand and supply curves.

To explain the buying and selling behavior of a firm under perfectly competitive, monopolistically competitive, or monopoly conditions, it has been assumed that the firm knows the demand curve it faces. Using that demand curve to obtain the firm's total revenue function and calculating the firm's total cost function (given prices in assumed perfectly competitive input markets) allows determination of its profit function. Then applying the assumption that the firm hires inputs and produces output so as to maximize profit leads to an explanation of the firm's output supply and input demand behavior.

However, this procedure cannot generally be followed for oligopoly. Since there are only a few large firms and since market demand is fixed, each firm knows that any price decision it makes will likely have a direct impact on its competitors. And it is the way its competitors react in terms of their own price decisions that determines the demand curve facing the firm. For example, if firm A, at one point on the demand curve it faces, were to lower the price of its product, sales would probably be drawn away from firm B. Were that to happen, firm B would likely respond by lowering its own price. Depending on how much firm B reduces it price, some or all of its sales would be restored, and B's lower price might even draw additional sales away from firm A. Firm A, then, cannot know what it can sell at the lower price, that is, it cannot know the point associated with the lower price on the demand curve facing it, until it sees what B will do. Thus, without knowing how B will react to its price changes, A cannot know the demand curve it faces. But there is a problem here that applies to all oligopolistic firms. In general, these firms do not know how their competitors will react to their price changes. And therefore, they cannot know the demand curves they face. In this sense oligopoly situations are characterized by uncertainty.

In light of this uncertainty, there are two ways to construct explanatory models of the oligopolistic firm and thereby explain its behavior:

1. Incorporate the uncertainty and include it as part of the explanation. This approach generally employs the notion of 'strategic behavior' – the idea that one firm's actions take into account possible measures and reactions taken by its competitors – as the basis for constructing explanatory models. This approach, however, will not be considered here.

2. Assume away the uncertainty. That is, make assumptions about the behavior of the firm and its competitors that allow determination of the demand curve it faces. Once the demand curve facing that firm is obtained, construction of an explanatory model of its behavior proceeds

as described for other forms of market structures in previous chapters. The type of assumptions made in this approach determines the nature of the explanation obtained.[1]

In the discussion that follows, three examples of explanatory oligopoly models based on assuming away the uncertainty will be considered. The first two apply to the case of standardized products; the third to differentiated products. In all cases firm profit-maximizing output quantity is inconsistent with efficiency or Pareto optimality since the marginal cost of that output is always less than the output price.

Before proceeding, a preliminary point should be mentioned. When a firm under perfect or monopolistic competition or monopoly is hiring inputs and producing outputs so as to maximize its profit, that firm is thought to be in equilibrium because it is at rest and has no desire to change anything. The firm is doing the best that it can against what the market as a whole presents to it (that is, against the market price or the demand curve it faces) and has no reason to change its output, input and, if it sets the price of its output, its output price. When a firm is at equilibrium under oligopoly in which an assumption that eliminates the uncertainty has been imposed, the equilibrium is referred to as a Nash equilibrium. At that equilibrium, the firm is doing the best it can against what its individual competitors present to it (their reactions to various prices set by the firm) and, again, there is no reason for it to change anything.

The first explanatory oligopoly model to be considered here arises in the context of short-run perfect oligopoly and focuses on an industry structure in which there is price leadership by a dominant firm. In particular, there is one dominant firm in the market with the remaining firms following the dominant firm's lead. The assumption that eliminates the uncertainty is as follows:

> The dominant firm sets the market price, lets the remaining firms sell all they want to sell at that price, and then sells to the demand that is left over.

Since the remaining firms take the price set by the dominant firm as fixed, this assumption implies that they behave as if they were perfect competitors reacting to a fixed price set by the market. As in the case of perfect competition, the supply curve of the remaining firms combined is the sum of their short-run marginal cost curves given their fixed quantity of capital (Chapter 14). For convenience that curve is taken to be a straight line starting from the vertical axis. To obtain the demand curve facing the dominant firm, at each price, the total quantity supplied by the remaining firms is subtracted from the total quantity demanded on the market. Using this demand curve and the dominant firm's cost function, the dominant firm's profit function is calculated. The dominant firm then chooses the profit-maximizing output and price as if it were a monopolist. The geometry appears in Figure 24-1.

In reference to Figure 24-1, the market demand curve is the downward sloping green line, and the supply curve of the remaining firms combined is the upward sloping green line. Both

[1] These approaches overlap in that identical analytical structures can emerge from their application to the same oligopoly phenomenon.

curved are assumed to be known to the dominant firm. The demand curve facing the dominant firm, the upper solid black line, is calculated by taking the quantity demanded of the dominant firm's output at each price, say p_x^d, as the difference, x^d, between the market quantity demanded at that price, x^m, and the sum of the quantities supplied by the remaining firms at that price, x^s, that is

$$x^d = x^m - x^s.$$

Therefore, if the dominant firm sets the price at p_x' (or higher), the remaining firms supply the entire market x' and the dominant firm sells nothing. If the dominant firm sets the price at p_x'' (or lower), then the remaining firms supply nothing, and the dominant firm supplies the entire market, at least x''. The demand curve facing the dominant firm follows the upper solid black line from the vertical axis at p_x' to point A and then continues along the green line from A to the x-axis. The dominant firm's marginal revenue curve, the lower solid lack line, is calculated in

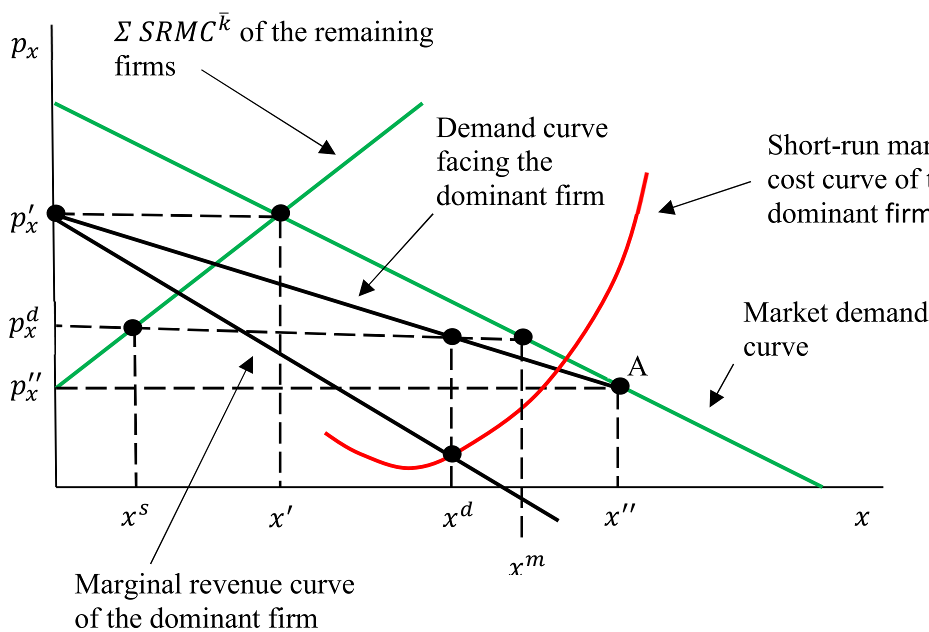


Figure 24-1

the usual manner by extending the black portion of its demand curve to the x-axis (Chapters 4 and 22, and supplemental Note B). The latter extension is not shown in the diagram. (The marginal revenue curve for the green portion of its demand curve facing the dominant firm is irrelevant for this diagram.)

Like the monopolist, the dominant firm maximizes its profit at x^d where its short-run marginal cost given its capital input or plant size equals its marginal revenue. Like the monopolist, it sets the market price at the level p_x^d at which, according to the demand curve it faces, it can sell that output. The quantity sold on the market is x^m. The remaining firms sell $x^s = x^m - x^d$.

The second explanatory model of perfect oligopoly behavior to be discussed here is that of the cartel. Cartels are illegal in the United States. But they exist elsewhere in the world. Perhaps the most well-known is the Oil Producing and Exporting Countries or OPEC cartel. When the cartel controls the entire industry, the assumption that eliminates the uncertainty is stated as:

> Each firm in the industry follows the price and output instructions of a central office which seeks to maximize the profits of the industry as a whole.

That is, the central office turns the industry into a monopoly facing the known market demand curve. To illustrate in the short run with two firms in the industry numbered 1 and 2, let x_1 and x_2 indicate their respective outputs, and $\bar{k}_1$ and $\bar{k}_2$ their respective capital inputs or plant sizes. Write $SRTC^{1\bar{k}_1}$ and $SRTC^{2\bar{k}_2}$ for their short-run total cost functions. Then with TR designating the combined revenue function based on the market demand curve, industry or cartel profit π would be written as

$$\pi(x_1, x_2) = TR(x_1 + x_2) - SRTC^{1\bar{k}_1}(x_1) - SRTC^{2\bar{k}_2}(x_2),$$

where $x_1 + x_2$ represents total market quantity demanded from which total revenue is calculated. Denote industry marginal revenue by $MR(x_1 + x_2)$. Assuming second-order conditions are satisfied, maximization of cartel profit requires x_1 and x_2 be selected so that[2]

$$MR(x_1 + x_2) = SRMC^{1\bar{k}_1}(x_1) = SRMC^{2\bar{k}_2}(x_2). \tag{24.1}$$

If the two short-run marginal costs were not equal, then transferring production from the firm with the higher marginal cost to the one with the lower marginal cost would reduce the total cost of producing the same output and increase profit. Hence cartel profit could not be at a maximum. Since the two marginal costs are equal, their value becomes the cartel short-run marginal cost. Cartel profit is then maximized by setting cartel marginal revenue equal to cartel short-run marginal cost. The verbal argument that at maximum profit marginal revenue should equal short-run marginal cost is the same here as that given following equation (13.6) in Chapter 13. The two marginal equations obtained from (24.1), namely

$$MR(x_1 + x_2) = SRMC^{1\bar{k}_1}(x_1) \quad \text{and} \quad MR(x_1 + x_2) = SRMC^{2\bar{k}_2}(x_2),$$

determine the maximizing values of x_1 and x_2 by solving them simultaneously. Based on these determinations, the cartel sets the output quantities that firm 1 and firm 2 produce. As with monopoly, the market price is also set by the cartel so that, according to the market demand curve, all of (and no more than) the maximizing quantities x_1 and x_2 are sold.

The last explanatory oligopoly model to be considered is that of the ‘kinked” demand curve. It has been observed in the past that under certain oligopoly situations, relatively long

---

[2] The mathematical derivation of these first-order equations is not presented here.

periods of time have occurred during which firm output price and output quantity have remained relatively constant in spite of input cost and output demand changes. The short-run kinked demand curve model was constructed to explain this phenomenon. In that context, there are several assumptions that need to be made to eliminate the uncertainty and construct the demand curve facing the firm:

a) Products sold in the market are differentiated. That is, the firm on which attention focuses is selling a unique product for which there are close substitutes.

b) There are two known demand curves facing the firm — one, D^1, based on the assumption that no competitor follows the firm's price changes and the other, D^2, based on the assumption that all competitors follow its price changes.

c) A price p_x^0 and output x^0 has been established in the market and the two demand curves it faces intersect at (p_x^0, x^0). The curves appear with D^1 (in black) flatter than D^2 (in green) in Figure 24-2 for the following reason: *I*f no competitor follows a cut in price from p_x^0, then the firm will sell more at the new price along D^1 than if all competitors follow along D^2. And if no competitor follows a price increase from p_x^0, then the firm will sell less at the new price along D^1 than if all competitors follow along D^2.

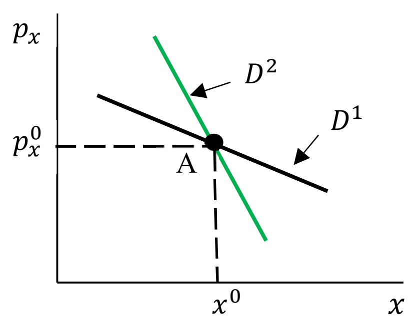


Figure 24-2

d) Finally assume that all competitors follow price reductions from p_x^0 but no competitor follows price increases.

Under these assumptions, the demand curve facing the firm is D^1 for prices above p_x^0 and D^2 for prices below p_x^0, thus producing a kink where the two parts of the demand curve meet at A.

The geometry of profit maximization under these assumptions is shown in Figure 24-3. In that diagram, the demand curve facing the firm (reproduced with the same colors on a larger scale from Figure 24-2) consists of the solid black and green lines labeled D^1 and D^2 respectively, that meet with a kink at point A. The marginal revenue curve for D^1 is calculated in the usual way after extending the solid line D^1 to the x-axis. (The extension appears as a dotted black line.) Only the solid portion of that marginal revenue curve, labeled MR^1, from the vertical axis to point B is relevant since the demand curve changes at output x^0 (point A). The marginal revenue curve for D^2 is obtained by extending D^2 to the vertical axis (dotted green line) and only the solid portion labeled MR^2 starting a point C is used. Because, when moving along demand curve its slope jumps across many values at point A, the marginal revenue there consists of the range of values between points B and C. Combining MR^1 and MR^2 with the vertical line connecting points B and C (colored red) gives the complete marginal revenue curve of the firm.

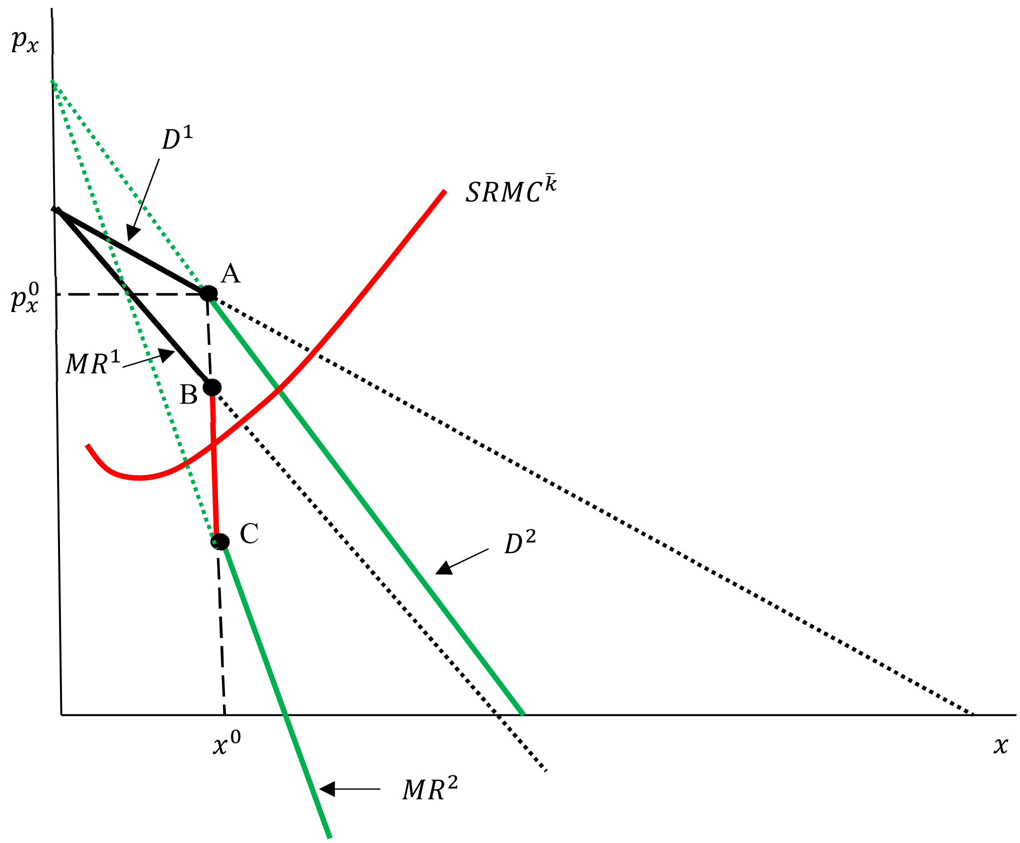


Figure 24-3

Profit maximization occurs at the output for which the marginal revenue curve intersects the short-run marginal cost curve specified for the firm's capital $\bar{k}$. In the diagram, this intersection occurs in the vertical portion between points B and C, and the firm is selling the established quantity x^0 at the established price p_x^0. Now if input prices change, then the marginal cost curve will change. As long as the input price variations (with the marginal revenue curve remaining fixed) are not large enough to move the intersection of marginal revenue and short-run marginal cost outside of the vertical (red) portion of the marginal revenue curve, there will be no change in the firm's output price and output quantity. Changes in demand might affect either or both of the two demand curves D^1 and D^2. Assume their slopes are altered while maintaining their intersection at (p_x^0, x^0). This, in turn, will affect MR^1 and MR^2 and hence the length of the red vertical portion of the marginal revenue curve. Once again, as long as points B and/or C (with the marginal cost curve remaining fixed) do not modify so that the intersection of marginal revenue and short-run marginal cost remains in the vertical (red) portion of the marginal revenue curve, then these changes in demand will have no effect on output price and output quantity. Thus the kinked demand curve model suggests that with sufficiently small changes in input prices and demand conditions, output price and quantity can remain constant for long periods of time.

Chapter 25
Externalities, Public Goods, and Imperfect Information

The last three chapters have focused on firm behaviors and efficiency issues that arise when markets fail due to imperfections in their competitive structures. It is now time to consider the problems of externalities, public goods, and imperfect information – market failures that arise for different reasons. These failures were defined and examples of them provided in Chapter 21.

Recall that an externality exists if the buying-selling actions of one or more persons or firms in a market impose costs on or provide benefits to other individuals or firms outside of the market. When an externality imposes costs, it is referred to as a negative externality; when it provides benefits, it is called a positive externality. Until now the possible presence of an externality in a perfectly competitive market has been ignored. The costs or benefits associated with it have not been accounted for in any of the concepts or equations of the Walrasian model described in previous chapters. Clearly, introducing them into the model necessarily leads to a different position of rest or equilibrium. Call the latter an externality-generated general equilibrium. In that context, it can be shown that an externality-generated general equilibrium is Pareto optimal.[1] But its distribution of inputs and outputs is different from that obtained in the original perfectly competitive general equilibrium before that model was modified to account for the externality. Once the modification is made, the before-modification distribution of inputs and outputs is no longer Pareto optimal and at that distribution it is now possible through redistribution with possibly modified output quantities to make at least one person better off (higher utility) without making anyone else worse off (lower utility). This may be understood to happen because in a market with an unaccounted-for externality, either market price no longer measures the actual worth or benefit of a unit of the good to society or (short-run or long-run) marginal cost no longer measures the true marginal cost of that unit to society. The market fails since the Pareto optimal outcome that now accounts for the costs or benefits of the externality is not achieved.

To be more specific, focus on a single firm in the short run. When externalities are not present, profit in the perfectly competitive firm producing good x is defined by the difference between total revenue and total cost (Chapter 13):

$$\pi(x) = TR(x) - SRTC^{\bar{k}}(x). \tag{25.1}$$

In this equation, $SRTC^{\bar{k}}(x)$ represents only the cost of the inputs used to produce output x. And with p_x thought of as what consumers are willing to pay per unit of quantity x (Chapter 8) and hence as a measure of the benefits per unit of quantity x accruing to them, $TR(x) = xp_x$ may be understood as total benefits to society received from the sale of output x.[2] Under this

[1] The argument is essentially the same as that given in Chapter 20 to establish the first fundamental theorem of welfare economics.

[2] Of course, this is not the same as consumer surplus. Referring to Figure 8-4 of Chapter 8, consumer surplus is the area of triangle ABC. Total revenue or total benefits to society is the rectangle below it.

interpretation, $\pi(x)$ becomes the benefit to society of producing output x net of the total cost of its production. Maximizing profit becomes maximizing net social benefit, and at the x associated with that maximum

$$SRMC^{\bar{k}}(x) = p_x, \tag{25.2}$$

(combined with the equations of the general equilibrium of which it is a part), Pareto optimality or efficiency prevails (Chapter 20).

Applying these ideas when both positive and negative externalities are present in the market, include in total revenue the benefits of output x to individuals and firms outside of the market for good x, that is beyond the revenue internal to the firm or xp_x. Call this expanded form of total revenue total social benefit (it remains a function of x). Similarly, in addition to the internal cost of the inputs, include in total cost all of the outside costs of the externality and rename it total social cost (also a function of x). Then equation (25.1) becomes

$$\text{net social benefit } (x) = \text{total social benefit } (x) - \text{total social cost } (x), \tag{25.3}$$

In parallel with the discussion of equilibrium and efficiency in the absence of externalities, to obtain the externality-generated general equilibrium and Pareto optimality for the full microeconomy that includes the externalities in the market for good x, equation (25.2) has to be maximized. In particular, assuming second-order conditions are satisfied, maximizing net social benefit by differentiating equation (25.3) and equating the derivative to zero gives a quantity x for which the derivatives or marginals of total social benefit and total social cost are equal:

$$\text{marginal social benefit } (x) = \text{marginal social cost } (x)$$

or, symbolically representing marginal social benefit (x) by $MSB(x)$ and marginal social cost (x) by MS$C^{\bar{k}}(x)(x)$,

$$MSB(x) = MSC^{\bar{k}}(x)(x). \tag{25.4}$$

Equation (25.4) replaces (25.2) for the firm producing good x. Combining (25.4) with the remaining equilibrium equations reflecting perfect competition elsewhere without externalities ensures Pareto optimality. This follows from the analogue of the first fundamental theorem of welfare economics when externalities are present in the market for good x.

Of course, if externalities are present but ignored by the profit-maximizing firm producing good x, the output of that firm still satisfies equation (25.2). But that is no longer good enough for Pareto optimality because the extra costs or benefits have been left out. What is needed in in this case to secure Pareto optimality is equation (25.4).

To clarify the argument with externalities present still further, let $SRMC^{\bar{k}}(x)$ be the original marginal cost concept, that is, the marginal cost of producing output x that includes only the cost of the inputs employed to produce it. Let ME$C^{\bar{k}}(x)$ represent the firm's marginal external or outside-of-the-market cost of the externality that is not included in $SRMC^{\bar{k}}(x)$. Then the internal and external parts of the marginal social cost concept become explicit in the equation

$$MSC^{\bar{k}}(x) = SRMC^{\bar{k}}(x) + MEC^{\bar{k}}(x). \tag{25.5}$$

If $MEB(x)$ denotes the marginal external benefits not included in the output price p_x, then splitting $MSB(x)$ into its internal and external parts yields

$$MSB(x) = p_x + MEB(x). \tag{25.6}$$

As previously indicated, to obtain Pareto optimality or efficiency that takes into account <u>all</u> costs and benefits requires the firm to produce an output for which equation (25.4) is satisfied. Substitution of equations (25.5) and (25.6) into equation (25.4) yields

$$SRMC^{\bar{k}}(x) + MEC^{\bar{k}}(x) = p_x + MEB(x)$$

or

$$p_x = SRMC^{\bar{k}}(x) + MEC^{\bar{k}}(x) - MEB(x). \tag{25.7}$$

And, to repeat the previous conclusion, if an externality is present and the firm producing good x does not take that externality into account, that leaves the firm producing where $SRMC^{\bar{k}}(x) = p_x$ which, in this situation, leads to inefficiency.

Figure 25-1 illustrates a few possibilities assuming for convenience that $MEC^{\bar{k}}(x)$ and

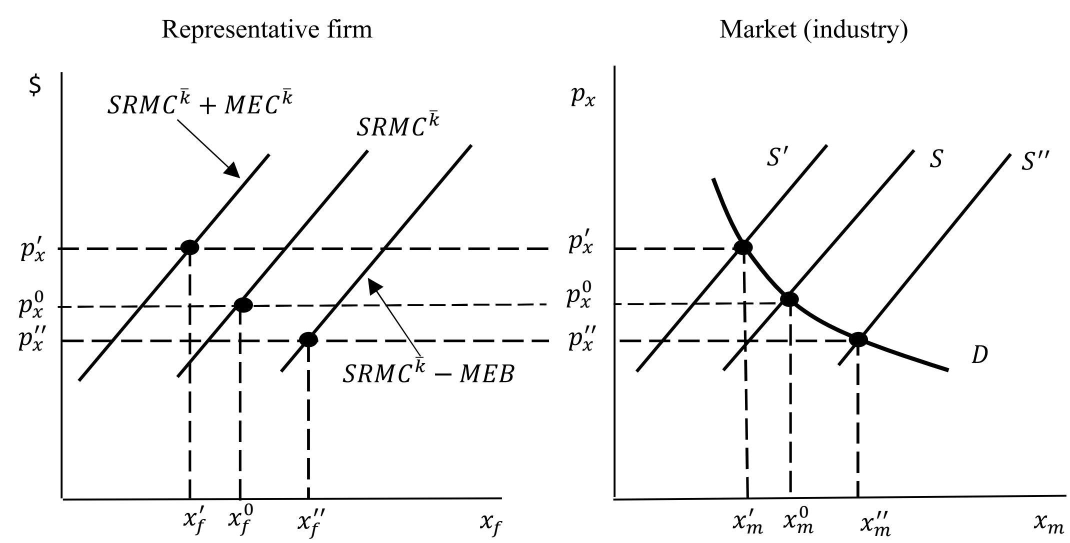


Figure 25-1

$MEB(x)$ remain constant as x varies and the marginal cost curve $SRMC^{\bar{k}}$ and hence the market supply curve S are upward sloping straight lines. The subscripts f and m on x indicate firm and market quantities respectively. Ignoring the presence of the externalities, market and profit-maximizing firm equilibrium occurs at (x_m^0, p_x^0) (right-hand diagram) where supply equals demand and at x_f^0 (left-hand diagram) where the representative firm produces at $SRMC^{\bar{k}}\left(x_f^0\right) = p_x^0$.

If the ignored externalities are only negative so that $MEB(x) = 0$, then using equation (25.7), the firm's marginal cost curve and the market supply curve shift up by the constant amount $MEC^{\bar{k}}(x)$ at each x (like the specific tax in Figure 14-5 of Chapter 14). The latter is designated by S'. The market equilibrium price rises to p_x', and the equilibrium quantities that equate supply to demand and lead to profit-maximization and efficiency are to the left of, x_m^0 and x_f^0 at, respectively, x_m' and x_f' where, in terms of the representative firm, $p_x' = SRMC^{\bar{k}}(x_f') + MEC^{\bar{k}}(x_f')$. Similarly, if the ignored externalities are only positive with $MEC^{\bar{k}}(x) = 0$, then again from (25.7), the representative firm marginal cost and market supply curves shift down, and at the revised equilibrium price p_x'', the quantities leading to profit maximization, market equilibrium, and efficiency lie to the right of x_m^0 and x_f^0 at x_m'' and x_f'' where $p_x'' = SRMC^{\bar{k}}(x_f'') - MEB(x_f'')$. Thus, ignoring a negative externality that is present, at x_m^0 and x_f^0 the market and firms are producing too much output for efficiency; neglecting a positive externality, they are producing too little.

As noted in Chapter 21, the market failure arising from an externality provides a justification for government intervention in the market. That intervention would move the market closer to what is required for Pareto optimality (where $MSB(x) = MSC^{\bar{k}}(x)$) and thereby make at least one person better off without making anyone else worse off. Several options available to the government are listed below:

1. Tax or subsidize[3] the production of output to change costs and revenues so that the curve representing $SRMC^{\bar{k}}(x)$ becomes more like that representing $SRMC^{\bar{k}}(x) + MSC^{\bar{k}}(x) - MEB(x)$. For example, assume $MEB(x) = 0$ and $MEC^{\bar{k}}(x) = t$ for all x where $t > 0$. Under these assumptions, requiring the firm to pay a specific tax of *t* dollars per unit of output sold (Chapter 14) forces the $SRMC^{\bar{k}} + MEC^{\bar{k}}$ curve in the left-hand diagram of Figure 25-1 to lie above the $SRMC^{\bar{k}}$ curve by the amount *t* at each x. That is, *t* becomes a substitute for $MEC^{\bar{k}}(x)$. Alternatively, with $MEC^{\bar{k}}(x) = 0$ and $MEB(x) = q > 0$ for all x, paying the firm a subsidy of q dollars per unit of output means that the $SRMC^{\bar{k}} - MEC^{\bar{k}}$ curve is everywhere below the firm's marginal cost curve by the quantity q, and q becomes a substitute for $MEB(x)$.

2. Impose legal constraints so that, for example, pollution can be stopped by government order or court injunction and those who pollute can be held liable for the damage they cause. The fact that General Electric had to clean out the PCBs it left in the Housatonic River of Massachusetts and C Connecticut after closing a manufacturing facility located in the former is a case in point.

[3] A subsidy is a negative tax.

3. Regulate directly by setting standards, for example, of what kind of and in what quantities pollutants can be discharged into the air and water. Government emission standards for automobiles is an illustration.

4. Give limited quantities of permits for, say. emitting pollutants to firms and allow the permits to be bought and sold. Those firms not needing all of the permits they receive because they are polluting less than the quantity of permits they have allow, can sell their excess permits to those who need more. In this way a market is created for pollutant permits. When maximizing profits, firms that have to buy permits now include some of the external costs of pollution as the costs of permits in $SRMC^{\bar{k}}(x)$. This raises the original $SRMC^{\bar{k}}(x)$ based only on input costs bringing it closer to $MSC^{\bar{k}}(x)$.

Turning to public goods, the original definition given in Chapter 21 can be refined and made more precise by separating it into two parts. A consumption good is

1. **nonrival** (in consumption) if one person's consumption or enjoyment of a unit of it does not detract from that of another's,[4] and

2. **nonexcludable** if no one can be prevented from consuming or enjoying the benefits of a unit of the good (without explicitly paying for it) once it is produced.

A public good is both nonrival and nonexcludable. The example of street lighting cited in Chapter 21 exhibits these two properties.

Since it is nonexcludable, consumers do not determine quantities demanded of a public good in response to prices set by the market. And without explicit demand, that is, without individual and hence market demand curves, a market price as determined by the interaction of the competitive forces of demand and supply cannot be generated and an outcome leading to Pareto optimality cannot be achieved. In this sense, all markets for public goods fail. Only governments can, often through a contract with a private firm, arrange for their production. (An exception is radio and TV broadcasting which is paid for by advertising.) But then the question arises of how much of a public good should be produced. Can the quantity that would yield Pareto optimality were the public good added to the Walrasian model representing the microeconomy be determined?

This question has no easy answer. Suppose consumers were asked how much they were willing to pay per unit of a public good at each quantity x. That is, each person is asked to reveal his/her willingness to pay curve as characterized in Chapter 8 (Public radio and TV seem to pursue this idea in asking individuals to contribute money to cover production costs of broadcasting.) As noted above, willingness to pay per unit at a quantity may be thought of as a measure of the marginal benefit per unit accruing to the individual when consuming that

[4] An apple, for example does not have this property. Once it is consumed by one person it cannot be consumed by another.

quantity. Summing these per-unit willingnesses to pay for each quantity over all persons consuming the public good gives an aggregate willingness-to-pay or marginal social benefit MSB curve for the good by society at large. (This involves a "vertical" sum of the willingness - to-pay curves as compared, when they are thought of demand curves, to their "horizontal" sum which yields the market demand curve.) A two-person example appears in Figure 25-2, where the willingness-to-pay curves for two persons, labeled D^1 and D^2, are vertically summed to

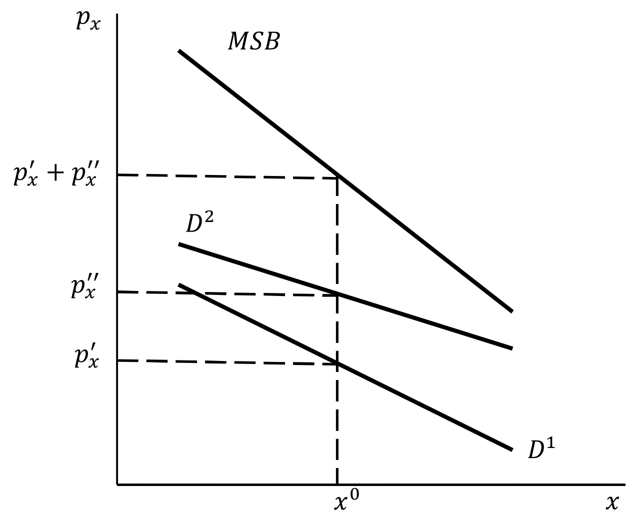


Figure 25-2

obtain the MSB^xcurve. Now, as pictured in Figure 25-3, the intersection of this aggregate MSB curve with the $SRMC^{\bar{k}}$ or short-run marginal cost curve of the firm that would be producing the public good (assuming no negative externalities) gives the quantity x^0 that would lead to Pareto

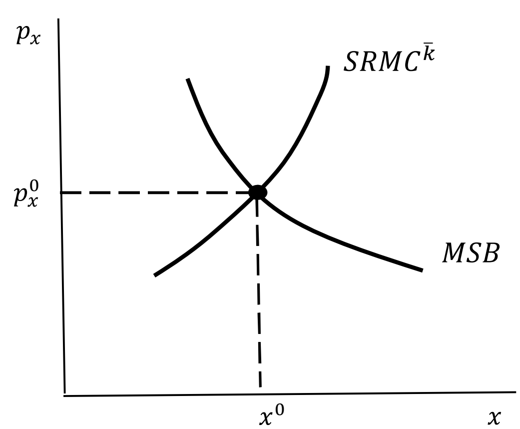


Figure 25-3

optimality since, at that quantity, the per-unit willingness of society to pay for the public good p_x^0, that is its marginal social benefit equals the marginal cost of producing it.

But to do this, it is necessary to know what consumers are willing to pay, and individual consumers have no incentive to reveal that information. Indeed, they have strong incentive to hide it for two reasons:

a. As previously suggested, because people can enjoy the benefits of a public good regardless of whether they pay for it or not, everyone is disinclined to pay for it. This is called the free rider problem.

b. The provision of a public good is usually so costly that its provision does not depend on whether one person pays or not. So why should an individual pay? This is known as the drop-in-the-bucket problem.

With the willingness-to-pay approach not working, it is appropriate to return to the original problem: How much of a public good should the government arrange to produce? Can the quantity that would lead to Pareto optimality or efficiency be determined? The answers to these questions depend on the answer to a related question. Can individual preferences be revealed and aggregated into preferences for society as a whole? This is often referred to as the problem of social choice. If societal preferences could be obtained, Pareto optimality could be re-expressed in terms of them.

One way of approaching the social choice problem is to have people vote. But this, too, has its difficulties. For example, suppose there are three alternative quantities of a public good, A, B, and C, from which to choose. Suppose society consists of three consumers, 1, 2, and 3. Let consumers express their preferences by voting for their preferred option in each option pair and let society's preference in each pair be determined by majority rule. Assume individual preferences are as follows:

Person 1 prefers A to B and B to C, and hence by transitivity, A to C.
Person 2 prefers B to C and C to A, and hence by transitivity, B to A.
Person 3 prefers C to A and A to B, and hence by transitivity, C to B.

Now for the pair (A,B), Persons 1 and 3 prefer A to B, while Person 2 prefers B to A. By majority rule society prefers A to B. A similar calculation applies to pairs (B,C), and (A,C). Thus voting determines that

Society prefers A to B and B to C, but C to A.

Because they are transitive, each person's preferences lead to the conclusion that one option is preferred by that person to the other two; Person 1 prefers A to B and C, Person 2 prefers B to A and C, and Person 3 prefers C to A and B. But society's preferences obtained through this voting procedure violate the transitivity property. Consequently, the vote does not reveal a preference for one option over the remaining two at the societal level. This is known as the voting paradox. It suggests that this does not seem to be a good way to answer the questions raised by the presence of a public good. Without having transitive societal preferences, the determination of the Pareto optimal or efficient amount of a public good would appear to be beyond reach.

In democratic societies, representatives are elected and bureaucrats are appointed to make and implement decisions for everyone. Such decisions presumably involve attempts to provide the quantities of public goods consistent with what might be acceptable to most people and that would come close to efficiency. But it is necessary to recognize that elected or appointed officials might, for one reason or another, make choices that do not appear to approximate very well either what people seem to want or Pareto optimality. When this happens market failure can be compounded by what in Chapter 21 has been called government failure to achieve or approach a situation in which it is not possible to make at least one person better off without making someone else worse off.

The fourth (and last) form of market failure, namely imperfect information, can arise in two different ways:

1. Adverse selection occurs when a buyer or seller enters into a transaction with another party who has more information than the buyer or seller about the product transacted. The difference in knowledge of the product is called asymmetric information and can lead to an absence of Pareto optimality in that the individuals are not receiving or giving up what they think they are when, respectively, buying or selling. The example of a used car dealer selling a ‘lemon’ to a customer who is unaware of the car’s defects illustrates the point. In that case, as described in Chapter 21, one of the equations, say that for person 1, needed to establish the first fundamental theorem of welfare economics (20.11) of Chapter 20), namely,

$$\frac{MU_x^1(x_1^0, y_1^0)}{MU_y^1(x_1^0, y_1^0)} = \frac{p_x}{p_y}$$

is invalid and the theorem cannot hold. One possibility of government intervention to eliminate this inefficiency is the exactment of so-called ‘lemon laws’ that enable purchasers buying a product with a defect of which they are unaware to have the defect repaired, or to receive a replacement or refund.

2. Moral hazard occurs when one party to a contract passes the cost of his/her/its behavior on to the other party. For example, after signing an automobile insurance contract, a person drives more recklessly than before because he/she does not have to pay the full cost of an accident. In this case, the insurer does not have the same information about the risk of covering the cost of an accident as the insured. This also results in an absence of Pareto optimality since, for the insurer, the true marginal cost of producing its output (automobile insurance), which now includes extra risk, no longer equals its output price. Government intervention here to reduce the inefficiency can take the form of enacting rules such as speed limits, the violation of which can result in a substantial fine.

Appendix:

Supplementary Notes

Supplemental Note A
The Relationship between Marginal Revenue and Elasticity

From the (4.2) equation-specification of arc-price elasticity ε_{AB} between points α and β on the demand curve shown in Chapter 4,

$$\varepsilon_{AB} = -\left(\frac{x' - x''}{p_x' - p_x''}\right)\left(\frac{p_x'}{x''}\right) = -\,p_x'\left[\frac{1}{\dfrac{p_x'x'' - p_x''x''}{x' - x''}}\right]. \quad \text{(A.1)}$$

Rewriting the denominator in the square brackets of (A.1),

$$\frac{p_x'x'' - p_x''x''}{x' - x''} = \frac{p_x'x'' - p_x'x' + p_x'x' - p_x''x''}{x' - x''} = \frac{p_x'x'' - p_x'x'}{x' - x''} + \frac{TR(x') - TR(x'')}{x' - x''}, \quad \text{(A.2)}$$

where $p_x'x'$ has been both added in and subtracted from the numerator of the fraction on the left, and $TR(x') = p_x'x'$ and $TR(x'') = p_x''x''$. The first term to the right of the second equality in (A.2) reduces to p_x', and the second term is the marginal revenue in moving between x' and x'' (equation (4.7) in Chapter 4). Using these facts, (A.2) becomes

$$\frac{p_x'x'' - p_x''x''}{x' - x''} = -\,p_x' + MR(x'). \quad \text{(A.3)}$$

Substituting the right-hand expression in (A.3) for the denominator between the square brackets in the fraction of (A.1) and solving for $MR(x')$, results in

$$MR(x') = p_x'\left(1 - \frac{1}{\varepsilon_{AB}}\right).$$

Supplemental Note B
The Geometry of the Marginal Revenue Curve with a Linear Demand Curve

As indicated in Chapter 4, the linear demand curve is described by

$$x = \alpha p_x + \beta, \tag{B.1}$$

where $\alpha < 0$ and $\beta > 0$ are fixed numbers. Solving for p_x,

$$p_x = \frac{x}{\alpha} - \frac{\beta}{\alpha}. \tag{B.2}$$

The graph of equation (B.2), the solid straight line labeled *D* in Figure B-1, is the same as that of

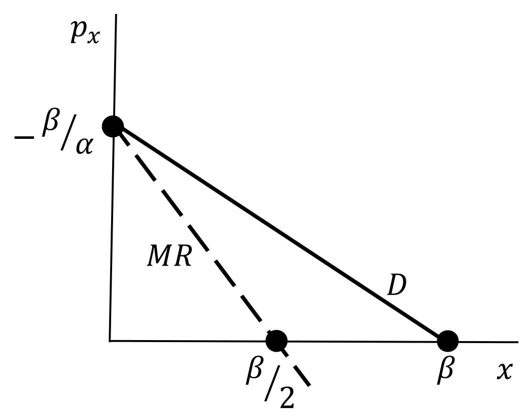


Figure B-1

(B.1) although, with respect to (B.2),. the axes are not reversed, and the dependent variable p_x remains on the vertical axis. To determine the marginal revenue line, it is first necessary to calculate total revenue as a function of x. Using (B.2),

$$TR(x) = xp_x = x\left[\frac{x}{\alpha} - \frac{\beta}{\alpha}\right] = \frac{x^2}{\alpha} - \frac{\beta x}{\alpha}, \tag{B.3}$$

and differentiation of (B.3) with respect to x yields the marginal revenue function

$$MR(x) = \frac{2x}{\alpha} - \frac{\beta}{\alpha}. \tag{B.4}$$

The graph of (B.4) is drawn as a dashed line in Figure B-1. Now for both (B.2) and (B.4), $p_x = MR(0) = -\beta/\alpha$ when $x = 0$, so that the demand and marginal revenue lines have the same p_x-intercept. On the other hand, when $p_x = 0$ along the demand curve, $x = \beta$. But when $MR(x) =$

0 along the marginal revenue line, $x = \beta/2$. Thus the x-intercept on the marginal revenue line is half the distance from the origin to the x-intercept on the demand line.

Supplemental Note C
Assumptions Made to Explain Consumer Behavior

1. The consumer has preferences and indifferences among baskets of commodities such that, for any two baskets either one is preferred to the other or the two are indifferent. (Completeness.)

2. Preferences and indifferences are transitive. (If basket *A* is preferred to basket *B*, and basket *B* is preferred to basket *C*, then basket *A* is preferred to basket *C*. Similarly for indifferences.)

3. Preferences and indifferences are represented by a utility function in the sense that (a) if one basket is preferred to another, then the preferred basket has a higher utility value than the other, and (b) if two baskets are indifferent, then they have the same utility value.

4. The utility function has the following four properties:

- 4a. It is continuous and all marginal utilities can be calculated.
- 4b. A larger basket of commodities is always preferred to, and therefore has a higher utility value than a smaller one.
- 4c. Indifference curves are strictly convex.
- 4d. Indifference curves do not touch the co-ordinate axes of the commodity space.

5. The consumer purchases or demands that basket from his/her budget set that is preferred over all other baskets in the budget set and therefore provides the most utility. In other words, the consumer demands that basket that maximizes his/her utility subject to the budget constraint.

Supplemental Note D

Assumptions Made to Explain Firm Behavior

1. The firm has a long-run production function $x = f(\ell,k)$ based on a given technology.

2. The long-run production function has the following properties:
   - 2a. Zero input produces zero output ($f(0,0) = 0$), and nonnegative input produces nonnegative output ($f(\ell,k) \geq 0$ for all $(\ell,k) \geq 0$).
   - 2b. It is continuous and all marginal products can be calculated.
   - 2c. If ridge lines exist, all marginal products are positive and all isoquants are strictly convex between the ridge lines up to an intersection point if there is one.
   - 2d. If ridge lines do not exist, all marginal products are positive and all isoquants ar strictly convex everywhere throughout the input space, and no isoquant touches the co-ordinate axes.

3. Long-run and short-run total cost curves appear as drawn in this book so that average and marginal cost curves can be determined and have the shapes attributed to them.

4. The firm hires (demands) inputs and produces and sells (supplies) output so as to maximize its profit.

(The short-run production function is obtained from the long-run production function by fixing k at some value k^0, that is, $x = f(\ell,k^0)$.)

Supplemental Note E
Summary of How the Model of the Perfectly Competitive Firm Arrives at Profit Maximization

1. If ridge lines exist in the input space, eliminate the regions outside of the area between them and beyond any intersection point if there is one. This reduces the input space to its relevant region. (If ridge lies do not exist, the relevant region is the entire input space excluding the co-ordinate axes.)

2. Using input price information and cost minimization (long run) or fixed capital information $\bar{k}$ (short run), calculate the appropriate expansion path in the relevant region and confine attention to it.

3. Using the production function, expansion path, and input price information, calculate all cost functions and curves expressing cost as a function of output.

4. Using output price information, calculate all revenue functions and curves.

5. Using cost and revenue information, calculate the profit function and the profit-maximizing output x^0.

6. From the intersection of the isoquant relating to the profit-maximizing output and the long- or short-run expansion path, calculate the profit-maximizing input quantities.

Supplemental Note F
Equivalence of the Short-Run, First-Order Profit Maximization Equations with Respect to Output and Labor Input

In the short run, let capital be fixed at $\bar{k}$ and let x^0 and ℓ^0 be the profit-maximizing quantities of output and labor input respectively.

Start with the first-order equation $p_\ell = VMP^\ell(\ell^0)$ for profit maximization with respect to labor input at ℓ^0, where p_ℓ is the price of labor. Then $x^0 = f(\ell^0, \bar{k}) = TP^\ell(\ell^0)$ is the output produced with labor input ℓ^0. Using the definition of $VMP^\ell(\ell)$ in Chapter 15 and that of $\mathrm{M}P^\ell(\ell)$ in Chapter 10, and the convention that when labor input ℓ^0 changes by $\Delta\ell$, output x^0 changes by Δx,

$$p_\ell = VMP^\ell(\ell^0) = \; p_x MP^\ell(\ell^0) = \; p_x \frac{TP^\ell(\ell^0 + \Delta\ell) - TP^\ell(\ell^0)}{\Delta\ell} = \; p_x \frac{\Delta x}{\Delta\ell},$$

where p_x is the price of output x and, by the above convention, $TP^\ell(\ell^0 + \Delta\ell) - TP^\ell(\ell^0) = x^0 + \Delta x - x^0 = \Delta x$. Therefore

$$p_l = p_x \frac{\Delta x}{\Delta\ell}. \tag{F.1}$$

Solving equation (F.1) for p_x and using the definition of total variable cost $TVC(x) = \; p_\ell \ell$ for appropriately chosen ℓ, the definition of $\mathrm{SRM}C^{\bar{k}}(x)$ from Chapter 12, and the same convention,[1]

$$p_x = \frac{p_l \Delta\ell}{\Delta x} = \frac{p_\ell(\ell^0 + \Delta\ell) - p_\ell \ell^0}{\Delta x} = \frac{TVC(x^0 + \Delta x) - TVC(x^0)}{\Delta x} = \mathrm{SRM}C^{\bar{k}}(x^0),$$

since $x^0$is the output produced using labor input ℓ^0. Thus

$$p_x = \mathrm{SRM}C^{\bar{k}}(x^0),$$

which is the first-order equation for profit maximization with respect to output at x^0.

Starting with the first-order equation for profit maximization with respect to output at x^0 and reversing these steps gives the first-order equation for profit maximization with respect to labor input at ℓ^0.

---

[1] Recall that, as described in Chapter 12, $\mathrm{SRM}C^{\bar{k}}(x)$ may be calculated using either $SRTC(x)$ or $TVC(x)$.

Supplemental Note G
Definitions of General Equilibrium and Pareto Optimality

General Equilibrium – short run:

The entire model of the micro-economy is at general equilibrium at

(i) quantities of final goods bought and factors sold by each consumer,
(ii) quantities of inputs bought and outputs produced and sold by each firm, and
(iii) market quantities and prices of each good, and

provided that

(a) each consumer is buying final goods and selling factors so as to maximize utility subject to his/her budget constraint,
(b) each firm is hiring inputs and producing and selling outputs so as to maximize its profit, and
(c) supply equals demand in every market.

General Equilibrium – long run:

Add

(d) all profits in all firms are zero.

Pareto Optimality:

A its distribution $(x_1^0, y_1^0, x_2^0, y_2^0)$ of (x^0, y^0) among consumers is Pareto optimal or efficient (in general) provided that:

(i) (x^0, y^0) lies on the transformation curve,
(ii) there is no other distribution of (x^0, y^0) and no distribution of any other pair of outputs on the transformation curve at which one person is better off (higher utility) without the other person being worse off (lower utility).

Supplemental Note H
Second-Order Condition Ensuring Maximum Monopoly Profit

In general, the second-order condition guaranteeing that $\bar{x}$ in equation (22.3) of Chapter 22 identifies maximum monopoly profit is the same as the second-order equation (13.8) from Chapter 13 for the perfectly competitive firm (adjusted for the short run):

$$\frac{d^2\pi(\bar{x})}{dx^2} = \frac{d^2TR(\bar{x})}{dx^2} - \frac{d^2SRTC^{\bar{k}}(\bar{x})}{dx^2} < 0. \qquad (H.1)$$

However, unlike the case of perfectly competition, it is not necessary that the marginal cost curve slope always upward around $\bar{x}$. For example, when the demand curve facing the monopolist is linear as in Supplemental Note B, the marginal revenue function is given by equation (B.4):

$$MR(x) = \frac{2x}{\alpha} - \frac{\beta}{\alpha}, \qquad (\text{H}.2)$$

where $\alpha < 0$ and $\beta > 0$. Using (H.2), the second-order derivatives to the right of the equals sign in (H.1) may be rewritten as

$$\frac{d^2TR(\bar{x})}{dx^2} = \frac{dMR(\bar{x})}{dx} = \frac{2}{\alpha} \quad \text{and} \quad \frac{d^2SRTC^{\bar{k}}(\bar{x})}{dx^2} = \frac{dSSRMC^{\bar{k}}(\bar{x})}{dx}.$$

And substituting these terms back into (H.1), the second-order condition for maximum profit becomes

$$\frac{d^2\pi(\bar{x})}{dx^2} = \frac{2}{\alpha} - \frac{dSRMC^{\bar{k}}(\bar{x})}{dx} < 0$$

or

$$\frac{2}{\alpha} < \frac{dSRMC^{\bar{k}}(\bar{x})}{dx}. \qquad (H.3)$$

Since $\alpha < 0$, inequality (H.3) is satisfied for negative values of $dSSRMC^{\bar{k}}(\bar{x})/dx$ greater (that is, closer to zero) than $2/\alpha$. In other words, it is possible for the marginal cost curve to slope downward around the profit-maximizing output $\bar{x}$.

Index