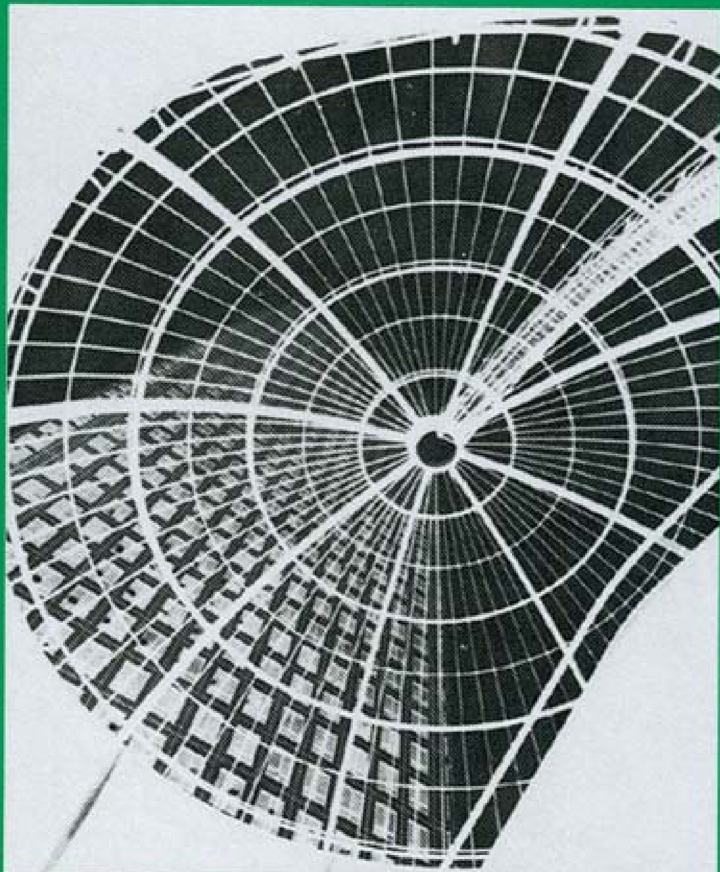
Microeconomics for Business



SAGE TEXTS

SATYA P. DAS

Microeconomics for Business

Microeconomics for Business

Satya P. Das

SAGE Publications
Los Angeles • London • New Delhi • Singapore

First published in 2007 by

Sage Publications India Pvt Ltd
B 1/I1, Mohan Cooperative Industrial Area
Mathura Road, New Delhi 110 044
www.sagepub.in

Sage Publications Inc
2455 Teller Road
Thousand Oaks, California 91320

Sage Publications Ltd
1 Oliver's Yard
55 City Road
London EC1Y1SP

Sage Publications Asia-Pacific Pte Ltd
33 Pekin Street
#02-01 Far East Square
Singapore 048763

Published by Vivek Mehra for Sage Publications India Pvt Ltd, typeset in 10/12 Palatino Roman by Quick Sort (India) Private Limited, Chennai and printed at Chaman Enterprises, New Delhi.

Library of Congress Cataloging-in-Publication Data

Das, Satya P.
Microeconomics for business/Satya P. Das.
p. cm.
Includes bibliographical reference and index.
1. Managerial economics. 2. Microeconomics. 3. Managerial economics--India. I. Title.

HD30.22.D37 33.8—dc22 2007 2007035314

ISBN: 978-0-7619-3592-6 (PB) 978-81-7829-753-8 (India-PB)

The Sage Team: Sugata Ghosh, Samprati Pani and Mathew P.J.M.

Dedicated to Sanghamitra, Arpita and Arup

Contents

List of Figures

List of Tables

Foreword

Among all major branches of social science, students treat economics with respect as well as fear. Economics commands respect since it facilitates careers in commerce, accounting, management and so on. Nevertheless, it is regarded as a tough subject on grounds of its abstract content, diagrams, use of mathematics, complexity of data and so on. To an average student, economics remains a dry and alien subject. We, the teachers of economics at various levels, are partly to be blamed for this alienation. The near-absence of standard textbooks, written exclusively for the students of this subcontinent, is the other reason for this sorry state of affairs. The present work by Dr Satya P. Das aims to fill this gap.

It is heartening to note that a person of Prof. Satya P. Das's stature and intellect has written a textbook for undergraduate students of management and related disciplines. As a classmate and close friend, I have known Satya for the past three decades. After teaching in various well-known universities of the United States for two decades, Dr Das settled down in India about a decade ago. Dr Das writes in a very simple, lucid style. His choice of examples (gleaned from day-to-day experiences) dispels the fear of alienation from the student. The detailed exercises, case-studies and assignments will make it much easier for both the students as well as the instructors not only to understand the concepts but also to apply it to management decision-making process. *Microeconomics for Business* would prove to be immensely useful for management graduates. I wish Dr Das writes another volume entitled *Macroeconomics for Business* as well very soon.

Dr J. K. Goyal
Director
Jagannath Institute of Management Sciences

Preface

This book is a product of teaching microeconomics to undergraduate students in various universities in the US. When I first started to teach this subject in the seventies, I immediately found a major difference between the method and style of teaching required in the US and what I was exposed to in India in my student days. In the US (and presumably in other developed countries), it was/is the job of the teacher to strongly motivate the subject material—by discussing many examples and applications—so that the students find the subject useful in a practical sense. This is what I have tried to do in this book. 'Business' is obviously a very applied area. Where can the principles of microeconomics be more fruitfully applied than business?

Although 'for business' appears in the title, it is not meant only for the students in business. It should also be suitable for commerce and economics students. On purpose, I have kept the formal mathematical tools of analysis to a minimum so that the basic ideas and concepts do not look burdensome or technically frightening.

I am grateful to my family and the Indian Statistical Institute, Delhi Centre for providing me the atmosphere and freedom to complete this work.

0 Setting Norms

When I was studying economics during my undergraduate days in the early seventies, we read books by both Indian and foreign authors. The former set of books had everything ‘point-wise’, without much attempt to build or illustrate concepts. Almost no topic would start without giving a bunch of definitions, the most glaring of which was a collection of definitions of economics itself, essentially quotations from great economists of the past. A chapter would typically end with a list of merits and demerits of a particular theory or approach. The whole material was largely geared towards memorisation without much of understanding.

On the other hand, books from foreign authors made real efforts to build concepts. They did not ‘defend’ their writing by overly citing what some great economist had said in the past. If one understood the grammar and sentences fully, one felt a degree of comfort and assurance that one has really understood something. However, I had problems with many foreign books too. A simple concept that could be conveyed effectively in two lines was sometimes explained in two paragraphs with all the caveats, exceptions, and so on. By the end of a huge sentence or a half-page-long monotonous paragraph, I would lose track of what the author really wanted to say. But, on the whole, I liked them insofar as the understanding of the subject matter was concerned.

Paul Samuelson’s *Principles* was, however, an extraordinary exception, a delight. The language was simple and direct. Any one having a fair command over English would perfectly understand the concepts that were being laid down. The main targeted audience was the students, not the teachers. Of course, there is always a vital role for a teacher: to explain things in a classroom even in a simpler manner with more illustrations and supplementary material. However, one natural limitation of a book such as that of even Samuelson was that the real-life examples were drawn from the US and other foreign countries, some of which I could not fully relate to.

Over 30 years have gone by. With modern word processing facilities, the fonts look nicer and the diagrams are clearer. There are more reader-friendly, foreign-authored books available in the market. These are good things. But most of the writing of college textbooks in economics by us, the authors in India, looks pretty much the same now as it was a generation ago. Mind you this is not my opinion only. Many college teachers I have talked to share the same view. This is rather unfortunate for our students.

In the modern world when things are changing fast we cannot hide behind the fear of breaking away from the past in a big way. There is little need to look up to the West for certification of what is to be there or not to be there in a book. This is not to say that the books 'from the West' are wrong or irrelevant. Make no mistake that most original ideas in economics still come from Western countries. Books originating from there are still generally very good.

But we, the Indian authors, can certainly do better and even aspire to do better than the best available. (This is how the best gets better over time.) In the process the East and West can jointly benefit. There is no more any *Lakshman-rekha* dictating us that we cannot cross a line in setting norms in economics teaching and writing, as long as they are reasonable and relevant. We, the teachers in India, can put our own stamp on textbook writing with confidence for the benefit of our students who are our future generation. We owe it to them.

This is the spirit in which this book is written. The current chapter is numbered '0' and I am the first to admit that this is highly unconventional. But is it unreasonable? No, because it just sets the stage for the nature of things to come. Was it not our great ancient scholar Arya Bhatt who discovered '0'? Should we not be proud to use it when appropriate?

Numbering a chapter uncommonly is merely symbolic. More substantially, this book is written almost entirely from my own teaching experience without following any particular established book in the market. Definitions and taxonomy have been de-emphasised consciously. Instead, a lot of emphasis has been put on simplicity of language, understanding of the subject matter and economic intuition. At the same time, rigour in terms of algebraic treatment is not spared, while unnecessary and silly algebra is not thrown in to create any image or impression. Furthermore, I have always believed that the best and most effective way to understand concepts is through examples. Hence, I have tried my best to illustrate concepts by bringing in hypothetical and real-life examples mostly from India and some from abroad. In other words, the book is 'India-centric.' This is one distinguishing feature of this book.

In Western countries, new concepts, themes and techniques are first introduced in textbooks by individual authors. If these are well received by instructors in various colleges and universities, then they make their entry into the syllabi of an institution, which are invariably minimal anyway, so as to offer considerable leeway to the instructors.

Once some new material is 'established' in the West, it is transmitted into syllabi in India and, typically, that too after a long lag. In writing this book, although I have followed some syllabi of business economics for a bachelor's degree in business and of economics for a bachelor's degree in commerce, I have not bound myself with them. I have taken the initiative of deleting some old material and concepts and adding some new ones. I wish to make it clear that in doing so, my purpose is not to 'defy' our syllabi but aid them to evolve faster. Economics is a fast-evolving science. New concepts, examples and applications must substitute some old ones. It is high time that we, the teachers, cease to almost blindly follow the tradition and, instead, consciously and selectively sense which material or concept can be abandoned and which ones have to be introduced.

More to-the-point and as another distinguishing feature, the book has two chapters, the material of which is largely ignored in any syllabus of elementary microeconomics for business and commerce that I have seen in India and elsewhere. One is Chapter 7, 'Demand for Assets', which parallels the earlier chapters on demand for goods. A student of business or commerce (even an economics student) must know some of the common financial assets offered in the market, their characteristics and the factors governing their demand. The other is Chapter 13 on game theory, which has made a huge impact on the science of economics over the last three decades. It is imperative that students at the undergraduate level know some of its basics and its applications. As it turns out, the basic concepts in game theory are not at all difficult for an undergraduate student to understand and appreciate.

These are not the only changes I have made. There are some other small changes here and there that my fellow teachers can easily see. But despite all this, most of the material should look familiar to an instructor.

Last but not least, I wish to bring out some real changes in the pattern of questions that are asked in college and university exams. Each chapter has a lot of questions at the end and the emphasis is on the applications of the concepts and definitions, not on the memorisation of these per se. There is a premium on precision demanded by the questions that really test the understanding of the material. Some of the questions are application-oriented, relating to real-life happenings. Brief answers to some of the questions are given chapter-wise at the end of the book.

To help the students, more in terms of learning, each chapter begins with the list of concepts developed in the chapter and ends with (before the questions) 'Economic Facts and Insights.' The latter is not a summary of results derived in the chapter but rather some of the main points and relevant facts, which the students may wish to carry into the future in their bag of knowledge, wisdom and understanding about the real world.

Finally a word on the style of the language is in order. I do not believe that a college text needs to be written in a 'serious-sounding' and dry language, as it is meant for adult students, not school children. I have consciously attempted to keep the language simple, direct and entertaining. Conversational English is my style.

I can only hope that the students and fellow teachers enjoy and benefit from this book, not just from the viewpoint of securing good marks (which is important), but also from the viewpoint of deeply understanding the material so as to apply it to the social and economic problems we all face.

1 Introduction

CONCEPTS

- Scarcity
- Choice
- Economics
- Opportunity Cost
- Production Possibility Curve
- Increasing Marginal Opportunity Cost
- Three Fundamental Problems of Economics: ‘What’, ‘How’ and ‘for Whom’
- Market Economy versus Command Economy
- Economic Agents
- Positive versus Normative Economics
- Microeconomics
- Macroeconomics
- Business/Managerial Economics

WHAT IS ECONOMICS ALL ABOUT?

Economics is usually taught in an 'arts' programme in colleges. Thus many think that economics is not a science. But it *is*, although not exactly like physics and chemistry, which are physical sciences. In a physical science, labs, equipment and such are used heavily. In economics, these are not used as much, but the goals and the broad methodologies are similar. A physical science attempts to explain various physical phenomena in and around us, or in the solar system and the universe. Economics is a study of economic phenomena around individuals, regions, countries or in the world as a whole. In physics, for instance, we inquire into the effects of throwing a stone with a given force and at a given angle along its trajectory over the space. In economics, we inquire into for instance, the effects of a policy change on the growth trajectory of an economy over time. In physical sciences, there are various laws. There are a few laws in economics too. The only distinction is that economics is a *social science*, dealing with the behaviour of individuals and organisations within a society.[1]

Many textbooks have striven to lay down a definition—sometimes a series of definitions—of economics by quoting celebrated economists of the past. In my view, this is rather silly and actually counter-productive because it encourages nothing other than memorisation. It is more important to have a clear sense of what economics is about, via examples. Many non-economists think that economics is all about how to make or manage money. This is not true; economics is much broader in scope. It is about making choices in the face of scarcity. If things were not scarce, there would not be a subject matter like economics, because then an individual or a nation could have anything that it wanted. Unfortunately, that is not the case. If nothing else, time is scarce. Even the richest person in the world today, Bill Gates, has to make a choice on a particular morning whether to do yoga for one hour or to have a video conference with a Microsoft's top executive who is travelling in India.[2] Look at yourselves: how hard-pressed for time you are when the final exams draw near. No sane individual needs any convincing that time is scarce. Not just time, but almost anything else, even water, is scarce in the sense that it is not always free. The degree of scarcity varies, however, from country to country. For instance, in a typically developed country, labour is relatively more scarce than technology and equipment, while in a developing country it is the opposite.

Because things are scarce, one has to make a choice; thus, the problem of choice arises due to scarcity. The study of such problems of choice at the individual, social, regional, national and international level is what economics is all about. See Clip 1.1 for a perspective on the goal of economics. In the chapters to come, you may not see the word 'scarce' or 'scarcity' used very much, but they always remain in the background.

[1] In general, science refers to a body of knowledge that leads to 'empirically verifiable hypotheses', that is, generating predictions (effects of some action on something else) that can be tested. If a prediction cannot be tested, it is not scientific. In economics, one derives predictions or hypotheses that can be tested by using data and statistical/econometric methods.

[2] This is only a hypothetical example.

CONSTRAINTS AND OPPORTUNITY COSTS

Scarcity implies that, at any given point of time, an individual, organisation or a country cannot have everything in any quantity that is desired. Some constraint must be present. If you want more of one thing, you have to forego some amount of some other thing. This is what 'cost' means in economics. A uni-ball pen costing Rs 10 means that if you want to have a uni-ball pen, you have to forego Rs 10 worth of other goods that you could have had otherwise.

More exactly, the term 'cost' in economics refers to what is called **opportunity cost**. The opportunity cost of a given activity is defined by the value of the next best alternative. To understand this, consider the following examples.

Clip 1.1: Reducing Wants or Reducing Scarcity to Solve an Economic Problem?

Human wants are unlimited. Therefore, scarcity poses a problem. Some elaborate definitions of economics explicitly mention the unlimitedness of wants. While there is nothing wrong with it technically, unfortunately, policy and decision makers sometimes seize upon it (knowingly or unknowingly) and emphasise 'reducing wants' as an acceptable method of solving economic problems. The logic is straightforward. If some things are scarce, advise people to reduce consumption, which will then match the limited amount available. But, this is a negatively oriented philosophy. For instance, think of electricity or drinking water. In most parts of India, their availability is limited, especially during summer months. Fundamentally, for the government or policy makers, there are two options to 'solve' this problem: increase supply by sufficiently investing in power and water projects or decrease demand by propagating a slogan that we should reduce consumption 'in the national interest'. Which one is the right approach? If the second option is emphasised year after year, there will be little incentive to focus on the first option of increasing supply. Over time, the public 'learns' to feel guilty about normally consuming power and water, and that tends to further reduce the intent of the government or the policy makers to invest sufficiently in power and water. The practical danger is that the (scarcity reducing) supply side continues to be neglected under one pretext or another. The shortages continue as if destined. Is this what we really want? Is this any indicator of our endeavour towards growth and development? The obvious answer is 'no'. Reducing wants to solve an economic problem may be justified only as a short run, emergency measure in an unanticipated situation like war, famine, drought and so on. But it has no rationale over a longer time period. The correct perspective of economics is to recognise—and salute—the unlimitedness of wants and yet try not to suppress or reduce them, unless they lead to other problems like seriously impairing the environment. The aim should be to fulfill the wants as best as possible—by investing in appropriate technology and by making sensible choices among alternatives. This is the spirit of economics.

NUMERICAL EXAMPLE 1.1

Mr Rajnish is a software engineer working for Wipro in Bangalore, earning Rs 10 lakh per year. There are, say, three alternative careers available for Mr Rajnish. He can work for Microsoft or IBM in Bangalore for Rs 9 lakh and Rs 9.5 lakh per year respectively. Still another alternative is that he can set up his own software firm, expecting to make a profit of Rs 8.75 lakh a year for himself. What is Mr Rajnish's opportunity cost of working in Wipro?

The next best alternative is to work for IBM, the value of which option is Rs 9.5 lakh. Hence this amount is the opportunity cost of working for Wipro.

NUMERICAL EXAMPLE 1.2

An auto-parts company of India wants to establish a plant outside India. The alternatives are Germany, Indonesia, Japan and the US. Given its financial situation, the company is constrained to set up only one plant outside India. Assume that the setting-up costs and the operating costs of a plant in any of these four countries are the same. The marketing research department offers a projection that if the plant is set up in Germany, Indonesia, Japan or the US, it will fetch a turnover (revenue) of $2.5 million, $2 million, $2.3 million and $2.8 million respectively. Of course, given these choices the company will opt to set up a plant in the US. What is then its opportunity cost?

Among the remaining alternatives, the maximum expected revenue is $2.5 million (if the plant is set up in Germany). Hence this amount is the opportunity cost.

Consider still another example, which is a little different in nature. Suppose that Mr Debraj Pattanaik owns a firm that assembles computers and attends service calls from customers who already have computers (bought from Mr Pattanaik and elsewhere). He has 10 employees. Each of them can assemble a computer or attend a service call. There is no difference in efficiency among the employees. Over a week's period, suppose that one employee can assemble 8 computers or attend 24 calls. How does Mr Pattanaik decide how many employees he should use for assembling and how many for attending calls?

A natural way is to first determine the various combinations of assembled computers and service call attendance that are feasible. Next, he can select a particular combination, depending on profitability of selling computers and attending service calls. Let us not be interested in the latter issue, but only in the former.

For example, if Mr Pattanaik uses all his employees to assemble computers, 80 computers are made but there is zero attendance to service calls. Instead, if he assigns 3 of them to the assembling job, 24 computers can be assembled and 168 calls can be attended. Of course, if all employees are asked to attend calls, there is zero production of computers and 240 calls are attended. Notice that, as Mr Pattanaik keeps increasing the production of assembled computers, the number of service calls attended falls.

In the example, there is an opportunity cost of computers assembled in terms of service calls completed and vice versa. Indeed, we can calculate the opportunity costs. Since one employee can either assemble 8 computers or attend 24 calls, we can say that 8 computers 'cost' 24 calls and thus the opportunity cost of one computer is equal to 3 service calls and, equivalently, the opportunity cost of one service call equals 1/3 computers.

PRODUCTION POSSIBILITY CURVE

Mr Pattanaik's problem illustrates the constraints facing a single individual. Interestingly, the entire economy of any country or region faces a similar constraint. This is illustrated by what is called a production possibility curve.

Realise that, at any point of time, the technologies available to produce various goods and services as well as the resources available to an economy—meaning the size of its working population, land, buildings, machinery and so on—are all given. Thus an economy cannot produce an arbitrary amount of any particular good or service. If all resources are used in producing a single good, say bicycles, only a given number of bicycles can be produced. Starting from a given allocation of resources to different sectors of an economy, if more resources are employed in one particular sector, the output of that sector increases and less resources are available for other sectors, reducing the output in the latter. In deciding which combination of goods may serve the economy the best, we first have to look at various combinations that can be made available to an economy.

Consider a hypothetical economy, in which two goods can be produced: motor bikes and *lehenga*s. All motor bikes are of the same quality and so are *lehenga*s. Suppose all resources of this economy (like land, and skilled and unskilled labour) are used in the *lehenga* sector and if they work efficiently, 150 lakh (1.5 crore) *lehenga*s can be produced (within, say, a year). Assume that the same resources can produce motor bikes also. If, instead, all resources are employed in making motor bikes, suppose that 10,000 motor bikes can be made. These are two production possibilities that are rather extreme. Most likely there will be other possibilities, which are in-between. For instance, it is possible that if the economy is producing 1 crore *lehenga*s, it can produce, say, 6,000 motor bikes.

Table 1.1 summarises the various production possibilities that are available to the economy. Not surprisingly, as the production of one good increases, that of the other declines. This is due to the scarcity of resources. If more resources go into

Table 1.1 Production Possibilities

	Motor Bikes (in thousands)	*Lehenga*s *(in crores)*
Possibility A	0	1.5
Possibility B	2	1.4
Possibility C	4	1.24
Possibility D	6	1.00
Possibility E	8	0.6
Possibility F	10	0

one sector and hence that sector's output increases, less must be available for other sectors and they will produce less than before.

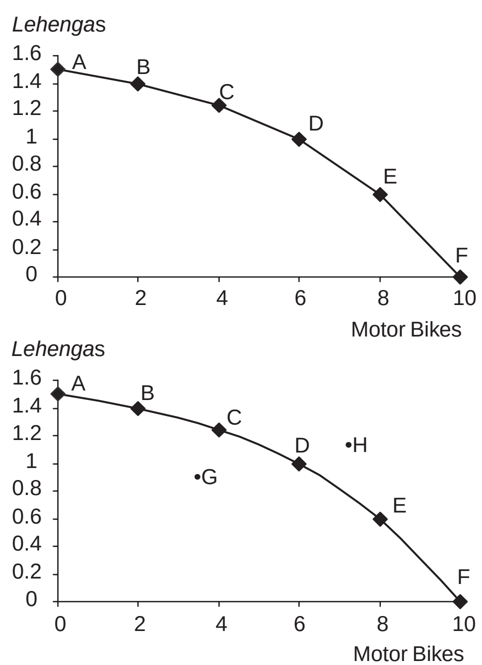
If we now plot these possibilities, namely, (0, 1.5), (2, 1.4) and so on and join the line segments, we get a curve as shown in the top panel of Figure 1.1. It measures motor bikes along the x-(horizontal) axis and *lehenga*s on the y-(vertical) axis. This is the **production possibility curve** of our hypothetical economy.

More realistically, if there are numerous production possibilities, not just six, then we get a smoother curve such as that in the bottom panel of Figure 1.1. This is how a production possibility curve (*PPC*) is normally exhibited.

Formally, the *PPC* is defined for a two-good economy, and it shows various combinations of the two goods that can be produced with available technologies and with given resources, which are fully and efficiently employed. In other words, a *PPC* shows the maximum amount that can be produced of one good, when the amount produced of the other good is given. It is downward sloping, since more production of one good is associated with less of the other.

Mark that the *PPC* does not indicate the actual point chosen in the economy. It only shows the possibilities. The economy may not be even operating on the curve if, for example, there is unemployment of labour (as is true for a country like India) or resources work inefficiently (for instance, when workers go on strike frequently). In that case, the economy will operate *strictly within* the *PPC*, for example, at a point

Figure 1.1 The Production Possibility Curve



like G. However, by definition, an economy cannot operate anywhere outside of the *PPC* (such as at point H). Moreover, assuming that the economy is operating on the curve, we cannot, without further information, indicate the exact point of operation. It depends on the preferences of individuals in the economy.

Realise that, although a *PPC* is defined in the context of a two-good economy, the idea behind it is general and holds for any number of goods, that is, if all resources are used efficiently in an economy, at any given point of time, more production of one good must imply less of some other good or goods.

Increasing Marginal Opportunity Cost and the Shape of the *PPC*

Along a *PPC*, as more production of one good implies some sacrifice of the other good, the rate of this sacrifice is the **marginal opportunity cost** of the expanding good. In Table 1.1, starting from possibility B, if bike production rises by two units (from 2 to 4), 140 – 124 = 16 lakh *lehenga*s are to be foregone. Hence, between the production possibilities B and C, the marginal opportunity cost of motor bikes (in thousands) is 16/2 = 8 lakh = 0.08 *lehenga*s (in crores). Similarly, between the production possibilities C and D, the marginal opportunity cost (per thousand bikes) is 0.12 crore *lehenga*s and so on. Thus, the marginal opportunity cost of a good on a *PPC* is the amount sacrificed of the other good per unit increase in the production of the good in question.

Note that 'marginal' means 'additional' and it is a very important notion in economics. You will see repeated use of it in later chapters.

Compared to Table 1.1, Table 1.2 has an additional column, listing the marginal opportunity cost of motor bikes. Observe that, as the output of motor bikes increases, its marginal opportunity cost increases (from 0.05 to 0.08, 0.08 to 0.12 and so on). Why does the marginal opportunity cost increase? The economic reason is that as more and more of a good is produced, the factors producing it become *marginally* less and less productive. Thus more and more of the other good has to be sacrificed to ensure a given increment of the former good.

Increasing marginal opportunity cost implies that the graph of *PPC* will look concave as in Figure 1.1. If, instead, the marginal opportunity cost were decreasing, you can check by constructing an example, the *PPC* will be convex. Finally, if

Table 1.2 Marginal Opportunity Cost along the *PPC*

	Motor Bikes (in '000)	*Lehengas (in crores)*	*Marginal Opportunity Cost of Bikes (in lehengas)*
Possibility A	0	1.5	–
Possibility B	2	1.4	0.05 (between A and B)
Possibility C	4	1.24	0.08 (between B and C)
Possibility D	6	1.00	0.12 (between C and D)
Possibility E	8	0.60	0.20 (between D and E)
Possibility F	10	0	0.30 (between E and F)

the marginal opportunity cost were constant, the *PPC* will be a straight line. However, typically the marginal opportunity cost of a particular good on the *PPC* increases and, therefore, the *PPC* is concave.

NUMERICAL EXAMPLE 1.3

A country named Spice produces two goods: chili and sugar. The production possibilities of these two goods are given in Table 1.3. Calculate the marginal opportunity cost of sugar in terms of chili and that of chili in terms of sugar at various points. Is the assumption of increasing marginal cost satisfied? Draw the *PPC* by plotting the production possibilities and joining them.

Between the sugar-chili production combinations (0,102) to (1,100), the marginal opportunity cost of sugar is equal to 102 – 100 = 2 units of chili, since 2 units of chili are sacrificed to obtain one extra unit of sugar. The marginal opportunity cost of chili is the inverse of that of sugar and thus equal to ½: between these combinations, 102 – 100 = 2 extra units of chili are obtained by sacrificing one unit of sugar and hence one extra unit of chili can be produced by giving up ½ unit of sugar. Likewise, the marginal opportunity costs at other combinations are calculated and listed in Table 1.4.

As we go down this table, sugar production increases and we see that the marginal opportunity cost of sugar also increases. Similarly, as we go up the table, chili production becomes higher and higher and its marginal opportunity cost

Table 1.3 Production Possibility Combinations (Numerical Example 1.3)

Production of Sugar	*Production of Chili*
0	102
1	100
2	95
3	85
4	73
5	55
6	30
7	0

Table 1.4 Marginal Opportunity Cost (Numerical Example 1.3)

Between Production Combinations	*Marginal Opportunity Cost of Sugar*	*Marginal Opportunity Cost of Chili*
(1,100) and (2,95)	100 – 95 = 5	1/5
(2,95) and (3,85)	95 – 85 = 10	1/10
(3,85) and (4,73)	85 – 73 = 12	1/12
(4,73) and (5,55)	73 – 55 = 18	1/18
(5,55) and (6,30)	55 – 30 = 25	1/25
(6,30) and (7, 0)	30 – 0 = 30	1/30

increases (from 1/30 to 1/25, from 1/25 to 1/18 and so on). Hence the assumption of increasing marginal cost is met for both goods.

Figure 1.2 depicts the *PPC* of Spice, and, notice that it is concave.

Shift of the *PPC*

Although along any given *PPC*, an increase in the output of one good is associated with a decrease in the output of the other, it does not mean that an economy can never produce more of both (or generally all) goods. The *PPC* defines the possibilities at a given point of time. But, over time, if technological improvements take place or more resources are available, the economy will be capable of producing more of both (or all) goods.

For example, in Figure 1.3, the *PPC* marked *AC* may represent an economy in 2006, while the outer *PPC* marked *FH* may represent it in 2010. That is, the *PPC* of

Figure 1.2 The *PPC* (Numerical Example 1.3)

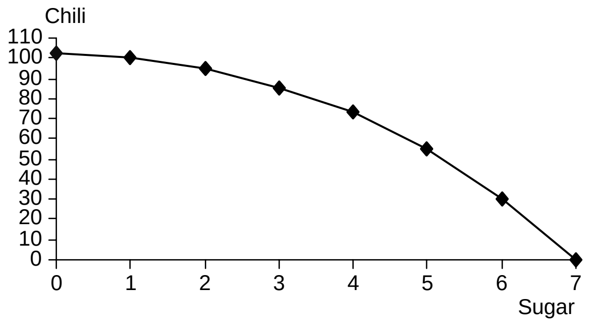
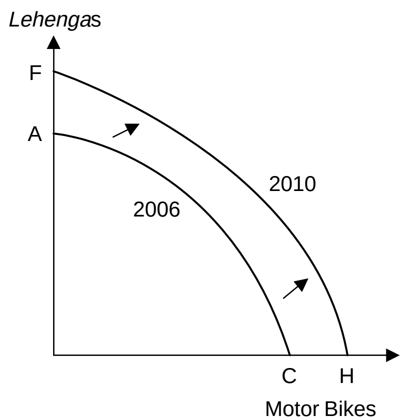


Figure 1.3 Shift of the *PPC*



a country can shift to the right from one period to another. This is indeed what we understand by economic growth.

NUMERICAL EXAMPLE 1.4

A country produces two goods: food (F) and manufactures (M). The good F is produced by labour and land, while the good M is produced by labour and capital. Suppose there is an increase in the availability of land for producing food. How would it shift the *PPC*? Illustrate it with a numerical example and plot the *PPC*s.

Suppose all labour is used in producing good M. As this sector does not use land, its production will not increase as land supply increases. This implies that the *PPC* will not shift to the right on the axis measuring good M. However, as long as there is some labour being used in producing good F, for any given level of production of good M, there will be an increase in the production of good F, since the same amount of labour, together with more land, will produce more of good F. The *PPC* will thus shift on the axis measuring good F. The overall nature of the shift of the *PPC* will be different from what is shown in Figure 1.3. It will look as Figure 1.4(a). A numerical example is given in Table 1.5. The original production possibilities are given along the first two columns, while the new production possibilities, after an increase in the availability of land, are given along the first column and the third column. The various points are plotted in Figure 1.4(b).

NATURE OF AN ECONOMIC PROBLEM

The production possibility curve is a good illustration of the problem of scarcity and choice. At any given point of time, an economy can only produce so much of goods. Therefore, choices are limited to points along or within the *PPC*. It also illustrates that

Figure 1.4 The *PPC*s before and after the Shifts (Numerical Example 1.4)

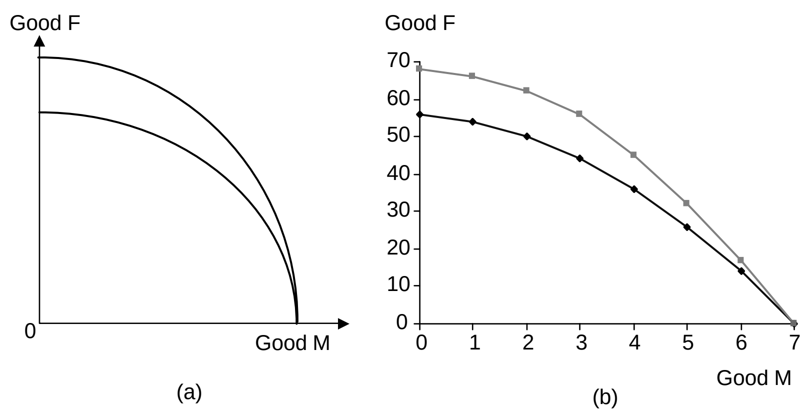


Table 1.5 Table for Numerical Example 1.4

Production of Good M	*Production of Good F (originally)*	*Production of Good F (after an increase in the availability of land)*
0	56	68
1	54	66
2	50	62
3	44	56
4	36	45
5	26	32
6	14	17
7	0	0

scarcity implies a positive cost of goods. If something is scarce, having more of it must involve sacrificing some amount of something else. This defines the opportunity cost.

Indeed we have illustrated the concept of economic costs in different ways. Any economic problem always involves costs. But it also has benefits of some kind. The combination of the two (costs and benefits) defines an economic problem.

The costs or benefits may or may not always be measurable in terms of money. As an example, suppose your friends are inciting you to ask your father for a ride in his imported, chauffeured car. You know that your father is a busy man and if you ask him he will surely get angry and admonish you, but eventually he will agree. Are you going to ask your father? You can think of it as an economic problem. There is a 'benefit' from asking your father, which is the enjoyment with your friends from the car ride. But there is a 'cost' in terms of the disutility or the pain of absorbing your father's harsh words. If you think that the benefits are greater than the cost, you go ahead and ask; otherwise, you do not.

Of course, you will see in the following chapters that the problem of identifying the benefits and costs and weighing them is not as simplistic as in the foregoing example. But they are not inordinately hard either.

CENTRAL PROBLEMS OF AN ECONOMY: 'WHAT', 'HOW' AND 'FOR WHOM'

Remember that economics deals with choice problems arising out of scarcity. There are indeed many choice problems that any particular economy solves within a given time period. During the fiscal year 2002–03, 23.2 million tons of potato was produced in India.[3] The output of any crop is not entirely driven by external factors like rainfall. It is partly influenced by how much of land is used to raise a particular crop and partly by the applications of fertilisers, power and so on. And these are consequences of individual choice as well as government policies. Hence, external factors held constant, India's production of potatoes in a given year is an outcome of choice.

[3]The source is *Economic Survey 2004–2005*, Ministry of Finance.

India, like many other countries, does not produce jet planes. But it produces helicopters, small aircrafts and some fighter planes. This is also a choice problem, given the constraints.[4]

Besides 'which goods are produced', being a problem of choice, 'how or in which method a good is to be produced' is also a choice problem. Usually, there is more than one method to produce a given commodity. For instance, electricity can be generated by hydro, coal or nuclear power. Agricultural activity can be undertaken by a relatively labour-intensive method (as in India on the average) or by a relatively capital-intensive method (as in a developed country like Germany).

Who earns how much is also a choice problem from an economy's viewpoint. There are differences in earnings across occupations within a country and across countries within the same occupation. The University Grants Commission (UGC) basic pay scale for assistant professors in India, for example, ranges from Rs 8,000 to Rs 13,500. If the dearness pay and allowance plus housing allowance are added, the gross monthly salary is nearly doubled, ranging from Rs 16,000 to Rs 27,000 per month, or Rs 1.82 lakh to Rs 3.24 lakh annually. As compared to this, in fields like business, economics or statistics, an assistant professor in an American university earns no less than $45,000 a year, which, at an exchange rate of $1 = Rs 43, is equivalent to Rs 19.35 lakh. Who earns how much is a choice problem in the sense that if on the average, computer engineers are earning more than historians, it is the economy or the market, which chooses to reward computer engineers more than historians.

The various economic problems facing an economy fall into three categories: 'what,' 'how' and 'for whom.'

What

What goods and services would be produced in an economy and in what quantities? For example, in the fiscal year 2002–03, the Indian economy produced 33 million tons of crude petroleum. Why is it 33 million tons and not, say, 40 million tons? In the same financial year, India produced about 11.5 million bicycles.[5] What factors determine these quantities? And so on.

How

By which methods would the goods and services be produced? Should garments in India be produced by relatively labour-intensive or machine-intensive methods? Cement can be produced by what is called a 'dry' process or a 'wet' process. Which technology is to be used in a new cement plant to be established in a given region? These are the questions that come under the 'how' problem.

[4]You may argue that India does not produce jet planes because it does not have the necessary technology. However, if a technology is not available domestically, it can certainly be acquired by paying for it. So acquiring a particular technology is also a choice problem. Many technologies can be purchased if we decide to pay for it. But we should not buy any available technology even if we can afford it. The benefits from having a technology must be weighed against the cost of acquiring it.

[5]The source is *Economic Survey 2004–2005*, Ministry of Finance.

For Whom

From the various goods and services that are available to an economy, who gets how much to consume? It depends on who earns how much or who has how much assets. For example, how much does an economics post-graduate earn as compared to a political science post-graduate or a chemical engineer? The 'for whom' question refers essentially to the issue of income and wealth distribution in an economy.

TWO BASIC ALTERNATIVE ECONOMIC SYSTEMS TO SOLVE THE CENTRAL PROBLEMS

In a market-oriented or capitalist economy, the three fundamental problems are solved by the 'market.' There is a price of any good or service, which is influenced by the forces of demand and supply. These forces guide which goods are to be produced and consumed and in what quantity. For example, *dosa* (the South Indian dish) is produced in the Indian economy because the technology of making *dosa* is available, the cost of producing and supplying it is not too high and there is a demand for it. This illustrates how the 'what' problem is solved.

Alternatively, in a centrally planned economic system, which was in practice in the former Soviet Union and other East European countries till the late eighties, these problems are addressed in a very direct manner by the government. There is a central planning authority that decides which goods and how much should be consumed and produced in the economy within a given span of time, say a year or five years. These are like targets. They are set according to the overall growth and development strategy for the economy that is considered 'desirable' by the members of the planning authority.

Factories are government-owned and the production methods are chosen by the planning authority. Salaries are also decided by the government.

In brief, all three central problems are essentially solved by direct command from the government. That is why a centrally planned economy is also called a **command economy**.

Which system is better in solving the fundamental problems of an economy? This is where sometimes emotions run high. But, the record of the world economy till today goes in favour of the market economy for the simple and stunning fact that the Soviet system has collapsed. Even the Chinese economy that used to be centrally planned is moving vigorously towards a market system.

ECONOMIC AGENTS IN A MARKET ECONOMY

In the chapters to come, we will study the **market economy**. Who are the 'players' or the **economic agents** in such an economy?

Essentially, there are three sets of players: consumers, producers and the government. Consumers demand goods and services, and supply factors of production like labour, land, machines and so on. On the other hand, producers supply goods and services and demand factors of production. Even in an economy in which there is no interference by the government in business in any manner, the government can play a critical role in the form of protecting property rights. This means that, by enforcing laws, the legal rights of consumers and producers are protected so that they are able to efficiently transact in the market. In reality, of course, the government of any country does more than just protect property rights; it affects production and consumption decisions by imposing taxes, granting subsidies and so on (some of these will be studied in this book).

POSITIVE VERSUS NORMATIVE ECONOMICS

Apart from the three broad questions, what kinds of more specific questions/issues are addressed in economics? There are several. For instance, what are the various factors that determine the demand for a particular good in the market? What are the various factors that determine the supply of a particular good in market? How would an increase in the price of coffee affect the demand for tea? What should be the government's policy in order to increase employment in the economy? Why does the announcement of a new central budget every year have some sort of an impact on the stock market? What should be India's best policy with respect to import of agricultural goods? There are of course many more questions of potential interest and relevance.

However, if you look at the sample of questions outlined above, you will see that all the questions can be categorised into two types. One type is concerned with the effect of something or some policy change on some variable. The other type is concerned with how a particular policy should be shaped in the best interests of the economy in a given context. The former is called **positive economics** or positive economic issues, whereas the latter is called **normative economics** or normative economic issues.

The answers to the positive economic questions form the foundation for the answers to the normative economic questions. For instance, if we want to know what should be the government's best policy towards, say, the software industry (which is a normative question), we have to first know how a certain policy will affect the consumers and the producers of the software and other related industries (which is a positive question).

MICROECONOMICS VERSUS MACROECONOMICS

Economics has many branches or sub-disciplines. Out of these, there are two core ones called **microeconomics** and **macroeconomics.** The former refers mostly, but not exclusively, to the analysis of scarcity and choice problems facing a single economic unit such as a producer or a consumer. Suppose you own a barber shop.

How many barbers should you hire? How many persons should you serve per day on an average? What price are you going to charge for a crew-style haircut? As another example, given your monthly pocket money, how many ice-creams and chocolates are you going to buy? These are questions of individual choice. Microeconomics deals with the principles behind such choices.

On the other hand, macroeconomics deals with the behaviour of aggregates such as real Gross Domestic Product (GDP), employment, interest rate and so on. What determines the real GDP or inflation rate in an economy? What policies can reduce the rate of unemployment in a developing country like India? And so on.

This book is meant to cover some basic principles of microeconomics from the viewpoint of students studying business/managerial economics and commerce.

MANAGERIAL/BUSINESS ECONOMICS

This leads to what managerial/business economics is all about. Here too, various people have given various definitions, which are exactly the same in spirit although different in terms of the words and phrases used. Going through these is to belabour the obvious without serving any useful purpose. Simply put, managerial/business economics refers to the study of economic principles that are relevant or useful in business and managerial decision-making.

For instance, any business survives because it is able to sell a product or service to consumers. Thus, it is pertinent to know how consumers make their choices regarding how much of different goods and services they should buy. Any business involves producing something, which may involve employing workers. How many workers should a firm hire, given the market conditions and technology? Is it profitable for a firm to buy a particular piece of equipment? What are the different kinds of costs that a firm has to incur over a particular horizon of time? These are the kind of issues studied in business or managerial economics.

The relationship between economics and managerial/business economics is somewhat similar to that between physics and mechanical engineering, chemistry and chemical engineering or statistics and econometrics. For example, econometrics is about statistical measurements in economics. It can be interpreted as statistics applied to economics. Thus, in a broad sense, econometrics is all statistics. Yet it is different because of the difference in its focus. Likewise, managerial economics can be thought of as economics applied to managerial or business decision-making.

Although a good manager should have a good understanding of the overall functioning of an economy in which the product or service she is associated with is sold, she is mostly concerned with decision-making with regard to a particular group of individuals within a company, a particular product or a particular market. Thus managerial economics typically refers to microeconomics rather than macroeconomics.

Economic Facts and Insights

- Economics is a social science.
- Economics deals with choice problems in the face of scarcity.
- At a given point of time, the production possibility curve is downward sloping, that is, a higher production of one good implies less production of some other good. But, over time, the production possibility curve can shift and a nation can produce more of all goods.
- Underutilisation of resources means that the economy is producing inside its production possibility curve.
- Economic problems involve taking into account the relative benefits and costs of some actions.
- There are three central problems of any economy: 'what', 'how' and 'for whom'. The 'what' problem refers to what and how much goods and services are to be produced in an economy. The 'how' problem refers to the methods of producing a good. The 'for whom' problem refers to income and wealth distribution in an economy.
- Positive economics deals with the effect of a change in a parameter of an economy on various entities in the economy.
- Normative economics is concerned with analysing what policies best serve the interests of an economy.
- Micro economics deals with individual choice problems, whereas macro economics studies the behaviour of economy-wide aggregates like GDP, price level, interest rate and so on.
- Managerial economics refers to the study of economic principles that are relevant or useful in business and managerial decision-making.

EXERCISES

1.1 Give two examples of 'how' problems, other than what is given in the text.

1.2 Give two examples of 'what' problems, other than what is given in the text.

1.3 'How does India's association with the WTO affect the production of agricultural goods in India?' Is this a positive or a normative question? Justify.

1.4 'Should India try to negotiate a free trade agreement with Japan?' Is this a positive or a normative question? Justify.

1.5 Consider an economy in which two goods are produced: (*a*) computers made with the help of labour and specialised machines that can be used in making computers only and (*b*) wheat produced with help of labour and land that again can be used in producing wheat only. In other words, while labour is used in producing both goods, specialised machines are specific to the computer sector and land is specific to the wheat sector. Suppose in such an economy the supply of specialised machines increases. How will this shift the economy's production possibility curve?

1.6 An economy always produces on, but not inside, a *PPC*. Defend or refute.

1.7 A country produces two goods: cell phones and shoes. Along its *PPC*, as more shoes are produced, its marginal opportunity cost in terms of cell phones increases. Using a numerical example, determine if increasing marginal opportunity cost holds for cell phones as well. In other words, in a two-good economy, does increasing marginal opportunity cost of one good vis-à-vis the other along the *PPC* imply increasing marginal opportunity cost of the other good in terms of the former?

1.8 With everything else constant, an increase in the rate of unemployment in the economy shifts the production possible curve to the left. Defend or refute.

1.9 You work in a call-centre. The following options are available if you quit this job: (*a*) work as a sales person in a computer store at Rs 5,000 per month, (*b*) a secretary to an M.P. at Rs 7,000 per month, (*c*) a part-time consultant for an exporting firm at Rs 9,000 per month and (*d*) a technician in a pharmaceutical company at Rs 8,000 per month. What is your opportunity cost of working in the call-centre?

1.10 Suppose refugees from a neighbouring country settle in a host country. How will it affect the *PPC* of the host country?

1.11 A *chaatwallah*, in every half an hour, can serve 10 plates of *chaat* or 30 *gol guppa*s. He works 5 hours a day. Draw his daily *PPC* of serving *chaat* plates and *gol guppa*s.

1.12 A country produces two goods: cricket bats and *sari*s. The production possibilities (in respective units) are given in the following table:

Cricket Bats	*Sari*s
0	64
5	56
10	48
15	40
20	32
25	24
30	16
35	8
40	0

Draw the *PPC*. In terms of the marginal opportunity cost, explain why it looks different from the *PPC*s drawn in the text.

1.13 Bribe taking and bribe giving are not moral, but people do give and take bribes to get various things done—it is an economic decision. Identify the nature of benefits and costs of (*a*) bribe giving and (*b*) bribe taking. (Note: This does not justify the presence of bribes.)

1.14 Give two examples of economic questions through which one can distinguish between micro and macroeconomics.

1.15 'The relationship between economics and managerial economics is like that between mechanical engineering and chemical engineering.' Do you agree or disagree? Give reasons.

1.16 The government should provide jobs to every adult capable of working. This is a noble economic objective for any government. But are there any constraints in meeting this objective? Discuss.

2 Demand

CONCEPTS

- Determinants of Demand
- Demand Schedule
- Demand Curve
- Demand Function
- Normal/Superior Good
- Inferior Good
- Substitute Goods
- Complementary Goods
- Veblen Effect
- Veblen Good
- Bandwagon Effect
- Snob Effect
- Network Externality
- Market Demand
- Change in Quantity Demanded versus Change in Demand
- Price Elasticity of Demand
- Arc Elasticity of Demand
- Point Elasticity of Demand
- Elastic Demand
- Inelastic Demand
- Unitarily Elastic
- Perfectly Inelastic
- Perfectly Elastic
- Income Elasticity
- Cross Price Elasticity
- Product Line
- Advertising Elasticity

Mr Goyal was thinking of selling two 100-square-yard plots that he owned in his hometown at Rs 10 lakh each, in order to pay for his son's education in America (costing Rs 20 lakh). Suddenly, some IT industries moved into the town as the state government began to develop it as an 'IT Park'. Demand for real estate skyrocketed. Within six months one plot fetched Rs 20 lakh. Mr Goyal had to sell only one plot to finance his son's education abroad.

In 1973, the OPEC (Organization of Petroleum Exporting Countries) collectively curtailed the production of oil drastically. The supply of oil in the world market fell so much that the price of oil quadrupled within one year.[1] The world economic scenario has since then changed permanently.

The preceding examples illustrate how market-forces work. In the first example, there was an increase in demand, which resulted in a land-price hike. In the second example, the price of oil soared because of a reduction in supply. 'Demand' and 'Supply' are very important concepts in economics, which help us to understand and explain simple as well as complicated situations. In this chapter and in Chapters 4, 5 and 6, we will learn about demand.

INDIVIDUAL DEMAND FOR A COMMODITY/SERVICE

Suppose each month, your parents give you Rs 1,500 as your stipend, which you spend on food, snacks, stationery, movies, clothes and so on. In this context, think of your demand for ice-cream in a month and ask yourself what factors determine how many ice-cream cones you would buy within a month. You may come up with the following factors:

PRICE OF ICE-CREAM CONES

Assume that all ice-cream cones are of the same kind and quality. Suppose they become more expensive than before—from Rs 15 to Rs 30 a piece. Unless you cannot live without eating so many ice-creams per month, the effect would be that you would consume less ice-cream than before. (Perhaps you will switch partly to a substitute product, say *kulfi*.)

YOUR MONTHLY STIPEND

Suppose, as a reward for having done very well in the exams, your parents raise your stipend from Rs 4,000 to Rs 5,000. You would probably spend more on clothing, entertainment and food (including ice-cream), that is, the effect would be that you would demand more ice-cream.

[1]A barrel of crude oil from Saudi Arabia was costing $2.59 in September 1973. In January 1974, it increased to $11.65.

PRICE OF *KULFI*

Suppose that for some very unusual reason, *kulfi* becomes very cheap—Rs 4 per plate instead of Rs 10. (Your stipend remains unchanged at Rs 4,000.) Would you buy as many ice-creams as before? Given that you do like *kulfi* (although perhaps not as much as ice-cream), you would probably eat more *kulfi* and less ice-cream.

WEIGHT CONSCIOUSNESS

Suppose you have become very weight conscious lately. Since ice-cream is 'fattening,' you may consume less ice-cream than before even when there is no change in prices or your stipend.

FUTURE PRICE EXPECTATION

Imagine a rumour saying that within a matter of one week, ice-cream will be doubly expensive because of supply problems. If such a rumour spreads, many of you would dash to buy ice-cream now and store them in the refrigerator. In other words, your *present* demand for ice-cream is affected by *future* price expectation, while there is no change in the stipend or in any price.

We have listed above most, if not all, of the factors that may affect your demand for ice-cream.[2] We can generalise from this. Interpret the price of ice-cream as 'own price' the price of *kulfi* as the 'price of a related good,' monthly stipend as 'income' and weight-consciousness as 'taste' in the sense that it has changed your 'effective' taste for ice-cream.

Now think of consumables like clothing, household items and other types of food. We can say that the demand for a product by an individual or a household depends on:

(*a*) own price,
(*b*) income,
(*c*) prices of related goods,
(*d*) taste and
(*e*) future price expectations.

The term 'income' refers to the total expenditure on all the different goods you may wish to purchase in a given time period. Note that such a characterisation applies not just to goods (commodities) but also services.[3]

The question now is how, in general, these factors affect the quantity demanded of a product or a service.

[2]There may be other factors. For instance, a change in weather may affect your demand for ice-cream.

[3]Goods or commodities like fish, jeans or a lipstick have physical dimensions. Services refer to consultation with a doctor, house repair jobs and such, which do not have a physical characteristic. Yet they command a price—just like a commodity.

Own Price

The effect of own price is captured through what is called the **law of demand** in economics. It states that, other things remaining unchanged, the quantity demanded of a product by a consumer falls as its own price increases. This is quite intuitive—saying that if a product becomes more (or less) expensive, you would buy less (or more) of it. Note that: (*i*) 'other things' refers to other factors affecting demand like income and prices of related goods, and (*ii*) the law applies in the context of a given area and within a given time period.

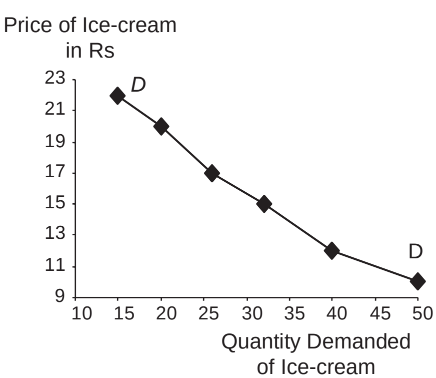
Table 2.1 presents a numerical example of this law and is interpreted as follows. If the price of an ice-cream cone were Rs 10, Sunita would buy 50 cones; if it were Rs 12, she would buy 40 cones and so on—although at any given point of time (and space) there is a uniform price of ice-cream. The law of demand presented in a table is called a **Demand Schedule**. Table 2.1 is a demand schedule.

The numbers in Table 2.1 are graphed in Figure 2.1(a), in which (own) price is measured on the y-axis and quantity demanded along the x-axis. This is the

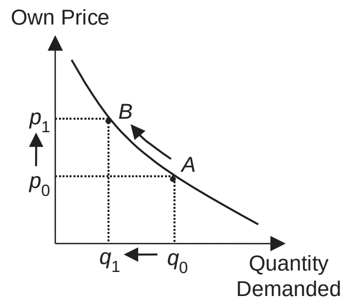
Table 2.1 Demand for Ice-cream by Sunita during Summer Months

Own Price in Rs	*Quantity Demanded of Ice-cream*
10	50
12	40
15	32
17	26
20	20
22	15

Figure 2.1 Demand Curves



(a) Demand Curve Corresponding to Table 2.1



(b) A Hypothetical, Smooth Demand Curve

demand curve corresponding to the demand schedule in Table 2.1. In general, **demand curve** is a graphical representation of the law of demand, measuring the inverse relationship between own price and the quantity demanded.[4] A demand curve is typically illustrated as a smooth downward sloping curve, as shown in Figure 2.1(b). For example, at the price p_0, the quantity demanded is q_0 and the consumer is at the point A on the demand curve. At a higher price p_1, the quantity demanded is less (equal to q_1) and the consumer is at point B. And so on.

The law of demand holds generally but not universally. There are exceptions and we will learn about these as we go along. But unless mentioned explicitly, it is presumed that the law holds and the demand curve is downward sloping.

Income

Suppose the income of a consumer rises and other things (including the own price) remain the same. Would she buy more or less of a particular good?

Normally one would buy more. But it may not be true for all goods. As an example, suppose your income (or stipend) is low. You are probably consuming a fair amount of low-priced snacks like peanuts and biscuits, which you do not like as much as, say, ice-cream, but you buy them because that is mostly what you can afford with your low income. However, if your income rises, it is likely that you will buy less of peanuts and biscuits and more of ice-cream because you are in a position to afford what you like more.

In this example, as income increases, the demand for ice-cream increases and that for peanuts falls. In general, if the demand for a product increases with income, it is called a **superior** or a **normal good**. But, if less of a good is demanded as income increases, it is called an **inferior good**. See Clip 2.1 for a discussion of inferior goods.

Clip 2.1: Which Goods are Inferior in Reality?

Typically, certain cereals are inferior goods. For instance, in most parts of the world, maize, also called corn, has been found to be an inferior good. Rosegrant, Agcaoili-Sombilla and Perez (1995) report that, in Philippines, a 1 per cent increase in income leads to a 0.15 per cent decrease in white maize consumed. Minot and Goletti (2000) find that in Vietnam, even rice is an inferior good given that an increase in income is associated with a decrease in the amount of rice consumed. According to Mckenzie (2002), tortillas in Mexico are inferior goods.

Are cereals inferior goods in India? Most studies, like Radhakrishna and Ravi (1994), Bhalla, Hazell and Kerr (1999) and others, find that they are not. However, according to Kumar (1998), they are; this result is an exception.

(continued)

[4]An inverse relationship between two entities means that, as one entity increases, the other decreases.

References

Bhalla, G. S., P. Hazell and J. Kerr. 1999. 'Prospects for India's Cereal Supply and Demand to 2020'. Food, Agriculture and Environment Discussion Paper 29, Washington, DC: International Food Policy Research Institute.

Kumar, P. 1998. 'Food Demand and Supply Projections for India'. Agricultural Economics Policy Paper 98-01, New Delhi: Indian Agricultural Research Institute.

Mckenzie, D. 2002. 'Are Tortillas a Giffen Good in Mexico?', *Economics Bulletin*, 15(1): 1–7.

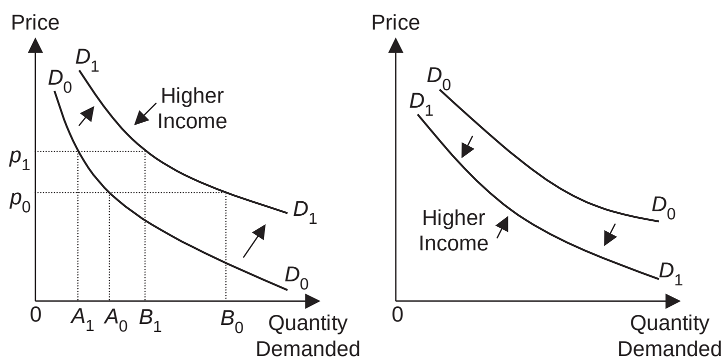
Minot, N. and F. Goletti. 2000. 'Rice Market Liberalization and Poverty in Vietnam'. Research Report No. 114, Washington, DC: International Food Policy Research Institute.

Rosegrant, M. W., M. Agcaoili-Sombilla, and N. Perez. 1995. 'Global Food Projections to 2020: Implications for Investment'. 2020 Vision Discussion Paper No. 5, Washington, DC: International Food Policy Research Institute.

The effect of income on quantity demanded is illustrated in Figure 2.2. Panels (a) and (b) respectively show a normal good and an inferior good. In both panels, D_0D_0 represents the original demand curve and D_1D_1 the new demand curve after an increase in income. Panel (a) shows that at price p_0, for instance, the quantity demanded originally was A_0. At a higher income but at the same price p_0, the quantity demanded is B_0, which is greater than A_0. Similarly, at price p_1, the original quantity demanded was A_1, whereas the new quantity demanded with higher income is B_1. The same pattern holds at any particular price—quantity demanded is greater with higher income—which exactly defines a normal or a superior good. We can then say that, for a normal or superior good, an increase in income shifts the demand curve to the right; by the same logic, a decrease in income shifts the demand curve to the left.

Panel (b) of Figure 2.2 shows the opposite for an inferior good—an increase (or a decrease) in income shifts the demand curve of such a good to the left (or right).

Figure 2.2 Effect of an Increase in Income on Demand



(a) Superior Good (b) Inferior Good

NUMERICAL EXAMPLE 2.1

Table 2.2 is the original market demand schedule of Mr Yamada for a particular kind of fish. Suppose that Mr Yamada's income has increased and for him the fish in question is a normal good. Fill in the blanks in the 3rd column, which will represent a shift of Mr Yamada's demand curve for this fish.

Since this fish is a normal good for Mr Yamada, an increase in income will induce him to demand more at each price. An example of this is given in Table 2.3. Notice that, along column 3, the quantity demanded at each price is higher as compared to column 2. (Yet along column 3 the quantity demanded falls as the price increases.) If we plot the old and the new demand schedules, the latter will be positioned to the right of the former—as in Figure 2.2(a).

PRICES OF RELATED GOODS

In taste, two goods can be related to each other in two ways: they may be substitutes for each other or one may be a complement to the other. In drinks, tea and coffee are substitutes for each other. In meat, chicken and mutton are substitutes. In staple food, rice and *chapatti* are substitutes. But consider tea and sugar. Sugar is a complement to tea—if you consume more tea, you consume more sugar (unless, of course, you are diabetic).[5] Substitutability or complementarity is not restricted to

Table 2.2 Original Market Demand Schedule of Mr Yamada (Numerical Example 2.1)

Price in Yen (the Japanese currency)	*Mr Yamada's Demand for Fish*	*His Demand for Fish after an Increase in Income*
100	10	?
200	7	?
300	6	?
400	5	?
500	3	?
600	1	?

Table 2.3 Mr Yamada's Demand after an Increase in Income (Numerical Example 2.1)

Price in Yen (the Japanese currency)	*Mr Yamada's Demand for Fish*	*His Demand for Fish after an Increase in Income*
100	10	13 (>10)
200	7	10 (>7)
300	6	9 (>6)
400	5	6 (>5)
500	3	4 (>2)
600	1	2 (>1)

[5]Similarly, *chaat masala* is complementary to *chaat*. That is why it is called '*chaat masala*' in the first place.

food or drinks only. For instance, the demand for auto mechanic services is complementary to the demand for automobiles.

If two goods are related in taste it is natural that a change in the price of one good will affect the demand for the other. This is called the **cross price effect**, defined as the effect of a change in the price of one good on the demand for another.

Consider the demand for tea. How will it be affected by an increase in the price of coffee? It would increase, since a substitute good is now costlier. This implies that, as the price of coffee increases, the demand curve for tea shifts to the right (that is, at any given price of tea, more tea will be demanded at a higher coffee price). Thus, an increase (or a decrease) in the price of a substitute good shifts the demand curve of a product to the right (or left). This is exhibited in Figure 2.3(a).

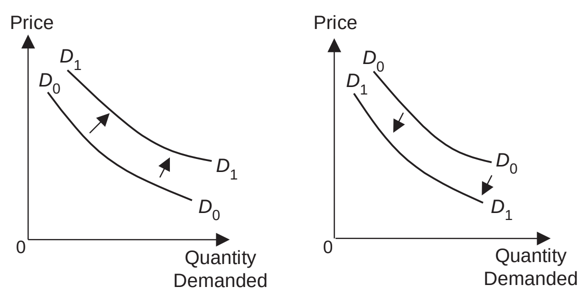
Now ask yourself how an increase in the price of tea would affect the demand for sugar. Tea consumption will fall and so will the demand for sugar. It is opposite of what happens when the price of a substitute good increases. Thus, an increase (or a decrease) in the price of a complementary good shifts the demand curve of a product to the left (or right). This is shown in Figure 2.3(b).

Taste

A person's taste for a product changes for various reasons. When younger, you very much disliked bitter guord; but now your natural dislike may not be that intense. A change in taste may occur even when there is no change in your natural liking. For instance, because of health problems, you may be consuming less chocolate and ice-cream, even though your natural liking for them remain as strong as ever. Taste changes can happen with respect to non-food products too.

How does a taste change affect the demand for a commodity? It is quite simple: a favourable (or an unfavourable) change in taste for a product increases (or decreases) the demand and hence shifts the demand curve to the right (or left).

Figure 2.3 Effect of an Increase in the Price of a Related Good



(a) Price Increase of a Substitute Good

(b) Price Increase of a Complementary Good

Expectations

Suppose there is an unexpected weather forecast that a huge cyclone is going to hit your town in the next three days. People then anticipate that there will be supply problems for many daily usable items during and immediately after the cyclone and hence these items will be much more costly. This induces them to rush to the market now and buy things like oil, bread and potatoes in bulk. Thus it is a situation where there is no change in income, own price, prices of related goods or tastes, yet there is an increase in the *current* demand for some products. We can then say that an increase (or a decrease) in the expected price of a product increases (or decreases) the current demand for it and shifts its current demand curve to the right (or left).

Other Factors

There are still other factors that can influence a person's demand for a commodity. For example, some people, in order to impress others (and possibly themselves), tend to buy more of a good when it becomes more expensive. You can think of this as a (strange) change in taste because of a change in price. This is called a **Veblen effect** and the good in question is called a **Veblen good**.[6] Veblen goods are 'conspicuous goods', which others may look upon with some envy—like a branded car or a fancy jacket. As the quantity demanded increases with price, the law of demand does not obviously hold for a Veblen good and for a person whose behaviour is subject to the Veblen effect.

There may be a **bandwagon effect,** that is, if others start to use a product in a big way, it becomes fashionable and so you go for it too without looking at its merit. In other words, you 'follow the crowd'.

There can also be a **snob effect**, referring to a situation where you want to be different from others for the sake of being different, that is, you 'go against the crowd'. If more people buy something, your demand for it falls. This is the opposite of the bandwagon effect.

Still another effect can be in the form of a **network externality**, defined as a change in the benefit to a person when the number of other individuals using it increases. For instance, if your friends, relatives and colleagues do not use e-mail, your demand for e-mail will be small. But if they all start to use e-mail, your demand for e-mail will increase. Thus a person's benefit increases with the size of the network and, therefore, an increase in the number of consumers in a network increases the individual demand for the product or service. Note that it is somewhat similar to but not the same as the bandwagon effect, which makes you 'go' for the product without looking at its merit.

The opposite can also happen—a person's benefit may decrease with the number of users in a network. Consider using the bus service in a city. If the number of bus-riders increases and the buses are overcrowded, it may lower your benefit from a

[6] It is named after Thorstein Veblen, an American economist and psychologist in the late 19th and the early 20th century.

bus-ride and induce you to use less of it. This is a case of negative externality, whereas the e-mail example was a case of positive externality. Unless specified, the convention is that network externality refers to the case of a positive externality.

Mathematically Speaking

The Demand Function

The dependence of demand for a particular good on several factors can be expressed algebraically in the form of a **demand function**. Consider the demand for, say, good A. Use the following notations: $q_A \equiv$ the quantity demanded of good A, $p_A \equiv$ the own price, p_B, $p_C \equiv$ prices of related goods (there can be other related goods), $M \equiv$ the income and $T \equiv$ other factors. The demand function for good A can be written as a mathematical relation:

$$q_A = f(p_A, p_B, p_C, M, T).$$

How different factors affecting the quantity demanded of a good determine the properties of the demand function f.

The law of demand implies that, other things remaining constant, as p_A changes, q_A changes in the opposite direction. Hence, the partial derivative $\partial f/\partial p_A$, which is same as $\partial q_A/\partial p_A$, is negative in sign. Suppose good B is a substitute for good A. Then, a change in p_B causes q_A to change in the same direction. Therefore, $\partial f/\partial p_B > 0$. Similarly, if good A is complementary to good C, $\partial f/\partial p_C < 0$.

The partial derivative with respect to M depends on whether good A is normal or inferior. By applying similar reasoning, $\partial f/\partial M$ is positive or negative if good A is normal or inferior respectively. The sign of $\partial f/\partial T$ depends on the nature of T.

Suppressing the 'other factors' T, an example of a demand function will be:

$$q_A = 100 - 2p_A + 0.5p_B - p_C + 1.5M.$$

It means, for example, that if there is a unit increase in own price, the quantity demanded will fall by 2 units. For instance, let p_A be measured in hundreds of rupees. Then, starting from an initial value, if p_A increases by Rs 100, then q_A will decrease by 2 units.

If we are only concerned with changes in own price, that is, p_B, p_C and M are unchanged, for notational simplicity, we can also suppress them and write the demand function very simply as:

$$q_A = f(p_A).$$

As an example, we can write:

$$q_A = 200 - 2p_A.$$

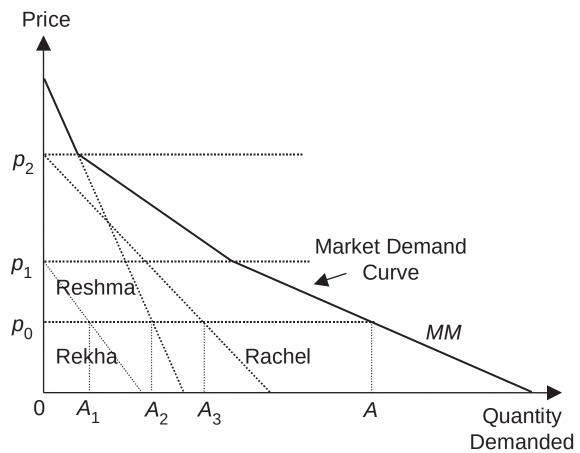
* * * * * *

MARKET DEMAND

So far we have analysed the demand for a particular good by a single individual, family or household. **Market demand** is the total demand by all individuals comprising the entire market.[7]

As a simple example, let there be only three individuals in the market: Rekha, Reshma and Rachel. Figure 2.4 depicts their individual demand curves for ice-cream, marked by the respective dotted lines. (Reshma's demand curve is partly hidden; it extends all the way up to the y-axis.) It is presumed for simplicity that the demand for ice-cream can be measured continuously as points on a line. For instance, at price p_0, Rekha demands $0A_1$, Reshma $0A_2$ and Rachel $0A_3$. Then the quantity demanded in the market at price p_0 equals $0A_1 + 0A_2 + 0A_3$. This sum is equal to $0A$, where the point A is marked on the curve MM. Other points on this curve are obtained by similar exercises at other prices.[8] The MM curve is then the market demand curve.[9]

Figure 2.4 Market Demand Curve



[7]By 'market,' one could mean the entire city, province, region, country or even the whole world. It depends on the context of the analysis.

[8]Note that at any price equal to or greater than p_2, the quantity demanded of ice-cream by Rekha and Rachel is zero. Thus the total quantity demanded is same as the quantity demanded by Reshma only. Similarly, at any price between p_1 and p_2, the quantity demanded by Rekha is zero. Hence the total quantity demanded equals the sum of quantities demanded by Reshma and Rachel. At any price below p_1, all three individuals demand positive quantities and these are added to obtain the total quantity demanded in the market.

[9]Note that if there were 10 or 10 lakh individuals in the market, we would have done the same thing in principle, that is, added up the quantities demanded by all individuals in the market at each price.

In obtaining the market demand curve, essentially, the quantities measured along the horizontal, x-axis are added up. Thus we can say that the market demand curve is the horizontal or lateral summation of individual demand curves.

Mathematically, if there are, say, H families, denoted by 1, 2, 3, …, H, and their demand functions are $f_1(p)$, $f_2(p)$ and so on, the market demand function is given by $f_1(p) + f_2(p) + \cdots + f_H(p)$.

What are the factors that can shift a market demand curve? Since it is based on individual demand curves and the number of individuals constituting the market, these factors are:

(*a*) income levels of individuals, that is, income distribution across consumers in the market,
(*b*) prices of related goods,
(*c*) tastes of individuals (that is, distribution of tastes) and
(*d*) the number of individuals or consuming units in the market.

NUMERICAL EXAMPLE 2.2

Suppose there are three individuals in the market: Ram, Rahim and Randall. Their individual demand schedules for apples are given in Table 2.4. For example, at the price of Rs 9, Rahim consumes (demands) 4 apples. Derive the market demand schedule and plot individual and market demand curves.

At each price, we sum up the individual quantities demanded. Thus at price equal to Rs 5, the total or market quantity demanded = 15 + 7 + 14 = 36. At price Rs 6, it is equal to 12 + 6 + 12 = 30, and so on. Table 2.5 is the market demand schedule.

Figure 2.5 depicts the individual curves and the market demand curve in this numerical example.

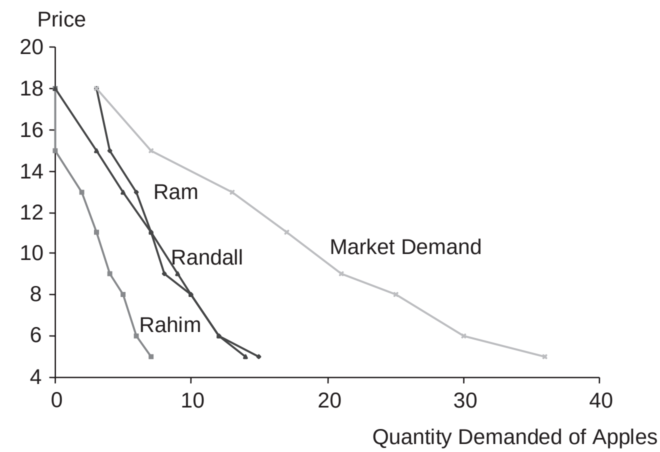
Table 2.4 Individual Demand Schedules (Numerical Example 2.2)

Price (Rs)	*Quantity Demanded by Ram*	*Quantity Demanded by Rahim*	*Quantity Demanded by Randall*
5	15	7	14
6	12	6	12
8	10	5	10
9	8	4	9
11	7	3	7
13	6	2	5
15	4	0	3
18	3	0	0

Table 2.5 Market Demand Schedule (Numerical Example 2.2)

Price (Rs)	*Total Quantity Demanded of Apples*
5	36
6	30
8	25
9	21
11	17
13	13
15	7
18	3

Figure 2.5 Market Demand Curve in Numerical Example 2.2



CHANGE IN QUANTITY DEMANDED VERSUS CHANGE IN DEMAND

Think for a moment about how we have organised the effects of various factors on demand for a particular commodity (by an individual or the entire market). First, the effect of a change in the own price is characterised by the law of demand and depicted as the demand curve. Second, the effects of changes in other factors like income, prices of related goods and taste are seen through a shift in the demand curve.

Having understood this, we can differentiate the two terms in the title of this section. Refer back to Figure 2.1(b): suppose the own price changes from p_0 to p_1, then the consumer 'moves', so-to-speak, from point *A* to point *B* on the same demand curve. This is called a **change in quantity demanded**—meaning a movement along a given demand curve due to a change in the own price. On the other hand, a **change in demand** refers to a change in the whole demand schedule or curve, which is same as a shift of the demand curve that occurs due to a change in other factors such as income, prices of related goods or taste. In other words, the first term refers to the movement *along* a demand curve and the second to a movement *of* it.

ELASTICITY

So far, we have learnt how different factors *qualitatively* affect the demand for a particular good—the *direction* of change in the quantity demanded of a good—as the own price, the prices of related goods, income or tastes change. We now learn about concepts of 'elasticity' that aim to *quantify* the responsiveness of the quantity demanded of a good to changes in the various factors.

Price Elasticity of Demand

It measures the response of the quantity demanded to a change in the own price and is defined by:

$$e_p = -\frac{\%\text{ change in quantity demanded}}{\%\text{ change in the own price}}. \tag{2.1}$$

We know from the law of demand that the own price and the quantity demanded change in opposite ways. Thus the percentage change in the quantity demanded bears a sign, which is the opposite of that in own price, implying that the ratio of the two percentage changes is negative. Hence the negative sign in front of the formula given in (2.1) makes e_p positive. Strictly speaking, e_p gives the *absolute* value of the price elasticity, which is commonly (and loosely) called 'price elasticity.'[10] By definition, the greater the responsiveness of the quantity demanded for the same proportionate change in price, the higher is the price elasticity.

MEASUREMENT

There are alternative methods to measure e_p. The percentage method is one of these.

This method directly applies the formula (2.1). Let, originally, the price of a product be p_0, at which the quantity demanded is equal to q_0. Let the price change to p_1 and, accordingly (along the demand curve), let the quantity demanded change to q_1. Then the percentage changes in the own price and in the quantity demanded have the following expressions respectively:

$$\frac{p_1 - p_0}{p_0} \times 100; \quad \frac{q_1 - q_0}{q_0} \times 100.$$

[10]Some textbooks define price elasticity as the formula given in (2.1), except for the negative sign.

The first one goes in the denominator (2.1) and the second in its numerator. Hence,

$$e_p = -\frac{(q_1 - q_0)/q_0}{(p_1 - p_0)/p_0}. \tag{2.2}$$

In economics, a change in something is typically denoted by the greek letter Δ (pronounced Delta). For example, Δq_d would denote the change in the quantity demanded, equal to $q_1 - q_0$, and, similarly Δp would denote $p_1 - p_0$. Using this notation, we can also write (2.1) as

$$e_p = -\frac{\Delta q_d / q_d}{\Delta p / p}. \tag{2.3}$$

A numerical example that uses (2.2) or (2.3) is given below.

NUMERICAL EXAMPLE 2.3

Renu used to consume 10 ice-creams a month when the price of an ice-cream was Rs 12. Ice-cream has become more expensive with Rs 15 a piece. Renu now consumes 8 ice-creams a month. What is the e_p of demand for ice-cream by Renu?

The original price and quantity are: $p_0 = 12$ and $q_0 = 10$. The new price and quantity are: $p_1 = 15$ and $q_1 = 8$. Thus the percentage change in price $= 100 \times (15 - 12)/12$. Similarly, the percentage change in the quantity demanded $= 100 \times (8 - 10)/10$. Applying (2.2) or (2.3),

$$e_p = -\frac{-2/10}{3/12} = 0.80. \tag{2.4}$$

Arc Elasticity

There is, however, a problem with the percentage method. Consider the above numerical example by regarding the new price and quantity as the old price and quantity respectively, and the old price and quantity as the new price and quantity respectively. Thus, let $p_0 = 15$, $q_0 = 8$, $p_1 = 12$ and $q_1 = 10$. Since the concept of elasticity refers to the magnitude of the response of the quantity demanded to a price change, the order between which is the old and which is the new price-quantity combination should not ideally matter. But, unfortunately, it does. Apply (2.2) or (2.3) to this example and you will find that e_p is not what is given in (2.4). It is equal to 1.25.[11]

It is not hard to see why this difference is arising. It is because the denominator of the percentage change of price or quantity is sensitive to the original price or quantity. To get around this problem, a compromise is done. Instead of p_0 in the

[11] We have $(q_1 - q_0)/q_0 = (10 - 8)/8 = 1/4$ and $(p_1 - p_0)/p_0 = (12 - 15)/15 = -1/5$. Thus $e_p = -(1/4) \div (-1/5) = 1.25$.

denominator of the percentage change of price, we take $(p_0 + p_1)/2$, that is, the average of the original price and the new price. Similarly q_0 in the denominator of the percentage change in quantity is replaced by $(q_0 + q_1)/2$. Then the formula becomes:

$$-\frac{(q_1 - q_0)(p_0 + p_1)}{(q_0 + q_1)(p_1 - p_0)}. \tag{2.5}$$

This is called the **arc elasticity**. According to this measure, the computed price elasticity is independent of the original price-quantity combination. In the Numerical Example 2.3, the arc elasticity is equal to 1.00.

NUMERICAL EXAMPLE 2.4

Arup was buying 20 music CDs when the price of a CD was Rs 200. Now the price has come down to Rs 150 and he is buying 30 CDs. (The quality of CDs and the number of songs in each CD are the same.) What is the arc elasticity of demand for music CDs by Arup?

Here the original price and quantity demanded are: $p_0 = 200$ and $q_0 = 20$. The new price and quantity demanded are: $p_1 = 150$ and $q_1 = 30$. Thus $p_0 + p_1 = 350$, $q_0 + q_1 = 50$, $p_1 - p_0 = -50$, and $q_1 - q_0 = 10$. Applying the formula given in (2.5), the arc elasticity is equal to 1.4.

Geometric Method and Point Elasticity

Arc elasticity is an approximate formula, appropriate when the price change is not small. When the price change is very small, we can use a graphical formula or a **geometric method**. This is illustrated in Figure 2.6.

Consider Panel (a). The demand curve is a straight line intersecting the axes at *A* and *B*. Suppose, initially the price is p_0. A consumer is demanding the amount q_0, that is, she is at point *C* on the demand line. Then, as shown mathematically in Appendix 2A, e_p is equal to BC/CA (the ratio of the lower segment to the upper segment of the line *AB*).[12]

If the demand curve is not a straight line, it is a bit more complicated but essentially the same. Suppose the demand curve looks like the downward sloping curve shown in panel (b) and we are to compute the price elasticity at the price p_0. Now, draw the tangent of the demand curve at the price where the elasticity is to be computed. At p_0, it is *GH*. For a very small price change around p_0, e_p turns out to be equal to HC/CG (again the ratio of the lower to the upper segment). This is also proven in Appendix 2A.

[12]Note that if we consider a price, which is higher, this ratio will be higher, which means that the higher the price, the greater is the point elasticity. This is economically intuitive, saying that if a product gets more expensive, the consumer becomes more sensitive about her demand for it in response to a further price change.

Figure 2.6 Point Elasticity

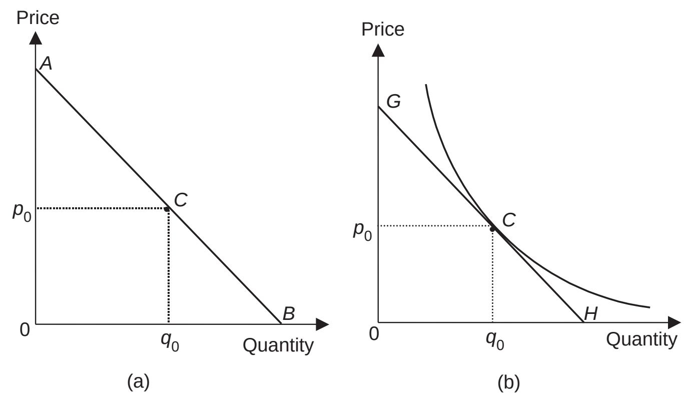
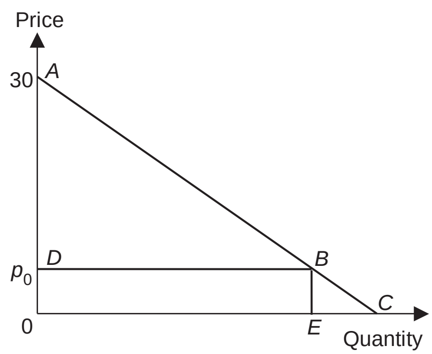


Figure 2.7 Computing Point Elasticity



This geometry-based formula is called **point elasticity**, wherein the word 'point' refers to a small variation. In general, the point elasticity at a given price is equal to the ratio of the lower segment to the upper segment of the demand line at that price or the same ratio on the tangent of the demand curve at that price.

NUMERICAL EXAMPLE 2.5

Suppose the demand curve for a product is a straight line as shown in Figure 2.7. Its intercept on the price axis equals 30. Calculate the point elasticity at the price p_0, where $p_0 = 9$.

Since $p_0 = 9$, we have $BE = 0D = 9$ and $AD = 0A - 0D = 30 - 9 = 21$. Hence $BE/AD = 9/21 = 3/7$. The triangles BEC and ADB are similar. Thus $BE/AD = BC/AB$. Since $BE/AD = 3/7$, we have the measure of point elasticity (BC/AB) equal to 3/7.

Elastic and Inelastic Demand

These concepts relate to the magnitude of the price elasticity. If $e_p > 1$, we say that the product demand is **elastic**. Typical examples are luxury items like branded shoes, jewelry and food in a posh hotel. Luxury goods are relatively 'dispensable', not essential. Therefore, the quantity demanded of such products can vary a lot, implying that the price elasticity of these goods typically exceeds one. On the other hand, if $e_p < 1$, we say that the demand is **inelastic**. Necessary goods like basic food and clothing items fall into this category. Since these are essential, the demand for them is relatively insensitive to price changes. In the intermediate case of $e_p = 1$, the demand for the product is said to be **unitarily elastic**; the demand is neither elastic nor inelastic. If $e_p = 1$, the demand curve has a particular mathematical shape called rectangular parabola, as shown in Figure 2.8(a), which is asymptotic to both axes.

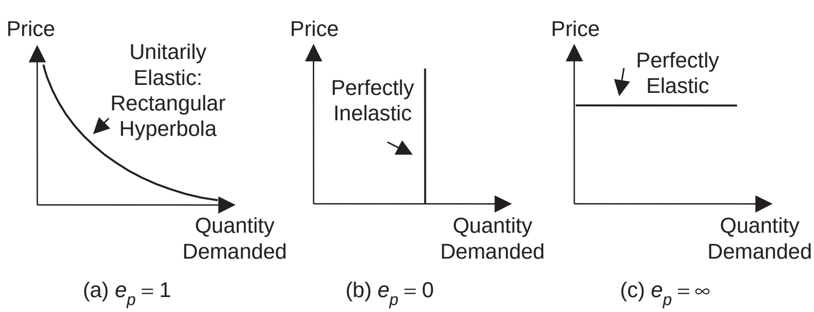
There are two other special cases. If $e_p = 0$, then the product demand is said to be **perfectly inelastic**. This holds when there is no change in quantity demanded as price changes, that is, the product is absolutely essential to the consumer. A very bad case of addiction is an example. If someone is severely addicted to smoking and has to smoke a pack a day, he/she may buy one packet a day irrespective of what the price of cigarettes are (so long as cigarettes are not too costly). This is shown in Figure 2.8(b). The demand curve is vertical.

Figure 2.8(c) depicts the opposite case when e_p is infinite and the demand curve is a flat (horizontal) line. In this case the demand is **perfectly elastic**. An example will be given in Chapter 10.

Determinants of Price Elasticity

A major determinant of price elasticity is *the availability of close substitutes*. If close substitutes are not available to a consumer, she is likely to be too 'dependent' on the product and, hence, her demand for it is likely to be relatively inelastic.

Figure 2.8 Unitarily Elastic, Perfectly Inelastic and Perfectly Elastic Demand



Essential items, by definition, do not have many close substitutes and, therefore, the demand for them is typically inelastic. On the other hand, luxury items have other substitutes and the demand for them is likely to be elastic.

Price elasticity depends also on whether the good in question is a narrowly defined product or a broadly defined one. For example, suppose the price of Levi's jeans increases, while prices of other brands like Lee and Wrangler are unchanged. This will lead to a large decline in the demand for Levi's jeans because Levi's jeans are a narrowly defined product and there are close substitutes available. But if we are looking at a product like a family's consumption of clothing in general, the demand for it is likely to be inelastic, since clothing is essential. Thus, with everything else the same, the price elasticity of a narrowly defined product is likely to be high and that of a broadly defined product is likely to be small.

Another determinant is *the proportion of total expenditure spent on the product*. If this proportion is large, a consumer's demand will be quite sensitive to a price change, that is, the elasticity will be high, as this product is very important in the budget. But if the proportion of total expenditure spent on the product is small, as in the case of salt, the elasticity is likely to be small.

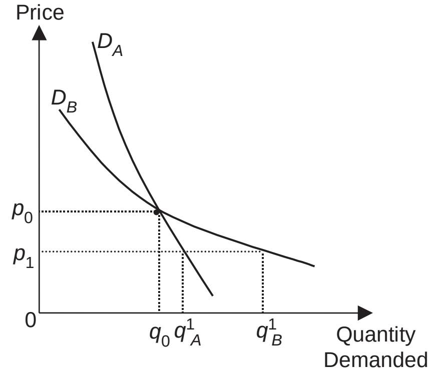
Individual habits also matter. A product may be very 'essential' for someone but not for others. You may not be a regular jogger and hence your demand for jogging shoes may be elastic. But your friend may be an avid jogger and her demand for jogging shoes may be inelastic.

All else the same, *time period* affects price elasticity. The greater the time span, the higher is the elasticity of demand for the same product. If the time period is short, it is hard to find or develop substitutes of a product, while it becomes easier over a longer period. The oil price shock experienced by the world economy in the early seventies is a prime example. Within a year, the price of oil increased by more than three-fold. The world economy was badly hit because almost all countries were extremely dependent on oil and their demand for oil was inelastic. However, over time other forms of energy were explored. The demand for oil is more elastic now than in the mid-seventies.

Comparing Two Demand Curves

So far we have focused on price elasticity on a given demand curve. Can we compare elasticities of two demand curves? In general, we cannot. But if the two demand curves intersect, we can rank the price elasticity at their point of intersection. Turn to Figure 2.9. The two demand curves, D_A and D_B, intersect at the price p_0. Suppose this is the original price. At this price, the quantities demanded on both curves are equal to q_0. Let the price fall to p_1. The quantity demanded increases to q_A^1 along D_A and q_B^1 along D_B. Note that the increase in the quantity demanded (Δq) along the flatter demand curve D_B is greater than that along the steeper demand curve D_A. On both demand curves, the original price and the original quantity demanded are the same and, moreover, the price change (Δp) is also the same. Hence price elasticity must be greater along the flatter curve D_B.

Figure 2.9 Comparing Demand Elasticities



If we had considered an increase in price, we can easily check that the magnitude of decrease in the quantity demanded is greater along D_B also. Therefore, of the two demand curves, the flatter demand curve is more elastic than the steeper demand curve at their point of intersection.

Price Elasticity, Total Expenditure and Total Revenue

The information on price elasticity is helpful not just in assessing the responsiveness of quantity demanded of a product to a change in the own price but also in determining how the total spending on a good, price × quantity demanded, may change as the product price changes.

Suppose (for some reason) the price of petrol increases from Rs 50 to Rs 60 per litre. Is your family going to spend more on petrol or less? In the short run (say within a week's time after the petrol price hike), the demand for petrol is likely to change only a little, if at all. As a result, the family's spending on petrol will be higher. But, over time, the family will look for substitutes. Perhaps the children in the family are going to use bus to go to school rather than the family car. Previously, the family almost always took the car to go to hill stations. Now it may decide to take the train. Since the quantity demanded falls when the price increases, it is not clear whether the total spending on petrol will increase or decrease with a price change.

As you can guess, the answer depends on the magnitude of the price elasticity of demand. If the demand for petrol is inelastic, there will not be much decline in the quantity demanded of petrol as its price increases and thus price × quantity demanded of petrol will increase. More precisely, if $e_p < 1$, the percentage change in price exceeds the percentage change in quantity demanded and, therefore, the direction of change in price will govern the direction of change in the total expenditure on the product. Thus, if $e_p < 1$, an increase (or a decrease) in the price will lead to an increase (or a decrease) in the total expenditure on the product.

Similarly, if the demand for a product is elastic, that is, $e_p > 1$, the percentage change in quantity demanded is greater than that of price. Hence, the change in the direction of quantity demanded will dictate the change in the direction of total expenditure. As the price increases (or decreases), the quantity demanded will increase (or decrease) and the total expenditure will increase (or decrease).

In the intermediate case, where $e_p = 1$, the percentage change in price is equal to the percentage change in quantity demanded and, hence, the total expenditure does not change with price.

In more precise terms, because total expenditure = price × quantity demanded, we have the following mathematical relationship:[13]

$$\begin{aligned}
&\text{\% increase in the total expenditure}\\
&\quad = \text{\% increase in price} + \text{\% increase in quantity demanded}\\
&\quad = (\text{\% increase in price})\left(1 + \frac{\text{\% increase in quantity demanded}}{\text{\% increase in price}}\right)\\
&\quad = (\text{\% increase in price})\left(1 - e_p\right).
\end{aligned}$$

Table 2.6 summarises the qualitative relationships between price change and its effect on total expenditure.

The impact of a change in price on the total expenditure on a good can also be useful from the perspective of a business or a seller. Realise that the total expenditure of consumers on a product sold by a seller is equal to the seller's total revenue or what is called the total turnover. Hence, if a seller is contemplating a price change of a product and wants to know whether total revenue or turnover will increase or decrease, the price elasticity of demand is the key. If this elasticity is greater than one, a price increase (or decrease) will lower (or raise) the total revenue. If it is less than one, a price increase (or decrease) will raise (or lower) the total revenue.

Table 2.6 Price Change and Its Impact on Total Expenditure

Price Change	*Elasticity*	*Total Expenditure*
↑	$e_p > 1$	↓
↓	$e_p > 1$	↑
↑	$e_p < 1$	↑
↓	$e_p < 1$	↓
↑↓	$e_p = 1$	No Change

[13] This relationship holds exact only if the price change is marginally small. Otherwise, it holds approximately.

Mathematically Speaking

Price Change and Total Expenditure

For a small price change, (2.3) reduces to

$$e_p = -\frac{p}{q_D}\frac{dq_D}{dp}.$$

Total expenditure on a good is equal to pq_D. Hence the change in the total expenditure due to a change in price is given by:

$$\frac{d(pq_D)}{dp} = q_D + p\frac{dq_D}{dp} = q_D\left(1 + \frac{p}{q_D}\frac{dq_D}{dp}\right) = q_D(1 - e_p),$$

which is positive, zero or negative accordingly as $e_p < 1$, $= 1$ or > 1. Thus, total expenditure increases, remains constant or decreases with an increase in price as the price elasticity is less than, equal to or greater than one. Table 2.6 is a statement of this result.

* * * * * *

Income Elasticity of Demand

So far we have focused on the responsiveness of the quantity demanded with respect to the own price. We can also ask how responsive the quantity demanded is towards income changes. This is captured by **income elasticity of demand**, defined by:

$$e_I = \frac{\text{\% change in quantity demanded}}{\text{\% change in income}}. \tag{2.6}$$

For a normal good, an increase in income leads to an increase in quantity demanded; hence e_I is positive. For an inferior good, income and quantity demanded are inversely related and thus $e_I < 0$. From now on let us consider only a normal good.

The magnitude of the income elasticity of demand for a product depends on how essential or non-essential the product is, that is, the nature of the need for the product. Since the demand for essential goods is not likely to vary much, the income elasticity (just as its price elasticity) is likely to be small. Otherwise, it is likely to be relatively large.

Further, there is a relationship between a change in income and the corresponding change in the share of total expenditure on a product, depending on whether e_I is greater or less than one. If we let p and q denote respectively the price and the quantity demanded, the total expenditure on the product is equal to $E = pq$. Thus, the share of total expenditure on the product equals E/M or pq/M, where M is the total expenditure on all goods. If $e_I > 1$, by definition, the percentage change in quantity demanded is greater than the percentage change in income. Hence, the price of the product remaining unchanged, the ratio of quantity demanded to income and, therefore, the ratio of total expenditure on the product to income (which is same as the share of total expenditure on the product) must increase. Similarly, if $e_I < 1$, an increase in income will lead to a decrease in the share of total expenditure on the product.

Income elasticity is important for targeting of sales or what is the same thing as formulating a marketing strategy. Suppose a company sells its product globally and is looking at two 'emerging markets,' meaning countries where the aggregate income is expected to grow at a decent rate. In one of these, the income elasticity of demand for its product is less than one and in the other it is greater than one. Then the company should promote its product more aggressively in the second market. Over the long run, income elasticity measures can be used for planning a firm's growth in a particular product or service.

Clip 2.2 provides a sample of price and income elasticities of demand that have been estimated by different authors for various products in various countries.

Clip 2.2: Price and Income Elasticity Estimates

There are innumerable econometric studies, which have estimated price and income elasticities for various products and services with respect to various countries. A few are reported here.

Commodity or Service	*Study*	*Estimates*
		Price Elasticity
Non-cereal food, India (urban + rural)	Meenakshi and Ray (1999)	0.804
Clothing, India (urban + rural)	Meenakshi and Ray (1999)	0.560
Electricity, India (urban)	Filippini and Pachauri (2002)	0.32 (winter) 0.39 (monsoon) 0.16 (summer)
High-speed Internet Access through Cable Modems, US (residential)	Kridel, Rappoport and Taylor (2000)	1.08 to 1.79
		Income Elasticity
Electricity, India (urban)	Filippini and Pachauri (2002)	0.69 (winter) 0.64 (monsoon) 0.66 (summer)
White Maize, Philippines	Rosegrant, Agcaoili-Sombilla and Perez (1995)	0.15

(continued)

References

Filippini, M. and S. Pachauri. 2002. 'Elasticities of Electricity Demand in Urban Indian Households'. Working Paper No. 16, Zurich: Centre for Energy Policy and Economics, Swiss Federal Institute of Technology.

Kridel, D. J., P. N. Rappoport and L. D. Taylor. 2000. 'The Demand for High-Speed Access to the Internet: The Case of Cable Modems'. Presented at the 13th Biennial Conference of the International Telecommunications Society, Buenos Aires.

Meenakshi, J. V. and Ranjan Ray. 1999. 'Regional Differences in India's Food Expenditure Pattern: A Complete Demand Systems Approach', *Journal of International Development*, 11: 53–81.

Rosegrant, M. W., M. Agcaoili-Sombilla and N. Perez. 1995. Global Food Projections to 2020: Implications for Investment. 2020 Vision Discussion Paper No. 5, Washington, DC: International Food Policy Research Institute.

Cross Price Elasticity

This elasticity measures the sensitivity of quantity demanded of one good with respect to a change in the price of a related good. **Cross price elasticity** is defined as:

$$e_c = \frac{\%\text{ change in quantity demanded}}{\%\text{ change in the price of the related good}}. \tag{2.7}$$

If the good in question is a substitute of the good whose price is changing, then $e_c > 0$ since the quantity demanded and the price of the related good change in the same direction. Otherwise, if the product in question is complementary to the good whose price is changing, $e_c < 0$.

A simple example of the usefulness of cross price elasticity is when a rival company has reduced the price of its product and you would like to know by how much it would affect the quantity demand of your product in the market, which is a substitute of the rival company's product.

Further, if a firm produces a **production line** instead of a single model of a product, knowing the cross elasticity can be quite useful. The following example illustrates this.

Consider the car market in India, which has different segments, A, B, C, D and so on. 'A' refers to the smallest-size cars like Maruti 800, Zen, Tata Indica and low to middle valued Santro. 'B' refers to the next level like high-end Santro, Getz, low-end Accent and low-end Esteem. Middle-sized cars in the 6 to 7 lakh range such as middle-end to high-end Esteem, Honda City and middle-end to high-end Accent constitute the 'C' segment and so on. Consider the decision making by Hyundai of India for example. It produces cars for different segments like other manufacturers. Suppose that Hyundai of India finds out that its revenues from new car sales in the B segment have remained stagnant for quite some time. Its management decides to slash the price of its B segment cars by Rs 50,000, expecting that it will increase its car sales and revenues from this segment (under the assumption that the demand for B segment cars is price elastic). Will this strategy necessarily increase the overall, total revenues coming from sales in different segments? Not necessarily, because there will be a cross-substitution effect. Some of the new customers who were thinking of buying an A segment Hyundai car before will switch over to a B segment car. Similarly, from the other end, some potential buyers of the C segment Hyundai car may also want to switch to a B segment car. In other words, Hyundai will face a decline in the revenue from sales in the A and C segments. This is where information on the cross price elasticity is critical. The higher the elasticity of demand for A and C segment Hyundai cars with respect to the price of B segment Hyundai cars, the greater will be the decline in revenues from A and C segments, and the less likely will it be that a price cutting strategy of this kind will boost the company's total revenues.

See Clip 2.3 for some estimates of cross price elasticity in different parts of the world.

Clip 2.3: Estimates of Cross Price Effects

Like own price and income elasticities, there are many estimates available for cross price elasticities for different pairs of commodities and for different countries. The following table gives a sample that covers three continents.

Demand for the Commodity or Service	*Substitute Good Whose Price Change is Considered*	*Study*	*Estimates*
Cable modem for high-speed Internet access	Dial-up access	Kridel, Rappoport and Taylor (2000)	0.15 (US)
Poultry	Beef	FAO Regional Office for Asia and the Pacific (1999)	0.10 (Indonesia)
Electricity	Firewood	Kebede, Bekele and Kedir (2002)	0.59 (Ethiopia)

References

Food and Agricultural Organization (FAO) Regional Office for Asia and the Pacific. 1999. *Livestock Industries of Indonesia Prior to the Asian Financial Crisis*. Bangkok: RAP Publication (FAO).

Kebede, B., A. Bekele and E. Kedir. 2002. 'Can the Urban Poor Afford Modern Energy? The Case of Ethiopia', *Energy Policy*, 30: 1029–1045.

Kridel, D. J., P. N. Rappoport and L. D. Taylor. 2000. 'The Demand for High-Speed Access to the Internet: The Case of Cable Modems'. Presented at the 13th Biennial Conference of the International Telecommunications Society, Buenos Aires.

Advertising Elasticity

By influencing people's tastes, advertising by business firms tends to change the consumers' demand for a particular product or a particular brand of a product. We define advertising elasticity as:

$$\text{Advertising Elasticity} = \frac{\%\text{ change in sales}}{\%\text{ change in the units of advertising}}.$$

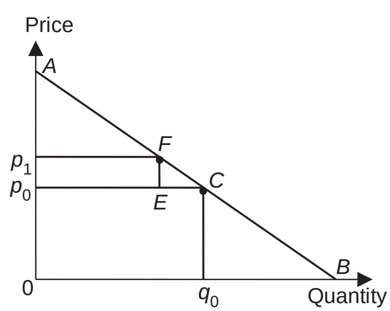
Several studies that have estimated advertising elasticity for various categories of products. For example, in the context of the US, a study by The Kaiser Family Foundation in 2003, *Impact of Direct-to-Consumer Advertising on Prescription Drug Spending*, found that in the medicine market during the period 1996 to 1999, a 10 per cent increase in direct-to-consumer advertising and physician promotion activity by medicine companies led to a 1 per cent increase in the medicine sales. Thus the advertising elasticity in this case is 0.10.

What are the factors influencing this elasticity? One is advertising by the rival firms. The more the advertising by rivals, the less will be the advertising elasticity facing a particular firm. Another factor is the stage of the product in the market. If the product is relatively new in the market, the advertising elasticity is likely to be high (as more and more people come to know about it and its quality), but as the product gets old, this elasticity is likely to become smaller.

APPENDIX 2A: DERIVATION OF THE POINT ELASTICITY FORMULA

Consider the straight line demand curve *AB* in Figure 2A.1: it is the same as Figure 2.6(a) except that there are more lines and points. The original price is $p_0 = Cq_0$, and the original quantity demanded is $q_0 = p_0C$. We are to prove that the point elasticity at the price p_0 is equal to BC/AC.

Figure 2A.1 Deriving the Point Elasticity Formula



Suppose there is an increase in price and the new price is p_1. The new quantity demanded is p_1F. Consider now the triangle FEC. The change in price and the change in quantity demanded are respectively equal to $\Delta p = -FE$ and $\Delta q = -EC$. The '–' sign for the quantity change means a decrease in quantity. Thus,

$$e_p = -\frac{p_0}{q_0}\frac{\Delta q}{\Delta p} = \frac{Cq_0}{p_0C}\frac{EC}{FE}.$$

As the triangles FEC and Cq_0B are similar, $EC/FE = q_0B/Cq_0$. Hence,

$$e_p = \frac{Cq_0}{p_0C}\frac{q_0B}{Cq_0} = \frac{q_0B}{p_0C}.$$

The triangles Cq_0B and Ap_0C are similar. Hence $q_0B/Cq_0 = BC/CA$. Thus $e_p = BC/CA$. This proves the point elasticity formula for a straight line demand curve.

If it is a demand curve as in Figure 2.6(b), for a small change in price, $\Delta p/\Delta q$ is simply the slope of the demand curve (as price and quantity are respectively measured vertically and horizontally). At the price p_0, the slope = $-0G/0H$. Thus, $\Delta p/\Delta q = -0G/0H$, or $\Delta q/\Delta p = -0H/0G$, and

$$e_p = -\frac{p_0}{q_0}\frac{\Delta q}{\Delta p} = \frac{Cq_0}{p_0C}\frac{0H}{0G}.$$

The triangles $G0H$ and Cq_0H are similar. Hence $0H/0G = q_0H/Cq_0$. Thus,

$$e_p = \frac{Cq_0}{p_0C}\frac{0H}{0G} = \frac{Cq_0}{p_0C}\frac{q_0H}{Cq_0} = \frac{q_0H}{p_0C}.$$

The last ratio is equal to HC/CG since the triangles Cq_0H and Gp_0C are similar. This proves that $e_p = HC/CG$.

Economic Facts and Insights

- The quantity demanded of a product is dependent on own price, prices of related goods, income, taste and future price expectations.
- An increase in the own price tends to reduce the quantity demanded of a product.
- An increase in income may increase the demand for a product (this is the case of a normal good) or decrease the demand for a product (this is the case of an inferior good). Accordingly, the demand curve shifts to the right (or left) for a normal (or inferior) good as income increases.
- In reality, there are very few goods, which are inferior.

(continued)

- An increase in the price of a substitute good tends to increase the demand for a given good, that is, it shifts the demand curve of a given product to the right. An increase in the price of a complementary good tends to do the opposite.
- The Veblen effect, bandwagon effect, snob effect and network externality refer to situations where other people's demand affects the demand behaviour of an individual.
- Market demand curve is the horizontal summation of individual demand curves.
- The price elasticity of demand measures the responsiveness of quantity demanded of a product with respect to a change in the own price.
- The magnitude of price elasticity depends on whether a good is a luxury or a necessity; availability of substitute goods; how narrowly defined a product is; and the proportion of total expenditure on a product.
- With everything else the same, the greater the time period, the higher will be the price elasticity of a product as people tend to discover more substitutes with time.
- If the demand for a product is price elastic (or inelastic), an increase in it leads to a decrease (or an increase) in the total spending on the product.
- Income elasticity of demand measures the responsiveness of quantity demanded of a product with respect to a change in income. It is useful for a firm in assessing which markets to emphasise on in its overall marketing strategy.
- Cross price elasticity refers to the responsiveness of quantity demanded of a product with respect to a change in the price of a related good. It is positive in case of a change in the price of a substitute good and negative in case of a change in the price of a complementary good. Cross price elasticity measures are important to estimate the overall revenue implication of a change in the price of a product for a firm producing a product line.
- Advertising elasticity refers to the responsiveness of quantity demanded of a product with respect to a change in its units of advertising.

EXERCISES

2.1 What does a demand curve tell? What does it assume as given?

2.2 Rajnish stays in the college hostel. He used to get a stipend of Rs 5,000 per month from his father. Recently, his father has become more generous and gives him Rs 7,000. We observe that Rajnish is going to a lot more movies than before but eating less at the nearby *dhaba*. There are no changes in the movie ticket prices or that of the food at the *dhaba*. From this, can we say that both movies and food at the *dhaba* are normal goods for Rajnish?

2.3 There are train and bus services between Delhi and Agra. Suppose the train fares between the two cities come down. How will it affect the demand for bus travel between the two cities? Which determinant of demand is at work here?

2.4 Name two pairs of non-food items (besides those cited in the text) such that one is complementary in consumption to the other.

2.5 How would a speculation that petrol price is going to be doubled in two days affect your demand for petrol today and why?

2.6 It is announced that a new steel plant will come up in an area within a year's time. How will it affect the demand for land in that area now? Which determinant of demand is at work here?

2.7 Explain the Veblen effect in demanding a good.

2.8 Differentiate between the bandwagon effect and the snob effect.

2.9 What is network externality?

2.10 Explain the determinants of a market demand curve.

2.11 Because movie tickets have become more costly, Zorba goes out to see movies less often and hence is also eating out less in restaurants. Can we infer from this that there has been a change in demand by Zorba for movies but a change in quantity demanded for restaurant food?

2.12 Onions were selling at Rs 10 per kg in Nagpur and the total quantity sold of onion per month was 20 tons. Recently, onion price doubled and the quantity sold decreased to 14 tons. Calculate the arc price elasticity of demand for onions in Nagpur.

2.13 Petrol price has increased, but Mr Gupta's family is spending less on petrol. What can we infer about the elasticity of demand for petrol by this family?

2.14 Suppose there are two straight line demand curves, originating from the same point on the price axis. Show that, at any given price, the price elasticities of the two curves are equal.

2.15 If there are two straight line parallel demand curves, at any price, the price elasticity is greater along the line closer to the origin. Is this statement true or false? Justify.

2.16 A private hospital, Get Well, was charging Rs 2 lakh for a double-bypass heart surgery and it was generating Rs 36 lakh revenue from it per month. It has increased its rate to Rs 3.5 lakh and now it is generating Rs 32 lakh revenue from it. What can we say about the price elasticity of demand for double-bypass surgery at Get Well?

2.17 Mr Ganesh and his family members are fond of *rasmalai*. Recently, the price of *rasmalai* increased by 20 per cent and the total spending by Mr Ganesh's family on *rasmalai* increased by 15 per cent. What is the price elasticity of demand for *rasmalai* by this family?

2.18 'The income elasticity of demand for a product must always be positive.' Defend or refute.

2.19 'The cross price elasticity of demand for a product is always positive.' Defend or refute.

2.20 If the advertising elasticity is 2, what will be the percentage increase in sales from a 10 per cent increase in the units of advertising?

3 Supply

CONCEPTS

- Determinants of Supply
- Supply Curve
- Business Taxes
- Change in Quantity Supplied versus Change in Supply
- Price Elasticity of Supply
- Point Elasticity of Supply
- Supply Schedule
- Supply Function
- Hoarding
- Market Supply
- Arc Elasticity of Supply

While consumers or households demand goods and services, it is the producers or firms who supply them. In this chapter we analyse what factors determine the supply of a particular good to a market within a given time period.

As an example, suppose that you have a factory in Chennai that produces football. What factors determine how many footballs you produce and supply to the market in a span of, say, a year?

One is the market price of a football itself (assuming that all footballs are exactly of the same size and quality). All else the same, if the market price of football increases, you would tend to supply more.

Another factor is the cost of raw materials, labour and so on that go into the production of footballs. If the leather and other synthetic material with which a football is made become more expensive, it is a less profitable business and you would produce less.

Technology also matters. If you find a more efficient way of combining labour, machines and raw materials in producing football, you would produce more.

Still another factor is the price of substitute goods in production. Suppose your factory makes volleyballs as well, and to our pleasant surprise, India wins a silver medal in women's volley ball. Volleyball will now be in great demand among women and men in sport and the price of volleyballs will rise. How would this affect your supply of footballs? It would decline, as you would divert more of your productive resources from football to volleyball.

It is now time to state these factors of supply generally and consider their implications.

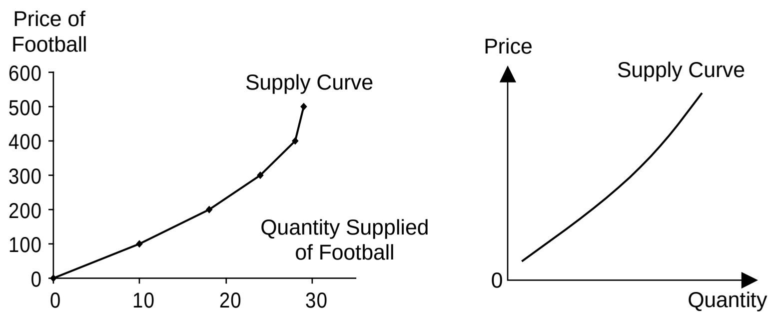
OWN PRICE AND THE SUPPLY CURVE

All other things remaining the same, the higher the price of a product, the more a firm will produce and supply to the market. This is the **law of supply**, analogous to the law of demand. Table 3.1 lists various quantities of football supplied at different prices. Such a table describes a supply schedule. Figure 3.1(a) graphs this table, measuring price and quantity along the y-axis and the x-axis respectively.

Table 3.1 A Supply Schedule

Price of Football (Rs)	*Quantity Supplied of Football (in '000)*
0	0
100	10
200	18
300	24
400	28
500	29

Figure 3.1 Supply Curve



(a) Supply Curve Based on Table 3.1

(b) A General, Hypothetical Supply Curve

The resulting curve is a **supply curve**, a graphical representation of the law of supply or a supply schedule. It is upward sloping or positively sloped because a higher quantity of supply is associated with a higher price. Figure 3.1(b) draws a smooth supply curve for analytical purposes.

SHIFTS OF A SUPPLY CURVE

Similar to the case of demand, factors other than the own price can cause a shift in a supply curve.

Input Prices

Consider Figure 3.2, in which S_0 is the original supply curve. At the price p_1, the quantity supplied is q_1; at p_2, the quantity supplied is q_2 and so on. Now suppose there is an increase in the price of an input or in the prices of several inputs that are used in making the product. This will raise costs. As this reduces profitability, at any given price, a firm will supply less. At p_1, the quantity supplied now is q_1', which is less than q_1, at p_2, it is q_2' and so on. The resulting supply curve is S', lying to the left of S_0. Hence, an increase (or a decrease) in input prices shifts a supply curve to the left (or right).

NUMERICAL EXAMPLE 3.1

A company produces steel *almirah*s of a given type, dimension and quality. The market price of steel rises by 20 per cent. Through a numerical example, illustrate how this would affect the company's supply curve of *almirah*s.

An increase in the price of steel reflects an input price increase. This will imply that at the same price for an *almirah*, the company will supply less. Table 3.2 offers

Figure 3.2 An Input Price Increase and the Supply Curve

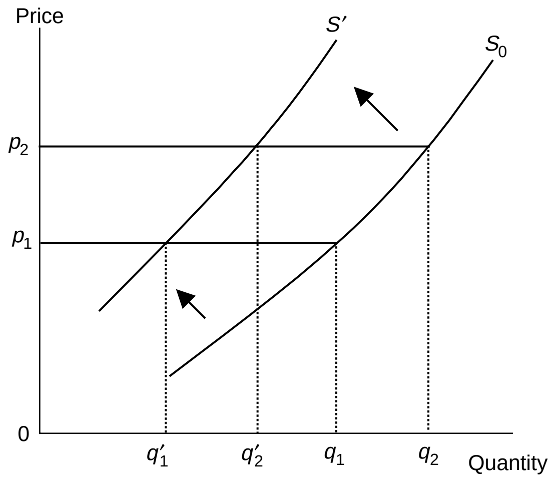


Table 3.2 Table for Numerical Example 3.1

Price of an Almirah (Rs)	*Quantity Supplied before the Steel Price Increase*	*Quantity Supplied after the Steel Price Increase*
2,000	100	90
2,200	110	99
2,400	115	107
2,600	122	114
2,700	128	120
3,000	132	123

an example. Column 2 enumerates the original quantities supplied at various prices. In column 3 each entry is less than the corresponding entry in column 2, meaning a decrease in the quantities at each price. Thus, in relation to column 2, column 3 illustrates the increase in the price of steel. We can verify that the graphical plot of column 3 against the price column lies to the left of that of column 2, representing a leftward shift of the supply curve as shown in Figure 3.2.

Technology

As discussed earlier, a technological improvement will induce a producer to supply more at any given price (the opposite of what happens when input prices increase). Thus it will shift a supply curve to the right, as illustrated in Figure 3.3. The old supply curve is S_0 and the new one after the technological improvement is S_1.

Figure 3.3 Technological Improvement and the Supply Curve

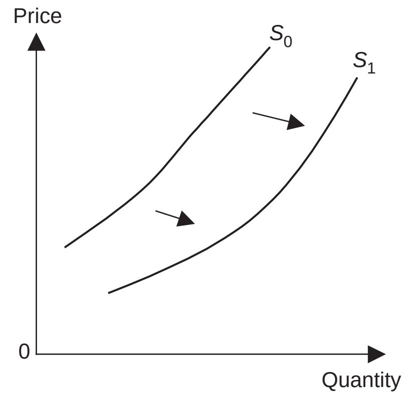
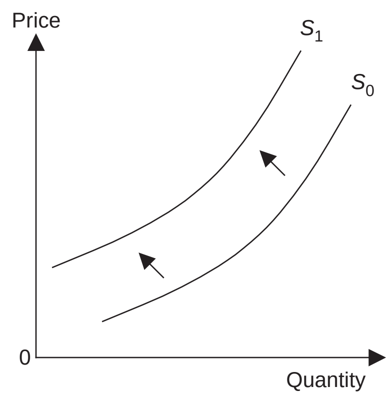


Figure 3.4 An Increase in the Price of a Substitute Good and the Supply Curve



Price of a Substitute Good in Production

In our football-volleyball example, an increase in the price of volleyball led to a decrease in the production of football. Generalising from this, we can say that an increase in the price of a substitute good in production will shift the supply curve of a given product to the left. This is illustrated in Figure 3.4. The original supply curve is S_0, whereas the new supply curve is S_1.

Business Taxes

Businesses not only incur the cost of production, but also pay a variety of taxes because of their production activity. For example, there are excise taxes (see Clip 3.1

Clip 3.1: Excise Taxes in India

Excise taxes are levied mostly by the central government for the sake of raising revenues. It varies from one sector to another. Some sectors may be exempt from this tax. For instance, in 2002–2003, producing kerosene carried an excise duty/tax of 16 per cent; for tea it was Re 1 per kilogram. In 2003–04, the excise duty rates on garments, passenger cars and computers stood at 10, 24 and 8 per cent respectively, down from 12, 32 and 16 per cent respectively in the previous financial year. State governments also impose excise duties but on a very limited number of products, mostly alcoholic.

for a brief discussion of these taxes). If a business buys inputs from abroad, it may have to pay a customs duty. If it is a service, then a service tax has to be paid.[1] Starting in the financial year 2005–06, many states use VAT (Value Added Tax).[2]

All these taxes that businesses pay can be understood as some tax on output and we can, in general, call them 'business taxes.' If such taxes increase, from a producer's point of view, it is like an increase in the price of an input. Thus, an increase (or a decrease) in business taxes shifts the supply curve to the left (or right).

Other Factors

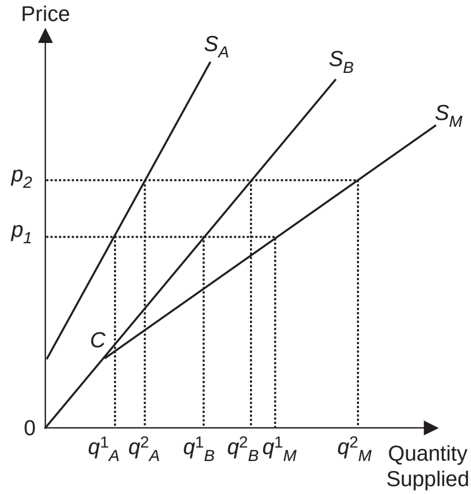
Other factors affect the supply curve of a product as well. For the production of an agricultural good, weather plays an important role. We all know how dependent our agricultural output is on the monsoons.

Another factor is price speculation which is same as future price expectation. If it is speculated that the price of a good will rise in the near future because of an external factor, suppliers may engage in product **hoarding** by not releasing the good to the market. They do so because they expect to make larger profits later by releasing the output when the market price is much higher. Hoarding behaviour tends to reduce the current supply of the product, that is, it tends to shift the current supply curve to the left. This happens typically in case of essential goods during emergencies like a war, flood or drought. In the same vein, if suppliers perceive that the price of the good is expected to fall in the future, they would tend to supply more to the market now. In general, an increase (or a decrease) in the expected future price would decrease (or increase) the current supply of a commodity to the market and, therefore, shift the supply curve to the left (or right).

[1]Service tax is a relatively new concept. It has been in existence since the mid-nineties only. Over time, the service tax rate has been increasing. In the financial year 2005–06, it was set at 10.2%.

[2]A VAT effectively means that businesses do not have to pay tax on inputs purchased from other sectors; taxes are paid only on the value added by the firm.

Figure 3.5 Market Supply Curve



MARKET SUPPLY CURVE

So far, we have discussed the supply curve of a single producer. For the entire industry we can talk about the market supply curve, analogous to the market demand curve. If there are 50 producers in a market, for example, the market supply curve is the horizontal summation of the supply curves of these 50 producers. What are the determinants of the market supply curve? The answer is: all factors that affect the individual supply curve plus the number of firms.

For illustration, Figure 3.5 assumes two firms in the market, A and B. Their respective supply curves are indicated as S_A and S_B. At the price p_1, firm A and firm B respectively supply the quantities, q_A^1 and q_B^1. The total quantity supplied in the market is equal to $q_A^1 + q_B^1 \equiv q_M^1$. Similarly, at the price p_2, the individual quantities supplied are q_A^2 and q_B^2. Hence the total quantity supplied equals $q_A^2 + q_B^2 \equiv q_M^2$. If we join the price-quantity combinations like (p_1, q_M^1), (p_2, q_M^2) and so on, we obtain the market supply curve, shown as the curve $0CS_M$.

CHANGE IN QUANTITY SUPPLIED VERSUS CHANGE IN SUPPLY

The distinction is similar to that between a change in quantity demanded and a change in demand. Change in quantity supplied refers to a movement along a given supply curve and this occurs because of a change in the own price. In contrast, a change in supply refers to a shift of an entire supply curve. For instance, a change in technology changes the supply.

Essentially these terms distinguish the effect of own price on the supply of a good from the effects of other determinants.

ELASTICITY

Parallel to demand, elasticity of supply measures the responsiveness of quantity supplied. The **price elasticity of supply** is defined by:

$$e_s = \frac{\text{\% change in quantity supplied}}{\text{\% change in own price}}. \tag{3.1}$$

This is the percentage formula. The numerator is equal to $(\Delta q_s/q_s) \times 100$, where Δ denotes the change and q_s denotes the quantity supplied. Similarly, the denominator equals $(\Delta p/p) \times 100$, where p denotes price. Hence, the percentage formula can be alternatively expressed as:

$$e_s = \frac{(\Delta q_s / q_s)\times 100}{(\Delta p / p)\times 100} = \frac{p}{q_s}\frac{\Delta q_s}{\Delta p}. \tag{3.2}$$

Arc Elasticity

For any change in price which is not very small, e_s is sensitive to an interchange between the original price and the new price and the corresponding interchange between the original quantity and the new quantity. This problem is avoided by using the **arc elasticity of supply** defined by:

$$e_s = \frac{p_0 + p_1}{q_0 + q_1}\frac{\Delta q_s}{\Delta p}, \tag{3.3}$$

where (p_0, q_0) is the old price-quantity pair and (p_1, q_1) the new one. This is analogous to the arc elasticity of demand.

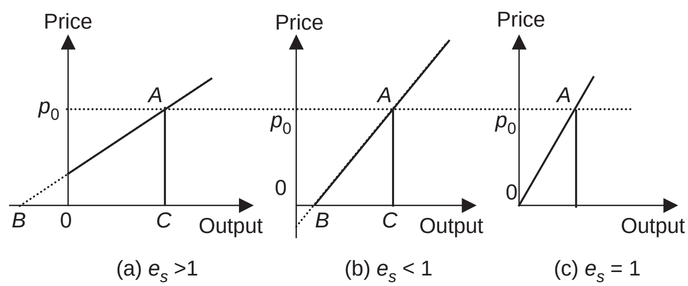
NUMERICAL EXAMPLE 3.2

A sugar manufacturer was supplying 10 quintals of sugar to the market when sugar was selling for Rs 10 per kg. The price of sugar has increased to Rs 14 per kg and the manufacturer is supplying 15 quintals to the market. What is the arc elasticity of supply?

The original price and quantity supplied are: $p_0 = 10$ and $q_0 = 10$. The new price and quantities are: $p_1 = 14$ and $q_1 = 15$. Thus, $\Delta p = 4$, $\Delta q_s = 5$, $p_0 + p_1 = 24$ and $q_0 + q_1 = 25$. Applying (3.3),

$$e_s = \frac{24}{25}\cdot\frac{5}{4} = 1.2.$$

Figure 3.6 Point Elasticity of Straight Line Supply Curve



Point Elasticity of Supply

For a very small change in price, there is also a **point elasticity** method. Suppose the supply curve is a straight line. Figure 3.6 shows three possibilities. The supply line intersects the price axis first, the quantity axis first or the origin as shown in panels (a), (b) and (c) respectively. As proven in Appendix 3A, in panels (a) and (b), the point elasticity is equal to $BC/0C$, the ratio of the horizontal segment to the quantity supplied. In panel (a) $BC > 0C$ and hence $BC/0C > 1$, that is, the point elasticity of supply is greater than one, or the supply is said to be elastic. In panel (b) it is the opposite: $BC/0C < 1$. The point elasticity is less than one or supply is inelastic.

Panel (c) can be thought of as a special case, where the point B coincides with the origin. Thus, $BC = 0C$ and $e_s = 1$. In other words, a straight line supply curve passing through the origin always has price elasticity equal to one (or 'unitarily elastic'), regardless of how flat or steep the supply curve is. For instance, in Figure 3.7, all the three supply curves, S_1, S_2 and S_3, have $e_s = 1$ at any price.

There are two special cases of straight-line supply curves: one where it is vertical and the other where it is horizontal. These are shown in Figure 3.8. A vertical supply curve has $e_s = 0$, since the quantity supplied is fixed. Typically, this holds when we look at the supply of a product at a given instant of time. For example, during a particular day, the supply of fish in the entire market of Kolkata is fixed. Assuming that fish cannot be stored, that amount of fish has to be sold in the market that day irrespective of how much price it fetches.[3] Along a horizontal supply curve, $e_s = \infty$, meaning that a little or a 'marginal' change in price will lead to an arbitrarily large change in supply.

If the supply curve is not a straight line, the point elasticity at a given price depends on the slope of or the tangent to the supply curve at that price. If this tangent intersects the price axis first, then, as in Figure 3.6(a), the elasticity exceeds

[3]But if we look at supply of a commodity over a period of time, it will be responsive to a change in price.

Figure 3.7 Unitarily Elastic Supply Curves

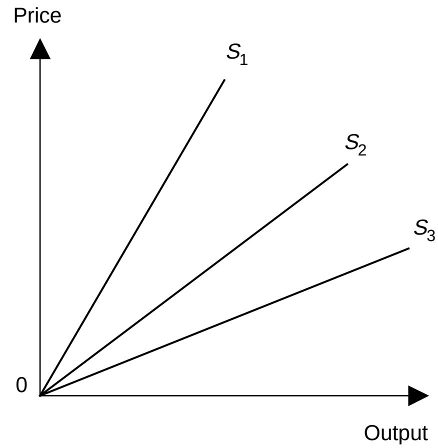
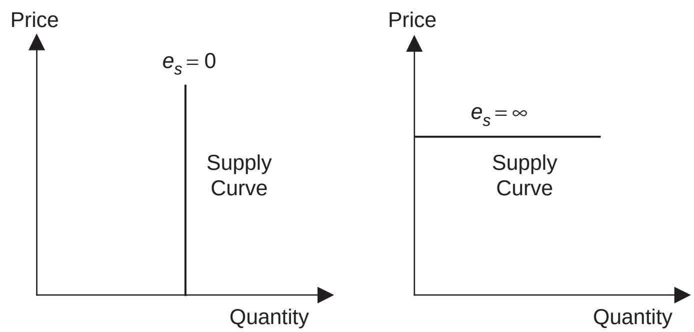


Figure 3.8 Special Cases



one. If it intersects the quantity axis first or the origin, it is less than or equal to one respectively. In Figure 3.9, for instance, at price p_0, the tangent intersects the quantity axis and, therefore, the price elasticity of supply is less than one.

How do we compare elasticities of two supply curves? Of course, if they are straight lines passing through the origin, we know that they will both have elasticity equal to one. We also know that if one (straight line) supply curve has the intercept on the price axis and the other on the quantity axis, the elasticity along the former is greater. But, barring these cases, we cannot compare in general except at their point of intersection. Turn to Figure 3.10. Just as in case of two intersecting demand curves, at the price p_0, the flatter supply curve S_B has higher elasticity than the steeper supply curve S_1 because the quantity adjustment due to a given price change is greater along the former. The reasoning is analogous to that

Figure 3.9 Point Elasticity of a General Supply Curve

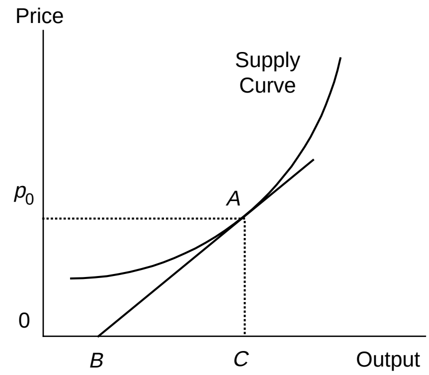
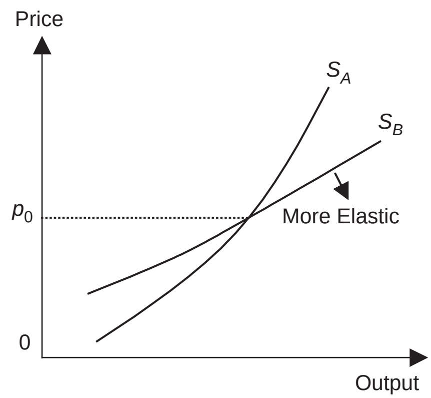


Figure 3.10 Comparing Supply Elasticities



behind the comparison shown in Figure 2.9 in Chapter 2. Thus, of two supply curves, the flatter supply curve is more elastic than the steeper supply curve at their point of intersection.

What about the elasticities of supply with respect to other factors like technology, input prices or taxes? In principle, one can define them. But there are several kinds of inputs and different types of technological improvements. Therefore, the standard practice is to not specify them as a general category. However, depending on what we wish to analyse, we can always define and discuss the relevant elasticity of supply. For instance, if we are considering producers who are manufacturing competing goods, we define a cross price elasticity of supply, for example, the elasticity of wheat production with respect to the price of sugar cane.

See Clip 3.2 for a sample of elasticities of supply.

Clip 3.2: Estimates of Elasticity of Supply

In general, compared to demand elasticities, supply elasticities are harder to estimate because of data availability problems and because firms operate in different types of market environments. Yet, some estimates are available and the following table presents a sample. Notice that it contains own price elasticities of supply as well as examples of elasticities with respect to input price changes.

Elasticity of the Supply of …	*With respect to the Price of …*	*Study*	*Estimate*
Rice	Own	Chand (1998–99)	0.25 (India)
Wheat	Own	Chand (1998–99)	0.85 (India)
Oysters	Own	Muth, Anderson, Karns, Murray and Domanico (2000)	1.97: halfshell oysters 2.30: shucked oysters 1.64: shellstock oysters (US)
Broiler Poultry	Own	Fabiosa, Jenses and Yan (2004)	0.22 (Indonesia)
Rice	Fertiliser (an input)	Chand (1998–99)	–0.16 (India)
Wheat	Fertiliser (an input)	Chand (1998–99)	–0.02 (India)

References

Chand, R. 1999. 'Supply Responsiveness in India's Rainfed Agriculture', in the *Annual Report 1998–99,* the National Centre for Agricultural Economics and Policy Research. New Delhi: Indian Council of Agricultural Research.

Fabiosa, J. F., H. H. Jensen and D. Yan. 2004. 'Output Supply and Input Demand System of Commercial and Backyard Poultry Producers in Indonesia'. Working Paper 04-WP 363, Center for Agriculture and Rural Development, Iowa State University.

Muth, M. K., D. W. Anderson, S. A. Karns, B. C. Murray and J. L. Domanico. 2000. *Economic Impacts of Requiring Post-Harvest Treatment of Oysters: Final Report*. Prepared for Interstate Shellfish Sanitation Conference, Research Triangle Institute, US.

Mathematically Speaking

The Supply Function

It is analogous to the mathematics of demand. Let q_A denote the output or the quantity supplied of good A, p_A its price and p_B the price of a substitute good in production. Suppose there are three inputs: labour, capital and raw material. Let their prices be respectively w_L, w_K and w_R. Finally, let C represent an index of the level of technology or efficiency of production. We can then write a **supply function** for good A as:

$$q_A = g_A(p_A, p_B, w_L, w_K, w_R, \mathrm{C}).$$

The law of supply implies that $\partial g_A/\partial p_A > 0$. Substitutability of production implies $\partial g_A/\partial p_B < 0$. An increase in the price of an input also has a negative effect on output. Hence $\partial g_A/\partial w_L < 0$, $\partial g_A/\partial w_K < 0$ and $\partial g_A/\partial w_R < 0$. A technology improvement induces a firm to produce more. Therefore, $\partial g_A/\partial C > 0$.

If we are concerned only with changes in the own price, we can suppress other variables and write the supply function as $q_A = g_A(p_A)$. A linear function like

$$q_A = F + vp_A,\ v > 0$$

is the simplest example.

Notice that the cases (a) $F < 0$, (b) $F > 0$ and (c) $F = 0$ correspond respectively to panels (a), (b) and (c) of Figure 3.6.

Finally, if there are N firms in the industry (indexed by 1, 2, …, N) with supply functions $g_A^1(p_A)$, $g_A^2(p_A)$,….$g_A^N(p_A)$ respectively, the market supply function can be stated as

$$G_A(p_A) \equiv g_A^1(p_A) + g_A^2(p_A) + \ldots + g_A^N(p_A).$$

* * * * * *

APPENDIX 3A: POINT ELASTICITY FORMULAS

Consider Figure 3.6. For a small price change from p_0,

$$e_s = \frac{p_0}{0C}\frac{\Delta q_s}{\Delta p}\bigg|_{\Delta p \to 0} = \frac{AC}{0C}\frac{dq_s}{dp}.$$

In panels (a) and (b), $dp/dq_s = \text{rise}/\text{run} = AC/BC$, implying

$$\frac{dq_s}{dp} = \frac{BC}{AC}.$$

Substituting this expression,

$$e_s = \frac{AC}{0C}\frac{dq_s}{dp} = \frac{AC}{0C}\frac{BC}{AC} = \frac{BC}{0C}.$$

In panel (c), $dp/dq_s = AC/0C$, or $dq_s/dp = 0C/AC$. Hence,

$$e_s = \frac{AC}{0C}\frac{dq_s}{dp} = \frac{AC}{0C}\frac{0C}{AC} = 1.$$

Economic Facts and Insights

- The quantity supplied of a product is dependent on own price, prices of related goods in production, technology, input prices, business taxes and future price expectations.
- An increase in the own price tends to increase the quantity supplied of a product.
- An increase in the price of a related good in production decreases the supply the product in question.
- A technology improvement shifts the supply curve to the right, while an increase in an input price shifts the supply curve of a product to the left.
- Business taxes like excise duties, VAT and service tax can be seen as a part of the overall costs of production. Thus, an increase in the business taxes shifts the supply curve to the left.
- An increase in the expected future price of a product tends to reduce its current supply.
- Market supply curve is the horizontal summation of the individual supply curves.
- The price elasticity of supply measures the responsiveness of quantity supplied to a change in the own price.

EXERCISES

3.1 A furniture manufacturer in Mumbai faces an increase in the price of teak wood. How will it affect the supply curve of furniture in the market?

3.2 Our furniture manufacturer has found a new mechanised way of polishing furniture by which costs are reduced. Will this result in a change in the quantity supplied or a change in supply by the furniture manufacturer? Give reasons.

3.3 A farmer has some land, a part of which is used for growing wheat and the remaining for corn. Starting with a given supply curve of corn, if wheat begins to fetch a higher price in the market, how will the farmer's supply curve of corn be affected and why?

3.4 Suppose the government reduces the excise duty imposed on a particular sector. How will this affect the supply curve?

3.5 A massive flood cuts off supplies of essentials like edible oil and onions to a particular region. All else the same, how would this affect the current supply of edible oil by local sellers and why?

3.6 How will entries of foreign firms into a particular sector shift the individual supply curve of local firms as well as the market supply curve?

3.7 A local fisherman catches fish from a nearby river early in the morning and sells all his catch by the same day (he does not have any storing facility). Consider his daily supply curve of fish to the market. Is it upward sloping, that is, will quantity supplied increase if there is an increase in the market price of fish?

3.8 'An increase in the market demand curve will shift the market supply curve as well.' Defend or refute.

3.9 A pencil manufacturer was supplying 400 dozen pencils when the price was Rs 10. Now the market price has risen to Rs 14 and the manufacturer is supplying 420 dozen of pencils to the market. Calculate the arc elasticity of supply.

3.10 'Among two straight line supply curves, the steeper one is less price elastic.' Comment.

3.11 A ladies-clothing manufacturer produces *sari*s and *shalwars*. How will an increase in the average price of *sari*s affect the supply curve of *shalwar*s?

3.12 In recent years there is an increase in the outsourcing of jobs, especially in the IT (Information Technology) sector, from developed countries like the US to developing countries like India and China. The companies that do this are called BPO companies, where BPO is the short form of 'Business Process Outsourcing.' Foreign companies engage in outsourcing because it lowers the overall cost of goods and services that they provide. How would an increase in such outsourcing affect the supply curve of the goods and services they produce?

4 Consumer Behaviour

CONCEPTS

- Total Utility
- Marginal Utility
- Marginal Utility of Money
- Law of Diminishing Marginal Utility
- Law of Equi-proportional Utility
- Indifference Curve
- Indifference Map
- Non-satiation
- Marginal Rate of Substitution
- Law of Diminishing Marginal Rate of Substitution
- Price or Budget Line
- Relative Price
- Consumer's Equilibrium Condition
- Utility Function
- Income Effect
- Normal or Superior Good
- Inferior Good
- Income Consumption Curve
- Engel Curve
- Substitution Effect (Hicks')
- Substitution Effect (Slutsky's)
- Price Effect
- Giffen Good
- Price Consumption Curve
- Composite Good Theorem
- Revealed Preference
- Weak Axiom of Revealed Preference

In Chapters 2 and 3 we studied about demand and supply. In the current chapter, we analyse the *basis* of demand. For example, why would a consumer buy less of a good when its price rises? You may think at this moment that it is a silly question because it is obvious that when the price of something rises one would buy less of it. But you will soon realise that there is a non-obvious answer for it. It depends on 'consumer behaviour'—more particularly, on certain characteristics of a consumer's preferences as well as her budgetary constraints.

There are three approaches to consumer behaviour or what is also called **consumer theory** or **demand theory**. These are: the marginal utility approach, the indifference curve approach and the revealed preference approach. In what follows, we will study them in turn.

MARGINAL UTILITY APPROACH

This is the simplest of all the approaches and was popularised by Alfred Marshall who is regarded as the father of microeconomics. It involves some concepts of utility.

Concepts

An individual buys something because consuming it provides her some pleasure or satisfaction. This notion is captured by **total utility,** defined as the total satisfaction—measured in some psychological unit—obtained from consuming a given amount of a good. Let us name this unit as 'util.' For example, if you consume 10 chocolates then the marginal utility theory assumes that you are able to quantify your satisfaction from it, such as 100 utils, 60 utils or whatever.

In comparison to total utility, **marginal utility** is defined as the additional or the extra utility obtained from consuming an *extra* unit of a product. To continue our example, suppose that 10 chocolates give you 70 utils of total utility and 11 chocolates give you 80 utils of total utility. We then say that the marginal utility of 11 chocolates is equal to 80 – 70 = 10 utils.

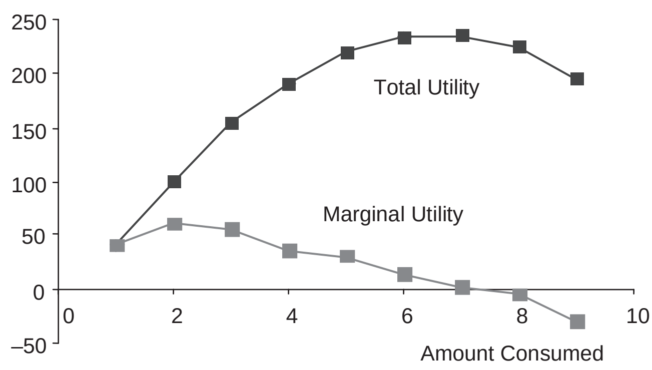
Suppose you are an old Hindi movie buff. You regularly watch them by renting CDs and Table 4.1 represents your total utility (TU) and marginal utility (MU) schedules from seeing these movies. Notice that, at each level of consumption, marginal utility is equal to the extra utility obtained from the last unit consumed. For instance, the MU of 3 movies is equal to 55 utils, because the third movie gives extra utility equal to 155 – 100 = 55 utils. Also, mark that if you consume too many movies, your marginal utility becomes negative, that is, after a certain point, seeing one more movie becomes a drag rather than an enjoyment.

While the MU schedule, by its definition, can be derived from a TU schedule, the TU schedule can also be derived from an MU schedule. Imagine that only the third column of Table 4.1 is given to you, but not the second. You can actually derive the second column easily from the third by adding up marginal utilities. For example, the total utility of 4 movies = the MU of one movie (40) + the MU of

Table 4.1 Total Utility

No. of Old Hindi Movies	*Total Utility*	*Marginal Utility*
0	0	–
1	40	40
2	100	60
3	155	55
4	190	35
5	220	30
6	233	13
7	235	2
8	225	–5
9	195	–30

Figure 4.1 Total Utility and Marginal Utility Curves



2 movies (60) + the MU of 3 movies (55) + the MU of 4 movies (35) = 190 utils. In other words, total utility is the sum of marginal utilities. Table 4.1 is graphed in Figure 4.1, showing the total utility curve and the marginal utility curve.

Notice an important pattern in the marginal utility column or curve—after the second unit of consumption, it continuously falls. This phenomenon is called the **law of diminishing marginal utility**, which states that after a certain level of consumption (possibly from the very first unit), the marginal utility falls as more of a product is consumed. The law says something natural—no matter how much you like something, after consuming a certain amount, the intensity of desire for it gradually falls as you consume more and more of it. As we shall see, this law implies why the demand curve is downward sloping, that is, why as price increases, the quantity demanded of a good falls.

Going back to our example, suppose that an old Hindi movie CD can be rented for Rs 15 (assume that you do not have cable connection in your TV and renting CDs is the only way for you to see these movies). The question is, how many movies are you going to consume or demand?

Table 4.2 Utilities in Terms of Rupees

No. of Old Hindi Movies	*Total Utility Rs*	*Marginal Utility Rs*
0	0	–
1	20	20
2	50	30
3	77.5	27.5
4	95	17.5
5	110	15
6	116.5	6.5
7	117.5	1
8	112.5	–2.5
9	97.5	–15

Marginal Utility of Re 1 = 2 utils

From Table 4.1 or Figure 4.1, it is obvious that you are not going to rent beyond 7 movies; why would you get movies when they give you negative satisfaction? However, below 7, there is a trade-off between benefits (utility) and costs. On the one hand, if you rent and see more movies, your total utility or benefit increases. On the other hand, your cost increases as you have to pay more for them in terms of opportunity cost—paying more on movies implies sacrificing other goods and services you could have bought with the money you spend on movies. Therefore, the benefits and costs must be compared.

In order to do this, the utility must be expressed in terms of money. Assume that, for you, one rupee is worth 2 utils. Then the TUs and the MUs in Table 4.1 can be converted in terms of money or rupees. Table 4.2 presents them.

NUMERICAL EXAMPLE 4.1

A person's marginal utility from consuming bananas is given in Table 4.3. Derive her total utility schedule. If the marginal utility of a rupee is 24 utils, derive the marginal utility schedule in terms of money.

Since total utility is the sum of marginal utilities, the total utility of two bananas is equal to the marginal utility of one banana (39) plus that of two bananas (36) = 75 utils. Table 4.4 calculates the total utilities for various units consumed.

By definition, the marginal utility (in terms of money) is equal to the marginal utility of the good divided by the marginal utility of rupee. Thus, we simply divide column 2 of Table 4.3 by 24. The results are given in Table 4.5.

Decision Rule

We are now in a position to answer how many movies you should rent. Refer to Table 4.2. Consider renting the first movie—compared to no movie, it gives extra utility equal to Rs 20 in terms of money. But you pay only Rs 15. Hence you should buy at least one movie. Similarly, the second movie gives utility worth Rs 30, while, again, you have to pay only Rs 15 for it. Thus the second movie is worth getting.

Table 4.3 Marginal Utility (Numerical Example 4.1)

Number of Bananas	*Marginal Utility*
0	–
1	39
2	36
3	33
4	24
5	12
6	0
7	–9

Table 4.4 Total Utility (Numerical Example 4.1)

Number of Bananas	*Total Utility*
0	–
1	39
2	39 + 36 = 75
3	75 + 33 = 108
4	108 + 24 = 132
5	132 + 12 = 144
6	144 + 0 = 144
7	144 – 9 = 135

Table 4.5 Marginal Utility in Terms of Money (Numerical Example 4.1)

Number of Bananas	*Marginal Utility in Terms of Money*
0	–
1	39/24 = 1.625
2	36/24 = 1.5
3	33/24 = 1.375
4	24/24 = 1
5	12/24 = 0.5
6	0/24 = 0
7	–9/24 = –0.375

Extending this logic, it is clear that you should buy at least 4 movies. Should you buy the fifth? This situation is a bit different. The extra gain in utility in terms of money (Rs 15) exactly equals how much you pay for it. Thus the answer is: you may buy the fifth unit or may not; it does not matter. But should you buy the sixth unit? Clearly no, because the extra utility gain is worth Rs 6.5 only, less than the rental price of a CD. Hence the answer is: you should rent either 4 or 5 movies. Note in particular that when you rent 5 units:

$$\text{Marginal Utility of the Product in terms of money} = \text{Its Price} \qquad (4.1)$$

Also realise that, if the units in which the good is consumed can be measured continuously as points on a straight line, there will be little difference between the fourth unit and fifth unit.

Thus, (4.1) is the general rule or the general principle of maximising the surplus of total utility over how much you spend on a particular good. It is the optimality rule for purchasing or demanding a good.

The condition (4.1) can be expressed in two alternative ways. By definition, the MU of a product in terms of money equals MU of the product divided by the marginal utility of money (rupee). Thus (4.1) can be rewritten as:

$$\frac{\text{Marginal Utility of a Product}}{\text{Marginal Utility of a Rupee}} = \text{Its Price.} \tag{4.2}$$

We can rewrite the above equation as:

$$\frac{\text{Marginal Utility of a Product}}{\text{Its Price}} = \text{Marginal Utility of a Rupee.} \tag{4.3}$$

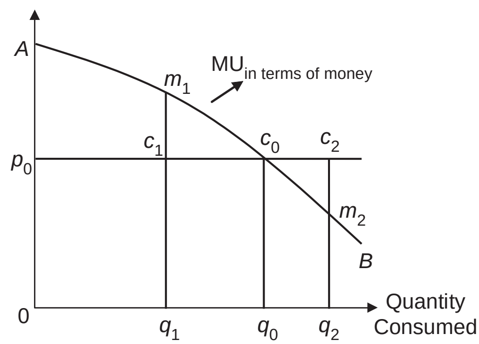
Any of the conditions (4.1), (4.2) or (4.3) is called the **consumer's equilibrium** condition. 'Equilibrium,' a commonly used term in economics, means a position of rest. In the present context, the consumer should 'rest' with regard to her purchase of a good as soon as the condition (4.1) is met. This is sometimes loosely called the 'marginal utility = price' principle, implicitly assuming that the marginal utility of a rupee is equal to one.

It is important to understand that the law of diminishing marginal utility underlies the condition (4.1). How? Compared to the level of purchase or demand at which (4.1) is satisfied, if you have more, the law implies that the extra benefit will be less than the extra cost (the price) and so the surplus cannot be increased. Similarly, if you buy less, the law implies that the marginal utility will be higher—the benefit foregone will exceed what is saved in terms of the cost (the price) and hence the surplus cannot be increased. Therefore, the surplus must be maximised at that point where the 'marginal utility = price' principle is met.

This principle is illustrated in Figure 4.2. Assume that the commodity can be measured on a continuous scale and the law of diminishing marginal utility holds right from the first unit of consumption. The curve AB denotes the marginal utility (in terms of money) curve. Since MU is the addition to TU, the area under this curve equals the total utility in terms of money. Thus at the consumption level q_1, the total utility in terms of money is equal to the area $0Am_1q_1$. At q_2, it is $0Am_2q_2$. At q_0, it is $0Ac_0q_0$ and so on. Let p_0 be the price of the product. The area under the p_0 line measures the total cost or the total expenditure on the good. If at this price, the amount q_1 is bought, the total expenditure = price × quantity = the area $0p_0c_1q_1$. Similarly at q_0, it is $0p_0c_0q_0$ and so on. Thus the area under the MU (in terms of curve), which lies above the p_0 line measures the surplus.

Where is the surplus maximised? The answer is at q_0, where the 'marginal utility = price' rule holds. The surplus at q_0 equals $0Ac_0q_0 - 0p_0c_0q_0 = Ap_0c_0$. Note that at any point to the left of this level such as q_1, the surplus = $0Am_1q_1$ − $0p_0c_1q_1$ =

Figure 4.2 Optimal Purchase Rule



$Ap_0c_1m_1 = Ap_0c_0 - m_1c_1c_0 < Ap_0c_0$. At q_2, which is to the right of q_0, similar calculation yields that the surplus = $Ap_0c_0 - m_2c_2c_0 < Ap_0c_0$. This proves that the surplus is maximised at q_0.

NUMERICAL EXAMPLE 4.2

Consider the figures given in the Numerical Example 4.1. Suppose bananas sell for Re 1 a piece. How many bananas will the consumer buy?

Compare the marginal utilities (in terms of money) given in Table 4.5 to the price of bananas. Condition (4.1) is met at 4 units of consumption. Thus the consumer will buy 4 units.[1]

Why is the Demand Curve Downward Sloping?

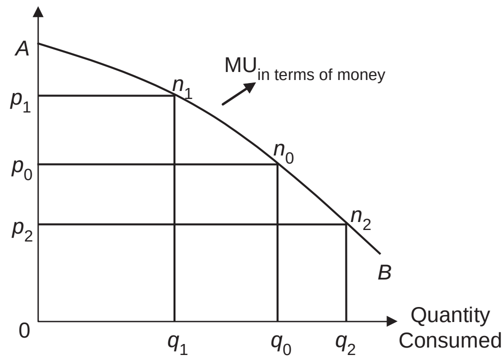
We can now answer why the demand curve is downward sloping. Turn to Figure 4.3: *AB* is the marginal utility (in terms of the money) curve. The 'marginal utility = price' rule implies that at the price p_0, the consumer will buy the quantity q_0. At price p_1, she will demand q_1; at price p_2, the quantity demanded by her will be q_2; and so on. This means that the points like c_0, c_1 and c_2 must lie on the demand curve. Hence the marginal utility curve itself is the demand curve. It then follows that the demand curve is downward sloping due to the law of diminishing marginal utility. Simply put, diminishing marginal utility implies that as the consumer buys more of a good, she is willing to pay less and that is why she buys more when the price falls. This is the essence of the marginal utility theory.

Many Goods

In reality, a consumer buys many goods and not just one. In this case, the principle of maximising the surplus is a simple extension of (4.3), that is, the consumer

[1]Another answer is 3 units, as bananas are not being measured on a continuous scale.

Figure 4.3 The Demand Curve



should allocate his spending so that the marginal utility of a rupee is same across all goods.

Algebraically, let different goods be named 1, 2, 3 and so on; their respective marginal utilities be equal to MU_1, MU_2, MU_3 and so on; and their respective prices be equal to p_1, p_2, p_3 and so on. From (4.3), the marginal utility of a rupee spent on good 1 equals MU_1/p_1, that on good 2 is equal to MU_2/p_2 and so on. Thus (4.3) extends to:

$$\frac{MU_1}{p_1} = \frac{MU_2}{p_2} = \frac{MU_3}{p_3} = ... \tag{4.4}$$

This goes by the name of the **law of equi-proportional utility**.

The economic reason is straightforward. If, for instance, one rupee brings more marginal utility from consuming, say, good A than from consuming good B, the consumer's total utility will be increased if more rupees are spent on good A and less on good B. This process should continue till there is no difference between the marginal utility of a rupee spent on different goods.

NUMERICAL EXAMPLE 4.3

Among many things, Sangeeta buys *alu bhujia* packets at Rs 10 a packet and *Amar Chitra Katha* picture books at Rs 15. Her total utility schedules with respect to these two goods are given in Table 4.6. The marginal utility of a rupee is worth 4 utils for her. She has Rs 85 to spend on these two goods. How many *alu bhujia* packets and *Amar Chitra Katha* books should she buy?

First we derive the marginal utility schedule of each good. These are additions to the total utility. Table 4.7 lists them in columns 2 and 4. Next we divide the marginal utilities by respective prices to obtain their ratio; these are given along columns (3) and (5). We now compare these two columns and see where the condition (4.4) is satisfied. We see that it is met at 4 packets of *alu bhujia* and 3 *Amar*

Table 4.6 Total Utility (Numerical Example 4.3)

Units Consumed	*Total Utility from* **Alu Bhujia**	*Total Utility from* **Amar Chitra Katha** *Books*
0	0	0
1	60	81
2	116	153
3	163	213
4	203	267
5	239	312
7	271	351
8	300	381

Table 4.7 Marginal Utilities (Numerical Example 4.3)

Units Consumed (1)	*Marginal Utility of* **Alu Bhujia** *(2)*	*Marginal Utility ÷ Price:* **Alu Bhujia** *(3)*	*Marginal Utility of* **Amar Chitra Katha** *Books (4)*	*Marginal Utility ÷ Price:* **Amar Chitra Katha** *Books (5)*
0	–	–	–	–
1	60	6.0	81	5.4
2	56	5.6	72	4.8
3	47	4.7	60	**4.0**
4	40	**4.0**	54	3.6
5	36	3.6	45	3.0
7	32	3.2	39	2.6
8	29	2.9	30	2.0

Chitra Katha books. The cost of this bundle is equal to Rs (4 × 10 + 3 × 15) = Rs 85, which Sangeeta can just afford. Thus the answer is that she will buy 4 packets of *alu bhujia* and 3 *Amar Chitra Katha* books.

Although we have got our answer, it is useful to understand the *process* of purchasing these goods sequentially. To start with, Sangeeta has no *alu bhujia* or *Amar Chitra Katha* book. Should she first buy an *alu bhujia* packet or an *Amar Chitra Katha* book? Looking at column 3 of Table 4.7, if she buys an *alu bhujia* packet, she gets pleasure equal to 6 utils per rupee. Instead, if she buys an *Amar Chitra Katha* book, she gets a pleasure equal to 5.4 utils per rupee (refer to column 5 of Table 4.7). So money is 'well spent' on *alu bhujia*, that is, she first buys one packet of *alu bhujia*. Her next choice is between buying the second packet of *alu bhujia* and the first *Amar Chitra Katha* book. These two choices offer 5.6 utils per rupee and 5.4 utils per rupee respectively. So Sangeeta will go for her second packet of *alu bhujia*. Consider her next choice, which is between the third *alu bhujia* packet (yielding 4.8 utils per rupee) and the first *Amar Chitra Katha* book (yielding 5.4 utils per rupee). Thus she will go for her first *Amar Chitra Katha* book, which offers higher marginal utility per rupee. Extending this logic, we see that the next two purchases will be an *Amar Chitra Katha* book and a packet of *alu bhujia*. So far she has bought 3 *alu bhujia* packets and 2 *Amar Chitra Katha* books. She has spent Rs ($3 \times 10 + 2 \times 15$) = Rs 60 and is left with Rs 25. In the next purchase, both *alu bhujia* and *Amar Chitra Katha* book offer utility equal to 4 utils per rupee. She buys both and spends the remaining Rs 25. Her total purchases are thus 4 packets of *alu bhujia* and 3 *Amar Chitra Katha* books.

A Critique

The marginal utility theory, or what is sometimes called the Marshallian theory, has some obvious limitations. First, it assumes that the marginal utility of any particular good depends on the amount consumed of that good only. In reality, however, liking or disliking a particular good may depend on how much of some other good is consumed. For example, how much satisfaction you get from buying an extra litre of petrol for your motor bike may depend, in part, on how many times you visit movie theatres near your college. A family's utility from consuming on extra kilogram of rice may depend on how much wheat it consumes. Second, in explaining how a consumer responds to a price change, it assumes that the marginal utility of money is constant. This is not realistic; a change in price would, in general, change a consumer's real income, which, in turn, would change the marginal utility (value) of money. Finally, the marginal utility theory assumes that the consumer is able to quantify, in absolute terms, the satisfaction obtained from consuming any given amount of a good. In more formal terms, we say that the consumer is able to measure utility *cardinally*. But this is unrealistic, since satisfaction or pleasure is psychological and not something that is observable.

Despite these weaknesses, the Marshallian theory is useful in many ways. First, the basic idea of utility from consuming or experiencing something is used in understanding various economic phenomena. Second, it clearly and simply illustrates benefits and costs when you want to demand something but have to pay for it. Third, note that what is critical in understanding demand behaviour is not total utility per se but the marginal utility. The 'marginal utility = price' rule essentially reflects **marginalism**, that is, weighing marginal benefit against marginal cost, which is critical in many kinds of economic decision-making and not just in consumer behaviour.

INDIFFERENCE CURVE APPROACH

This is the modern approach to consumer theory. It precisely rectifies the three weaknesses of the Marshallian theory outlined in the earlier section. It takes into account a consumer's preference for various goods consumed together. It also considers how the quantities demanded for various goods are interrelated because of the budgetary constraints facing the consumer. For instance, if she has Rs 110 as her income and there are two goods available, say, *kulfi* at Rs 15 a piece and Kit-Kat chocolate bar at Rs 12 a piece, then this consumer can afford 3 *kulfis* and 5 chocolates but not 5 *kulfis* and 3 chocolates. In short, how much of different goods a consumer buys, depends on her preferences and her budget. In other words, the indifference curve approach looks at the problem of simultaneously choosing optimal amounts of several goods.

Furthermore, the indifference curve approach does not assume that the marginal utility of money is constant.

Finally, it is not necessary to measure satisfaction or utility cardinally. All that is needed is that the consumer should be able to compare utility obtained from any two bundles of goods, without having to assign a number for the absolute amount of satisfaction obtained from consuming a particular bundle. In formal terms, in the indifference curve approach, utility is measured *ordinally*.

We first analyse consumer preference, next the budget and then her decision making regarding which combination of goods to buy.

Indifference Curve

Suppose you consume only two goods: *kulfi* and chocolate. Start with any combination of these goods which you (the consumer) may want to consume without any consideration for whether you can afford it or not. One possibility is, say, zero *kulfi* along with 40 chocolates. In short, write this bundle as (0, 40). Consider another bundle with 1 *kulfi* and some chocolates. Ask yourself how many chocolates along with 1 *kulfi* will give you the same satisfaction as the bundle (0, 40)?

Assume that you always prefer more of a commodity to less, which, formally, is called the assumption of **non-satiation**, or greediness.[2] This assumption helps in answering the question posed above. Since you are getting more *kulfi* than before (1 instead of 0), you will be willing to give up some chocolate so as to enjoy the same level of satisfaction as before. Suppose it is 30 chocolates for you, that is, you are 'indifferent' between the bundles, (0, 40) and (1, 30). Similarly, suppose that, if you consume 2 *kulfis*, you need 22 chocolates to make you indifferent between this bundle and the previous two. Table 4.8 lists various bundles, giving you equal satisfaction, that is, you are indifferent between any pair of bundles in the list.

Table 4.8 Bundles between which the Consumer is Indifferent

Kulfi	*Chocolates*
0	40
1	30
2	22
3	15
4	9
5	5
6	2
7	0

[2]It amounts to saying that the marginal utility is positive.

Figure 4.4 Indifference Curve

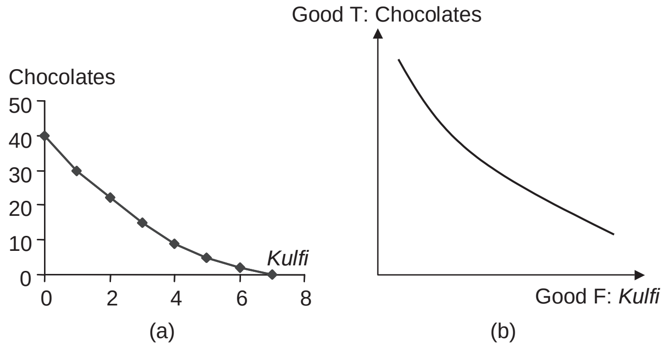


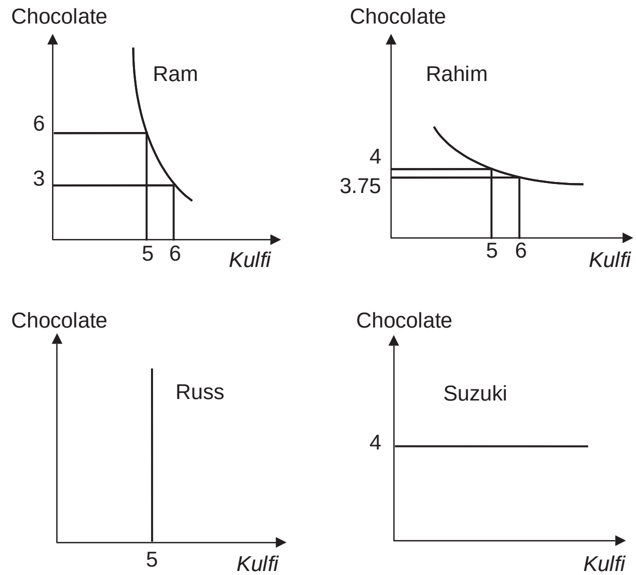
Figure 4.4(a) graphs these bundles and joins them. The resulting curve is an **indifference curve**, generally defined as a locus of commodity bundles along which a consumer is indifferent. If goods are measured on a continuous scale as points on a real line and not just as integers 1, 2, 3 and so on, this curve will look very smooth, as drawn in Figure 4.4(b). In what follows, we will, on many occasions, use the abbreviation 'IC' for indifference curve.

Individual tastes can be different. Your tastes may be similar to your best friend's but not the same. Your brother's tastes may be very different from yours. Figure 4.5 shows ICs, showing the different tastes of four boys. First, look at those of Ram and Rahim. These are downward sloping like Figure 4.4(b). But, the difference is that Rahim's is much flatter than Ram's. What does this mean? It implies that Ram has a strong preference, in some sense, for *kulfi* and Rahim has a strong preference for chocolate. How? Look at Ram's IC. He is indifferent between the bundles (5 *kulfi*s, 6 chocolates) and (6 *kulfi*s, 3 chocolates). Compared to the first bundle, if he wants one extra *kulfi*, he is willing to give up 3 chocolates. Likewise, notice that if Rahim wants one extra *kulfi*, he is only willing to give up 0.25 chocolates. That is, Ram is willing to give up more chocolates than Rahim, which is same as saying that Ram has a weaker preference for chocolate and Rahim has a stronger preference for the same. Put differently, Ram has a strong preference for *kulfi*, while Rahim has a strong liking for chocolate. The point is that steepness or flatness of indifference curves indicates the relative strength of preference towards goods.

Now look at Russ's IC. It is vertical, that is, its steepness is absolute. Hence, Russ is an extreme case of Ram. His satisfaction depends on only on how much *kulfi* he consumes. He is willing to give up any amount of chocolate to get one more *kulfi*. This is why his indifference curve is vertical. Similarly, Suzuki is an extreme case of Rahim. His satisfaction depends only on chocolate and he is ready to give up any amount of *kulfi* for an extra bit of chocolate.

From now on, however, we will analyse downward sloping indifference curves like Ram's or Rahim's and will ignore the special cases like Russ's or Suzuki's.

Figure 4.5 Various Indifference Curves



Indifference Map

Remember that in attempting to introduce an indifference curve, we started with an arbitrary bundle (0, 40). Had we started with a different one, say (2, 40), we would have obtained a different indifference curve—showing bundles which offer the same utility as does (2, 40). A consumer's taste is represented by not a single indifference curve but many indifference curves. A collection of indifference curves is called an **indifference map**. Each indifference curve in an indifference map stands for some level of satisfaction or utility relative to some other level of satisfaction. Figure 4.6 shows a typical indifference map.

Nowhere in the indifference curve analysis is the consumer assumed to tell or 'assign' one particular number to the level of satisfaction obtained from consuming a bundle. She is only required to compare or rank utility between any pair of bundles. If both yield the same level of satisfaction, then both lie on the same indifference curve; otherwise, one bundle lies on a higher indifference curve compared to the other. In this sense, utility is measured 'ordinally', not cardinally.

SPECIAL CASES

While an *IC* is normally a nice looking, smooth curve as in Figure 4.4(b) or Figure 4.6, it can look also like the ones drawn in Figure 4.7. In panel (a), they are rectangular. This is the case where the consumer always consumes the two goods in a given proportion. For instance, if there are only two goods, say tea and sugar,

Figure 4.6 An Indifference Map

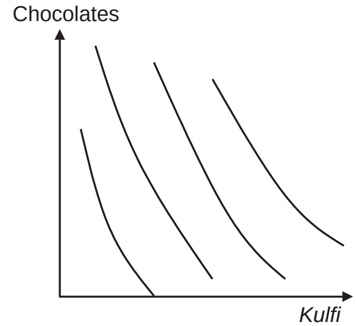
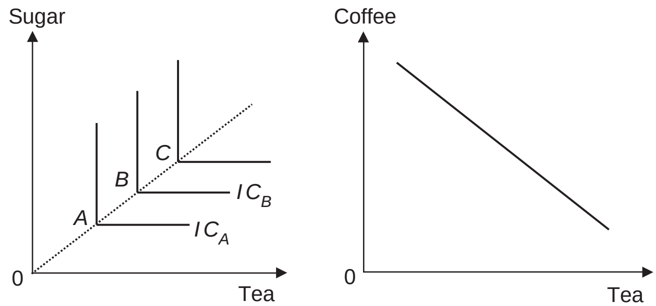


Figure 4.7 Special Cases



(a) Perfect Complements (b) Perfect Substitutes

and you are extremely particular about the proportion of sugar to how much tea you drink, then with the amount of tea being given, an increase in sugar per se will not add to your utility, or with the amount of sugar being given, an increase in tea per se will not make you happier. In panel (a), this proportion is given by the slope of the ray *OC*. For example, if you are initially consuming at point *A* and given more tea (but no more or no less sugar), you get the same amount of satisfaction as at *A*. This explains the horizontal segment of the indifference curve IC_A. Similarly, if you are given more sugar but there is no change in tea, you are as well off as at *A*—this explains the vertical part of the same *IC*. Of course, if you get more sugar and more tea in the same proportion, you are happier. For instance, IC_B with more tea and more sugar at *B* than at *A* represents higher utility than IC_A. Panel (a) depicts the case where the two goods are **perfect complements** of each other.[3]

Now consider Figure 4.7(b), where the indifference curves are straight lines. Since the slope of a straight line is fixed, it means that a fixed amount of one good

[3]Such indifference curves represent zero substitutability in consumption between the two goods.

is given up for acquiring one extra unit of the other so that the consumer's utility is unchanged. Thus the two goods are **perfect substitutes** of each other. It is as if one unit of one good is being sacrificed in order to have one more unit of the other good, when units are defined appropriately. Put differently, the two goods are virtually the same good in the perception or taste of the consumer. It can happen for products like tea and coffee or, more realistically with, say, white coloured pens and pink coloured pens of exactly the same quality.

However, from now onwards, like Russ's and Suzuki's indifference curves earlier, we will ignore the special looking indifference curves in Figure 4.7.

Marginal Rate of Substitution

As you must have realised by now, substitutability between two goods underlies a normal looking, downward sloping indifference curve. The rate of sacrifice of a good for the other along a given indifference curve is called the **marginal rate of substitution**. For instance, if you go back to Table 4.8, starting with the bundle (0, 40), one unit increase in *kulfi* requires a sacrifice of 10 chocolates. Hence the marginal rate of substitution (briefly MRS) of *kulfi* for chocolates is 10. That of chocolate for *kulfi* is the inverse of this, equal to 1/10. Similarly, between the bundles (1, 30) and (2, 22), the MRS of *kulfi* for chocolate is 8 and that of chocolate for *kulfi* is 1/8, and so on.[4]

NUMERICAL EXAMPLE 4.4

Consider the indifference schedule given in Table 4.9. Between different adjacent bundles, compute the marginal rate of substitution of one good vis-à-vis the other.

As the consumer moves from bundle A to bundle B, one unit increase in good Y entails a sacrifice of 91 – 87 = 4 units of good X. Hence the MRS of Y for X is 4 units of good X. The MRS of X for Y is the inverse of that of Y for X; this is equal to ¼. Table 4.10 lists the MRSs between various bundles.

There is a relationship between MRS and marginal utility. Let the marginal utility of *kulfi* and that of chocolate be respectively denoted as MU_{kulfi} and $MU_{\text{chocolate}}$.

Table 4.9 An Indifference Schedule

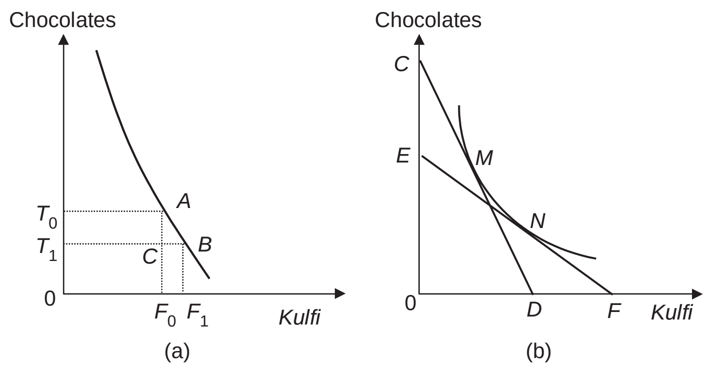
Bundles	*Good X*	*Good Y*
A	91	1
B	87	2
C	75	4
D	61	6
E	53	7
F	23	10
G	1	12

[4]If you look at Figure 4.7, in panel (a) the MRS = 0 and in panel (b) it is constant.

Table 4.10 Marginal Rate of Substitution (Numerical Example 4.4)

Between Bundles	*MRS of X for Y*	*MRS of Y for X*
A & B	1/4	4
B & C	1/6	6
C & D	1/7	7
D & E	1/8	8
E & F	1/10	10
F & G	1/11	11

Figure 4.8 Marginal Rate of Substitution



Suppose that the MRS of *kulfis* equal to m (chocolates). Then, by the definition of MRS, the utility from m chocolates must be equal to MU_{kulfi}. By the definition of marginal utility, the utility gained from consuming m chocolates = m.MU$_{\text{chocolate}}$. Hence, $m.MU_{\text{chocolate}} = MU_{\text{kulfi}}$. We can write as $m = MU_{\text{kulfi}}/MU_{\text{chocolate}}$, that is, MRS equals the ratio of marginal utilities.

MRS also relates to the slope of an *IC*.[5] To see this, consider the *IC* drawn in Figure 4.8(a). Starting with the bundle (F_0, T_0), consider another bundle (F_1, T_1), which has a slightly higher amount of *kulfi*. The MRS of *kulfi* is the ratio $T_0T_1/F_0F_1 = AC/BC$. If you now imagine sliding F_1 closer to F_0 so that the increase in *kulfi* is very small and the distance BC tends to zero, then the ratio AC/BC is nothing but the slope of the indifference curve at (F_0, T_0). Hence, MRS is equal to the slope of the indifference curve.[6] Figure 4.8(b) depicts another *IC* with M and N as two bundles on it. The MRS of *kulfi* for chocolate at M is the slope of the tangent line *CMD* and at N it is the slope of the tangent line *ENF*.

[5]The slope of a curve at a point is the slope of the tangent to the curve at that point.

[6]Strictly speaking, it is equal to the *absolute value* of the slope of the indifference curve.

If, in general, there are two goods *a* and *b*, we have:

$$\text{MRS}_{a \text{ for } b} = \frac{MU_a}{MU_b} = \text{slope of the } IC \text{ on the axis measuring} \tag{4.5}$$

Properties of Indifference Curves

1. Indifference curves are downward sloping, which follows from the assumption of non-satiation.

   However, suppose that one good is an economic 'bad' in the sense that the more you consume of it, the worse you feel, for example, pollution. In this case, the marginal utility is negative. Non-satiation obviously does not hold—indeed, its opposite holds. However, let another good (say *kulfi*) satisfy non-satiation. How would an indifference curve with *kulfi* and pollution look like? It will be upward sloping, as shown in Figure 4.9, because if you are given more of the 'bad', you need to be compensated with more of the 'good' to be indifferent.

   But, unless any specific mention is made, we continue to assume that the assumption of non-satiation is met and the indifference curve is downward sloping. Indeed, since the opposite of 'bad' is 'good', if we want to analyse the pollution issue with an indifference curve, we can always think of environment as the opposite of pollution and the indifference curve with environment on one axis and *kulfi* on the other will be downward sloping. Thus, essentially, the generalisation that an indifference curve is downward sloping holds.
2. Any two indifference curves for a particular consumer cannot intersect. We can prove this by the method of contradiction. Suppose, to the contrary, two indifference curves, IC_1 and IC_2, intersect, as shown in Figure 4.10(a). Mark bundles *A*, *B* and *C*. The bundle *A* is common along both indifference curves. The bundles *B* and *C* have the same amount of *kulfi* (F_1), but *C* has

Figure 4.9 Indifference Curve with an Economic 'Bad'

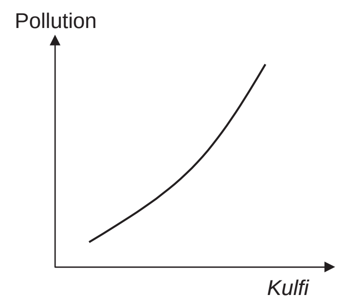
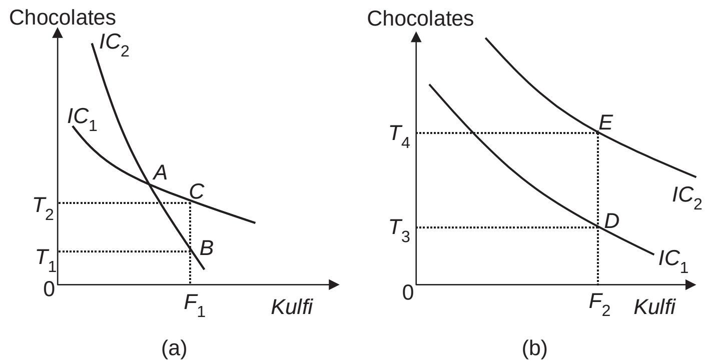


Figure 4.10 Some Properties of Indifference Curves



more chocolates (T_2) than B (T_1). The last fact implies that the consumer prefers C to B. Note also that the consumer is indifferent between A and B (since both lie on the indifference curve IC_1) and she is also indifferent between A and C (as both lie on the indifference curve IC_2). The last two statements imply that the consumer must be indifferent between B and C. This, however, contradicts that C is preferred to B. Hence, our supposition that two indifference curves intersect cannot be true.

3. An indifference curve to the right of a given indifference curve represents higher utility or satisfaction. This is simple and intuitive. In Figure 4.10(b), IC_2 is to the right of IC_1. Mark bundle D on IC_1 and bundle E on IC_2. Both have the same amount of *kulfi* (equal to F_2), but E has more chocolate (T_4 as compared to T_3). Hence E is preferred to D and, therefore, IC_2 must represent higher utility than IC_1.
4. Going back to Figure 4.8(b), notice that as you go down an IC, that is, as the quantity of *kulfi* increases, the slope of the indifference curve falls, that is, the MRS of *kulfi* decreases. This is called **diminishing marginal rate of substitution,** meaning that as you want to have more and more of a good, you are willing to sacrifice less and less of the other good (as the other good becomes increasingly scarce). It is quite intuitive and constitutes an important assumption in the indifference curve theory.[7] As the MRS equals the slope of an IC, diminishing MRS implies that the slope of the IC falls with the quantity of the good measured on the horizontal axis. In turn, this implies that an indifference curve is convex to the origin.[8]

[7]It is the counterpart of the law of diminishing marginal utility in the marginal utility theory.

[8]If we had assumed increasing MRS, the slope would have increased gradually and the curve would have looked like the production possibility frontier, concave to the origin.

Mathematically Speaking

Indifference Curve

The indifference curve is based on a **utility function**, relating total utility to different commodity bundles consumed. Denote the quantities consumed of *kulfi* and chocolate as F and T respectively and the total utility as V. Then a general utility function can be specified as $V = U(F, T)$, where U denotes the utility function.[9] Three examples are given below.

Example 4.A	Example 4.B	Example 4.C
$V = \sqrt{F} + \sqrt[3]{T}$	$V = \ln F + T$	$V = F^{0.6} + T^{0.4}$

Recall the definition of marginal utility. In terms of the utility function here, realise that it is nothing but a partial derivative of the utility function. In Example 4.A, the marginal utility of *kulfi* is given by $\partial U / \partial F = 1/(2\sqrt{F})$. In Example 4.B, the marginal utility of chocolate is equal to $\partial U/\partial T = 1$. The utility function in Example 4.C is called a Cobb-Douglas utility function.[10] In this utility function, the marginal utilities of F and T have the respective expressions: $(0.6)(T/F)^{0.4}$ and $(0.4)(F/T)^{0.6}$.

How is an indifference curve defined? Keep the total utility V unchanged, say at $\bar{V}$ ($\bar{V}$ could be 5, 1.3 or whatever), and write $\bar{V} = U(F, T)$. Given $\bar{V}$, this implicitly defines the quantity consumed of one good as a function of the quantity consumed of the other and this function is the equation of the indifference curve. For instance, in Example 4.A, $\bar{V} = \sqrt{F} + \sqrt[3]{T}$ yields

$$T = (\bar{V} - \sqrt{F})^3,$$

which is the equation of the indifference curve. You can see that as F increases, T falls, that is, the indifference curve is downward sloping.

If we totally differentiate the utility function, we have $dV = (\partial U/\partial F)dF + (\partial U/\partial T)dT$. Recall that MRS refers to the relationship between the changes in the quantities consumed (dF and dT) such that the total satisfaction is unchanged (that is, $dV = 0$). Setting $dV = 0$, we can write the above equation as:

$$-\left.\frac{dT}{dF}\right|_{V=\bar{V}} = \frac{\partial U / \partial F}{\partial U / \partial T}.$$

[9]More generally, if there are many goods whose quantities are x_1, x_2, x_3 and so on, we write the utility function as $V = U(x_1, x_2, x_3, \ldots)$.

[10]It is named after Charles Cobb and Paul Douglas, who discovered this functional form in 1928 in relation to production analysis. Later it also came to be used in demand theory. Cobb was a mathematician. Interestingly, Douglas who started his career as an economist in academia, left academics later and became a senator in the United States.

The left-hand side is simply the definition of MRS. The right-hand side is the ratio of marginal utilities. Hence it is proved that MRS equals the ratio of marginal utilities. The assumption of diminishing MRS means $\dfrac{d(\text{MRS})}{dF} < 0$, that is,

$$-\frac{d}{dF}\left(-\frac{dT}{dF}\bigg|_{V=\bar{V}}\right) < 0, \text{ or, } \frac{d^2T}{dF^2} > 0.$$

Check that $d^2T/dF^2 > 0$ for the three utility functions introduced above.

* * * * * *

More Fundamentally

Structure of Preferences

At a more fundamental level, the definitions of indifference curve and indifference map and the properties of indifference curve are based on the following assumptions about a consumer's preferences.

Comparability: This states that, given any two bundles, say *A* and *B*, the consumer can always compare them—either she prefers *A* to *B*, or prefers *B* to *A* or she is indifferent between the two. In other words, she cannot say that 'I do not know how to rank these two bundles'. What does this assumption imply in our indifference curve analysis? It implies that for any two bundles, one must have an indifference curve passing through each, and by comparing the indifference curves, the consumer can always rank the two bundles.

Consistency: It states that if one bundle is preferred to another, the latter cannot be preferred to the former. This may appear too obvious but, nonetheless, it is an assumption, meaning that we cannot have a situation where in a diagram like Figure 4.10(a), IC_2 represents a higher utility compared to IC_1 and vice versa.

Non-Satiation: We already know what this is. It is merely stated here in more general terms. That is, if there are two bundles *A* and *B*, and if compared to *B*, *A* has more quantities of some goods but not less of any good, then the consumer prefers *A* to *B*. In terms of the indifference map, it means that an indifference curve to the right represents higher utility.

Transitivity: If there are three bundles, say *A*, *B* and *C*, such that *A* is preferred to *B* and *B* is preferred to *C*, then *A* must be preferred to *C*. Recall that this assumption is used in proving that two indifference curves cannot intersect.

Convexity: Suppose that the consumer is indifferent between two bundles, *A* and *B*. The bundle *A* has 6 *kulfi*s and 10 chocolates, while the bundle *B* has 4 *kulfi*s and

16 chocolates. Take the average of *kulfi*s in the two bundles; it is 5 *kulfi*s. Similarly, take the average of chocolates in the two bundles; it is 13. Consider a bundle *C* having the respective average amounts, that is, 5 *kulfi*s and 13 chocolates. Convexity of preferences means that the consumer prefers *C* to *A* or *B*. More generally, convexity means that if the consumer is indifferent between two bundles, then the bundle having the respective average amounts is always preferred.

* * * * * *

Budget

Recall our earlier statement that how much of different goods a consumer would buy depends on her preferences and budget. We are done with characterising preferences. Now we will look at her budget, meaning what different bundles she can buy. Common sense suggests that it should depend on two factors: (a) prices of various goods she wants to consume and (b) her income.[11]

Suppose that the price of *kulfi* is Rs 15 per piece and that of a chocolate bar is Rs 12. Denote these prices respectively as $p_F = 15$ and $p_T = 12$. Suppose Ravi's income (say M) is Rs 300. What different combinations (bundles) of *kulfi* and chocolates can Ravi buy with his income and these prices? For simplicity, assume that both these goods can be measured on a continuous scale. Table 4.11 lists some of these bundles. Note that the total cost of each bundle is Rs 300 = M.

Figure 4.11(a) graphs the bundles in Table 4.11. Their joining line is the **price line** or the **budget line**, defined as the locus of bundles of the two goods the consumer can buy by fully spending her income, given the prices and her income. Figure 4.11(b) has three (arbitrary) price lines. Of these, for the moment, ignore the two dotted lines.

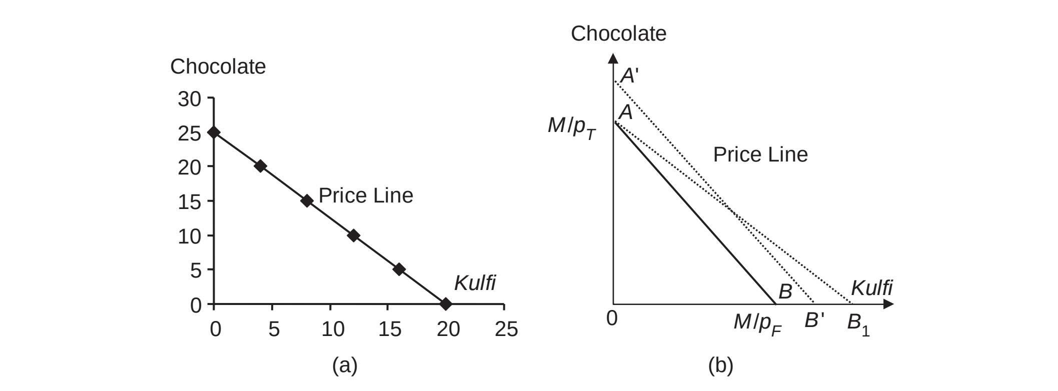
Note that a consumer can afford to buy any combination of goods on or inside the price line. The area inside or on the price line is called the **budget set**. If a

Table 4.11 Bundles that Ravi can Buy, Given Prices and His Income

Kulfi	*Chocolates*
0	25
4	20
8	15
12	10
16	5
20	0

[11] In demand theory, income typically refers to the total amount of money allocated for spending. More accurately, income means a consumer's earnings, one part of which is saved and the other spent or consumed. It is the second part that has traditionally been called income in demand theory.

Figure 4.11 Price Line



bundle is on the line then buying that bundle exhausts all income. A bundle strictly inside the line means that the income is not fully spent. The price line has the following properties.

1. On the price line, if a consumer wants more of one good, she must have less of the other (because income and price are fixed). Hence, the price line is downward sloping.
2. It is a straight line. Because the price of a good does not vary with how many units of it is purchased (for example, there are no discounts on volume), if the consumer wants to have one extra unit of a good, she must pay a fixed amount of extra money (equal to its price), which she can have by foregoing only a fixed amount of the other good. In our example, if she wants one extra *kulfi*, she would need Rs 15. In order to get Rs 15 she would have to give up 15/12 = 1.25 chocolates. In other words, one extra *kulfi* costs 1.25 chocolates no matter how much of *kulfi* and chocolates are being purchased initially. This constant rate of sacrifice implies that the price line is a straight line.
3. If the whole income is spent on one good, say *kulfi*, the amount of *kulfi* consumed must be equal to the income divided by the price of *kulfi*, that is, M/p_F. This is the maximum amount of kulfi the consumer can have (within his budget). That is, 0 of chocolate and M/p_F of kulfi must be on the price line. This bundle is nothing but the intercept of the price line on the '*kulfi* axis'. Similarly, the intercept of the price line on the 'chocolate axis' must represent the bundle having 0 *kulfi* and M/p_T chocolates. In Figure 4.11(b), the intercepts are marked M/p_F and M/p_T. In general, we can say that the intercept of the price line measures the maximum amount that can be purchased of the good (when all income is spent on that good).
4. The last property says something about the slope of the price line. We have the 'rise' equal to M/p_T and the 'run' equal to M/p_F. Hence, slope = rise/run = p_F/p_T. The ratio of prices is sometimes called the **relative price** (of the

good whose price is in the numerator). Therefore, in general, the slope of the price line is equal to the price ratio or the relative price of the good being measured along the x-axis.[12]

5. An increase (or a decrease) in income leads to a parallel shift of the price line to the right (or left). This happens as both intercepts shift proportionately, that is, M/p_F and M/p_T change proportionately when there is a change in M. The effect of an increase in income is shown by the dotted line $A'B'$ parallel to AB in Figure 4.11(b).
6. An increase (or a decrease) in the price of a good decreases (or increases) the intercept of the price line on the axis measuring that good. Suppose the price of *kulfi*, p_F, falls. This will increase the maximum amount of *kulfi* that can be purchased. Equivalently, M/p_F rises, that is, the intercept on the *kulfi* axis is greater.[13] This is illustrated by the (new) dotted price line AB_1 in Figure 4.11(b).

Mathematically Speaking

Price Line

As before, denote quantities consumed (demanded) of *kulfi* and chocolate as F and T respectively. The price line is defined by the relation $p_F F + p_T T = M$. We can rewrite this as:

$$T = \frac{M}{p_T} - \frac{p_F}{p_T} F,$$

which is the equation of the price line: T as a function of F. Note that its slope is equal to $dT/dF = -p_F/p_T$, which is negative and constant. Hence the price line is a downward sloping straight line, with the absolute value of its slope equal to the price ratio. The intercept on the T (vertical) axis is M/p_T. What is the intercept on the F (horizontal) axis? Set $T = 0$ and solve F; it is equal to $(M/p_T) \div (p_F/p_T) = M/p_F$.

* * * * * *

NUMERICAL EXAMPLE 4.5

A consumer has an income of Rs 100, which is to be spent on goods A and B. The prices are: p_A = Rs 4 and p_B = Rs 5. Fill in the blanks in Table 4.12 in terms of

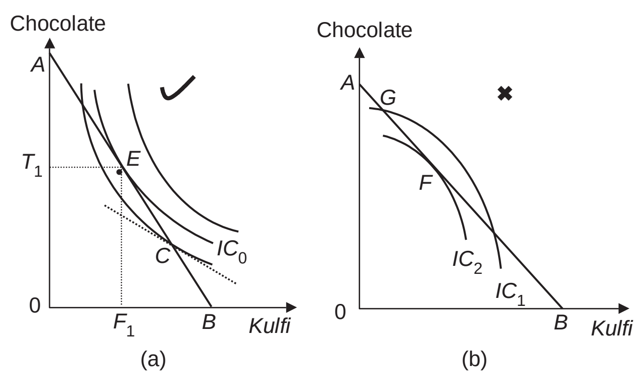
[12] Relative price is a general concept. It is defined as the amount of the other good that needs to be foregone in order to have one extra unit of the good in question. You can check that, by definition, it must be equal to the price ratio. In our numerical example, if the consumer wants to have one extra unit of *kulfi*, she would have to give up 15/12 = 1.25 chocolates. Thus, 1.25 chocolates are equal to the relative price of *kulfi*. What is the price ratio? This is also equal to 15/12 = 1.25.

[13] It means that even when income is unchanged, if the price of a good falls, the consumer can buy some bundles that cannot be bought before.

Table 4.12 Table for Numerical Example 4.5

Good A	*Good B*
–	20
10	–
–	8
20	–
–	3
25	–

Figure 4.12 Consumer's Equilibrium



how much of the respective good the consumer can buy given the prices and her income.

If the consumer is buying 20 units of good *B*, she spends Rs 100 on good *B*. Nothing is left for good *A*. Hence the amount of good *A* that can be purchased is 0. If 10 units of good *A* is bought, the amount Rs 10 × 4 = Rs 40 is spent on A. Rs 60 is left for good *B*, which will fetch Rs 60/Rs 5 = 12 units of good *B*. Likewise, as we go down the table, the respective numbers are—4 units of good *B*, 21.25 of good *A* and 0 of good *B*.

Consumer's Equilibrium

Once the consumer knows her preferences regarding various consumption bundles (that is, her indifference map) and which bundles can or cannot be afforded (that is, her budget line), the decision problem is straightforward in principle—first, confine to only those bundles which can be afforded and, second, among those, pick the one that offers maximum satisfaction or utility.

Refer to Figure 4.12(a), depicting the price line and the indifference map (ignore panel (b) for now). The consumer can purchase any bundle on or inside the price line AB, whereas any bundle to the right of the price line is not attainable. Among the bundles that are attainable, if you pick any bundle *strictly inside* the price line, you can always find a bundle *on* the price line, which has more of both goods. Hence, as long as she prefers more to less of a good, that is, the non-satiation assumption is met, the latter will be preferred to the former. Thus the 'best' bundle must lie *on* the price line.

Next, ask yourself, what is the highest indifference curve that the consumer can reach on the price line? The answer is IC_0. It is attained at the point E and this point is such that the price line is tangent to the indifference curve. Consider any point on AB other than E, say C. The indifference curve passing through that point lies to the left of IC_0, representing a lower level of satisfaction. Indeed, this is true for any point to the left or right of E on the price line. The bundle E is thus the best or the optimal bundle. We say that the consumer attains equilibrium at this point.

Why bundle E is the optimal choice can be understood in a different way. Start with a bundle like C, which is below and to the right of bundle E. At C, the slope of the indifference curve is less than that of the price line, that is, the MRS of *kulfi* for chocolate is less than the relative price of *kulfi*. This means that, psychologically, *kulfi* is worth less in terms of chocolate than what the consumer has to pay for *kulfi* in terms of chocolate foregone. Hence the consumer will benefit from having less *kulfi* and more chocolate. Similarly, it can be argued, starting with any bundle above and to the left of the point of tangency, the consumer will benefit from having more *kulfi* and less chocolate. Thus the consumer is most benefitted at the point E.

We now say that consumer's equilibrium is attained with that bundle on the price line at which the price line is the tangent to the indifference curve. Put differently, the optimal bundle is given by a point on the price line, where the slope of the indifference curve is equal to the slope of the price line.

We have seen earlier that the slope of the indifference curve is the MRS, which is equal to the ratio of marginal utilities, and the slope of the price line is equal to the ratio of prices. Hence we can write the **consumer's equilibrium condition** in general (over any two goods x and y) as

$$\mathrm{MRS}_{x \text{ for } y} = \frac{MU_x}{MU_y} = \frac{p_x}{p_y}. \tag{4.6}$$

Now turn to the Figure 4.12(b). The price line is the same as in panel (a), but the indifference curves are drawn concave to the origin, that is, the marginal rate of substitution of a good is *increasing* in it. Note that the condition (4.6) is satisfied at point F. But the total satisfaction is *not* maximised at this point. Consider any other bundle on the price line, say G. It is on a higher indifference curve (IC_1) than the one passing through the point F (IC_2). Therefore, for the condition (4.6) to characterise consumer's equilibrium, it is *essential* to assume diminishing MRS (that is, convex indifference curves). As discussed earlier, is economically reasonable also. Therefore, the condition (4.6) and diminishing marginal rate of substitution together constitute the necessary and sufficient conditions of consumer's equilibrium. Having noted this, from now on we will ignore panel (b).

The consumer's equilibrium condition can be approached in a different but an equivalent way. Denote the marginal utility of a rupee by $MU_{\text{Re 1}}$. Then, if for example, good x costs p_x rupees, an extra unit purchased of good x involves giving up p_x rupees and hence giving up utility equal to $p_x.MU_{\text{Re 1}}$. On the other hand, the extra unit purchased (and consumed) yields extra pleasure equal to the marginal utility of the good, that is, MU_x. A rational consumer would allocate his spending on a good such that these *marginal* utility losses and gains at the margin are equal, that is, $p_x.MU_{\text{Re 1}} = MU_x$. It is because, if, for example, $p_x.MU_{\text{Re 1}} < MU_x$, the consumer will gain by consuming a bit more of good x, as the extra utility cost of acquiring it is less than the extra gain in utility from consuming the good. Similarly, if $p_x.MU_{\text{Re 1}} > MU_x$, the consumer is spending too much on good x and will be better off by spending less on this good. We can express $p_x.MU_{\text{Re 1}} = MU_x$ as $MU_x/p_x = MU_{\text{Re 1}}$. This condition must hold for any commodity consumed. Hence, in general, if the consumer consumes many goods, x, y, z…, we can write the consumer's equilibrium conditions as

$$\frac{MU_x}{p_x} = \frac{MU_y}{p_y} = \frac{MU_z}{p_z} = ... = MU_{\text{Re 1}}. \tag{4.7}$$

Note that (4.7) is a generalised version of (4.6) and same as (4.4), obtained in the marginal utility theory. It states, in essence, that the consumer's equilibrium is attained when the marginal value of a rupee is the same across all goods consumed. This is quite intuitive. If, for instance, the consumer allocates her spending such that the marginal utility of one rupee spent on ice-cream is greater than that spent on *kulfi*, she will be better off by buying more of ice-cream and less of *kulfi*.

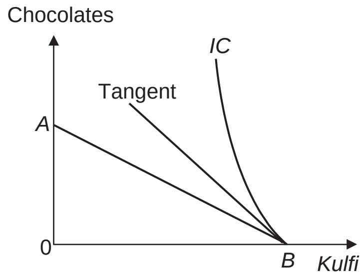
We can now link our theory here to that of 'demand' studied in Chapter 2. Referring to the equilibrium point E in Figure 4.12(a), we can say that F_0 and T_0 are respectively the quantities demanded of *kulfi* and chocolate at those prices and income that define the price line AB.

CORNER SOLUTION

Although Figure 4.12(a) shows that the consumer demands a positive amount of both goods, it need not be always the case. It is possible that the highest indifference curve on the price line is reached at a corner point. Figure 4.13 shows the case of a consumer who attains her equilibrium at point B, consuming zero chocolate and all the *kulfi* that can be afforded. A major difference with Figure 4.12 is that if the optimal bundle is a corner solution so to speak, the price line may not be equal to the slope of the indifference curve. This is indeed the case in Figure 4.13—the tangent to the IC has a higher slope than does the price line.

We note however that it is a very special case, very unlikely to occur when we consider a consumer's or a family's spending on broad aggregates of goods like

Figure 4.13 Corner Solution



food, clothing, furniture, entertainment, household services, health care and so on.[14] This is why, from now on, we will ignore this possibility and by consumer's equilibrium simply refer to the situation depicted in Figure 4.12(a).

Mathematically Speaking

Derivation of the Consumer Equilibrium Conditions (4.7) and Demand Functions

Suppose there are n goods, 1, 2, 3, …, n, the quantities of which are denoted as x_1, x_2, x_3, …, x_n. Let the utility function be $V = U(x_1, x_2, x_3, \ldots, x_n)$. Recall that the partials of this function are the respective marginal utilities; for example, the marginal utility of good 3 (MU_3) is equal to $\partial U/\partial x_3$. The consumer maximises his total utility $V = U(x_1, x_2, x_3, \ldots, x_n)$, subject to the budget equation:

$$p_1x_1 + p_2x_2 + p_3x_3 + \ldots + p_nx_n = M,$$

by appropriately choosing x_1, x_2, x_3, …, x_n.

We convert the above (mathematical) problem into the following one. Rewrite the budget equation as:

$$x_1 = \frac{M - p_2x_2 - p_3x_3 - \ldots p_nx_n}{p_1}.$$

Next, substitute this expression into the utility function and eliminate x_1 in it. We have:

$$V = U\left(\frac{M - p_2x_2 - p_3x_3 - \ldots p_nx_n}{p_1}., x_2, x_3, \ldots, x_n\right) \equiv F(x_2, x_3, \ldots, x_n),$$

a function of x_2, x_3, …, x_n.

[14]These are aggregates because each category can be further broken into sub-categories. For instance, food can be divided into cereals, vegetables, fruits, egg, milk and milk products, meat, fish and so on.

Maximising total utility subject to the budget equation is same as maximising the function F with respect to $x_2, x_3, \ldots, x_n$. Assuming that the optimal solutions are positive, that is, 'interior', in view of the General Appendix, the 'first-order' conditions of this maximisation are that the partials of F with respect to $x_2, x_3, \ldots, x_n$ are zero. Looking at how the function F is related to utility the function U, we have

$$\frac{\partial F}{\partial x_2} = \frac{\partial U}{\partial x_1}\left(-\frac{p_2}{p_1}\right) + \frac{\partial U}{\partial x_2}$$

Hence,

$$\frac{\partial F}{\partial x_2} = \frac{\partial U}{\partial x_1}\left(-\frac{p_2}{p_1}\right) + \frac{\partial U}{\partial x_2} = 0$$

can be expressed as

$$\frac{\partial U/\partial x_1}{p_1} = \frac{\partial U/\partial x_2}{p_2}, \text{ or}$$

$$\frac{MU_1}{p_1} = \frac{MU_2}{p_2}.$$

Similarly, $\partial F/\partial x_3 = 0$ gives

$$\frac{MU_1}{p_1} = \frac{MU_3}{p_3}.$$

We can do the same for x_4, x_5, up to x_n. With respect to x_n, $\partial F/\partial x_n = 0$ gives

$$\frac{MU_1}{p_1} = \frac{MU_n}{p_n}.$$

We now collect these equations and write them together as:

$$\frac{MU_1}{p_1} = \frac{MU_2}{p_2} = \frac{MU_3}{p_3} \ldots = \frac{MU_n}{p_n}.$$

This is same as our consumer's equilibrium conditions (4.7).

To illustrate how the demand functions are derived, consider the Cobb-Douglas utility function as in EXAMPLE 4.C. More generally, let the utility function be: $U = x_1^{\alpha_1} x_2^{\alpha_2}, \alpha_1, \alpha_2 > 0$. We have the budget equation: $p_1x_1 + p_2x_2 = M$, which we can write as $x_1 = (M - p_2x_2)/p_1$. Substituting this into the utility function, we have

$$U = \left(\frac{M - p_2x_2}{p_1}\right)^{\alpha_1} x_2^{\alpha_2} = F(x_2).$$

This is maximised with respect to x_2. We have

$$\frac{dF}{dx_2} = \frac{1}{p_1^{\alpha_1}}\left[-\alpha_1 p_2 (M - p_2 x_2)^{\alpha_1 - 1} x_2^{\alpha_2} + \alpha_2 (M - p_2 x_2)^{\alpha_1} x_2^{\alpha_2 - 1}\right].$$

The first-order condition is: $dF/dx_2 = 0$. Hence, setting the above expression to zero, we obtain the equation:

$$-\alpha_1 p_2 (M - p_2 x_2)^{\alpha_1 - 1} x_2^{\alpha_2} + \alpha_2 (M - p_2 x_2)^{\alpha_1} x_2^{\alpha_2 - 1} = 0.$$

This simplifies to

$$(\alpha_1 + \alpha_2) p_2 x_2 = \alpha_2 M_2$$

and solves x_2:

$$x_2 = \frac{\alpha_2}{\alpha_1 + \alpha_2} \frac{M}{p_2}.$$

This is the demand function for x_2. If we substitute this function into the budget equation, eliminate x_2 and rearrange, we get the demand function for x_1:

$$x_1 = \frac{\alpha_1}{\alpha_1 + \alpha_2} \frac{M}{p_1}.$$

Check for yourself that both goods are normal, that is, as M increases, both x_1 and x_2 increase. Furthermore and interestingly, the cross-price effects are zero, that is, $\partial x_1/\partial p_2 = \partial x_2/\partial p_1 = 0$.

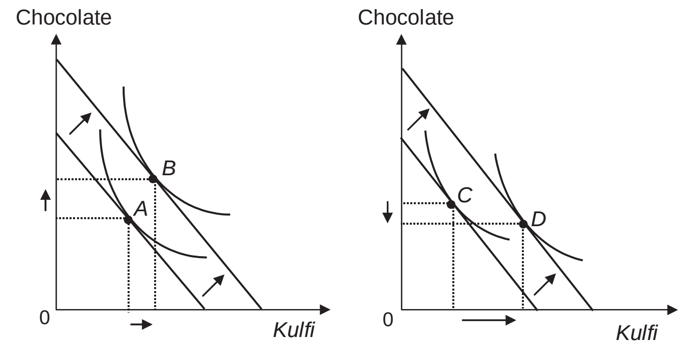
* * * * * *

Change in Income

Having considered the consumer's equilibrium, we can analyse how income and price changes would generally affect the consumer's equilibrium and thereby quantities demanded of goods. Here, we will analyse the effects of a change in income.

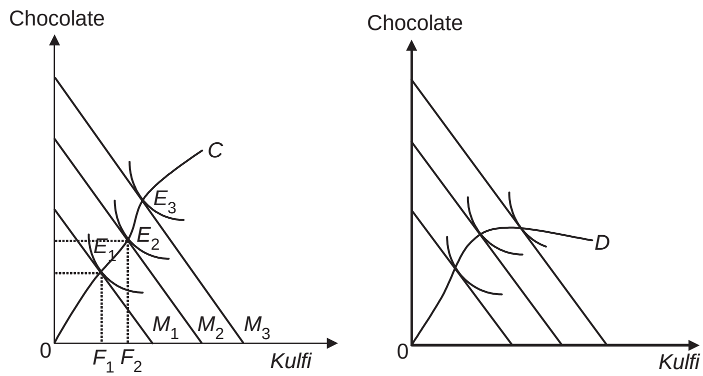
Suppose the consumer's income increases but prices remain the same. We have already seen that this will shift the price line to the right in parallel. This is illustrated in Figure 4.14. In panel (a), the original equilibrium point is A. As income increases, the price line shifts outward. The new equilibrium point is B where, compared to A, the consumption of both goods is higher. That is, as income increases, the consumer demands more of both goods. Hence, both goods are normal or superior. In panel (b), the original and the new equilibrium points are C and D respectively. Comparing these points we see that, with higher income, the consumer is buying more of *kulfi* but less of chocolate. Thus, for this consumer, *kulfi* is a normal good but chocolate is an inferior good. You can easily draw a diagram similar to panel (b) showing that *kulfi* is an inferior good but chocolate is a normal good.

Figure 4.14 Income Effect



(a) Both Goods Normal (b) One Good Inferior

Figure 4.15 Income Consumption Curves

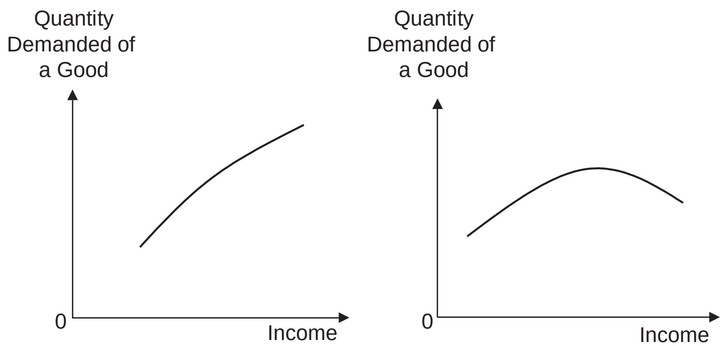


(a) Both Goods Normal (b) *Kulfi* Normal; Chocolate Inferior

The effect of a change in income on quantity demanded of a good is called the **income effect**. Our preceding discussion implies that this effect can be positive, defining a normal good, or negative, defining an inferior good.

While Figure 4.14 depicts a single shift of income, Figure 4.15 shows multiple income levels and the associated consumer equilibrium points. Joining these points we obtain a line called the **income consumption curve** (ICC). In panel (a), E_1, E_2 and E_3 are the equilibrium bundles at three different levels of income M_1, M_2 and M_3. (Ignore for now the markings F_1 and F_2.) Therefore, the line $0C$, joining such points, is the ICC. It depicts the case where both goods are normal. Put differently, if both goods are normal, the ICC is upward sloping.

Figure 4.16 Engel Curves



(a) Normal Good (b) Inferior Good

However, if, over some range of income, one good were inferior (both cannot be), then as income increases, the quantity demanded of this good would decrease and that of the other good would increase. As a result, the ICC curve will have a negatively sloping part, as shown in Figure 4.15(b).

From the ICC, we can mark the quantities demanded of a particular good at various levels of income. For instance, in Figure 4.15(a), F_1 of *kulfi* is demanded when the income is M_1; F_2 of *kulfi* is demanded at income M_2 and so on. If we graph F_1 against M_1, F_2 against M_2 and so on, we get what is called an **Engel curve**, a graph showing quantities demanded of a good at various levels of income.[15] Clearly, the Engel curve for a normal good would be upward sloping and that of an inferior good would have a downward sloping portion. These are illustrated in Figure 4.16.

Note that (*i*) the Engel curve is derived from the ICC and (*ii*) the elasticity of the Engel curve, by definition, equals the income elasticity of demand for a product.

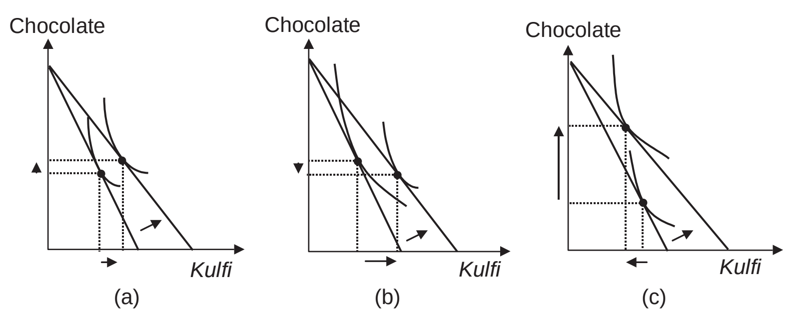
We next analyse the effects of a change in price.

Change in Price

Suppose the price of *kulfi* falls. The price line shifts outward on the *kulfi* axis. As shown in Figure 4.17, there are three possibilities—the consumer demands (*i*) more of *kulfi* and chocolate, (*ii*) more *kulfi* and less chocolate or (*iii*) less kulfi and more chocolate. Thus, the law of demand holds in cases (*i*) and (*ii*) but not in (*iii*). In relation to Chapter 2, we learn here that the law of demand may not necessarily hold. Put differently, the own price effect may not be always negative (that is, a decrease

[15]It is an important tool in applied consumer theory, named after a 19th century German statistician Ernst Engel, who made an extensive study of the expenditure patterns of families with different levels of income. For example, he (probably the first ever) found that the income elasticity of demand for food was relatively low.

Figure 4.17 Price Effect



or an increase in own price may not lead respectively to an increase or a decrease in the quantity demanded).

How do we explain this? The income effect and what is called a substitution effect together characterise the overall effect of a price change. Considering both effects, we will understand why that the quantity demanded of a good may or may not decrease with its price.

But before the substitution effect is introduced, it is important to know that the 'perverse' possibility that the quantity demanded may increase (or decrease) with an increase (or a decrease) in price is rather unusual in practice. For most purposes and applications the law of demand does hold.

Substitution Effect and Explaining the Price Effect

When the price of a product falls, the price line shifts outward and a consumer can afford some bundles, which could not be purchased earlier. You can interpret this as the consumer getting 'richer', that is, her real income or real purchasing power increasing—even though her income has not changed. This is why we see in Figure 4.17 that she is always on a higher indifference curve, that is, better off than before—irrespective of the change in quantities demanded.

Now consider the following hypothetical experiment. As the price of a product falls, take away some income from the consumer such that she is as well off as before. Similarly, if there is an increase in a product price, give some extra income to the consumer so that she is as well off as before. Substitution effect refers to the quantities consumed in this situation compared to the original situation before the price change.

In other words, the **substitution effect** is the effect of a price change on quantity demanded when income is 'compensated' such that the consumer's utility is unchanged. More precisely, it is called the **Hicksian substitution effect**, named after John Hicks, a Nobel Laureate in economics from England, who analysed this concept.

However, note that no one actually compensates the consumer when prices change in the market place. But, as will be seen in a moment, this 'thought experiment' is important in understanding how a price change affects consumer behaviour.

Refer to Figure 4.18. Let *AB* denote the original price line, E_0 the equilibrium point on it and IC_0 the associated level of utility. Suppose the price of *kulfi* falls so that *AC* is the new price line, E_1 the new equilibrium point and IC_1 the new (higher) level of utility. The price effect on quantity demanded of *kulfi* is given by the movement, $F_0 \rightarrow F_1$. Now, at the new prices, how much of income should we take away from this consumer so that she is back on the old indifference curve IC_0? The answer is $C'C$ in terms of *kulfi* (or equivalently $A'A$ in terms of chocolate). It is because, with the new price line $A'C'$ (parallel to *AC*), the consumer is able to attain the old indifference curve IC_0 by consuming at E'. Comparing E' with the original equilibrium at E_0 gives the substitution effect. This effect on *kulfi* is given by $F_0 \rightarrow F'$ or F_0F'.

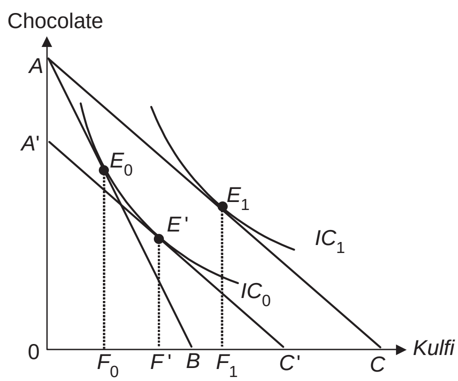
The substitution effect captures the *relative price effect* (in this example, as the price of *kulfi* falls, it becomes relatively cheaper), without any change in real income. Moreover, the own substitution effect is always negative—as a good becomes relatively more expensive, the consumer (along the same indifference curve) substitutes away from this good and hence consumes (demands) less of it.

We are now in a position to understand the price effect, which contains a substitution effect and an income effect. In our example, as the price of *kulfi* falls, by the substitution effect, the consumer buys more of *kulfi*, measured by $F_0 \rightarrow F'$. A decrease in price, increases the real income (by $A'A$ in terms of chocolate). This leads to an income effect, measured by $F' \rightarrow F_1$. Both effects add up to $F_0 \rightarrow F_1$, the price effect. Hence, in general,

$$\text{Price Effect} = \text{Substitution Effect} + \text{Income Effect.}$$

This is the key behind understanding the price effect.

Figure 4.18 Hicksian Substitution Effect

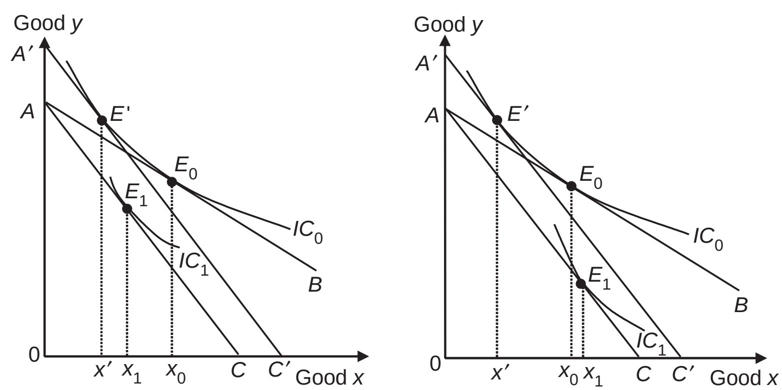


Note that, for a normal good, as its price changes, both the substitution effect and the income effect on the quantity demanded work in the same direction and, therefore, the law of demand holds. For instance, if there is a decrease in price, the quantity demanded of the product increases by the substitution effect (as its relative price falls). In addition, a fall in price means an increase in real income and, if the good is normal, the quantity demanded of it increases as real income increases. Hence both the substitution effect and the income effect lead to an increase in the quantity demanded. This holds in the example illustrated in Figure 4.18.

If the good is inferior, the income effect is 'perverse' or negative (that is, an increase in income leads to a decrease in the quantity demanded) and thus it works opposite to the substitution effect. The law of demand may or may not hold. If the perverse income effect is smaller in magnitude than the substitution effect, the latter dominates and the law of demand still holds. But if the perverse income effect is sufficiently strong to outweigh the substitution effect, the law of demand fails to hold. Such a good is called a **Giffen good** and such an unusual situation is called a **Giffen paradox**, named after an English economist and statistician R. Giffen, who studied such goods. We say that a good is a **Giffen good** if an increase (or a decrease) in its price leads to an increase (or a decrease) in its quantity demanded.

The inferior-good situation is illustrated in Figure 4.19. Both panels illustrate the effect of an increase in price of good *x*, such that the price line shifts from *AB* to *AC*. The original consumption bundle and the new one are respectively E_0 and E_1. In terms of demand for good *x*, the substitution effect is the reduction in quantity demanded from x_0 to x'. The income effect is the increase in quantity demanded from x' to x_1. Being an inferior good, the income effect is negative (in both panels)—the real income declines (due to an increase in the price of good *x*) and by the income effect, the quantity demanded of good *x* increases. However, the difference between the two panels is that in panel (a), the magnitude of the income effect is

Figure 4.19 Inferior Good



(a) Inferior but not a Giffen Good

(b) A Giffen Good

less than that of the substitution effect (the length $x'x_1$ is less than the length $x'x_0$), whereas the opposite is true in panel (b) with the length $x'x_1$ exceeding the length $x'x_0$. The consequence is that there is a difference in the price effect. In panel (a), the quantity demanded decreases from x_0 to x_1 (and hence the law of demand holds), while in panel (b), the quantity demanded increases from x_0 to x_1 and hence the law of demand does not hold.

In summary, panel (a) depicts the situation in which the good is inferior, but the substitution effect is greater than the income effect so that the law of demand holds. The good in question is not a Giffen good. Panel (b) shows a situation where the good is, again, inferior, but the income effect exceeds the substitution effect so that the law of demand fails to hold; this is a situation of Giffen paradox and the good is a Giffen good.

Giffen goods are thus a subset of inferior goods. In reality, however, there are not many examples of inferior goods and even fewer examples of Giffen goods. Jensen and Miller (2002) report that in north and south China, rice and noodles are Giffen goods.[16]

SLUTSKY'S SUBSTITUTION EFFECT

A Soviet economist, Eugen Slutsky had proposed an alternative definition of the substitution effect, similar to the Hicksian substitution effect. While in the latter, the income is adjusted so that the consumer is able to be on the original or the old indifference curve at the new set of prices, Slutsky's substitution effect is derived when the income adjustment is such that the consumer is able to buy the old bundle at the new set of prices.

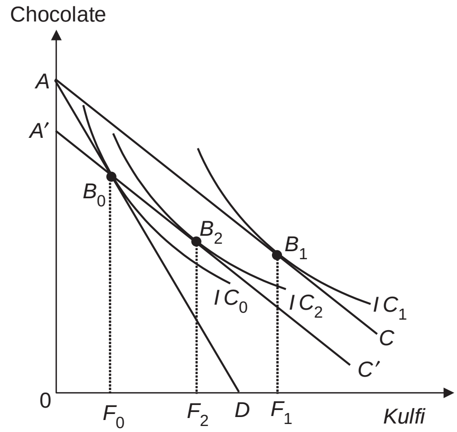
Turn to Figure 4.20. The original price line is *AD*. The consumer's equilibrium point is B_0 and she enjoys welfare indicated by IC_0. When the price of *kulfi* falls, *AC* is the new price line, B_1 is the new consumption point and the consumer is on a higher indifference curve IC_1. Let income be adjusted such that at the new set of prices the consumer is able to buy the old bundle B_0. This is indicated by the price line *A′C′*, which is parallel to the new price line *AC* and which passes through the old bundle B_0. The amount of income adjusted equals *AA′* in terms of chocolate. With such income adjustment, the consumer would attain equilibrium at B_2. The Slutsky's substitution effect refers to the movement from B_0 to B_2. In terms of demand for *kulfi*, this substitution effect is the increase in consumption from F_0 to F_2.[17]

However, Hicks's and Slutsky's substitution effects share the same important property—the own substitution effect is unambiguously negative, that is, by the substitution effect, as the price of a commodity increases, the consumer demands less of it. Further, the concept of price effect as the sum of a substitution effect and an income effect remains the same.

[16] R. Jensen and N. Miller. 2002. 'Giffen Behavior: Theory and Evidence'. Working Paper No. RWPO2-014, John F. Kennedy School of Government, Harvard University.

[17] The corresponding income effect is from F_2 to F_1.

Figure 4.20 Slutsky's Substitution Effect



Price Consumption Curve and the Demand Curve

Analogous to the income consumption curve, if we continuously vary the price of a good, we obtain **price consumption curve** (PCC). It is the locus of equilibrium consumption bundles at various relative prices when income is held constant.

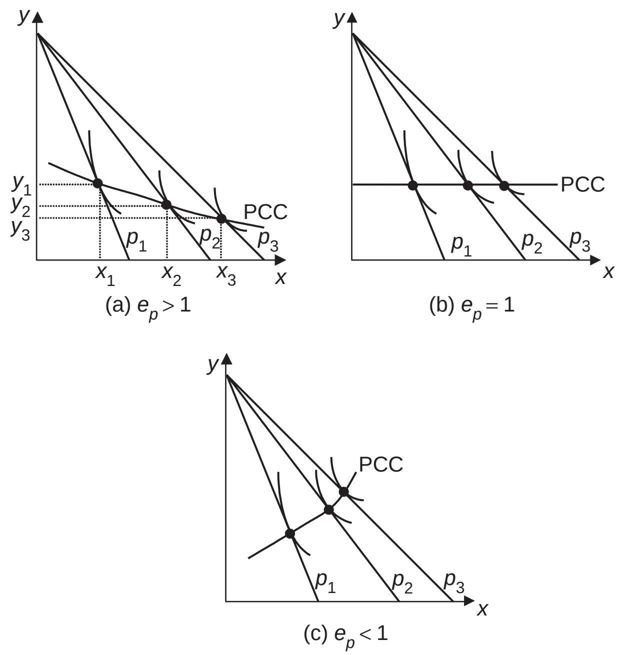
How does it look like? It depends on (*i*) whether a good is a Giffen good or not and (*ii*) how responsive it is to a change in price, that is, the price elasticity (even if it is not a Giffen good).

Suppose there are two goods x and y, and the price of only one good, say good x, changes. Further suppose that good x is not a Giffen good.[18] Consider various cases, depending on the magnitude of the price elasticity e_p.

Suppose $e_p > 1$. Then, if the price of good x falls, the quantity demanded of good x increases more than proportionately and because of that the total expenditure on good x rises. With income unchanged, this implies that the expenditure on the other good, y, must fall. With the price of good y unchanged, this means that the quantity demanded of good y falls too. In summary, as the price of good x decreases, the quantity demanded of good x increases and that of good y declines. The PCC in this situation is illustrated in Figure 4.21(a). The price of good x is falling gradually from p_1 to p_2, p_2 to p_3 and so on (each price line is indicated by the price of good x). The quantity demanded of good x increases from x_1 to x_2, x_2 to x_3 and so on. The quantity demanded of good y decreases from y_1 to y_2, y_2 to y_3 and so on. The PCC, as a result, is downward sloping.

[18]It is not difficult at all to draw the PCC for a Giffen good.

Figure 4.21 Price Consumption Curves



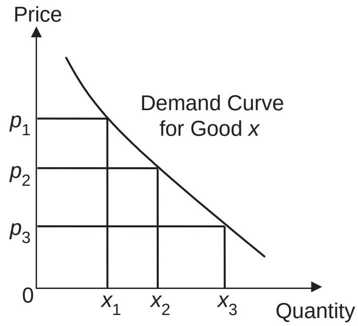
If e_p = 1, as the price of good x changes, the quantity demanded of it changes in the opposite direction and proportionately. The total expenditure on good x remains the same and that means the total expenditure on good y also remains the same. The price of good y being unchanged, there is thus no change in the quantity demanded of good y. This case is illustrated in Figure 4.21(b). As the price of good x keeps falling, the quantity demanded of it keeps increasing, while that of good y is unchanged. In this case, the PCC is a flat line.

Suppose e_p < 1. Applying similar reasoning, you can deduce that, if the price of good x decreases, the quantity demanded of both goods increases. Hence the PCC is upward sloping, as shown in Figure 4.21(c).

Once we know the PCC, we can derive the demand curve (which we simply had assumed in Chapter 2). How? Look at Figure 4.21(a) for instance. At price p_1, x_1 of good x is demanded. At price p_2, the quantity x_2 is demanded and so on. Mark that the income and the price of good y are fixed. Hence, 'all other things' in the law of demand are given. It follows that if we graph the price-quantity combinations (p_1, x_1), (p_2, x_2), (p_3, x_3) and so on, it will give us the demand curve for good x. Figure 4.22 does exactly that and shows the demand curve, derived from the PCC in Figure 4.21(a).[19]

[19]The PCCs in Figures 4.21(b) and 4.21(c) will give rise to different demand curves.

Figure 4.22 Deriving Demand Curve from the Price Consumption Curve



Mathematically Speaking

The Shape of the PCC

The price elasticity of good x is given by

$$e_p = -\frac{p_x}{x}\frac{dx}{dp_x},$$

where x is the quantity demanded of good x and p_x is its price. From the budget constraint

$$y = \frac{M - p_x x}{p_y},$$

where y is the quantity demanded of good y. Totally differentiating, we have

$$\frac{dy}{dp_x} = -\frac{1}{p_y}\left(x + p_x\frac{dx}{dp_x}\right) = -\frac{x}{p_y}\left(1 + \frac{p_x}{x}\frac{dx}{dp_x}\right)$$
$$= -\frac{x}{p_y}(1 - e_p).$$

Thus dy/dp_x is negative, zero or positive, that is, when p_x increases, the demand for good y decreases, remains unchanged or increases, accordingly as e_p is greater than, equal to or less than one. This result governs the shape of the PPCs drawn in Figure 4.21.

* * * * * *

Why the Demand Curve is Downward Sloping—A Summary

In the indifference curve analysis, this question refers to the nature of the price effect. Other prices remaining unchanged, when the price of a particular good falls, it becomes relatively cheaper compared to other goods. Hence, by the substitution effect, the quantity demanded of it increases. In addition, the decrease in the price increases the real income. If the income effect is positive (that is, if the good is a normal good), the quantity demanded of a product increases by the income effect also. That is, quantity demanded increases and the law of demand holds on account of both the substitution effect and the income effect. But if the income effect is negative (that is, if it is an inferior good), the quantity demanded falls with a decrease in price by the income effect. This is opposite of the substitution effect. If the magnitude of the income effect is small enough and does not exceed the substitution effect, the net effect of a decrease in price on the quantity demanded is positive and the law of demand holds.

How does this explanation of the law of demand relate to that in the marginal utility theory? Remember, this theory assumes that the marginal utility of money or income is constant, which essentially ignores the income effect. In other words, the marginal utility theory only captures the substitution effect of a price change.

Many Goods

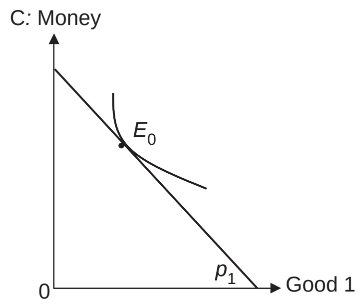
Of course, in real life, a consumer or a family buys many goods and not just two. If we want to analyse the demand behaviour of a consumer with respect to many goods at a time, the two-dimensional geometry cannot help. But this does not mean that the indifference curve analysis is irrelevant.

The indifference curve analysis can be easily generalised to any arbitrary number of goods and services and the main results are generalisations of what we have already learnt in the 'two-good case'. We have already seen the consumer equilibrium conditions in the 'many-good case', given in (4.7). Also, the income effect, the substitution effect and the price effect, all have the same meaning as in our two-good indifference curve model.

Interestingly, there is a special case where many goods can be effectively represented by two goods and the indifference curve analysis is directly applicable. Suppose that out of many goods we are only interested in the demand for one good *relative to the rest*. For example, we may want to study the food consumption pattern of a family in comparison to its consumption of all goods like housing, clothing, entertainment and so on. Realise that in this case we are essentially looking at two goods: food and 'others.'

There is a result in the demand theory (which you may learn in detail at the post-graduate level) stating that a group of goods can be lumped together and treated as a single (composite) good for analytical purposes as long as the prices of individual commodities within the group move proportionately. It is called the **composite good theorem**.

Figure 4.23 Composite Good



More particularly, this theorem implies that if the prices of a group of goods remain constant, the total expenditure on them can be treated as a composite good. We can make this more precise. Let there be n goods, namely, 1, 2, …, n. Let $p_1, p_2, \ldots, p_n$ be their respective prices and let $x_1, x_2, \ldots, x_n$ be the respective quantities demanded. Then, if the price of good 1 can change but prices of other goods cannot, the total expenditure on other goods, given by $C = p_2x_2 + p_3x_3 + \ldots + p_nx_n$, can be treated as a single good.

The composite good theorem also implies that we can draw an indifference curve measuring good 1 on one axis and C along the other. This is shown in Figure 4.23. Sometimes the expenditure on all other goods (C) is named 'money'—in the sense that C is equal to the total amount of money left to buy the goods other than the one we are focusing on. This is why C is indicated as money in Figure 4.23. We can also define a price line in terms of good 1 and good C. In the equation form it is: $p_1x_1 + C = M$, where M is the total income as before. Figure 4.23 also draws a budget line. Notice that the way good C is defined, its price is equal to one. Thus the slope of the price line must equal p_1. The consumer equilibrium is attained at the tangency point E_0. As before, we can analyse the income effect, the substitution effect and the price effect.

REVEALED PREFERENCE APPROACH

In both the marginal utility and the indifference curve approaches, the starting point is an individual's preferences. In the indifference curve analysis, in particular, her indifference map essentially gives an ordering or ranking of various bundles in terms of taste. Next, the budget set defines those bundles that the consumer can afford to buy at the given prices and income. Among these bundles, she chooses the one that she likes the best. In turn, this defines the consumer's demand for various goods. As price or income changes, the budget set changes and the consumer again chooses the bundle in the new budget set, which yields her the highest satisfaction. One definite—and central—result is that as a product price changes, by the substitution effect (that is, if income is compensated), quantity demanded changes in the opposite direction. Both approaches are founded on the notion that of 'preferences leading to choice.'

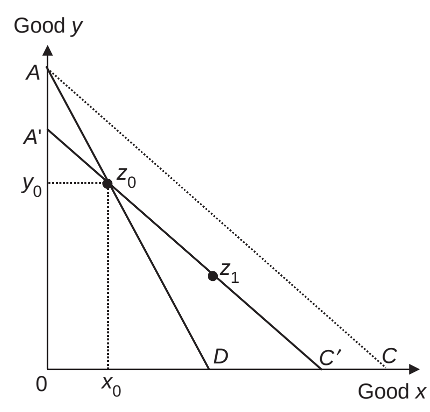
Revealed preference approach to demand, on the other hand, is built on a simple (yet beautiful) idea that we do not need to start from preferences.[20] If we simply observe what the consumer is choosing (given prices and income) and assume that the consumer's chosen bundle itself tells (reveals to) us that this bundle is the most preferred among all bundles that are available in the budget set, this is sufficient to imply that the substitution effect on product demand is negatively related to own price. How does this work?

Turn to Figure 4.24 (somewhat similar to Figure 4.20). Suppose that originally the price line is AD, that is, the budget set is $0AD$. In this situation, the consumer chooses a bundle, say z_0, with x_0 of good x and y_0 of good y. We do not ask how she chooses the bundle z_0; we simply observe it.

But, given that a bundle is chosen, we say that it is 'revealed preferred' to any other bundle in the budget set (otherwise the consumer would not have chosen z_0). In other words, revealed preference is defined as follows—a bundle z_1 is said to be **revealed preferred** to a bundle z_2 if both belong to some budget set in which z_1 is chosen (and z_2 is not). In other words, in this theory, preferences are induced from what the consumer chooses rather than vice versa.

Apart from this definition, there is an assumption called the **weak axiom of revealed preference**, that is, if a bundle z_1 is revealed preferred to another bundle z_2, then z_2 cannot be revealed preferred to z_1. It is like a consistency assumption. This axiom by itself implies that the own substitution effect must be negative and its proof is given below.

Figure 4.24 Substitution Effect in the Revealed Preference Approach



[20]Paul Samuelson, a famous American economist and a Nobel Laureate, is the originator of this theory.

Turning to Figure 4.24, suppose that the price of good x falls such that the new budget set is $0AC$. However, let the consumer's income be adjusted such that he is just able to buy the old bundle z_0, that is, let $0A'z_0C'$ be the new budget set. (This is same as Slutsky's income adjustment.) The new chosen bundle must lie in this budget set.

We now argue that the new chosen bundle must have more (at least no less) of good x than x_0 (which will prove that the substitution effect is negative). First, divide the new budget set $0A'z_0C'$ into two areas: $0A'z_0D$ and the triangle z_0DC'. If the new chosen bundle lies in the area $0A'z_0D$, it must be no other than z_0. It is because, if some other bundle, say z_1, is the chosen bundle, by definition, it is revealed to z_0, since both z_0 and z_1 are in the budget set $0A'z_0C'$ and z_1 is chosen. However, z_1 lying in the area $0A'z_0D$ implies that it is in the old budget set $0AD$ and since z_0 is the chosen bundle in this budget set, it is revealed preferred to z_1. Thus we have a situation where z_1 is revealed preferred to z_0 and z_0 is revealed preferred to z_1, contradicting the weak axiom of revealed preference. Hence the new chosen bundle cannot lie in the area $0A'z_0D$ unless it is same as z_0.

The remaining possibility is that the new chosen bundle lies in the triangle z_0DC'. But note that any bundle in this triangle has no less of good x than x_0; indeed, any bundle in the triangle z_0DC' other than z_0 has more of good x than x_0.

Combining the two possibilities about where the new chosen bundle may lie, it follows that it cannot have less of good x than x_0. This proves that the substitution effect on product demand is (weakly) negatively related to own price.

The fundamental point of this approach/theory is that the central result of the consumer theory, namely, the substitution effect, need not start with preferences over all possible consumption bundles. As long as we maintain that the actual choice behaviour of a consumer reveals her preferences over the bundles available in a budget set, it is enough to imply the substitution effect.

We know that, in general, the demand for any particular good may increase, decrease or remain unchanged as income changes, that is, there is no systematic sign of the income effect—it can be positive, negative or zero. Thus, although the income effect is not explicitly looked at in a typical analysis of the revealed preference theory, it is 'understood' that it can always be directly added to the substitution effect. The end product is a complete demand theory, not just the substitution effect.

Economic Facts and Insights

- A consumer attains equilibrium with respect to buying a good when the marginal utility of the product in terms of money equals the price of the product.
- The demand curve is downward sloping because of the law of diminishing marginal utility.

(continued)

- A consumer attains equilibrium with respect to buying many goods when the marginal utility of one rupee spent on each good is same across all goods.
- In the indifference curve analysis, a consumer attains equilibrium when the indifference curve is tangent to the price/budget line or, equivalently, the marginal rate of substitution of one good in terms of another is equal to the respective price ratio.
- An increase in income may increase, decrease or may not affect the demand for a good.
- An increase (or a decrease) in the price of a product increases (or lowers) its relative price and lowers (or raises) the real income of a consumer—in that his budget line shifts outward (shifts inward). This leads to a substitution effect and an income effect of a price change.
- The effect of a price change on the demand for a product, which is same as the price effect, is the sum of a substitution effect and an income effect.
- The own substitution effect of a price change on the quantity demanded is always negative in the sense that if the price of a product increases—and the consumer's income is compensated such that he is as well off as before or he is able to buy the old bundle—the quantity demanded of it decreases unambiguously.
- As the price of a product increases, the demand for a normal good decreases on account of both the substitution effect and the income effect. Therefore, the quantity demanded goes down unambiguously.
- For an inferior good, as its price rises, the quantity demanded may decrease or increase since the substitution effect and the income effect exert opposite movements on the quantity demanded.
- A Giffen good is an inferior good such that the perverse income effect outweighs the substitution effect. Hence the law of demand does not hold for a Giffen good.
- If the relative prices among a group of goods remain unchanged, then by virtue of the composite good theorem, these goods can be lumped together as one good.
- The revealed preference approach does not begin with preferences leading to inference on choice. A consistency assumption on the choice behaviour, called the weak axiom of revealed preference, is enough to imply that the own substitution effect is negative.

EXERCISES

4.1 How are total utility and marginal utility curves related? 'If we know a total utility schedule, we can derive a marginal utility schedule and vice versa.' Is this correct? Give reasons.

4.2 Atal's total utility schedule from buying phone card for international calls is the following.

Minutes of Call	*Total Utility*
0	0
1	50
2	95
3	135
4	170
5	200
6	225
7	245
8	260

The phone card costs Rs 8 per minute and Atal's marginal utility of a rupee is 3. How many minutes of call will he buy?

4.3 Carrot *halwa* sells for Rs 40/kg. A family consumes 3.5kg of carrot *halwa* (over a month) and its marginal utility is 100 utils. Assume that carrot *halwa* can be measured continuously. If for this family the marginal utility of rupee is 2 utils, should it consume more, less or is its consumption level of carrot *halwa* optimal?

4.4 Explain in terms of diminishing marginal utility, the logic behind the law of demand.

4.5 Among many other goods, Mumtaz regularly buys oil for her hair (L) and hand-lotion (N). Let MU_L and MU_N denote the respective marginal utilities and p_L and p_N their prices. Both goods are measured continuously. She is spending Rs 100 on hair oil and Rs 70 on hand-lotion and the marginal utilities are such that $MU_L/MU_N > p_L/p_N$. Is her spending on these two goods optimal? Or should she reallocate her spending? Give reasons.

4.6 An avid eater has Rs 205 to spend on chicken roll and ice-cream, selling at Rs 25 and Rs 20 respectively. The total utility schedules are given in the following table. How many chicken rolls and ice-creams will this person consume?

Units Consumed	*Marginal Utility: Chicken Rolls*	*Marginal Utility: Ice-cream*
0	0	0
1	625	480
2	550	400
3	500	360
4	425	300
5	375	240
6	250	180

4.7 An indifference schedule is given as follows. What is the marginal rate of substitution of good Y for good X between bundles C and D?

Bundle	*Good X*	*Good Y*
A	1	40
B	2	30
C	3	22
D	4	16
E	5	12

4.8 Briefly explain how the marginal utility analysis assumes cardinal utility, whereas the indifference curve analysis assumes ordinal utility.
4.9 Which assumption on preferences implies that the indifference curve is downward sloping and why?
4.10 Which assumption on preference implies that the indifference curve is convex to the origin and why?
4.11 Suppose that, of two goods, non-satiation holds for one good, whereas for the other good, the consumer does not care how much she consumes (that is, the marginal utility of this good is zero). How would the indifference curve look like?
4.12 What is the relationship between marginal rate of substitution and marginal utilities?
4.13 If two goods are perfectly complementary to each other in consumption, how would the indifference curve look like and why?
4.14 Suppose there are two goods and each is an economic 'bad'. How would the indifference curve look?
4.15 Prove that, as long as a consumer's preferences satisfy non-satiation, (a) all goods she consumes cannot be inferior and (b) if the price of a good falls, the demand for all goods cannot decrease.
4.16 Explain why the marginal rate of substitution of one good for another should decline as more of the good is consumed?
4.17 Why is the price line a straight line?
4.18 Suppose there are two goods *A* and *B*. Their respective prices are $p_A = 10$ and $p_B = 15$. Draw the budget line for a consumer whose income is Rs 300.
4.19 What is meant by 'relative price'? Explain it via an example.
4.20 How will a decrease in the price of a good shift the price line and why?
4.21 Suppose income and prices of all goods increase by the same percentage. How will it affect the budget line and why?
4.22 How will a decrease in income, coupled with an increase in the price of one good, shift the price line and why?
4.23 Suppose a family has an income of Rs 150. It spends this income on two goods: sugar and potato. The price of potato is Rs 15/kg. Sugar can be bought from the ration shop and the open market. The ration-shop price of sugar is Rs 10/kg and a consumer can buy a maximum of 5 kg of sugar. In the open market sugar can be bought at Rs 20/kg. Draw the consumer's price line.
4.24 On a price line choose a point above and to the left of the optimal bundle (the point of tangency between the price line and the indifference curve). Argue

in terms of marginal rate of substitution and relative price, how the consumer will benefit as he chooses a bundle closer to the tangency point.

4.25 Suppose preferences do not satisfy diminishing marginal rate of substitution. As a result, the indifference map is as shown in Figure 4.12(b). Prove that (in this case) the consumer will maximise satisfaction by consuming only one of the two goods, that is, all income is spent on one good.

4.26 What is an income consumption curve? Is it always upward sloping?

4.27 What is an Engel curve? Is it always upward sloping?

4.28 Explain the difference between Hicks' and Slutsky's substitution effect.

4.29 Suppose the price of an inferior good falls. Outline how this will affect the quantity demanded of this good via the substitution effect and the income effect.

4.30 'The own substitution effect is always negative, that is, when the price of a product falls, by the substitution effect the consumer will demand more of it'. Defend or refute.

4.31 'An inferior good is a Giffen good, but a Giffen good is not necessarily an inferior good'. Defend or refute.

4.32 Explain in terms of substitution and income effects, why the demand curve for a product is downward sloping.

4.33 Explain the notion of the price consumption curve.

4.34 Trace the shape of the PCC if the price of elasticity of demand for a good is zero.

4.35 Explain how the shape of the price consumption curve depends on the price elasticity of demand for the product whose price is changing.

4.36 'The marginal utility theory considers both the income and the substitution effect of a price change'. Agree or disagree. Give reasons.

4.37 Briefly describe what the composite good theorem means.

4.38 What is the weak axiom of revealed preference?

4.39 'In the revealed preference approach to demand, we begin with assumptions about preferences and then arrive at what a rational consumer will choose'. Defend or refute.

4.40 In equilibrium a consumer was buying 5 units of good A and some of good B. His income was Rs 100 and the prices were p_A = Rs 8 and p_B = Rs 5. The price of good A falls to Rs 5. By how much does his income need to be compensated so that he is able to buy the (old) bundle at the original equilibrium?

4.41 Refer to the previous question. Using the revealed preference approach and assuming that income is compensated, the amount purchased of good B at the new equilibrium is less than or equal to _________. Fill in the blank.

4.42 Suppose the utility function is $V = \exp(F) + T$. Show that it does not satisfy the assumption of diminishing MRS (and therefore if you plot the indifference curve, it will look concave to the origin).

4.43 Derive the consumer's equilibrium condition (4.6) by using the Lagrangean method outlined in the General Appendix.

5 Applications of Consumer Theory

CONCEPTS

- Total Willingness to Pay
- Consumer Surplus
- Cash Subsidy
- Kind Subsidy
- Direct Tax
- Indirect Tax
- Intertemporal Rate of Substitution
- Margin Propensity to Consume
- Marginal Propensity to Save

Consumer theory has numerous applications. In this chapter we consider four of them. The first one is an application of the marginal utility theory, while the remaining three are applications of the indifference curve analysis.

CONSUMER'S SURPLUS

We have learnt that if the price of a product falls (or rises), all else the same, the consumer is always better off (or worse off). It is useful to express such 'welfare' change in terms of money (or rupees). Why? Suppose that the government slashes the duty on printers imported from abroad. Obviously, those who purchase printers are better off. We would like to know how much money or purchasing power they gain, because of this policy change, in terms of their command to buy goods and services in general. As another example, suppose that the government increases the entertainment tax on movies and, as a result, the movie tickets become costlier by 20 per cent. Those of you who are avid movie goers would not like it. It is a loss of purchasing power to you. One may want to know how much purchasing power the movie goers are losing because of this tax policy change.

There are a few measures that quantify the welfare change in terms of money due to a price change. Of these, the most commonly used is called the **consumer surplus**, defined as the total willingness to pay for a product minus the total payment for the product.

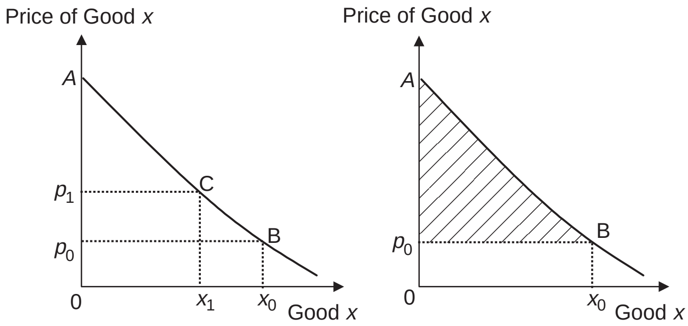
If we are talking about good x, for example, and p_x and x stand respectively for the price and the quantity demanded of this good, the total payment for good x is simply price × quantity demanded = $p_x x$.

The total willingness to pay for good x is defined as the total utility obtained from consuming x, measured in terms of money. It is the maximum amount that the consumer is willing to pay for the product. From the marginal utility theory, remember that this is measured by the area under the marginal utility (in terms of money). Further, this curve is same as the demand curve for the product. Hence the total willingness to pay is equal to the area under the demand curve. Turning to Figure 5.1(a), if $p_x = p_0$ and the consumer is demanding x_0, her total willingness to pay is equal to the area $0ABx_0$. Similarly, if $p_x = p_1$ and the consumer demands x_1, the total willingness to pay equals the area $0ACx_1$.

We are now ready to compute consumer surplus. If the price is p_0, for instance, the total willingness to pay = $0ABx_0$, while the total payment is equal to the area $0p_0 \times 0x_0 = 0p_0Bx_0$. Thus the consumer surplus = $0ABx_0 - 0p_0Bx_0 = p_0AB$. By similar calculation, at the price p_1, the consumer surplus is equal to p_1AC. In general, we then say that the consumer surplus equals the area under a demand curve over and above the line representing the price. For example, in panel (b) of Figure 4.23, the shaded area shows the consumer surplus at the price p_0.

Using consumer surplus, we can measure the change in welfare due to a price change. If for example the price of good x increases from p_0 to p_1, the change in consumer surplus is simply the difference between the consumer surpluses associated with p_0 and p_1. In Figure 5.1(a), this is equal to $p_0AB - p_1AC = -p_0p_1CB$.

Figure 5.1 Consumer Surplus



(a) Total Willingness to Pay (b) Consumer's Surplus

The negative sign reflects a welfare loss due to a price increase, which is expected. In absolute value, the change in the consumer surplus equals the area between the two respective price lines.

See Clip 5.1 for a sample of estimates of consumer surplus.

Clip 5.1: Consumer Surplus Estimates

Many estimates of consumer surplus are available for various commodities and services in India and other countries. For instance, Watal (2000) provides estimates of loss in consumer surplus in the demand for patentable pharmaceutical products if intellectual right protections are granted in India following our WTO commitments. Assuming that prices of such products would increase anywhere from 26 per cent to 242 per cent, the consumer surplus loss is evaluated at 50 million to 140 million US dollars. Peck, Chaloupka, Jha and Lightwood (1999) study the demand for tobacco products in India and China (and in other countries). They estimate that, in per capita terms and on 1990 prices, the consumer surpluses in demanding tobacco products are 12 US dollars and 9 US dollars in India and China respectively. Mitra (1999–2000) has examined the demand for four tourist spots in Arunachal Pradesh. Using local transport cost as the price of touring these spots, the consumer surplus per visit by Indians is estimated at Rs 995; that by foreigners is estimated at Rs 1, 232.

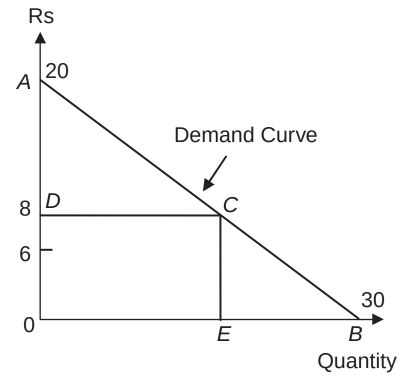
References

Mitra, Amitava. 2000. *Environmental Conservation and Demand for Nature-Based Tourism in Arunachal Pradesh*: Final Report. Submitted to Environmental Economics Research Committee, IGIDR, funded by World Bank Aided: Environmental Management Capacity Building Technical Assistance Project.

(continued)

Peck, Richard, Frank J. Chaloupka, Prabhat Jha and James Lightwood. 1999. 'A Welfare Analysis of Tobacco Use,' in Prabhat Jha and Frank J. Chaloupka (eds), *Curbing the Epidemic: Governments and the Economics of Tobacco Control*. Washington D. C.: World Bank.
Watal, J. 2000. 'Pharmaceutical Patents, Prices and Welfare Losses: Policy Options for India under the WTO TRIPS Agreement', *The World Economy*, 23(5), 733–52.

Figure 5.2 Consumer Surplus Calculation (Numerical Example 5.1)



NUMERICAL EXAMPLE 5.1

The demand curve facing a consumer is a straight line as shown in Figure 5.2. What is the consumer's total willingness to pay for the product if the price is equal to 8? Derive the consumer surplus at price equal to 6 and price equal to 8. Compare the two surpluses. Which one is greater and why?

At price = 8, we have $AD = 20 - 8 = 12$. The triangles ADC and $A0B$ are similar. Thus $DC/AD = 0B/0A$. The last ratio equals $30/20 = 3/2$. Thus $DC = (3/2)AD = (3/2)12 = 18$. This is the quantity demanded at price equal to 8. The total willingness to pay at this price is equal to the area under the demand curve, which, in turn, equals the areas of rectangle $0DCE$ plus the triangle ADC. These areas are respectively equal to $8 \times 18 = 144$ and $(1/2) \times 12 \times 18 = 108$. Thus the total willingness to pay = 144 + 90 = 234. The consumer surplus is the area of the triangle under the demand curve, equal to 108.

Similar calculation yields the consumer surplus at price 6 equal to 147. This is higher than the consumer surplus at price equal to 8, because the price is lower.

CASH VERSUS KIND SUBSIDY

The government offers many types of 'kind subsidy' to poor sections of the population or to its employees. For instance, food in the canteens of many government organisations is highly subsidised. A decent meal with 3 *chapatis, dal,* a *subji* and curd may cost an employee Rs 5, whereas the market price of that meal is Rs 20. Another example is LTC (leave travel concession) given to central and state government employees. This is again a subsidy in kind; the government is subsidising its employees' 'consumption' of travel.[1]

Note that, in each of these examples and in general, providing subsidy in kind costs something to the government (and ultimately to tax-payers). The objective behind providing a kind subsidy is to improve the welfare of the recipients.

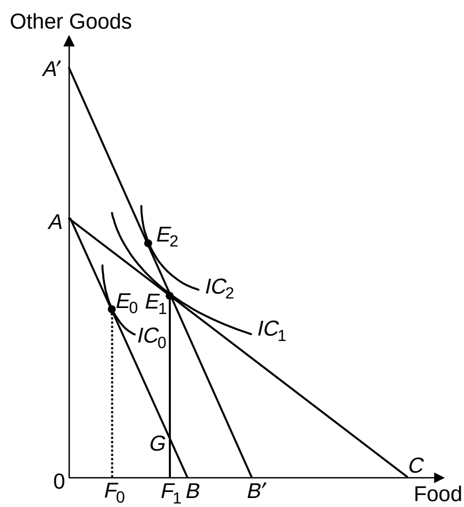
There is, however, an alternative, which will cost the same to the government and also please its recipients—that is, let the government pay cash to the recipients (called a cash subsidy) equal to the amount it was spending on the subsidy programme in kind. The issue is—which programme is better for the consumer given that both programmes cost the same to the government?

We analyse this by referring to Figure 5.3. It measures food on one axis and 'other goods' (by virtue of the composite-good theorem) along the other. Suppose that without any subsidy on food, the budget line of the consumer is AB.

The slope of AB measures the relative price of food. The consumer attains his equilibrium at E_0. He consumes F_0 of food and E_0F_0 of other goods. His welfare is indicated by the indifference curve IC_0.

Suppose the government subsidises the consumption of food (that is, food becomes cheaper to the consumer) such that the new budget line is AC. The person would now consume F_1 of food and E_1F_1 of other goods. Mark the point G, where E_1F_1 intersects AB. We can now argue that E_1G is the cost of this kind-subsidy

Figure 5.3 Cash versus Kind Subsidy



[1]In the US, the poor are given 'food stamps' by the government, which can be used as cash only if the person is buying food items. It is nothing but a subsidy on food consumption.

programme. How? Suppose the consumer was buying the same amount of food he is consuming now (F_1) at the unsubsidised (old) price, that is, along the old budget line AB. He would then have been left with GF_1 of other goods. But with the subsidy programme in place, he is left with E_1F_1 of other goods. Since the prices of other goods are unchanged, the difference between E_1F_1 and GF_1, equal to E_1G, must be what the government is paying as the subsidy.

Consider now the alternative programme of giving the amount E_1G by cash to the consumer. Mark the line $A'B'$ drawn parallel to AB and passing through E_1. It has the property that $E_1G = AA'$. With the cash subsidy in place, in terms of the other goods the consumer's disposable income (inclusive of the cash subsidy) is now $0A + E_1G = 0A + AA' = 0A'$. Since there is no subsidy on food, the new budget line must have the original slope AB. Thus $A'B'$ is the budget line. What is then the consumer's optimal point of consumption? It is E_2. Note that the optimal consumption point in the subsidy-in-kind programme, which is E_1, is still available to the consumer in the cash-subsidy programme (as E_1 lies on $A'B'$ too). But, E_1 is no longer the equilibrium point. Importantly, at E_2 the consumer is on a higher indifference curve (IC_2) as compared to that at E_1 (IC_1).

The conclusion is that the consumer is better off in the cash-subsidy programme than in the kind-subsidy programme. The underlying economic reason is the following. The cost of the kind-subsidy programme (E_1G) depends on the equilibrium choice of the consumption bundle in that programme; hence in a cash-subsidy programme whose cost is equal to the kind-subsidy programme, the above consumption bundle is also available (that is, the bundle E_1 is available along the budget line $A'B'$). This implies that the consumer cannot be worse off in the cash-subsidy programme as compared to the kind-subsidy programme because he always has the option of choosing E_1 in the cash-subsidy programme. Moreover, since the relative price of the good in question facing the consumer is different between the two schemes, he will choose a different bundle in the cash-subsidy programme (E_2) than the one in the kind-subsidy programme (E_1) and be better off in choosing so. Put differently, moving from the kind-subsidy programme to the cash-subsidy programme provides 'trading opportunities' to the consumer in terms of choosing his consumption bundle. This opportunity to trade enhances his utility.[2]

Our conclusion then raises a question, that is, why do we observe kind-subsidy in practice? One explanation is that the consumer, in accordance with his own preferences, may spend the cash-subsidy on items that the government does not want him to spend on, for example, alcohol. Many poor people, especially men, are indeed addicted to drinking and if any extra cash is given to them, intended to be spent on food, it is likely to be spent instead on alcohol. Our indifference curve model and our conclusion that a cash-subsidy programme is better than a kind-subsidy programme implicitly assume that the pattern of preferences of the targeted recipients is not 'perverted' from the viewpoint of the government.

[2] To carry forward the argument, the equilibrium bundle in the cash-subsidy programme is not available in the kind-subsidy programme (that is, E_2 is not available along the budget line AC).

DIRECT VERSUS INDIRECT TAX

Taxes are of two kinds. One is a **direct tax**, referring to a tax on an individual or an organisation. Direct taxes include personal income tax, corporate income tax, wealth tax, property tax and gift tax. In India, only the central government imposes personal income tax. There are no state income taxes. In some countries like the US, some states also levy income tax over and above the central (federal) income tax. Corporate income tax is levied on the profits of private companies. After this tax is paid, a company may invest a part of the remaining profits in building assets for the company or it is paid out to shareholders as dividends. Dividend income, as wage income, is subject to personal income tax. Direct taxes are those that cannot be shifted to other parties.

The other is an **indirect tax**, referring to taxes, which *alter* the cost or price of a good or service for either producers or consumers. There are many types of indirect taxes used in India (and other countries), such as import or customs tariff, excise tax, sales tax, service tax and value added tax (VAT).

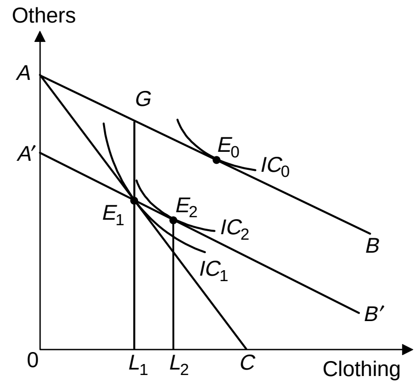
Suppose the government wants to generate a given amount of revenue from consumers or households for spending on national defence, building a dam, subsidies for the poor and so on. There are two options: an income tax as a direct tax and an indirect tax on a commodity or service. Which is the better option for the sake of consumer's welfare?

In many ways this issue is analogous to the earlier one of kind versus cash subsidy, which involved the method of disbursing among a class of people a given amount of tax revenue already raised rather than raising revenue itself. Note that an indirect tax increases the price of a product for the consumer and hence is the opposite of a kind subsidy. Further, an income tax is the opposite of a cash subsidy. But, interestingly, the answer to both the issues is similar. Income tax is a better option in terms of consumer's welfare, just as cash subsidy was shown to be better for the consumer than kind-subsidy.

Turn to Figure 5.4. There are two goods: clothing and 'others'. Suppose there is no tax to begin with. The price line is AB. The consumer's equilibrium is at E_0. First consider a sales tax on clothing. The price of clothing is higher for the consumer because of this tax. Let AC be the new price line. The consumer chooses the tangency point E_1. She consumes L_1 of clothing and E_1L_1 of 'others'. The indifference curve passing through E_1 is at a lower level than the one passing through E_0—as you would expect, the consumer is worse off than before. How much revenue is the government collecting? The answer is GE_1. It is because if the consumer did not have to pay the sales tax, she would have had GL_1 of 'others' but she actually has E_1L_1 of 'others'. Therefore, the difference, equal to GE_1, must be the sales tax proceeds.

If, instead of the sales tax, the amount GE_1 is taken away from the consumer as income tax, her budget line is $A'B'$, parallel to AB. The point E_2 is the equilibrium point. Notice that the consumer is on a higher indifference curve (IC_2) in comparison to the sales tax programme. Thus, for the consumer, an income tax is preferable to an indirect tax.

Figure 5.4 Direct versus Indirect Taxes



The economic reason behind this is similar to that in the case of kind versus cash subsidy. In the income tax programme, the equilibrium bundle with the indirect tax programme (E_1) is available to the consumer. Thus the consumer cannot be worse off than in the indirect tax programme. The relative price of clothing is different between the two programmes. Hence, in the income tax programme, the consumer will pick a bundle (E_2) other than (E_1) and be strictly better off compared to the indirect tax programme.

Why do governments then use indirect taxes? The prime reason is that an income tax is politically unattractive in the sense that it is visible and affects an individual's pocket directly. On the other hand, the cost of an indirect tax is less visible to the public and hence politically more convenient. Moreover, the authority to impose a direct tax like an income tax lies with the central government only, not with state governments. Hence, state governments, in order to raise revenues for their expenditure and commitments, resort to indirect taxes.

CONSUMPTION-SAVINGS DECISIONS

The last chapter dealt with the consumer's decision-making on how much to consume of different goods and services. More generally, we can think of the following decision problems facing an individual or a family. After it receives its income, a part of it is spent and the remainder is saved. How much to spend and how much to save is one problem. The second problem is to allocate the total spending on different goods and services (which we studied in the last chapter).[3]

[3]Remember that the term 'income' in the last chapter actually meant total spending.

Interestingly, the method we undertook to analyse the second problem can be applied to understand the first problem. This is the objective here: to analyse a consumer's decision with regard to consumption and savings.

Recall the composite-good theorem which says that if the relative prices among any bunch of goods remain unchanged, the total spending on these goods can be interpreted as a single good. Assuming that the relative prices among all goods the consumer buys do remain constant, we can them lump the total spending on them as one good, say 'consumption' or 'present consumption'. Now realise that people save so that they can use their savings towards consumption in the future. In other words, we can say that a consumer uses his income to meet present consumption as well as future consumption.

Assume for simplicity that a consumer lives only two periods, present (period 1) and future (period 2), that is, there is only one future period. This will imply that such a consumer will only save in period 1 to be used for consumption in period 2. (As the world ends after period 2, so to speak, there is no point in saving anything in the terminal period 2.)

In this scenario we can now frame the choice problem of a consumer in a precise way. Let C_1, S and Y_1 denote a person's consumption, savings and income in period 1. By definition, $S = Y_1 - C_1$. Note that savings can be positive, zero or negative. Negative savings mean that $C_1 > Y_1$, that is, the excess of consumption over income is financed by borrowing. Similarly, positive savings mean lending. If you keep your savings in a bank, it is like lending your savings to the bank. Let r denote the market interest rate at which one can borrow or lend. If a person saves amount S_0, the principal plus interest to be received or paid in period 2 is equal to $S_0(1 + r)$, depending on whether S_0 is positive or negative.

Budget Line

Let C_2 and Y_2 denote consumption and income in period 2. What is the relationship between C_2 and Y_2? It is

$$C_2 = Y_2 + S(1 + r).$$

Note that $S(1 + r)$ is positive or negative, as S is positive or negative. It means that if the consumer is a lender in period 1 (that is $S > 0$), he is able to consume more than his income in period 2. Similarly, if he is a borrower in period 1, he has to pay back the loan in period 2 and thus his future consumption will be less than his future income.

Remembering that $S = Y_1 - C_1$, we can write the above equation as:

$$\begin{aligned} &C_2 = Y_2 + (Y_1 - C_1)(1 + r), \quad \text{or} \\ &C_1 + \frac{C_2}{1+r} = Y_1 + \frac{Y_2}{1+r}. \qquad (5.1) \end{aligned}$$

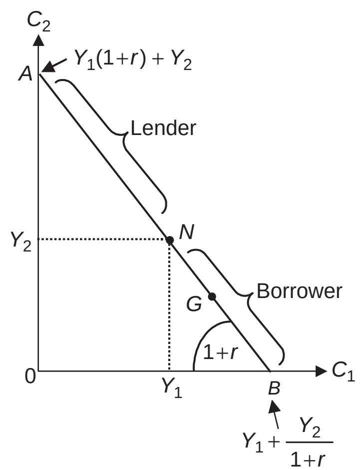
This is the intertemporal budget equation facing the individual.[4] The left-hand side is the discounted value of present and future consumptions, whereas the right-hand side is the discounted value of present and future incomes. The word 'discounted' means that future consumption and future income are adjusted to make them comparable to present consumption and present income. For instance, while C_2 is the future (period 2) consumption it is worth only $C_2/(1 + r)$ in terms of the period 1. That is, if you forgo $C_2/(1 + r)$ amount of consumption today (period 1) and save, it will fetch you $(1 + r)[C_2/(1 + r)] = C_2$ of consumption in period 2.

We assume that the present and future incomes are given to the consumer.[5] In this scenario, the consumer faces a choice problem involving C_1 and C_2. Indeed, you can think of C_1 and C_2 like two goods in the last chapter. In view of the budget equation (5.1), the price of present consumption is $p_1 = 1$ and the price of future consumption is $p_2 = 1/(1 + r)$.

Figure 5.5 depicts the budget line AB. If $C_2 = 0$, then $C_1 = Y_1 + Y_2/(1 + r)$. Thus $Y_1 + Y_2/(1 + r)$ is the intercept of the budget line on the x-axis measuring C_1. Likewise, you can determine that the intercept on the y-axis measuring C_2 is equal to $Y_1(1 + r) + Y_2$. The absolute value of the slope of AB is then:

$$\frac{\text{rise}}{\text{run}} = \frac{Y_1(1+r)+Y_2}{Y_1+Y_2/(1+r)} = 1+r.$$

Figure 5.5 Budget



[4]The word 'intertemporal' means 'over time.'

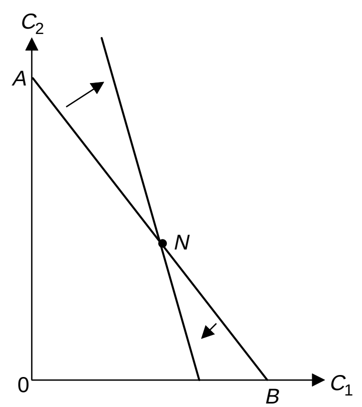
[5]For example, Y_1 can be your present salary in period 1 and Y_2 your expected salary in period 2. Or if you own some rental property Y_1 is your rental income in period 1 and Y_2 is the expected rental income in period 2.

The term $1 + r$ is indeed the relative price of present consumption in terms of the future consumption, as $p_1/p_2 = 1/[1/(1 + r)] = 1 + r$. This has a straightforward interpretation—if the interest rate is r, as you sacrifice one unit of present consumption (that is, save one unit), you get future consumption equal to $1 + r$.

There is a particular point on *AB* which is of special significance. Look at the point N, where $C_1 = Y_1$ and $C_2 = Y_2$, that is, the present consumption is exactly equal to the present income and future consumption matches with future income. This is the point at which there is no borrowing or lending, that is, savings are zero. If the consumer chooses a point on *AB* and below the point *N*, like *G*, then $C_1 > Y_1$. This means that he is a borrower. Similarly if he chooses a point on *AB* lying above the point *N*, then he is a lender.

However, before looking at which point the consumer will choose (in principle), let us consider how the budget line shifts when the current income (Y_1) future income (Y_2) or the interest rate (r) changes. Suppose Y_1 increases. Then the intercept on the x-axis moves to the right and that on the y-axis moves up. Thus the budget line shifts to the right. The same holds for an increase in Y_2. But an increase in the interest rate has a different kind of effect. Note three things. (*i*) A higher r means a lower value of $Y_1 + Y_2/(1 + r)$, implying that the intercept on the x-axis shifts towards the origin. (*ii*) This means a higher value of $Y_1(1 + r) + Y_2$ and thus the vertical intercept moves up. (*iii*) Think about the point *N*—since it is the zero borrowing-lending point, it must remain unaffected by any interest rate change. Put differently, the choice of $C_1 = Y_1$ and $C_2 = Y_2$ always satisfies the budget irrespective of what the interest rate is. That is, the point *N* does not move. These three features imply that the budget line shifts neither to the right nor to the left entirely. Instead, an interest rate increase moves the budget line clockwise, pivoting on the point *N*. This is shown in Figure 5.6.

Figure 5.6 Effect of an Interest Rate Increase on the Budget Line



Preferences and Consumer's Equilibrium

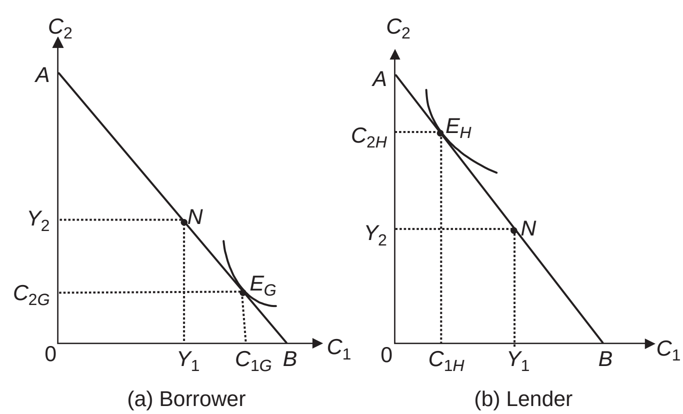
Which point on the budget line will be chosen by the consumer? It would partly depend on the preferences about present and future consumption. All else the same, if he is very impatient he would have a lot of present consumption and a little of future consumption. If he is very patient, he would choose the opposite. Indeed, C_1 and C_2, can be thought of as two goods (like ice-cream and chocolate). Accordingly, we can draw indifference curves for C_1 and C_2, and define the marginal rate of substitution of present consumption for future consumption as the units of future consumption the consumer wants to forego in order to have one extra unit of present consumption. This is called the **intertemporal rate of substitution.** Denote this as MRS_{12}. Recall that, in general, MRS is equal to the slope of the indifference curve. Hence MRS_{12} is equal to the slope of the IC in present and future consumption.

Once all this is understood we can readily characterise the consumer's equilibrium. Turn to Figure 5.7. Given his budget, the consumer maximises his utility from present and future consumption at the point of tangency between the budget line and the indifference curve. In panel (a) this is indicated at the point E_G. Panel (b) represents a different individual with a generally flatter indifference curve with the equilibrium point being E_H. In both cases, the tangency point between the indifference curve and the budget line is the equilibrium point. What is the general condition of consumer's equilibrium? The tangency point has the property that the slope of the IC = the slope of the budget line. Or

$$MRS_{12} = 1 + r.$$

There is, however, an important qualitative difference between the two panels. In panel (a), the consumer is a borrower (as E_G lies below N on the line AB); his

Figure 5.7 Consumer's Equilibrium



savings are negative. In panel (b), the consumer saves a positive amount and is a lender (as E_H lies above N on the line AB).

Realise now that what we have outlined so far is an economic analysis of consumption and savings as an application of the indifference curve analysis. We can now further apply it to understand how income and interest changes affect these decisions.

Changes in Incomes and the Interest Rate

Suppose there is an increase in the present income. This shifts out the budget line and the situation is exactly analogous to income effect studied in the last chapter. It is reasonable to assume that both current consumption and future consumption are normal goods (since these are broad aggregates rather than specific goods). This means that as the present income increases, current consumption and future consumption both increase. What about savings? Since future consumption also increases with present income, only a fraction of an increase in present income goes towards present consumption. This implies that savings increase.

In macroeconomics you must have come across the concepts **marginal propensity to consume (***MPC*) and **marginal propensity to save** (*MPS*). In the present context they are defined respectively as the increase in present consumption and savings per unit increase in present income. What we are saying then is that $MPC > 0$ and $MPS > 0$. Further, both MPC and MPS being positive, and because $MPC + MPS = 1$, it follows that $0 < MPC, MPS < 1$.

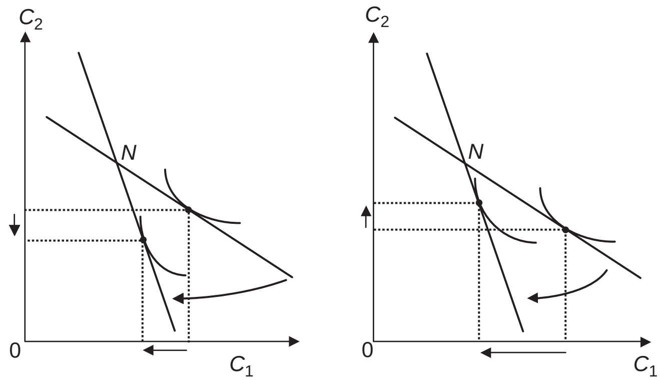
We thus have a 'microeconomic foundation' here as to why MPC and MPS are positive and less than one. The important underlying assumption is that both current consumption and future consumption are normal goods.

Consider now an increase in future income. This also shifts out the budget line and given our assumption of 'normality', it means an increase in present and future consumption also. However, as the present consumption increases while the present income is unchanged, it implies a decrease in the savings.

Since the budget line shifts out in each case, the consumer is always on a higher indifference curve. That is, a consumer always benefits from an increase in the present income or the future income.

Finally we analyse an increase in the interest rate. As we have seen before, the budget line rotates clockwise, pivoting on the point of no-borrowing-no-lending. How present and future consumptions change depends on whether the consumer is a borrower or a lender. Consider Figure 5.8. Note that in both panels, the equilibrium points, before and after a change in the interest rate, lie below the point N, showing that the consumer is a borrower. Observe that in panel (a), both present and future consumptions fall as the interest rate increases, whereas in panel (b), the present consumption falls but the future consumption rises. Combining the two, we can say that, if a consumer is a borrower, an increase in the interest rate leads to a decrease in the present consumption, while future consumption may increase or decrease.

Figure 5.8 Interest Rate Increase Affecting a Borrower



(a) Present and Future Consumptions Fall

(b) Present Consumption Falls But Future Consumption Increases

This result can be explained in terms of substitution and income effects. An increase in the interest rate increases the relative price of present consumption in terms of future consumption (that is, present consumption is now more 'expensive' relative to future consumption than before). Thus, by the substitution effect, the consumer will reduce current consumption and increase future consumption. Furthermore, since he is a borrower, an increase in the interest rate reduces his real income. Assuming normality, this would tend to reduce both present and future consumptions. Hence the net effect is that present consumption decreases unambiguously whereas future consumption may increase or decrease.

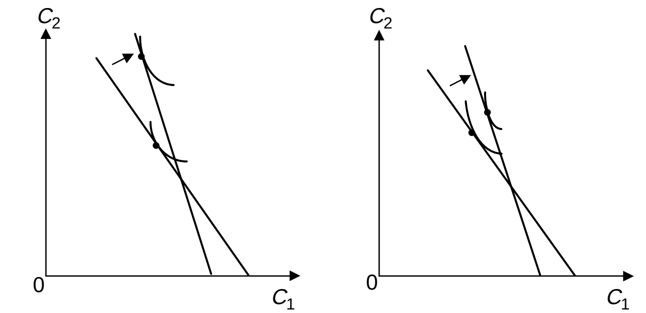
How does an increase in the interest rate affect savings? Mark that present consumption falls, while there is no change in the present income. This implies that savings, equal to the excess of present income over present consumption, increase.[6] Does the consumer benefit or lose, in terms of welfare? See that in both panels the consumer moves to a lower indifference curve. Thus he is worse off. Why? Because he is a borrower, an increase in the interest rate tends to reduce his real income.

The lender's case is shown in Figure 5.9. Combining the effects shown in both panels we see that as the interest rate rises, present consumption may increase or decrease while future consumption increases unambiguously. It is because, by the substitution effect, present consumption falls and future consumption rises as in the case of a borrower. But, being a lender an increase in the interest rate tends to increase real income. Given normality, this would tend to increase consumption in both periods. Therefore, the net effect on present consumption is ambiguous, while that on the future consumption is positive.

How are savings affected? Since present consumption may increase or decrease while the present income remains the same, unlike for the borrower, the lender's

[6]That is, the amount borrowed decreases.

Figure 5.9 Interest Rate Increase Affecting a Lender



(a) Present Consumption Falls but Future Consumption Increases

(b) Present and Future Consumptions Increase

savings may increase or decrease. The reason lies in the underlying substitution and income effects. Finally, we notice from Figure 5.8 that a lender is always better off with a higher interest rate. This is expected, since an increase in the interest rate tends to increase the real income of a lender.

Economic Facts and Insights

- Consumer surplus measures the level of welfare associated with a particular price of a commodity.
- An increase (or a decrease) in the price of a product leads to a decrease (or an increase) in the consumer surplus.
- Equilibrium consumption bundle in a kind-subsidy programme is always available in a cash-subsidy programme but not vice versa (presuming that both programmes involve the same amount of subsidy). Thus the latter programme offers more trading opportunities compared to the former and hence is preferred by the consumer.
- Equilibrium consumption bundle in an indirect tax programme is always available in an income tax programme but not vice versa (presuming that both programmes yield the same amount of revenues). Thus the latter programme offers more trading opportunities compared to the former and hence is preferred by the consumer.
- A person's savings can be interpreted as his future consumption.
- The interest rate reflects the relative price of present consumption in terms of future consumption.

(continued)

- Assuming that both present consumption and future consumption are normal goods, an increase in the present or future income implies more present and future consumption. Moreover, 0 < *MPC*, *MPS* < 1.
- An increase in the interest rate tends to increase (or decrease) the real income of a lender (or borrower) and, therefore, has a positive (or negative) effect on present consumption. Together with the substitution effect, it implies an effect of an ambiguous increase in the interest rate on savings by a lender and a positive effect on savings of a borrower. This means that the amount lent by a lender may increase, decrease or remain unchanged as the interest rate rises, while the amount borrowed by a borrower falls unambiguously.

EXERCISES

5.1 What is the definition of consumer surplus?

5.2 Your house-maid asks you for her *Diwali* gift. You offer her either Rs 100 or some item of Rs 100 that she can use. Which will she prefer and why?

5.3 Do the programmes of cash-subsidy and income tax shift the budget line of a consumer in a similar way? Explain.

5.4 Do the programme of kind-subsidy and an indirect tax shift the budget line of a consumer in a similar way? Explain.

5.5 Give examples of direct taxes.

5.6 Give examples of indirect taxes.

5.7 Suppose we compare an income tax programme with an indirect tax on a good, which a consumer does not consume. Which programme would he prefer?

5.8 Explain why the equilibrium bundle chosen by a consumer under a cash-subsidy or a direct tax programme is always available under a kind-subsidy or an indirect tax programme, given that the budgetary implications of the two subsidies or tax programmes are same to the government.

5.9 Intuitively explain why an individual would prefer a cash-subsidy programme to a kind-subsidy programme.

5.10 If individuals prefer cash-subsidy to kind-subsidy and both programmes cost the same to the government, why in reality do we see examples of kind-subsidy?

5.11 Intuitively explain why an individual would prefer a direct tax like income tax to an indirect tax on the consumption of a good.

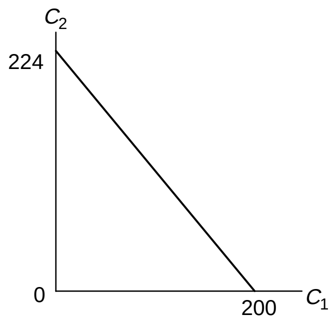
5.12 In reality an individual or a household consumes many goods. Then how can the present consumption or the future consumption be treated as a single good in the consumption-savings decision model?

5.13 Consider the consumption-savings model. The present and future incomes are Rs 10,000 and Rs 15,000 respectively. The interest rate is 10 per cent. Draw the intertemporal budget line indicating the intercepts and the slope.

5.14 Referring to the previous question, suppose that the consumer's present consumption is Rs 12,000. What is then his future-period consumption?

5.15 What is meant by intertemporal rate of substitution? Does it increase with an increase in the interest rate? Explain.

5.16 The intertemporal budget line is given as follows. What is the interest rate?



5.17 'Increases in present income and future income have a similar impact on a person's current savings'. Defend or refute.

5.18 Explain how a decrease in the interest rate would change the intertemporal budget line of an individual.

5.19 Explain in words how a decrease in the interest rate would affect the amount lent by a lender through income and substitution effects.

5.20 Explain in words how a decrease in the interest rate would affect the amount borrowed by a borrower through income and substitution effects.

6 Demand Estimation and Forecasting

CONCEPTS

- Regression Method
- Linear Regression
- Independent Variable
- Dependent Variable
- Slope Coefficient
- Simple Regression
- Demand Forecasting
- Medium-term Forecasting
- Expert Opinion
- Consumer Polls and Surveys
- Time Series Plot/Graph
- Seasonal Component
- Randomness or Error
- Scatter Diagram
- Regression Line
- Explanatory Variable
- Intercept Coefficient
- Least-Squares Estimates
- Multiple Regression
- Short-term Forecasting
- Long-term Forecasting
- Delphi Method
- Time Series Analysis
- Trend Component
- Cyclical Component
- De-seasonlising

We have learnt about how various factors like own price, prices of related goods and income affect the demand for a particular commodity or service. The demand function introduced in Chapter 2 is a symbolic (mathematical) expression of such cause-effect relationships. However, what is the exact (quantitative) relationship, say, between own price and quantity demanded of a product in an *actual* market situation? For instance, what is the exact shape and position of the household demand curve for electricity in the state of Karnataka? What does the demand curve for petrol look like for the whole of India?

It is very useful to know the quantitative nature of a demand curve because, if we know it, we can compute various elasticities like price elasticity and income elasticity. In Chapter 2, we have already discussed the potential applications of elasticities in managerial, business and other environments. Essentially, elasticities can be utilised to predict demand or changes in demand, based on assumptions about prices, income and so on as well as the changes in these. One of the objectives in this chapter is to briefly outline the procedure of 'estimating' or quantifying a demand function by using data on quantity demanded, prices and income and so on.

In a business environment, in particular, the forecasting of demand is quite important. A steel manufacturer in India would be interested to obtain a forecasting on the demand for steel in India, given the extent of competition from imports of steel from Korea and other countries. To make production plans, Maruti Udyog should be interested in the forecasting of demand for autos in the next few years to come. Forecasting is important not just at the national level but also at a local or regional level in deciding, for example, whether to open a local or regional office. In this chapter, we study demand forecasting as well.

DEMAND ESTIMATION BY REGRESSION METHOD

Suppose that a local small-scale garment company makes jeans of its own unique brand and sells these in various outlets in several cities. The general sales manager of this company has collected data from different outlets on the price charged and the number of jeans sold of its brand in a month. This is given in Table 6.1.

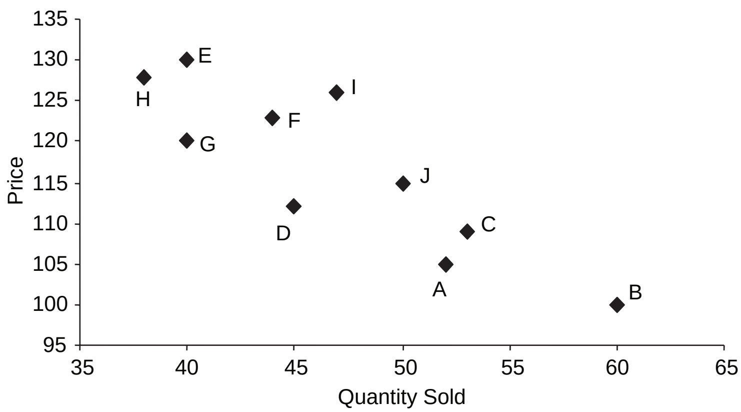
The various price–quantity combinations are plotted in Figure 6.1, which is a **scatter diagram**, generally defined as the plot of data on various combinations of two variables. Here the two variables are price and quantity sold. This diagram indicates the nature of the relationship between the two variables. For instance, Figure 6.1 roughly tells that lower prices are associated with higher quantities demanded, that is, the relationship between quantity demanded and the own price is negative.

How does the demand relation for this particular brand of jeans look? According to the regression method, it is the line or the curve 'that best fits the data in the scatter diagram' in some mathematical sense.

Table 6.1 Price–Quantity Sold Data

Outlet	*Price (Rs)*	*Number of Jeans Sold*
A	105	52
B	100	60
C	109	53
D	112	45
E	130	40
F	123	44
G	120	40
H	128	38
I	126	47
J	115	50

Figure 6.1 Scatter Diagram

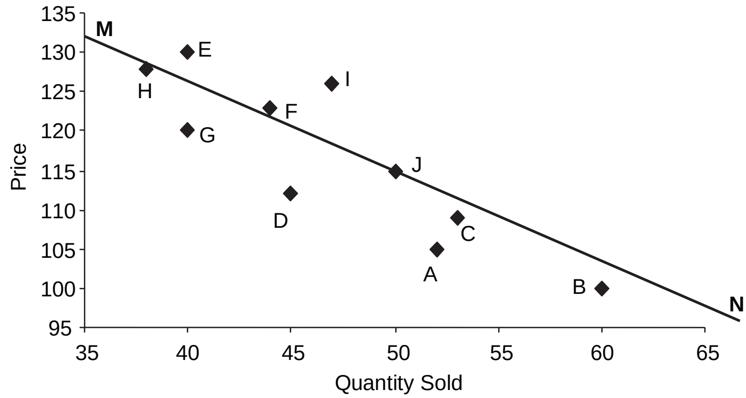


The regression analysis typically assumes a straight line relationship between the two variables involved. If so, it is called **linear regression**. Thus, in our example, assuming linear regression, we look for a straight line that best represents the points A, B, C, …, J. In Figure 6.2, MN is shown to represent this line, and it is the estimated demand function. In terms of regression, MN is called the **regression line**.

More exactly and algebraically, suppose that we want to estimate the effect of a variable X on another variable Y. In the regression analysis, X is then the **independent** or the **explanatory variable** and Y the **dependent variable**, meaning that the changes in X govern the changes in Y not but vice versa. For example, in estimating a demand function, X will be the own price and Y will be quantity sold/demanded since a demand relationship describes how a change in price affects a change in the quantity sold.

Let the various observations in the data on X and Y be denoted as (X_1, Y_1), (X_2, Y_2) and so on. In our example, in Table 6.1, (X_1, Y_1), (X_2, Y_2) and so on refer to

Figure 6.2 The Regression Line



entries A, B, C, ... up to J. Next, let the assumed straight line or the linear relationship between X and Y be generally stated as:

$$Y = a + bX,$$

where a and b are constants; typically a is called the **intercept coefficient** and b (the multiplicative term with the independent variable) is called the **slope coefficient**.[1] The regression analysis estimates these coefficients a and b and this completely defines the linear regression equation.

Intuitively, you can see that the regression line must be the one that passes through the scatter points such that its 'aggregate distance,' in some sense, from these points is minimised. More precisely, a and b are chosen such that the sum of the squares of the distance, along the axis measuring the dependent variable, between the scatter points and the regression line is minimised. This is why the estimates of a and b are called the **least-squares estimates**.

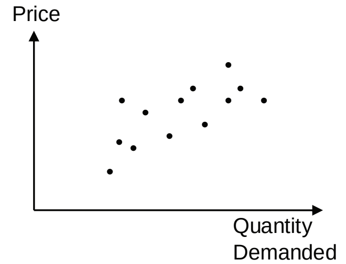
There are interpretations of a and b. Since b is the coefficient of X, its magnitude and sign capture the marginal effect of the explanatory variable X on Y. However, Y may also be affected by factors other than X. The coefficient a represents the effects of other possible explanatory variables outside of the regression analysis.

In our example given in Table 6.1, the regression or the least-squares estimates turn out to be $a = 115.645$ and $b = -0.588$.[2] Note that b is negative, reflecting a negative relationship between the price and the quantity demanded, as can be visually inspected in Figure 6.1. If, instead, we obtained a scatter diagram for a

[1] This is because if we graph X and Y respectively on the x-axis and y-axis, a is the intercept on the y-axis and b is the slope of the line.

[2] Any standard textbook on statistics will have the least-squares formulae for computing a and b.

Figure 6.3 Scatter Diagram Showing a Positive Relationship



demand relationship as shown in Figure 6.3, b would be positive, indicating the case of a Giffen good. In general, in the regression analysis, a positive or a negative relation between two variables will lead to a slope coefficient having a positive or negative sign respectively.

In the above, we have described what is called a **simple regression**, defined as one that has one explanatory variable. More desirably, in estimating a relationship, other relevant explanatory variables should be included—for example, the consumers' income, the prices of related goods and so on—in estimating a demand function. Of course, which variables are included in practice depends on the availability of data. In any case, in principle when more than one explanatory variable is included in a regression, it is called **multiple regression.**

Algebraically, let there be n explanatory variables available, namely, $X_1, X_2, ..., X_n$. Then a multiple, linear regression can be stated as:

$$Y = a + b_1X_1 + b_2X_2 + \ldots + b_nX_n,$$

where b_1 captures the marginal effect of X_1 on Y, b_2 captures the marginal effect of X_2 on Y and so on. The task is to estimate a, b_1, b_2, et cetera.

Although multiple regression is more complicated, the general principle behind it is the same as that of simple regression. The idea is to choose the various coefficients such that the regression equation best fits the data points.

In summary then the regression analysis has the following elements:

(*a*) It requires data on the dependent variable and some of the explanatory variables.

(*b*) A mathematical equation describing the relationship between the dependent variable and the available explanatory variables is postulated. Typically, a linear relationship is assumed. If so, it is called linear regression.

(*c*) In the linear regression, the intercept coefficient and the slope coefficients are estimated on the criterion that the sum of squares of the difference between the actual values of the dependent variable and its values according to the regression equation is minimised (In simple regression, it reduces to the sum of squares of the distance, along the axis measuring the dependent variable, between the scatter points and the regression line.) Such estimates of the coefficients are called the least-squares estimates.

(*d*) The sign and the magnitude of a slope coefficient show the sign and magnitude of the marginal effect of the corresponding explanatory variable on the dependent variable. In case of a demand estimation, the slope coefficient of the own price should turn out to be negative (unless it is a Giffen good).

Does the regression method give a 'perfect' estimation of what the true relationship (for example, demand relationship) is? The answer is no and never. However, there is a general question of how good a fit the regression line is in relation to the data. That depends on the explanatory variables chosen (based on available data) as well as variations in the values of the explanatory variables. For example, if a relevant explanatory variable is missing (due to data problems or because it has not occurred to the researcher that the particular variable is pertinent), then the fit will be poor. This does not, however, mean that regression is a bad method. (Do not lose sight of the fact that nothing is perfect.) If the fit of the regression equation is good (and there are statistical measures to estimate what is called the 'goodness of fit'), then regression estimates can be considered reliable.

DEMAND FORECASTING

Earlier, we gave examples of how demand forecasting can be useful to managerial decision-making within a firm. It can also be useful for the government in policy making. For example, a forecasting on the demand for petrol may guide the government's policy toward petrol pricing, building of reserves and so on. A good demand forecasting for HIV vaccine is considered to be important for developing countries where AIDS is a growing phenomenon.

There are different kinds of demand forecasting depending on (*a*) the time horizon and (*b*) the method involved. In the business world, quantity decisions that ought to change periodically in response to demand will benefit from a short-run forecasting. Examples include forecasting on inventory demand and demand for various raw materials. As a rough approximation, a forecasting up to six months ahead can be regarded as a **short-term forecasting**. A **medium-term forecasting** on demand (say 2–3 years hence) will be useful for decisions like a plant expansion or contraction and entering a new market. A **long-term demand forecasting**, over say 5 years, will be critical in decision-making such as setting up of R&D (research and development) infrastructure for the introduction of a new product or product line or setting up a new major plant.

Of course, whether it is a short-term, medium-term or long-term demand forecasting, in order for a forecasting to be useful to a manager, the results of it must be provided with a sufficient lead time.

There exist various demand forecasting techniques or methods. In what follows, we will discuss three of them: expert opinion, consumer polls and surveys, and time series analysis.

Expert Opinion

Many businesses often rely on expert opinion. Experts are presumably knowledgeable about the pattern of demand for certain products or industry in a certain region. In giving an opinion, an expert may use his intuition only or some data analysis. If the manager who seeks the forecasts is constrained to obtain one expert opinion, then that expert's opinion is most likely the forecast to be used by the manager.

But a manager may be in a position to ask more than one expert. Suppose different experts give very different opinions. How does the manager use this information? There is something called a **Delphi technique** that can be used—it essentially involves asking the experts repeatedly. It can be understood through a simple example. Suppose that there are three experts A, B and C. Over the next five years, according to A, the market demand for a product will grow at 10 per cent. B says that this growth rate will be 0, while C thinks that the market will grow at 6 per cent. There are thus huge differences in the expert opinion. Delphi method involves asking each expert again to submit revised estimates by providing him what others have predicted. For instance, A will be told that the two other experts have projected the growth rate at 0 per cent and 6 per cent. Given a chance to reevaluate, A may lower his projection taking into account what others have said. Similarly C may revise his projection upwards. Suppose that in the next round, the numbers obtained are 7 per cent, 4 per cent and 5 per cent. These numbers are closer to each other than the earlier numbers. This technique involves obtaining reevaluations until the opinions converge or nearly converge. In the current example, a manager may think that the numbers 7, 4 and 5 are close enough for the purpose and he may decide the number say 5 or 5.5 (something very close to the average). Or the manager may decide to go for one more round.

There are, however, problems with the Delphi method. First, good experts are typically very busy people and their fees are very high. This may be a major constraint for a manager. Second, it is not necessary that experts may want to revise their estimates. Because if they do it frequently, it may reflect that they are not very confident about themselves and, thus, that they do not possess a high expertise. They may not wish to hamper their reputation.

Despite these limitations, expert opinions on forecasts are much sought after in businesses.

Consumer Polls and Surveys

Various consumer surveys are sometimes used for forecasting. For example, National Council of Applied Economic Research (NCAER) based in Delhi conducts an annual survey called Market Information Survey of Households (MISH). A sample of 3 lakh households is used covering all districts in India. The survey asks the households about their income and their purchase of various products, both durable and non-durable, among other things. As the website of the NCAER indicates, this survey data is used, among others, by businesses like Bajaj Auto, Godrej and Bank of America.

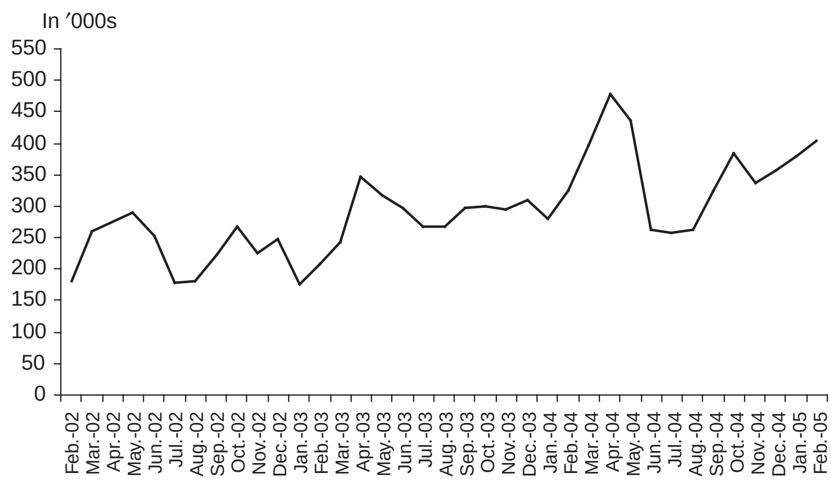
There are market research companies which conduct consumer polls and surveys on specific products and sell their forecasting to their clients.

Time Series Analysis

It is a statistical method. In regression, the dependent variable (for example, quantity demanded of a product) is estimated on the basis of how it is influenced by 'other factors' or independent variables (for example, own price, consumers' income). In the time series method, the idea is not to emphasise the relationship of one variable to some others, but to rely on the existing behaviour of the variable itself over time and extrapolate from that the likely future values of the variable.

The existing pattern of change in a variable over time can be seen from a **time series plot** or **graph**, which measures calendar time along the x-axis and the variable itself on the y-axis. For instance, Figure 6.4 shows the time series plot of monthly data on the number of refrigerators sold/consumed in India.[3]

Figure 6.4 Time-series of Total Number of Refrigerators Sold in India



[3]This is calculated as the total domestic production, plus imports and minus exports of refrigerators.

In general, let time be denoted by t. Suppose that we have data on variable Y (for example, quantity demanded of a product) for $t = 0, 1, 2, \ldots, T$. The question is: what is the forecast of X for $t = T + 1$, $T + 2$ and so on.

The time series analysis assumes that the value that a variable takes at any given point of time has four components.

One is a **trend** component, capturing any long-run trend of growth or decline in the variable over time. For instance, if we have a time series plot of the total number of cell phones sold in India in the recent years, we are certain to see an upward, that is, a positive trend. But if we look at the total number of 14-inch black and white TV sets sold, we expect to see a negative trend.

Another in the **seasonal** component. For many products and services, all other things remaining the same, the quantity demanded depends on the prevailing season. If we consider the demand for bottled mineral water, the demand is much higher in the summer than in the winter months. The demand for visits to general physicians for common sickness like cold, cough and flu is higher in the winter than in the summer months. Refrigerator sales are likely to reach their peak in months approaching or during summer. Indeed, Figure 6.4 shows that the sales are generally very high during March, April and May as compared to other months.

The third component is **cyclical**, relating to business-cycle reasons. For example, if the economy is passing through a recession, the demand for real estate is typically slack.

The last component is **randomness** or **error** resulting from unanticipated and unexplained events.

If we denote these four components respectively by T, S, C and R, then the time series analysis typically postulates:

$$Y = T' \, S' \, C' \, R.$$

Forecasting involves methods which first attempt to neutralise the seasonal component by various averaging methods (that is, smoothing techniques). It is called **de-seasonalising** the data, which essentially smoothens the data. Many computer softwares are available that de-seasonalise a time series data.

Once the data is de-seasonalised, there are several methods available to filter out the T (trend) and the C (cyclical) components separately or jointly. Whatever left is, by definition, the R (random) component.

The forecasting obtained in an actual situation depends on the sophistication of the methodology used, tools, computer software, time available and, not to mention, the skill of the individual entrusted with the task of forecasting.

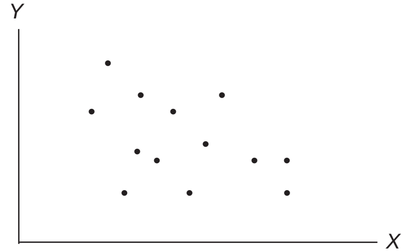
As an illustration, a very simplistic procedure will involve fitting a linear trend line to a de-seasonalised time series data by the regression method and then extrapolating from it to a future date. That is, a regression equation of the form $Y = a + bt$ is assumed, where t denotes calendar time (the independent variable) and Y is a de-seasonalised variable under consideration (for example, quantity demanded of a product). Using the least-squares estimates of a and b, we can set t to any future date (for example, month or quarter) and compute the corresponding value of Y, which is its forecast for that date.

Economic Facts and Insights

- The scatter diagram indicates the nature of relationship between two variables.
- In the regression method, a demand function is estimated by using data on quantity demanded of a product and its determinants such as own price, prices of related goods and income.
- Under linear regression, the slope coefficient measures the marginal effect of the respective explanatory variable on the dependent variable.
- If the slope coefficient is positive (or negative), it indicates a positive (or negative) relationship between the independent variable and the dependent variable.
- The intercept coefficient estimates the effect of other variables not included in the regression.
- The regression method does not yield the perfect estimation of the true relationship between the dependent variable and the set of independent variables.
- Demand forecasting is achieved by soliciting expert opinion, using consumer polls and surveys or by using time series techniques.
- In the time series method, the idea is not to emphasise the relationship of one variable to some others, but to rely on the existing behaviour of the variable itself over time and extrapolate from that the likely future values of the variable.
- In the time series method, the data is de-seasonalised (by using smoothing techniques) before the trend or the cyclical component is extracted.
- Linear trend line is obtained by regressing the de-seasonalised time series data with respect to time.

EXERCISES

6.1 Consider the following scatter diagram. If a regression line is fitted to this data, would it be positively sloped or negatively sloped? Explain.



6.2 What is linear regression? Which coefficient of linear regression indicates whether there is an increasing or a decreasing relationship between the two variables and why?
6.3 Suppose the data on price and quantity demanded for a commodity are plotted as a scatter diagram and the fitted regression line has a positive slope. Does it necessarily imply that the regression is yielding illogical result?
6.4 What is the criterion on which the least-squares regression coefficients are estimated?
6.5 In linear regression, what is the interpretation of the 'intercept coefficient'?
6.6 In linear regression, what is the interpretation of the 'slope coefficient'?
6.7 'In a multiple linear regression there are several intercept coefficients and several slope coefficients'. Comment.
6.8 In the context of a multiple linear regression of estimating a demand function, the data on quantity demanded, own price, the price of a substitute good and income are used. Which sign will we expect of the slope coefficient of the variable representing the price of a substitute good and why?
6.9 'The regression method does not give a perfect fit of the estimated line for the dependent variable. Therefore it is highly unreliable'. Comment.
6.10 How is demand forecasting useful to a manager? Substantiate your answer with suitable examples.
6.11 Briefly describe the Delphi method of forecasting demand.
6.12 What is the essential difference between the regression method and the time series method?
6.13 The following table gives monthly time series data on the number of colour televisions sold in India (in lakhs). Plot this data. Show that there is some seasonality, that is, during some months sales are either very high or low relative to the rest of the months in a year. Give a probable reason for the presence of seasonality.

Month/Year	*Sales*	*Month/Year*	*Sales*	*Month/Year*	*Sales*
May 02	5.8	May 03	4.3	May 04	7
June 02	5.7	June 03	4.61	June 04	6.8
July 02	4.9	July 03	5.74	July 04	7
Aug 02	5.62	Aug 03	7.07	Aug 04	7.25
Sep 02	5.2	Sep 03	9.82	Sep 04	10.5
Oct 02	8.82	Oct 03	10.77	Oct 04	11
Nov 02	5.7	Nov 03	6.2	Nov 04	7.5
Dec 02	6.8	Dec 03	6.8	Dec 04	6.68
Jan 03	7.86	Jan 04	7.05	Jan 05	7.07
Feb 03	7.9	Feb 04	7.3	Feb 05	7.36
March 03	5.37	March 04	8.27	March 05	5.65
Apr 03	4.18	Apr 04	6.75	Apr 05	6.97

6.14 Briefly describe what is meant by the 'trend component' of a time series data.
6.15 What are the different components of a time series data?
6.16 Briefly describe any simple procedure to obtain the trend line fitting a time series data.
6.17 Briefly describe various ways of demand forecasting.

7 Demand for Assets

CONCEPTS

- Liquidity
- Rate of Return
- Risk
- Present Value or Present Discounted Value
- Fixed Deposits
- Initial Public Offering (IPO)
- Stock Exchange
- Securities
- Portfolio
- Capital Gains and Losses
- Yield
- Discounting
- Current and Savings Accounts
- Shares or Stocks
- Dividends
- Bonds
- Mutual Funds
- Diversified Portfolio

In the last chapter we analysed how a rational individual allocates his earnings between consumption and savings. The amount earmarked for consumption is spent on different goods and services. What does a person do with her savings? She keeps it in cash, bank accounts, real estate, financial instruments and so on. This is the same as saying that she 'acquires' or 'demands' assets or makes investments. Put differently, a person's or a family's assets are acquired from the past and present savings. (Some assets may be obtained from others as gifts.)

What are the factors that determine how much an individual should demand or invest in a particular asset? In a sense, it is similar to asking, in terms of consumer theory, what factors determine how much an individual should spend on a particular commodity. There are, however, differences between the principles behind the demand for goods and those behind the demand for assets.

ASSET CHARACTERISTICS

In demanding a good (for consumption), the benefit is measured in one dimension, namely, the psychological satisfaction from consuming the good (which we called utility). In contrast, assets do not generate utility directly but via their capacity to generate purchasing power that can be used for consumption. The benefit from having an asset depends on three dimensions or characteristics of the asset: liquidity, returns and risk.

Liquidity

It means the ease with which a particular asset can be exchanged for money at the face value. Liquidity is important because people need money on a short-term basis to meet unforeseen needs. Obviously, your money invested in a bank in the form of a current or a savings deposit is more liquid than money invested in real estate. You can walk to your bank and cash a check in a matter of minutes or you can withdraw money from your deposit by using ATM machines. But if you have a piece of property and you want cash immediately by selling it instantly, you are likely to lose a whole lot relative to its market price. That is, you do not get back its face value. To get its worth, you need to advertise or spread the word that your property is for sale; then people come; some people may make offers, and this whole process can take weeks, sometimes months and even years.

Rate of Return

An investment is capable of generating earnings only after a certain period of time. Hence time is an important element. Suppose you have put Rs 1,000 in your savings account, the annual interest rate on the savings deposit is 6 per cent and the bank pays interest after one year.[1] Then the interest income of Rs 60 accrues to

[1]In India, banks pay interest once in every six months.

you after one year. The principle behind someone or some institution paying you an interest for the use of your money is that there is a *time value of money*—if you are not using your money for a given length of time, someone else can borrow it, use it for some purpose and, therefore, be willing to pay you something for the use of that money.

Going back to the example, if you continue to maintain your deposit, the interest earnings accrue year after year. This is called the **yield** of an asset, meaning periodic income or earnings generated by an asset. In our example, the yield is Rs 60 every year.[2] The yield and the initial amount invested determine the **rate of return** associated with an asset. In our example, the rate of return equals 60/1000 = 0.06 or 6 per cent, which is same as the interest rate.

Consider another example. Suppose Mr Ram Lal took out Rs 10 lakh from his savings to buy a flat for the purpose of renting out. It is given on a rent of Rs 6,000 per month. Assume that Rs 1,000 per month is spent on its maintenance. The net rental income is then Rs 5,000 per month, which translates to an annual yield of Rs 60,000. It is expected that after one year the flat will be sold at Rs 10.5 lakh. In this example the initial cost of investment is Rs 10 lakh, and over the year, the return on this investment is calculated as:

$$\frac{60,000}{10,00,000}+\left(\frac{10,50,000}{10,000,00}-1\right)=0.06+0.05=0.11=11\%.$$

The first term is the ratio of yield to the cost of the investment, as in the earlier example. The term inside the brackets is new and equals what is called **capital gains**. In general, if the expected sale value in the future exceeds (or falls short of) the cost, there is a capital gain (or capital loss). We have the formula,

$$\text{Rate of return of an asset} = \frac{\text{yield + expected sale value}}{\text{cost of the asset}} - 1.$$

There is an intuitive way to look at the return from an asset, namely, as an internal or implicit interest rate received. To understand this consider the same example. The cost of investment is Rs 10 lakh whereas after one year the asset pays back the rent (Rs 60,000) plus the sale value (Rs 10,50,000), equal to Rs 11.1 lakh. Ask yourself the following question. Suppose you deposit Rs 10 lakh in a bank. At what interest rate would you obtain Rs 11.1 lakh in a year (as principal plus interest)? That is, you set up the equation: $10,00,000(1 + r) = 11,10,000$ and solve for r, which is the internal or the implicit interest rate. We obtain $r = 0.11$, which is same as the rate of return calculated by taking into account the yield and the capital gains. Indeed, this way of calculating the rate of return of an asset can be generalised to more than one period (see the Numerical Example 7.2 below).

In layman's terms, the rate of return from holding an asset is the ratio between the excess money or purchasing power it generates per period and the initial cost

[2]This answer assumes that Rs 60 is taken out for consumption every year, so that the principal remains the same. Otherwise, the interest earnings will be added to the principal (that is, these will be compounded) and the yield will keep increasing over time even if the interest rate is fixed at 6 per cent.

of the asset (investment). An asset's rate of return is one of its most important characteristics. Ceteris paribus, the higher the rate of return the greater will be demand for an asset.

NUMERICAL EXAMPLE 7.1

Ahmed is in the business of buying and renting out state of the art cell phones to tourists from abroad. A new cell phone costs Rs 10,000. The rental income for a year is Rs 6,000. By the end of a year, Ahmed expects to sell this phone in the secondhand market at Rs 7,000. Having a cell phone and using it in this business can be thought of as an asset. What is Ahmed's rate of return from this asset?

The cost of the asset = Rs 10,000. Its yield = Rs 6,000 and the expected sale price = Rs 7,000. Applying the above formula, the rate of return is equal to:

$$\frac{6000+7000}{10000}-1=0.3=30\%.$$

NUMERICAL EXAMPLE 7.2

Consider the example of Mr Ram Lal investing Rs 10 lakh in a rental property. Suppose he sells the property two years after he bought it at Rs 10.84 lakh and during these two years it generated a net rental income of Rs 60,000 each year. What is the rate of return of this investment by Mr Lal?

Let r denote the implicit interest rate. Since the asset is held for two years, consider a time horizon of two years. A deposit of Rs 10 lakh in the bank at the interest rate of r would grow into Rs 10,00,000 $(1 + r)^2$ at the end of two years. On the other hand, if this asset is held, the net rental income of Rs 60,000 obtained after one year becomes Rs 60,000$(1 + r)$ after two years. Add this to the net rental income and sale value received by the end of the second year (Rs 60,000 + Rs 10,84,000). Altogether it amounts to Rs 60,000$(1 + r)$ + Rs 11,44,000. Hence set up the equation:

$$\text{Rs } 10{,}00{,}000(1 + r)^2 = \text{Rs } 60{,}000(1 + r) + \text{Rs } 11{,}44{,}000.$$

It yields the solution, $r = 0.10$. Hence the rate of return is equal to 10 per cent.[3]

Risk

Unless we are talking about deposits in a current/savings account or in a fixed deposit, almost any other investment carries some uncertainty on its return. In the last example, the rental income was given and known to be equal to Rs 60,000 per year. But it may not be known for sure before the flat is bought. Similarly, the price at which it will be sold next year is only 'expected', not known for sure. On most

[3]As you will realise, such a calculation of rate of return is intimately related to 'discounting' and 'present value analysis' to be introduced later.

investments, there is some degree of uncertainty and risk associated with its return.

The concept of **risk** can be understood further via the following example. Suppose that you have a given amount of money for investment and you are looking at two options: Investment A and Investment B. The former provides a return of 0.10 (10 per cent) with a probability of 50 per cent and of 0.06 (6 per cent) with a probability of 50 per cent. We can then calculate the *expected return*, defined as the weighted average of possible returns. For Investment A, it is equal to $.10 \times .50 + .06 \times .50 = 0.08$ (8 per cent). Investment B offers a return of 0.13 with probability 50 per cent and 0.03 with a probability of 50 per cent. This also has an expected return of 8 per cent. Mark, however, that while both have the same expected return, the variability (measured by the statistical formula for variance) of return is greater with Investment B. We then say that investment B is riskier than Investment A. The higher the degree of variation of returns relative to the average, the greater is the risk.

Obviously, risk is an important consideration in acquiring an asset. All else the same, the higher the risk associated with an asset, the less will be the demand for it.

NUMERICAL EXAMPLE 7.3

Consider two stocks C and D. C gives a return of 10 per cent with a probability of 40 per cent and of 15 per cent with a probability of 60 per cent. D gives a return of 12 per cent with a probability of 50 per cent and of 14 per cent with a probability of 50 per cent. Calculate the expected return of each of these assets and determine which one is the riskier asset.

The expected returns of stocks C and D are respectively equal to:

$$E(C) = .1 \times .4 + .15 \times .6 = .13 = 13\% \text{ and}$$
$$E(D) = .12 \times .5 + .14 \times .5 = .13 = 13\%.$$

That is, both stocks give the same expected or mean return. Hence, the stock whose variance of returns is higher is the riskier one. Denoting variance by V, and applying the statistical formula for the variance,

$$V(C) = (.1 - .13)^2 \times .4 + (.15 - .13)^2 \times .6 = .0006$$
$$V(D) = (.12 - .13)^2 \times .5 + (.14 - .13)^2 \times .5 = .0001.$$

Hence stock C is the riskier asset.

FURTHER UNDERSTANDING THE RATE OF RETURN: DISCOUNTING AND PRESENT VALUE ANALYSIS

Putting money in an interest-earning safe deposit like a bank account or a government bond is considered a benchmark investment in the sense that it guarantees a minimum return on your investment. If you are thinking of some other forms of investment, then, apart from the risk factor, you must compare its return to the minimum return (the interest rate). The comparison is not trivial, however, if the yield on the asset varies from one period to another. A method called the 'present discounted value' or simply a 'present value analysis' is typically used.

In what follows, this method is illustrated via an example.

Suppose you are an established businessman in your home town. You can borrow Rs 33 lakh from a bank at an annual interest rate of 15 per cent, which must be paid in three years from now. You are contemplating to use this money to organise a week-long trade show in two years. You expect that contacting business firms, making local arrangements and so on will take two years. At the end of the second year, you expect to hold the first trade show and at the end of the third year the second show. When the show is on, you expect to get payments from businesses who commit to participate and from potential customers who would come to sample the products. Let us say that these payments, that is, receipts, will be Rs 6 lakh in the second year and (after the news of the success of the first trade show spreads) Rs 43 lakh in the third year. For simplicity, suppose that there is no uncertainty or risk—you know for sure that you would be able to raise these amounts. Also suppose that liquidity is not an issue and your only consideration is returns. Holding the trade show can be seen as an asset, which is going to provide yields for two years. You have to weigh the rate of return of this asset or investment against the cost of borrowing.

The benefits and costs of acquiring this asset are given in Table 7.1, where you can notice that interest cost of Rs 33 lakh at 15 per cent is Rs 4,95,000 per annum. If we just add up the costs over the three years it comes to Rs 47,85,000. The benefit at the end of the third year is higher, equal to Rs 49,00,000. Does it mean that it is a profitable investment?

Unfortunately, this simplistic comparison is flawed. It is because a rupee one year from now is not the same as a rupee two years from now as long as there is an interest payment involved. Hence a simple adding up of figures over the two or more years is something like adding apples and oranges. To obtain a correct comparison between benefits and costs, we must convert them to equivalent rupees in the same year. This is done through **discounting** and a **present value analysis**.

Discounting means converting a given sum of money to be paid or received in some future date to an equivalent sum of money in an earlier period. For instance,

Table 7.1 Benefits and Costs

	End of the First Year (Rs)	*End of the Second Year (Rs)*	*End of the Third Year (Rs)*
Benefits	**0**	**6,00,000**	**43,00,000**
Costs			
Interest	4,95,000	4,95,000	4,95,000
Principal	0	0	33,00,000
Total Repayment	4,95,000	4,95,000	37,95,000

consider a sum of Rs 1,21,000 to be paid or received in 2012. What is its discounted value as of 2010? To answer this, we first need to know how much one rupee in 2010 will be worth in 2012. Suppose that the annual interest rate, expected to prevail between 2010 and 2012, is r. Then one rupee lent in 2010 will fetch $1 + r$ rupees in 2011 and $(1 + r)^2$ in 2012. That is, one rupee in 2012 is worth $1/(1 + r)^2$ in 2010. Thus Rs 1,21,000 in 2012 discounted back to 2010 is equal to Rs $1,21,000/(1 + r)^2$. If the interest rate is 10 per cent, that is, $r = 0.10$, this sum is equal to Rs $1,21,000/(1.1)^2$ = Rs 1,00,000.

We define **present value** or the **present discounted value** (PDV), as the sum of the discounted values of payments to be made or received in future periods when the sums are discounted back to the present period. Consider now the example given in Table 7.1 and ask, what is the PDV of the costs of investing in the trade show? Note that the first-period payment of Rs 4,95,000 discounted back to the current period at the interest rate of 15 per cent is equal to Rs 4,95,000/(1.15) = Rs 4,30,434. Similarly, the discounted value of the payment of Rs 4,95,000 in the second period as of the present period is equal to Rs $4,95,000/(1.15)^2$ = Rs 3,74,291. The discounted value of the third year payment = Rs $37,95,000/(1.15)^3$ = Rs 24,95,274. Hence the PDV of the costs = Rs 4,30,434 + Rs 3,74,291 + Rs 24,95,274 = Rs 33,00,000 (after taking into account rounding errors), the amount of the loan taken. This should not be surprising because the loan is incurred in the current period and the interest rate associated with the loan is being used in the discounting process.

Let us now calculate the PDV of the benefits. In the first period, the benefits are zero and thus its discounted value is zero. The discounted value of the second-period benefit of Rs 6,00,000 as of the current period = Rs $6,00,000/(1.15)^2$ = Rs 4,53,686. That of the third year benefit = Rs $43,00,000/(1.15)^3$ = Rs 28,27,320. Therefore, the PDV of benefits = Rs 0 + Rs 4,53,686 + Rs 28,27,320 = Rs 32,81,006. In summary,

The PDV of Costs = Rs 33,00,000

The PDV of Benefits = Rs 32,81,006.

The latter is less than the former. We, therefore, conclude that it is not a profitable investment.

What does it mean in terms of the rate of return? It simply means that the rate of return on this investment falls short of the interest rate (= 15 per cent), which is the cost of borrowing.

Note that we would have arrived at the same conclusion even if no money had been borrowed and, instead, the businessman had used Rs 33 lakh from his own funds—as long as he would have obtained 15 per cent interest from lending this amount in the market.

NUMERICAL EXAMPLE 7.4

A truck, if bought today, is expected to generate a rent of Rs 1 lakh this year and every year over the next ten years. Beyond the 10th year, it will cease to operate and the scrap or the *kabadi* value of the metal in the truck will be Rs 10,000. Find the PDV of owning this truck if the interest rate is 8 per cent. Find the same if the interest rate is 11 per cent. In which case is the PDV less and why?

Choose a lakh as the unit. Then the discounted sum of values of the yields over the first year and the next years at 8 per cent interest rate is equal to:

$$1+\frac{1}{1.08}+\frac{1}{(1.08)^2}+....+\frac{1}{(1.08)^{10}}=7.71.$$

The discounted value of the scrap (*kabadi*) value = $(.1/1.08)^{10}$ = 0.046. Adding these,

$$\text{PDV at } 8\% = 7.71 + 0.046 = 7.756 \text{ lakh.}$$

Doing exactly the same calculation at the interest rate of 11 per cent,

$$\text{PDV at } 11\% = 6.924 \text{ lakh,}$$

which is less than the PDV at 8 per cent interest rate. It is less because the higher the interest rate, the smaller is the present discounted value of any future year's yield and the smaller is the present discounted value of the future scrap (*kabadi*) value as well.

NUMERICAL EXAMPLE 7.5

A multinational firm A has a patent for the technology of a product, which it wants to sell. There are two potential buyers from India: firm B and firm C. Firm B is willing to pay Rs 17.5 lakh for it upfront (instantly), while firm C is willing to pay 4 lakh each year for five years including the current year. For company A, selling the technology also involves sending its technical team at its own cost to set up the operations, which is the same whether firm B or firm C buys the technology. There is no uncertainty about the future payments from firm C. The interest rate is expected to prevail at 10 per cent. Furthermore, suppose that A is not in a position to sell to both firms. To whom would it sell its technology?

Selling technology can be regarded as an asset. The cost of both assets is the same, equal to the cost of sending the technical team. Since firm B has agreed to offer Rs 17.5 lakh upfront, firm A will sell to firm B or firm C accordingly as the PDV of payments from Firm C falls short of or exceeds Rs 17.5 lakh. At 10 per cent rate of interest, the PDV of payments (in lakh) from firm C equals:

$$4+\frac{4}{1.1}+\frac{4}{(1.1)^2}+\frac{4}{(1.1)^3}+\frac{4}{(1.1)^4}=16.679.$$

This is less than Rs 17.5 lakh. Hence firm A will sell the technology to firm B.

ASSETS OFFERED IN THE MARKET

We discussed above the profitability of holding a particular asset vis-à-vis holding a loan, that is, a safe asset. We did not take into account the fact that in reality there are many assets available in the market. Thus even if the holding of an asset is profitable in the manner demonstrated above, it does not necessarily mean that a person should invest in it because there may be more than one profitable investment prospect. Profitability is a necessary, not sufficient condition of making a rational investment. Also, we ignored the issues of liquidity and riskiness. The market place offers various assets with various degrees of liquidity, expected return as well as riskiness.

Ceteris paribus, the less liquid an asset, the higher is its return. For instance, a five-year fixed deposit offers a higher interest than a savings account deposit. Why? It is because if a savings account deposit did offer the same interest as a five-year fixed deposit, no one would invest in the latter—such a fixed deposit scheme would not exist. The market cannot sustain a low-liquidity and low-return asset together with a high-liquidity and high-return asset. This means that, from an individual investor's viewpoint, there is a trade-off between liquidity and returns. Similarly, there is a trade-off between expected return and risk. Among assets of equal liquidity, the market will only sustain high-expected-return-cum-high-risk assets along with low-expected-return-cum-low-risk assets.[4]

There are both real and financial assets. **Real assets** refer to real estate, machines (for example, trucks), metals (for example, gold, silver), antiques, paintings and so on, while **financial assets** refer to fixed deposits, stocks, bonds and such. Here are some of the common financial assets (or instruments) that are offered in the market.

Cash: This is the simplest form of a financial asset. You keep some of your savings in cash locked up in an *almirah* or a safe deposit box. This is the most liquid of all assets. What is the return? It is zero.[5]

Current and Savings Accounts: These refer to the money kept in your bank account. There are minor differences between the two: the former is slightly more liquid than the latter but less liquid than cash. You have access to cash kept at home seven days a week, 24 hours a day; this may not be true with your bank. Generally, these accounts pay a modest rate of interest, which is their return.

Fixed Deposits: They do not offer any yields on a periodical basis (for example, annually). Money is locked for a longer period. Compared to a bank account, these are less liquid and carry a higher interest rate. The principal and the accumulated interest earnings are paid together at the end of the maturity date.

[4]It is because if there are two assets, one of which offers a higher expected return combined with a lower risk compared to the other, no one would demand the latter and hence it would not exist in the market in the first place.

[5]It is assumed implicitly that the price levels remain unchanged over time. Otherwise, we need to take into consideration the expected rate of inflation.

Shares or Stocks: These are ownership certificates issued by firms or companies. Suppose a company *X* offers 2,00,000 shares to the public for the first time. This is called **IPO**, standing for **initial public offering**. (An IPO by a company is advertised through banks and sometimes through the print media.) Suppose that, through your bank or broker, you buy 10,000 shares at a given price (say Rs 10 each). Physically, it involves you writing a cheque to the company's bank for the said purpose and in return you are issued a certificate saying that you own 10,000 shares of company *X*. This means that you have 5 per cent (= 10,000/2,00,000) ownership in the company. If, in a year, the company earns a profit of say 40 lakh rupees and all of it is paid back to the shareholders, you will receive 5 per cent of it, that is, a cheque of two lakh rupees as **dividend** income. If, instead, the management of the company decides that 20 lakh rupees is to be ploughed back to the company's operations and the remaining 20 lakh is to be distributed to the shareholders, your dividend income will be one lakh rupees. In general, dividend payments constitute the yield to holding a stock. See Clip 7.1 for an account of a massive, unprecedented dividend payment scheme for the shareholders of Microsoft Corporation in 2004.

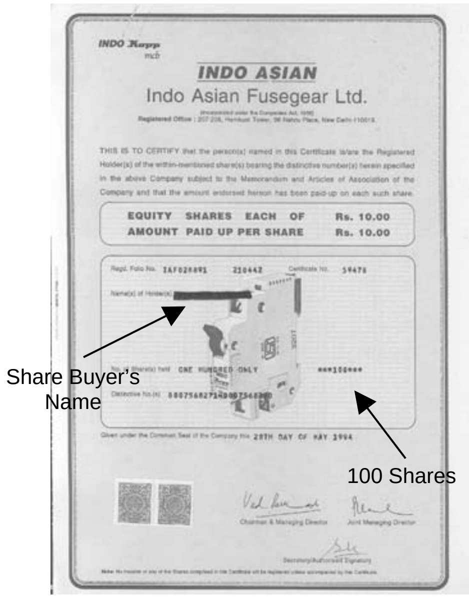
Figure 7.1 exhibits a copy of a real stock certificate, which shows an ownership of 100 stocks of Indo Asian Fusegear Ltd. The original purchase price from the company is Rs 10. As it is traded in the market, the share certificates are transferred to the names of new owners. There are columns at the back of the share certificate (not shown), which indicate the names of the transferees.

What happens if the company makes a loss? A shareholder gets nothing. In an unusual situation where the company has a large outstanding payment that

Clip 7.1: Massive Dividend Payments by Microsoft in 2004

By March of 2004, Bill Gate's Microsoft Corp. had amassed a cash of 60 billion US dollars after plough back. Typically, high-tech companies do not offer much dividend income (shareholders still benefit from holding its stock via capital gains) unless there are prospects of hostile takeover bids. There was a lot of pressure from the Wall Street to use this money for some purpose other than just earning a low interest rate. Various groups of investors in Microsoft's stocks offered different alternatives. Finally, the company decided to pay out \$3 per share as dividend. As of March 2004, there were 10.79 billion Microsoft shares floating in the market, which amounted to a total dividend of 32 billion dollars by Microsoft in the year 2004. This is something unprecedented in the history of the financial world. Interestingly, Bill Gates himself, who owns a lot of Microsoft stocks and stood to get 3.3 billion dollars, announced that this entire amount would go towards his Bill Gate's Foundation. This Foundation provides massive assistance towards the containment of AIDS in India.

Figure 7.1 A Share Certificate



cannot be met from its revenues or sale of assets, then a stockholder is personally liable to the extent of its shareholding in the company. For instance, let us say that you own 1 per cent of all shares in a tobacco company. In the court of law, this company is sued by a group of families over the deaths of and the health hazards faced by their relatives due to consuming the company's product. The court rules that the company must pay Rs 60 crore as compensation to the families of the victims.[6] However, the company is able to raise only 50 crore from selling its assets. Then the remaining 10 crore must be paid by the stockholders and your share comes to Rs 10 lakh, which you have to pay. In other words, holding a stock is accompanied by an exposure to risk.

There is also another source of risk associated with stocks, namely, capital gains or losses. This is because stocks are heavily traded in the market and their prices can fluctuate a great deal over time. But at the same time price fluctuations also provide the opportunity of making large gains (or a 'killing' so to speak) and indeed this is sometimes the motivation behind holding stocks.

Some companies list their stocks in a stock exchange. A **stock exchange** is where listed shares of various private and public limited companies are traded. See Clip 7.2 for an account of stock exchanges in India.

[6]This kind of a situation has never happened in India, but it has in a country like the US.

Clip 7.2: Stock Exchanges in India

There are six stock exchanges in India: Bangalore Stock Exchange, Bombay Stock Exchange (BSE), Calcutta Stock Exchange, Delhi Stock Exchange, Madras Stock Exchange and National Stock Exchange. The Bombay Stock Exchange and the National Stock Exchange (NSE) are both located in Mumbai. Among these the Bombay Stock Exchange is the oldest (started in 1875), the largest and the most cited in the media. National Stock Exchange is the newest, which started in 1993. BSE has over 400 branches across different cities and towns in India. Around 3,500 companies and 6,000 stocks are listed with BSE. (A large company may have different companies producing different products, with each issuing its own shares.) In terms of transactions, it is among the top five stock exchanges in the world. SENSEX is an index based on changes in the stock prices of 30 well established and financially sound companies listed with BSE. Movement of SENSEX reflects the general trend of stock prices over time. (BSE also constructs other indices and other stock exchanges have their indices too.)

In order to be listed in a stock exchange, a company must pay an initial fee and annual fees, and signs a contract with the Exchange to abide by its rules. Most rules pertain to transparency. For example, a listed company must disclose its shareholding pattern and profits on a regular basis. In return, the company's performance is seen nation-wide, which can then attract funds from numerous potential investors. The company also obtains various market-related analyses which, because of its membership, it can use for its benefit.

All stock exchanges are regulated by Securities and Exchange Board of India (SEBI), which ensures that the stock exchanges operate in an orderly manner and with transparency and fairness to traders.

The price of a particular share can change from one hour to the next depending on market expectations and how much of it was transacted. Financial columns of newspapers quote the highest, the lowest and the closing (at the end of the day) price of shares on the previous working day as well as its average over a given number of past months.

Bonds: These are certificates of indebtedness by a firm (for example, bonds floated by Tata industries, Reliance and so on) or a financial organisation (such as RBI and ICICI).[7] Bonds offer a guaranteed interest rate, which is typically higher than that offered in a savings account. Like stocks, bonds are traded in the market.

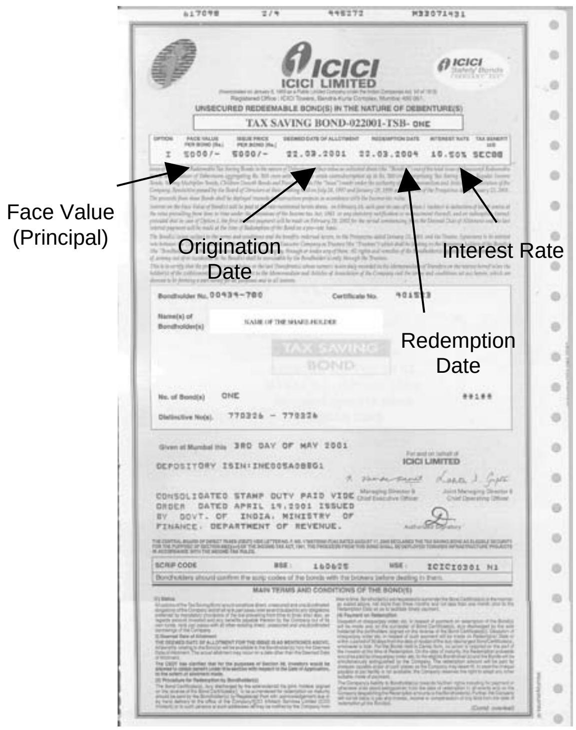
A bond is issued with a face value, that is, the principal (for example, Rs 10,000), a coupon (for example, Rs 1,000) and a redemption or maturity date. A coupon promises the amount that the firm will pay you periodically till the date of maturity.

[7]When the government borrows, it does it through a financial institution.

When the bond matures the amount of the face value is returned to the bond owner. The face value and the coupon value define the interest rate. In this example, the interest rate is equal to $(1{,}000/10{,}000) \times 100 = 10$ per cent.[8] Physically, like a stock, a bond is also a piece of paper stating the company's or the bank's name, your name as an owner of the bond, the face value, the coupon value and the maturity date. See Figure 7.2 to get an idea about how an ICICI bond looks like.

Compared to stocks, bonds are much less risky. Some bonds have virtually zero risk, such as those issued by the government. Bonds issued by private firms may not be totally riskless although the firms are supposed to honour bonds whenever the payment is due. There is a critical difference between a stock and a bond issued by a firm—unlike a stock, a bond does not offer any ownership of the company. See Clip 7.3 for an account of riskiness of bonds issued by private firms.[9]

Figure 7.2 A Bond Certificate



[8]In some cases, the interest rate is explicitly written on the bond in place of the coupon value.

[9]As said earlier, bonds are a debt instrument. There are various other debt instruments such as debentures, commercial papers, certificate of deposits and treasury bills based on who issues them, how secure they are and the maturity period (short term or long term).

Clip 7.3: Risk in Holding Private Bonds

In principle, bonds carry no or little risk. But in reality it is not that simple. The ability to honour bonds would depend on the firm's prevailing financial situation and its past history of paying out to its bond holders in time. Not just any firm can issue bonds and expect the public to buy its bonds. Even in the case of bonds from those firms who do sell them in the market, the risk varies. There are international companies, like Standard & Poor's, Moody's Investors Service and Fitch Investors Service, that rate bonds issued by publicly traded companies on the basis of market surveys. Moody's Investors Service classifies bonds into 'A-type', 'B-type' and 'C-type', among which those in the first category are the safest. Even among the A-type bonds, those rated as 'Aaa' are the best, followed by 'Aa1,' 'Aa2' and 'Aa3'. For example, in 2003, two Indian companies, Tata Power and Indian Oil Corporation, were rated 'Ba1' by the Moody's Investors Group. As a company's performance and record changes over time, its bond rating may also change.

Bonds whose rating exceeds a certain threshold are called *Investment Grade bonds*. (Bonds offered by the two Indian companies cited above belong to this category.) Those whose credit rating is less than the threshold are the so-called *junk bonds*.

There is also another source of risk associated with holding a bond, which has nothing to do with the performance record of the concerned firm. This has to do with inflation. The interest rate associated with a bond is fixed in nominal terms. But if there is an unanticipated high inflation, the real interest rate or the real return on bonds would become low. To the extent that price-level movement over time is uncertain, the return from bonds is also uncertain.

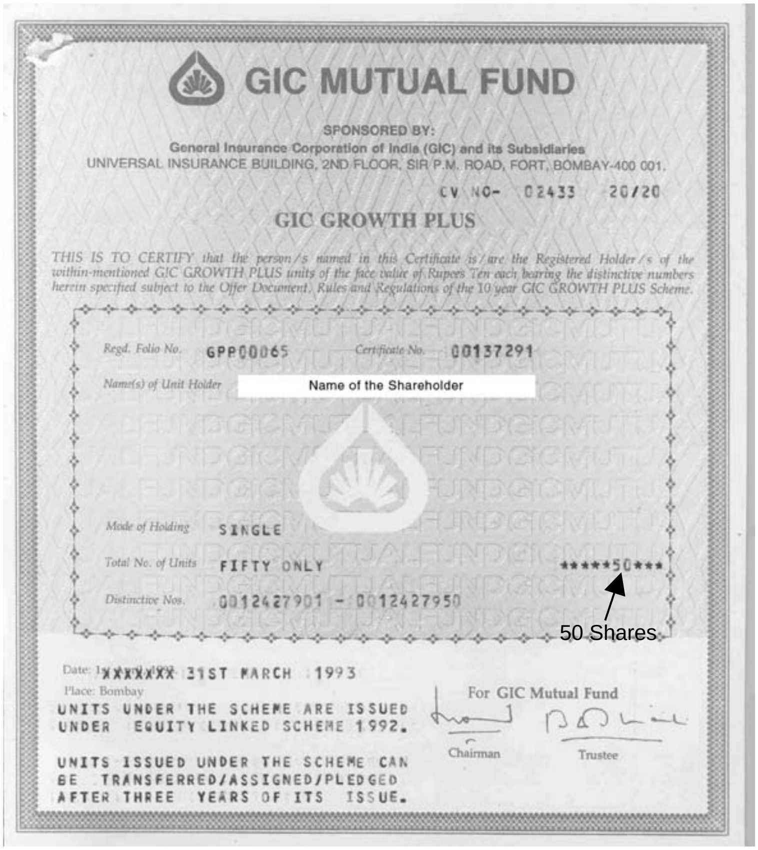
The overall point here is that bonds are not 100 per cent risk free.

Stocks and bonds both are referred by a term called **securities**.

Mutual Funds: A mutual fund is essentially an investment company or trust, which pools funds from small investors and invests the total sum in various individual stocks and bonds. The fund is, therefore, an intermediary, which invests the funds of small investors in a large number of securities. This is indeed quite convenient for a small investor, who does not have the time or resources to research on the merits of different securities available in the market. The fund's management does this for the investors. The key thing is that the fund's own investment is diversified so that investing in a mutual fund is not as risky as investing in a limited number of individual securities.

The objective of a mutual fund is pre-specified—which sectors or kinds of securities it holds and whether it targets relatively high-risk, high-return securities or low-risk, low-return securities. You can get this information from its prospectus or website. A mutual fund company typically floats a number of mutual funds of different categories.[10] Technically, an investor in a mutual fund becomes one of its shareholders. The price of a share is simply calculated as the fund's assets minus liabilities divided by the number of shares. Hence if the fund's investments are doing well, it gets reflected positively in the fund's assets, which translates into a higher price of the fund's share. Also, incomes generated from investments by the funds are passed on to the shareholders (which is like dividend payments to shareholders of a particular company). Figure 7.3 exhibits a mutual fund certificate of having 50 shares issued by GIC mutual fund company; the particular fund's name is GIC Growth Plus.

Figure 7.3 A Mutual Fund Certificate



GIC MUTUAL FUND

SPONSORED BY:
General Insurance Corporation of India (GIC) and its Subsidiaries
UNIVERSAL INSURANCE BUILDING, 2ND FLOOR, SIR P.M. ROAD, FORT, BOMBAY-400 001.

CV NO- 02433 20/20

GIC GROWTH PLUS

THIS IS TO CERTIFY that the person/s named in this Certificate is/are the Registered Holder/s of the within-mentioned GIC GROWTH PLUS units of the face value of Rupees Ten each bearing the distinctive numbers herein specified subject to the Offer Document, Rules and Regulations of the 10 year GIC GROWTH PLUS Scheme.

Regd. Folio No. GPP00065 Certificate No. 00137291

Name(s) of Unit Holder Name of the Shareholder

Mode of Holding SINGLE

Total No. of Units FIFTY ONLY *****50***

Distinctive Nos. 0012427901 - 0012427950

Date: XXXXXXX 31ST MARCH 1993
Place: Bombay

UNITS UNDER THE SCHEME ARE ISSUED UNDER EQUITY LINKED SCHEME 1992.

UNITS ISSUED UNDER THE SCHEME CAN BE TRANSFERRED/ASSIGNED/PLEDGED AFTER THREE YEARS OF ITS ISSUE.

For GIC Mutual Fund

Chairman Trustee

[10]For example, a US-based company called Franklin Templeton Investments has more than 30 branches in India. According to its website, its Indian operation caters to more than 10 lakh customers (investors). As of 2005, it has more than 20 funds. For example, one of them is called Franklin India Blue Chip Fund that invests in large blue chip companies. It is aimed towards obtaining a steady return over a long run. Another is Franklin Infotech Fund. It also seeks a relatively stable return, while investing primarily in the IT industry.

OPTIMAL PORTFOLIO

From an individual investor's perspective, the central question is, which assets and how much of an asset she should hold? A **portfolio** refers to the distribution of asset holdings by an entity. As a simple example, you may be holding 30 per cent of your investments in bank accounts, 60 per cent in real estate and 10 per cent in stocks, whereas I may be holding 50 per cent of my investment in bank account and 50 per cent in stocks. Then you and I hold different portfolios. The question posed above can be framed as, what is the best or optimal portfolio for an individual?

The answer is that there is no unique portfolio that is the best for everyone. It depends on an individual's needs for liquidity and preference towards risk. A full analysis of this problem, analogous to the demand for goods in the consumer theory, is beyond our scope. However, there are some general observations.

To generally protect ('hedge') oneself against too much of risk, an investor must hold a **diversified portfolio**, meaning a portfolio having assets with varying degrees of returns and risks.

If a person is relatively poor or has urgent needs for immediate consumption, she should invest most of her savings in relatively liquid and less risky assets. Otherwise she can invest in more illiquid and riskier assets.

Now consider assets that are equally liquid. How much should one invest in relatively high-risk, high-return assets and in relatively low-risk, low-return assets? That depends on an individual's attitude towards risk. Bill and Prem may be two billionaires, having equal amount of funds to invest. However, Bill, by nature, may be more averse to risk that Prem. Then Bill would invest relatively more in low-risk, low-return assets as compared to Prem. The attitude towards risk not only varies across individuals but also over time for the same individual. If you are in the early forties for example, you can afford to take chances and park a large portion of your savings in high-risk, high-return assets. But as you approach your retirement age, you would like to keep your investments in low-risk, low return assets.

Economic Facts and Insights

- Apart from gifts, a person's (or a family's) assets are built on his (family's) past and current savings.
- An asset has three characteristics: liquidity, return and risk.
- Rate of return of an asset equals the sum of yield to cost ratio and capital gains/losses to cost ratio.
- Variability of return associated with an asset indicates the risk in holding it.

(continued)

- Present value analysis compares the profitability of holding an asset vis-à-vis holding a risk free loan as an asset. It is also used to compare profitability across two assets (without considering risk).
- Future earnings and costs are measured in terms of the current period by means of discounting.
- The market offers real assets like real estate and machines as well as financial assets like stocks, bonds and mutual funds.
- Cash as well as current and savings accounts are very liquid assets.
- A stock holder is personally responsible for the liabilities facing the company to the extent of his sharingholding (that is, ownership) in the company.
- Many companies list their shares in a stock exchange.
- The return from holding a stock carries risk, depending on the performance and profitability of the firm which issues the stock.
- Dividend payments constitute the yield from a stock.
- Stock price fluctuations imply risk as well as an opportunity to make large gains.
- Large companies typically list their stocks in a stock exchange.
- Holding a bond is less risky than holding a stock in the sense that the company is legally obligated to pay interest and principal to its bondholders irrespective of whether the company is making profits or incurring losses.
- Investing in a mutual fund is less risky than investing in an individual stock because the mutual fund itself invests in a large number of securities using the investments (funds) it receives from investors.
- For a general protection against too much of risk, an individual must hold a diversified portfolio.
- There is no such thing as a portfolio that is the best for everyone. Each one's best portfolio depends on her needs for liquidity and preference towards risk.

EXERCISES

7.1 What information do we need in order to calculate the rate of return of an asset?

7.2 What is the difference between yield and capital gains of an asset?

7.3 In the Numerical Example 7.1, is there any capital gain or loss involved? If so, how much is it?

7.4 Mr Mitchell Abrahim invested \$2,000 in gold bullion, which he bought at \$520 per ounce. Two years later, he sold them all at \$566 per ounce. What is his annual rate of return from this investment?

7.5 Mahesh is looking into the prospects of buying one stock out of two, A and B. Stock A yields a return of 5 per cent with probability of 60 per cent and a return of 9 per cent with a probability of 40 per cent. Stock B offers a return of 4 per cent with probability of 30 per cent and a return of 10 per cent with a probability of 70 per cent. What is the expected return of each of these stocks?

7.6 This year your job is paying you Rs 9 lakh per year. You expect to get a promotion next year with a salary of Rs 10.7 lakh per year. If the interest rate is 7 per cent, what is the discounted value of the next year's expected salary?

7.7 An investment in a project gives a yield of Rs 4 lakh for four years starting the next year. (At the end of the project period, there is no return of the principal or any sale value of the investment made.) At 10 per cent rate of interest, what is the PDV of this investment?

7.8 'In general, bonds are riskier than stocks'. Comment.

7.9 'In general, individual shares are riskier than investment in a mutual fund'. Comment.

8 Types of Firms, their Goal and Production

CONCEPTS

- Individual Proprietorship
- Partnership
- Corporation
- Unlimited Liability
- Limited Liability
- Factors of Production
- Inputs
- Primary Factors of Production
- Labour
- Capital
- Land
- Production Function
- Cobb-Douglas Production Function
- CES Production Function
- Total Product or Total Physical Product
- Marginal Product or Marginal Physical Product
- Average Product or Average Physical Product
- Total, Marginal and Average Returns
- Total Product Curve or Total Physical Product Curve
- Marginal Product Curve or Marginal Physical Product Curve
- Average Product Curve or Average Physical Product Curve
- Law of Diminishing Marginal Product, Law of Diminishing Marginal Returns or Law of Diminishing Returns
- Returns to Scale
- Increasing Returns to Scale
- Constant Returns to Scale
- Diminishing Returns to Scale
- Linearly Homogeneous Production Function
- Isoquant or Production Indifference Curve
- Economic Region of Production
- Technique (as opposed to Technology)
- Marginal Rate of Technical Substitution
- Leontief Technology
- Input Coefficient
- Input-Output Analysis
- Iso-cost Line
- Cost Minimisation Problem
- Normal Input
- Inferior Input
- Expansion Path

Chapters 2, 4 and 5 dealt with demand for commodities by households. It is firms who supply goods and services to the market. While in Chapter 3 we learnt about the supply curve, in this chapter and the next few we will analyse the behaviour of firms with respect to how much they produce and supply to the market.

However, before getting into various concepts and analyses, it will be useful to know what types of firms operate in an economy in terms of ownership, management and liability, and the general goal of a firm. Surely, the business of a road-side ice-cream vendor is different from that of firms like Mother Diary, Amul, Kwality, Vadilal, Baskin-Robins or Nirulas that produce ice-cream.

TYPES OF FIRMS

There are mainly three types of firms: individual proprietorship, partnership and corporation. **Individual proprietorship** refers to any single-owner or single family-owner business. Most small businesses are of this kind. If you own a small stationery shop, then you are an individual proprietor. Management typically lies with the owner or the entrepreneur. If anything goes wrong with the business in terms of defaulting on loans or negligence in paying taxes, the individual owner is entirely responsible or 'liable.' For a business to operate legally (whether it is an individual proprietorship or not), it has to be registered with the Registrar of Companies (ROC).[1]

A **partnership** is one in which a group of people run a business as partners. They are co-owners and mostly manage the firm directly. A partnership typically involves a small business. Many professional groups, for example, lawyers and chartered accountants, form partnerships. The key feature of a partnership firm is that each partner is totally liable, or has **unlimited liability,** as does the owner in an individual proprietorship. It means, for example, that if there is an outstanding payment that needs to be made and the other partners cannot pay for some reason, then you as a partner are entirely responsible.

In a **corporation**, there is typically a difference between ownership and management. The shareholders are the owners and they appoint a board of directors, among whom some (like the CEO, Chief Executive Officer) are directly responsible for the operations of the enterprise. **Limited liability** is a crucial feature of a corporation. It means that a shareholder is responsible/liable for payments only to the extent of his ownership share in the company. For example, if some outstanding payment has to be made, say Rs 50 lakh, and there is no other way to finance it other than collecting it from the owners, then if you own 10 per cent of shares in the company, your liability cannot exceed Rs 5 lakh. Among corporations, there are private limited companies and public limited companies. In the

[1]There are many businesses, that are unregistered, for example, road-side vendors of various kinds.

former, the shares are held among a few and any new shareholder has to be approved by the existing shareholders. Shares are not traded in the stock market. A private limited company can 'go public' and become a 'public limited company' by offering an IPO (initial public offering). The general public can buy its shares. All big private companies (for example, Reliance, Infosys) fall into this category. Some public sector firms like BHEL, MTNL and Indian Oil are also public limited companies. In public sector firms, the government holds a large percentage of shares and appoints its own people to manage the company. Unlike private limited companies, public limited companies are supposed to follow more norms and they are subject to more regulatory scrutiny.[2]

GOAL OF A FIRM

In consumer theory, the basic behavioural assumption is that consumers maximise their surplus or utility. What is the objective of a firm? In economics, the prevailing assumption is that a firm's goal is to maximise its profit. This is quite intuitive since the entrepreneur's or the owner's benefit lies in the profit income the firm generates.

However, for corporations in which ownership is separated from management, it is argued that the management may not be necessarily maximising the shareholders' profits. Typically the shareholders are large in number and do not have adequate information about the true merits and demerits of important decision-making. Thus, the management can survive and do well for itself as long as it provides some minimum acceptable returns to the shareholders, thereby 'keeping them happy.' Rather, the argument goes, managers make decisions so as to maximise sales or revenues—to which their remuneration, perks and reputation are linked—subject to some minimum profit constraint.

However, the assumption one should make for analysing a firm's behaviour must rest ultimately not on how intuitive or reasonable the objective of the firm appears to an individual or an expert, but whether the assumption leads to good predictions about the firm's behaviour in terms of the observables. By this criterion, it turns out that the profit maximisation assumption is as good as any other alternative assumption that has been proposed in the literature on the theory of the firm.

Therefore, profit maximisation remains the dominant assumption regarding the goal of a firm and this will be assumed in this chapter and elsewhere in the book.

While the goal of profit making is certainly pertinent for a private firm, how about a public sector firm? The very adjective 'public' and its ownership (at least a

[2]There are other entities like some cooperatives, non-profit organisations and non-governmental organisations (NGOs) that operate like businesses.

large fraction of it) by the government strongly suggest that such a firm is supposed to work in the 'public interest' and not necessarily to maximise profits. Whether such firms in India have truly worked in the public interest or not is debatable, but it is a fact that a majority of these were incurring heavy losses consistently over decades. It is an open secret that a significant proportion of employment in these sectors was politically rather than economically motivated. Surely, these firms were not operating on the principle of profit maximisation.

But as you know, many losing public sector enterprises in India have been disinvested (sold to private parties). Further, the remaining firms are now under increasing pressure to show results in terms of revenues raised and profits earned. Hence, of late, profit maximisation appears as a good working hypothesis for the functioning of public sector firms as well.

FACTORS OF PRODUCTION OR INPUTS

We now begin the analysis of production. Our starting point will be to understand technology—not in an engineering sense but in terms of the patterns of relationship between various inputs used in the production process and output.

Of course, different goods have different technologies of production. Wheat and computers are not produced by the same technology. Even the same good is not produced by a single technology. Rice or paddy is produced in many parts of India by relatively labour intensive methods whereas highly mechanised methods of agriculture are used in many developed countries.

Even though specific technologies may differ, there are, however, common qualitative elements across the production processes of various goods. For instance, whether it is *kulfi*, hair-cutting, bicycles or rockets, the amount produced within a given time period depends, among other things, on how many workers are employed and how many hours they work. Similarly, it may depend on machinery, equipment, building, land, raw materials and so on. Various things that go into a production process are called, in general, **factors of production** or **inputs**.

Among the inputs, raw materials are a bit different from others in the sense that they are purchased from some other sectors of the economy as other 'goods' and they are completely used up in the production process. For instance, a bicycle industry uses steel as a raw material. Fertilisers are used as raw material in agriculture. But this is not the same for labourers in agriculture or employees in a factory. When someone works for a firm, he or she is not 'used up' like lumber or coal. A worker provides her service to a firm. The same is true for land. If you are running a software firm in a ground floor flat you have taken on rent, you are using the service of some space or land. It is not totally used up in the production process. Keeping this in mind, economists define three **primary factors of production** (or primary inputs) whose services are used in production: namely, labour, capital and land. **Labour** refers to hours of work by workers or employees.

Various equipment and machines are termed as **capital**. **Land** refers to areas on the ground (as in agriculture as well as spaces like rooms, flats and buildings.[3] Here and elsewhere we shall use the terms 'factors' and 'inputs' interchangeably.

PRODUCTION FUNCTION

The technical process that links inputs to output of a good is called the **production function** of that good. It is defined as the technological relationship between inputs and output—giving the maximum output that can be produced from various input combinations.

Suppose a firm employs two inputs, say, labour and land, measured respectively in hours and thousand square feet. Various input-to-output combinations are given in Table 8.1. Labour of 1 hour and 2,000 sq.ft of land produce, at the most, 4 units of output; 5 hours of labour and 7,000 sq.ft of land produce, at the most, 19 units of output; and so on. Table 8.1 describes a production function.

Note the following.

(*a*) It is normally assumed that inputs work to the best of their efficiency. Hence, instead of 'maximum output', we just say output.
(*b*) The notion of production function is not limited to two inputs. There can be many, including several types of capital or several different skills of labour.
(*c*) Table 8.1 lists only some, not all, possible combinations of inputs and output.
(*d*) In fact, very generally, one can write a production function as a mathematical function. If we denote various inputs as x_1, x_2, x_3, …, x_n, and the output by y, then, algebraically, we can write the production function as $y = f(x_1, x_2, \ldots, x_n)$. Indeed, this is analogous to the utility function introduced in Chapter 4. As an example, suppose that there are two inputs,

Table 8.1 Production Function

Labour (in hours)	*Land ('000 sq. ft)*	*Output (in units)*
0	0	0
1	2	4
1	3	6
3	6	14
4	6	16
5	7	19
6	10	21

[3]Some economists include a fourth primary factor of production: entrepreneurship, referring to the owner(s) of a firm, who undertake the effort/skill of organising production and the risk of the business itself. However, the reward to entrepreneurship is in the form of (residual) profit or loss, not in terms of explicit cost to a firm as in the case of hiring labour, capital or land.

1 and 2 and $y = A(\sqrt{x_1} + x_2)$. If $A = 10$, $x_1 = 9$ and $x_2 = 5$, then we have output, $y = 10(\sqrt{9} + 5) = 80$. The two often used production functions in applied economic analysis are the Cobb-Douglas and CES production functions.[4] For two inputs, say L (labour) and K (capital), these technologies are given by,

$$\text{Cobb-Douglas: } y = AL^{\alpha}K^{\beta},\ A > 0,\ \alpha > 0,\ \beta > 0$$

$$\text{CES: } y = A[\delta L^{\theta} + (1-\delta)K^{\theta}]^{\nu/\theta},\ A > 0,\ 0 < \delta, \theta < 1,\ \nu > 0.$$

RETURNS TO AN INPUT

A production function given in tabular form as in Table 8.1 or in algebraic form does not say much about how any single input contributes to production. This is measured by varying one input while keeping other inputs fixed. There are three important concepts here.

Total Physical Product, Marginal Physical Product and Average Physical Product

The **total product** or **total physical product** of an input, denoted by *TPP*, is defined as the total output associated with a given level of employment of the input in question. The **marginal product** or the **marginal physical product** (*MPP*) is defined as the increase in the total physical product per unit increase in the employment of an input. In both these definitions, it is presumed that other input levels are unchanged. Let the input in question whose employment is varying be called a 'variable input.'

Finally, the **average product** or the **average physical product** (*APP*) is defined as the *TPP* per unit of the variable input, that is, *APP* = *TPP*/*L*, where *L* stands for the variable input.

The three notions just defined are also respectively called total**,** marginal and average **returns** to an input. These notions are not unrelated to the production function; they do characterise it.

Table 8.2 shows a *TPP* schedule, where the variable input, *L*, is called labour. The graph of a *TPP* schedule gives the **total physical product curve**. Figure 8.1 shows the *TPP* curve for the *TPP* schedule given in Table 8.2. Why the *TPP* curve shapes this way will be discussed later.

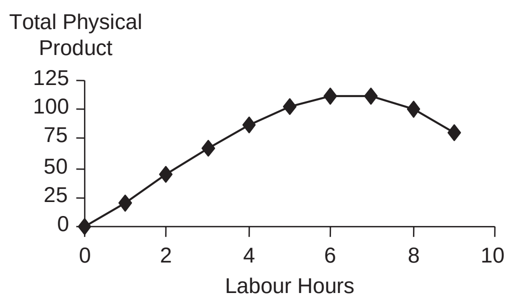
The marginal physical product, *MPP*, can be obtained from the total physical product, *TPP,* just as marginal utility is derived from total utility. For example, the

[4]Recall that in Chapter 4 we talked about the Cobb-Douglas utility function. In fact, the Cobb-Douglas utility function is an adaptation of the Cobb-Douglas production function discovered by Cobb and Douglas. The CES production function is due to four economists: Kenneth Arrow, Hollis Chenery, Baghicha Singh Minhas and Robert Solow. Sometimes it is also called ACMS or SMAC production function.

Table 8.2 A Total Physical Product

Labour Hours Employed (L)	*Total Physical Product (TPP)*
0	0
1	20
2	44
3	66
4	86
5	102
6	112
7	112
8	100
9	80

Figure 8.1 The Total Physical Product Curve Corresponding to Table 8.2



MPP at *L* = 2, which is 24, is equal to the difference between *TPP* at *L* = 2, which is 44, and *TPP* at *L* = 1, which is 20. The *MPP* schedule corresponding to Table 8.2 is given in column (2) of Table 8.3. The *APP* schedule is given in column (3) of Table 8.3; for each value of *L*, it is equal to *TPP* divided by *L* in Table 8.2.

We see that both *MPP* and *APP* initially increase and then decrease. Further, after a certain point, *MPP* becomes negative. The graphs of an *MPP* schedule and an *APP* schedule are respectively called the **marginal physical product curve** and the **average physical product curve**. These graphs based on Table 8.3 are shown in Figure 8.2. Both are inverse U-shaped curves since *MPP* and *APP* initially increase and then decrease.

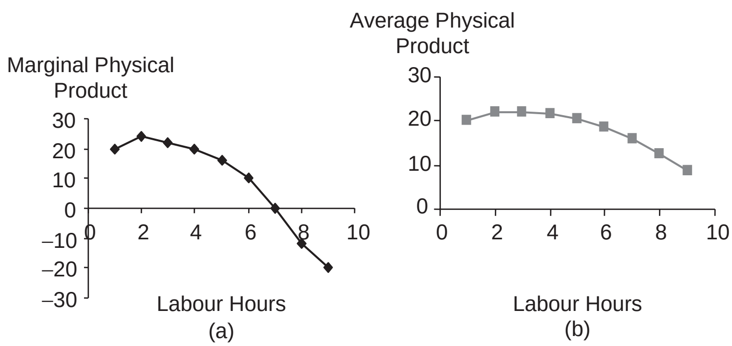
If the variable input can be measured continuously like decimal points on the real line—rather than discrete numbers like 1, 2, 3, … and so on—the *TPP*, *MPP* and *APP* curves will look smooth. Figure 8.3 depicts such curves.[5]

[5]Ignore for now the tangents drawn on panel (a), the shaded area in panel (b) or the superimposition of the *MPP* curve in panel (c).

Table 8.3 Marginal Physical and Average Physical Schedules

Labour Hours Employed (L) (1)	*Marginal Physical Product (MPP)* (2)	*Average Physical Product (APP)* (3)
0	–	–
1	20	20
2	24	22
3	22	22
4	20	21.50
5	16	20.40
6	10	18.67
7	0	16
8	–12	12.50
9	–20	8.89

Figure 8.2 Marginal Physical Product Curve and Average Physical Product Corresponding to Table 8.3



Note that the concepts of *TPP*, *MPP* and *APP* apply to all inputs, not just labour. But the important point is that they measure how output changes with respect to any particular input, 'one at a time,' when the other input levels are fixed.

Interrelationships

(*i*) Although both *MPP* and *APP* are derived from *TPP* by definition, there are a few interrelationships. Given any one of these three, the other two can be derived. Suppose *MPP*s are known to us. Since these are additions to *TPP*, we can get *TPP* by adding *MPP*s, that is, *TPP* is the sum of *MPP*s. In Table 8.3, at $L = 3$, the *MPP*s add up to $20 + 24 + 22 = 66$. Check from Table 8.2 that at $L = 3$, $TPP = 66$. If the *MPP* curve is smooth, then the sum of *MPP*s is nothing but the area under the *MPP* curve and it equals *TTP*.

For instance, as shown in Figure 8.3(b), the *TPP* of employing L_1 units of labour is equal to the area $0ABL_1$. Once we get *TPP*, we can readily obtain *APP* by applying its definition.

Similarly, if the *APP*s are known, we get *TPP* by multiplying *APP* with the level of employment. Then *MPP*s are obtained by applying its definition.

(*ii*) *APP* is equal to the slope of the ray from the origin to the point on the *TPP* curve. For instance, in Figure 8.3(a), at $L = L_2$, $TPP = y_2$. Thus $APP = L_2y_2/0L_2$, which is simply the slope of the ray $0y_2$ from the origin.

(*iii*) By definition, *MPP* is the addition to the *TPP* as the variable input increases by one unit. Hence, if the *TPP* curve is smooth, *MPP* is the derivative of *TPP*, equal to the slope of the *TPP* curve, as shown in Figure 8.3(a). It then follows that *TPP* increases, remains constant or decreases as *MPP* is positive, zero or negative.

(*iv*) Note in Figure 8.3(c) that the *MPP* curve cuts the *APP* curve at the maximum point, that is, as *APP* rises (to the left of the maximum point), *MPP* > *APP* and as *APP* falls (to the right of the maximum point), *MPP* < *APP*. The reason behind this is mathematical not economic and the mathematical logic is spelt out below through an example.

Suppose you are a very enthusiastic college teacher, very eager to know how your students did in a challenging test you gave last week. After checking each script you enter its score (marks) into the computer and the computer gives you the average score of all scripts you have checked so far. You have already checked 9 scripts and the average is 85 (out of 100).

Suppose that the next (10th script) receives 91. Without entering this number in the computer, can you say if the new average will be higher or lower than 85, the old average? Yes, you can—the average will be higher. So far the average was 85 and thus if the last ('marginal') student's score is greater than the old average, it will push up the average. This means that (*i*) if the 'average' is increasing, the 'marginal' must be greater than the average.

If instead, the 10th student had scored anything less than 85, by the same logic, it would have pulled down the average. This implies that (*ii*) if the 'average' is decreasing, the 'marginal' must be below the average.

Figure 8.3 Smooth *TPP*, *MPP* and *APP* Curves

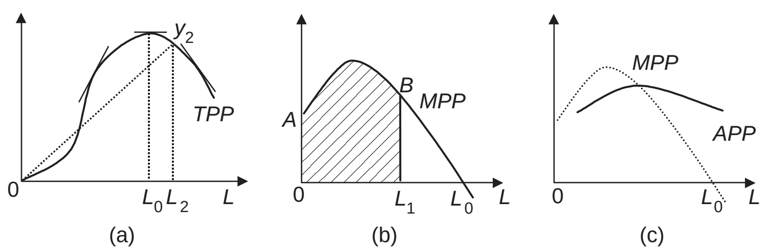
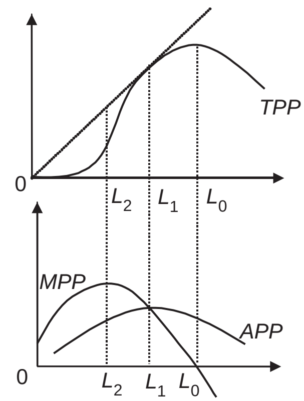


Figure 8.4 *TPP*, *MPP* and *APP* Curves in a Single Graph



Now you join statements (*i*) and (*ii*), substitute 'average' for *APP* and 'marginal' for *MPP*. It follows that if *APP* is rising, *MPP* > *APP*, and if *APP* is falling, *MPP* < *APP*. This, in turn, implies that the *MPP* curve must cut the *APP* curve at the latter's maximum point.

Various relationships among *TPP*, *MPP* and *APP* are summarised, so to speak, in Figure 8.4. The top part graphs the *TPP* curve and the bottom part the corresponding *MPP* and the *APP* curves. Between the levels of employment from 0 to L_2, *TPP* is increasing at an increasing rate (that is, the slope is rising). Accordingly, in the bottom panel *MPP* is increasing. From L_2 to L_1, *TPP* increases at a decreasing rate (that is, the slope decreases) and accordingly, in the bottom panel the *MPP* falls. At $L = L_0$, the slope of the *TPP* curve, equal to the *MPP*, is zero. Beyond L_0, *TPP* slopes downwards and thus *MPP* < 0. The *APP* being equal to the slope of the ray from the origin to the point on the *TPP* curve, we see in the top part that it increases between 0 to L_1 and falls thereafter. We also see that at $L = L_1$, *APP* is maximum and equal to the slope of the *TPP* curve (*MPP*); in other words, *MPP* = *APP*, where *APP* achieves its maximum.

LAW OF DIMINISHING RETURNS

Now we discuss the economic reason behind why various curves are shaped the way they are. As it turns out, the pattern that *MPP* initially increases, then diminishes and finally becomes negative is the key. This pattern has a name: the **law of diminishing marginal product** or the **law of diminishing marginal returns** or, briefly, the **law of diminishing returns**. According to this law, keeping the employment of other inputs fixed, as more of a particular input is used in production, after a certain level, its marginal physical product decreases. The logic behind this law is the following. When the employment of a single input gradually increases while all other inputs are kept unchanged, the factor proportions among other inputs and the input in question become initially more suitable for production because the scarcity of the (needed) variable input gradually falls. This tends to increase the marginal product. But, after a certain level, the variable factor can work with other given inputs only less efficiently (as other inputs become relatively scarce), that is, factor proportions become increasingly unsuitable for production. This implies that the marginal product of a factor would fall as more of it is used.

The *MPP* initially increasing and then falling implies an inverse U-shaped *MPP* curve, that is, the law of diminishing returns explains this shape of the *MPP* curve. In turn, this explains the particular shape of the *TPP* curve and the inverse U-shape of the *APP* curve.

Two points need to be noted with respect to the law of diminishing returns.

First, diminishing returns to a factor may set in right from the beginning, that is, *MPP* may be diminishing throughout. In that case, the point L_2 in Figure 8.4 essentially coincides with the origin.

Second, a profit-maximising firm will never hire so much of any input such that its *MPP* is negative. Why? Because, if it does, it always can employ a bit less of the input in question, which would increase output and revenue from sales as well as save costs (as less of the input is employed and paid) and thus increase profits. A firm will continue to do this until the *MPP* is no more negative. However, see Clip 8.1 for statistical evidence on the existence of negative marginal product in a public sector enterprise.

Clip 8.1 Evidence on Negative Marginal Product

It is an open secret that there is or used to be excess employment in public sector enterprises run by the government. The reason behind this is not hard to seek—some jobs are/were handed out on the basis of political rather than profit considerations. Can such excess employment be so high that the marginal product of a certain kind of labour employed is negative? It is very easy to conjecture on this but very difficult to back it up with 'hard' evidence.

However, an econometric study by Das and Sengupta (2004) finds evidence on negative marginal product of certain type of workers in the government steel sector in steel plants under SAIL (Steel Authority of India Limited). On the basis of data on employment and productivity in 16 blast furnaces over the period 1991–97, the authors have found the evidence of negative marginal product for non-executive workers. The output elasticity of such workers was –0.17, that is, 10 per cent increase in such workers would decrease output by 1.7 per cent.

The good news, however, is that, lately, the government steel sector has become profitable and casual empiricism tells that it has become more efficient.

Reference:

Das, Sanghamitra and Ramprasad Sengupta. 2004. 'Project Pursuit Regression and Disaggregate Productivity', *Journal of Applied Econometrics*, 19(3):397–418.

Mathematically Speaking

Returns to an Input and the Relationship between APP and MPP

As we have seen earlier, we can write a general production function as $y = f(x_1, x_2, ..., x_n)$, where y is the output and $x_1, x_2, ..., x_n$ are the employment levels of inputs 1, 2, 3, ..., n. Let the inputs 2, 3, ..., n be fixed, while x_1 may vary. That is, let the employment of only input 1 vary. Then the functional relationship between x_1 and y is the *TPP* of input 1, which we can write as $y = f(x_1)$, where the 'dot' means that the other inputs are fixed. *APP* is defined as $y/x_1 = f(x_1)/x_1$. We can write this simply as f/x_1. Similarly, *MPP* is the partial derivative of y with respect to x_1, that is, $\partial f/\partial x_1$. Positive *MPP* means $\partial f/\partial x_1 > 0$. Diminishing marginal product means that the second derivative is negative, that is, $\partial^2 f/\partial x_1^2 < 0$.

Differentiate the *APP* with respect to x_1. We have,

$$\frac{dAPP}{dx_1} = \frac{d[f(x_1)/x_1]}{dx_1} = \frac{\partial f/\partial x_1 - f/x_1}{x_1} = \frac{MPP - APP}{x_1}.$$

This proves that if *APP* is increasing (or decreasing), that is, $dAPP/dx_1$ is positive (or negative), *MPP* > (or <) *APP*.

* * * * * *

NUMERICAL EXAMPLE 8.1

The *MPP* schedule of a variable input is given in Table 8.4. Assuming that the *TPP* at zero level of the input is zero, derive the *TPP* and the *APP* schedules.

Since *TPP* is the sum of *MPP*s, *TPP* at the employment level equal to 1 is equal to *TPP* at zero level of employment (which is 0) plus the *MPP* of input at the employment level 1 (which is 40); that is, it is equal to 0 + 40 = 40. Likewise, the *TPP* at the employment level of 2 units is equal to 40 + *MPP* at this level of employment (which is 35) = 75. *TPP*s at higher levels of employment are calculated

Table 8.4 *MPP* Schedule (Numerical Example 8.1)

Employment Levelof an Input	*MPP*
0	–
1	40
2	35
4	29
5	21
6	11
7	0

Table 8.5 *TPP* and *APP* Schedules (Numerical Example 8.1)

Employment Level of an Input	*TPP*	*APP*
0	0	–
1	40	40
2	75	37.5
4	104	26
5	125	25
6	136	22.267
7	136	19.428

sequentially and listed in the middle column of Table 8.5. Once we know the *TPP*s, *APP*s can be obtained by dividing *TPP*s by the respective levels of employment. These are listed along the last column of Table 8.5. (At zero level of employment, *APP* is not defined.)

RETURNS TO SCALE

We have studied the concepts that capture the contribution of any single input towards production. This does not mean at all that a firm can vary only one input at a time. When market conditions or technologies change, a firm would very well adjust many inputs. For instance, if a garment factory finds a cheaper source of getting cotton, it would like to expand its production by hiring more workers (labour) and buying more sewing machines (capital) as well. There are concepts that capture the joint contribution of inputs towards output.

Assume that a firm is not employing any input so much that its marginal product is negative (a profit-maximising firm would not do it anyway). Suppose that it increases the employment of all inputs by a given proportion (for example, by 10 per cent). The output will surely increase because the marginal product of each input is positive. But by how much? If it increases more than proportionately (for example, by more than 10 per cent), we say that there are increasing returns to scale. If the output rises but less than proportionately, there are decreasing returns to scale. Finally, if it increases exactly proportionately, there are constant returns to scale. Formally, we say that there are **increasing returns to scale**, **constant returns to scale or decreasing/diminishing returns to scale** when**,** as all inputs increase proportionately, the output increases more than, exactly or less than proportionately respectively.

Note the following:

(*a*) The words 'diminishing' and 'decreasing' used above do not mean that output decreases when all inputs increase.

(*b*) These concepts relate again to the production function. A production function may exhibit one type of returns to scale at all input levels or may exhibit any combination.

(*c*) In particular, if a production function satisfies constant returns to scale throughout, it is also called a **linearly homogeneous production function**. For example, suppose there are two inputs, labour (L) and capital (K), and the production function is Cobb-Douglas, given by $y = L^{0.6}K^{0.4}$. Starting with any initial values of L and K, check that if L and K are doubled, y is also doubled.

Mathematically Speaking

Returns to Scale

Suppose all inputs increase by a proportion λ. Then the output is equal to $f(\lambda x_1, \lambda x_2, ..., \lambda x_n)$. Increasing, constant or decreasing returns to scale are differentiated on the basis of whether $f(\lambda x_1, \lambda x_2, ..., \lambda x_n)$ is greater than, equal to or less than $\lambda f(x_1, x_2, ..., x_n)$ respectively. The three types of returns to scale are defined as follows:

Increasing Returns to Scale if $f(\lambda x_1, \lambda x_2, ..., \lambda x_n) > \lambda f(x_1, x_2, ..., x_n)$
Constant Returns to Scale if $f(\lambda x_1, \lambda x_2, ..., \lambda x_n) = \lambda f(x_1, x_2, ..., x_n)$
Decreasing Returns to Scale if $f(\lambda x_1, \lambda x_2, ..., \lambda x_n) < \lambda f(x_1, x_2, ..., x_n)$.

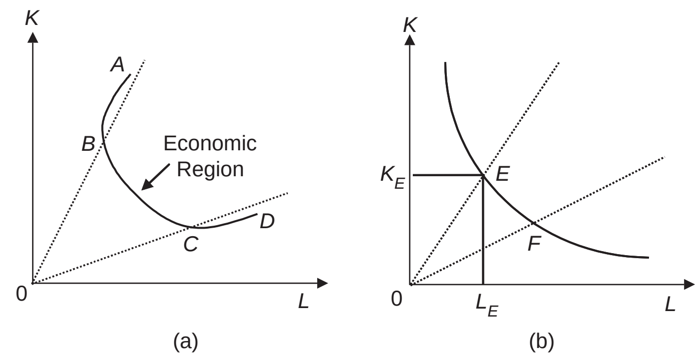
* * * * * *

ISOQUANTS

If there are two inputs, the production function can be shown graphically by an **isoquant**, defined as the locus of combinations of inputs, which give rise to the same level of output. If, for example, the two inputs are labour and capital, various combinations of L and K that produce, say, 60 units of output define the isoquant for output = 60. In principle, it is like an indifference curve. That is why an isoquant is also called a **production indifference curve.**

How does an isoquant look like? Suppose the *MPP*s of both inputs are positive. Then, starting from a given combination of inputs producing a given amount of output, if the employment of one input is increased, there is more output. Thus, if the output has to be brought down to its original level, the amount used of the other input has to be reduced. In other words, an increase in one input must be matched with an appropriate decrease in the other so that the output is unchanged. This implies that the isoquant will have a negative slope. If we consider a very high level of employment of an input such that its *MPP* is negative, then an increase in one input must be associated with an increase in the other such that the output remains the same. Accordingly, the isoquant will have a positive slope. Overall then, an isoquant will look like the line *ABCD* drawn in Figure 8.5(a). However, as discussed earlier, a profit-maximising firm will never use an input such that its *MPP* is negative. Thus the segments *AB* and *CD* on the isoquant *ABCD* are not

Figure 8.5 An Isoquant



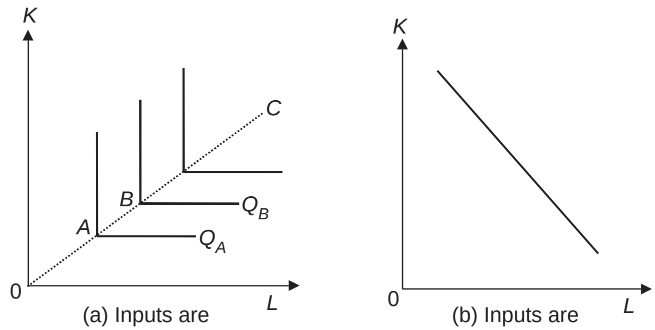
operational for such a firm. The segment BC is the only **economic region of production**. For this reason, from now on, an isoquant will be assumed to be a downward sloping curve, as shown in Figure 8.5(b).

A whole isoquant reflects technology whereas a **technique** refers to the proportion of inputs used at a *given* point on the isoquant. Referring to Figure 8.5(b), at point E, for example, $0L_E$ amount of labour and $0K_E$ amount of capital are used. The capital/labour ratio is $0K_E/0L_E$. Since $0K_E = EL_E$, this ratio is equal to $EL_E/0L_E$, which is the slope of the ray from the origin $0E$. Similarly at point F, the capital/labour ratio is the slope of the ray $0F$. Clearly, the ratio is higher at point E than at point F (as $0E$ is steeper than $0F$). We say that, compared to F, E involves a relatively more capital-intensive or a relatively less labour-intensive technique.

Once you see the similarity between an isoquant and an indifference curve, you realise that the economic analysis of production has some similarity with that of consumer behaviour. A collection of isoquants in a single graph is called an isoquant map. An isoquant to the right of a given isoquant represents a higher output. Two isoquants cannot intersect. Remember the definition of marginal rate of substitution of one good for another. The equivalent concept here is the **marginal rate of technical substitution** (*MRTS*), referring to the rate of substitution between inputs so that the output is unchanged. For instance, the *MRTS* of input 1 for input 2 equals the amount of input 2 that needs to be foregone per unit increase in input 1 so that the output remains unchanged. *MRTS* is equal to the ratio of marginal products just as the *MRS* in consumer theory is equal to the ratio of marginal utilities. If, for example, there are two inputs, labour and capital, then the *MRTS* of labour for capital is equal to the *MPP* of labour divided by the *MPP* of capital.

Remember the assumption of diminishing *MRS* in consumer theory, which implied that an indifference curve is convex to the origin. Here we assume diminishing *MRTS* and it implies that an isoquant is convex to the origin, as shown in Figure 8.5(b).

Figure 8.6 Special Cases of Isoquants



(a) Inputs are Perfect Complements

(b) Inputs are Perfect Substitutes

Also remember the special cases of indifference curves along which the goods were perfect complements or perfect substitutes. Similar looking isoquants reflect inputs being perfect complements and perfect substitutes of each other. These are shown in panel (a) and panel (b) respectively in Figure 8.6. Indeed, they are analogous to Figure 4.7 in Chapter 4. Perfect complementarity means that the inputs are used in a fixed proportion with each other. This implies that if the employment of one input is increased and yet output is to be kept unchanged, no amount of any other input can be sacrificed. Thus, $MRTS = 0$. In Figure 8.6(a) this proportion is indicated by the ray $0C$. Any deviation from this proportion leads to no increase in output. This explains the vertical and flat portions of the isoquants. Such a technology is called the **Leontief technology**, named after Wasily Leontief, a Nobel laureate in economics, who extensively used the assumption of this technology in his research. Perfect complementarity also means that the ratio of an input to output, called the **input coefficient**, is constant. The study of Leontief technology is also sometimes called an **input-output analysis**. On the other hand, when an isoquant is a straight line, as shown in panel (b), its slope is fixed. Thus *MRTS* is positive and constant. Since, irrespective of how much of each input is being used initially, a given amount of one input needs to be sacrificed per unit increase in the other input so as to keep the output unchanged, the inputs are perfect substitutes of each other.

However, from now on, we assume diminishing *MRTS* and accordingly assume downward sloping, convex isoquants as in Figure 8.5(b).

NUMERICAL EXAMPLE 8.2

In Table 8.6, the input bundles in the columns *U* and *V* define isoquants for two different levels of output. Derive the marginal rates of technical substitution of capital for labour along both isoquants. Is the assumption of diminishing marginal rate of technical substitution satisfied?

Table 8.6 Production Functions (Numerical Example 8.2)

	U		*V*	
Input Bundles	*Capital (K)*	*Labour (L)*	*Capital (K)*	*Labour (L)*
A	1	100	1	110
B	2	70	2	100
C	3	45	3	85
D	4	25	4	65
E	5	10	5	35
F	6	0	6	0

Table 8.7 *MRTSs* (Numerical Example 8.2)

Input Bundles	*MRTS of Capital for Labour along Isoquant U*	*MRTS of Capital for Labour along Isoquant V*
A to B	30	10
B to C	25	15
C to D	20	20
D to E	15	25
E to F	10	30

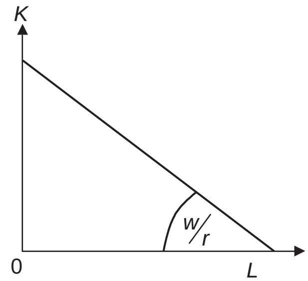
MRTS measures the rate of sacrifice of an input when the employment level of another input increases. The *MRTS* of capital for labour is the amount of labour sacrificed per unit increase in capital as we go down the input bundles in Table 8.6. Along isoquant U, from input bundle A to B, it is equal to 100 – 70 = 30 units of labour. From the input bundle B to C, it is 70 – 45 = 25 units of labour, and so on. Table 8.7 lists the *MRTS*s. We observe that *MRTS* is decreasing along the isoquant U but not decreasing along the isoquant V. (In the analysis of this chapter we disregard isoquants of the type V.)

ISO-COST LINE

We now characterise costs through what is called an iso-cost line (although there are a lot more concepts on costs to be introduced in the next chapter). Suppose labour and capital are the two inputs. Labour costs Rs w per hour and renting capital (for example, a truck) costs Rs r per unit. Thus, w and r are the respective factor or input prices for labour and capital. We define an **iso-cost line** as the locus of combinations of inputs whose total costs are the same.

In our example, the total cost = wL (cost of labour) + rK (cost of capital). Consider the iso-cost line for total cost equal to, say, Rs 1,000. Let $w = 50$ and $r = 40$. Various combinations of L and K such that $wL + rK = 1{,}000$ define the iso-cost line for total cost = Rs 1,000.

Figure 8.7 An Iso-cost Line



Clearly, along an iso-cost line, if the firm hires more of one input, it must hire less of the other. Indeed you can see that the iso-cost line is identical to the price line studied in consumer theory, except that total cost substitutes income, factor prices substitute product prices and we are talking about demand for inputs rather than demand for goods. Thus, like the price line, the iso-cost line is a downward sloping straight line. Figure 8.7 shows an arbitrary iso-cost line. Just as a slope of the price line equals the ratio of product prices, the slope of the iso-cost line is equal to the ratio of factor prices. In Figure 8.7 where labour is measured along the x-axis and capital along the y-axis, it is equal to the wage/rental ratio, that is, w/r.

While an iso-cost line is defined on the basis of a given total cost, the total cost itself may change, depending on how much a producer is willing to spend on inputs. In other words, iso-cost lines can shift just as the price line in consumer theory does. Different levels of total cost imply different iso-cost lines.

Consider an increase in the total cost. This will lead to a parallel rightward shift of the iso-cost line. Similarly, a decrease in total cost will lead to a parallel leftward shift of the iso-cost line. A change in a factor price can also shift an iso-cost line. Suppose there is an increase in the wage rate, w. If the firm hires labour only, with the same amount of total cost, it can hire less of it (as labour is more expensive). This will shift the intercept of the iso-cost line on the axis measuring labour towards the origin. In Figure 8.7 the iso-cost line will become steeper.[6]

COST MINIMISATION

By using the various concepts and assumptions at our disposal, we now analyse a decision-making problem facing a firm. Suppose a firm has committed itself to produce a given amount of output.[7] Which combination of inputs should it choose?

[6]Similarly, an increase in r will shift the intercept of the iso-cost line on the axis measuring capital and in Figure 8.7 the iso-cost line will become flatter.

[7]In Chapters 10, 14 and 15 we will study how a firm decides its output.

Turn to Figure 8.8(a). Suppose the output to be produced is y_0. The firm can choose any point on the isoquant representing y_0. Suppose it selects the input combination at point *A*. The total cost of this is indicated by the iso-cost line passing through this point, namely *aa*. Consider another combination *B*. Its total cost is shown by *bb*. Since the line *bb* lies to the left of the line *aa*, the combination *B* costs less than the combination *A*. Similar calculation can be made from any other combination on the isoquant.

As stated earlier, the firm's objective is to maximise profits. Profits, by definition, are equal to total sales revenues minus the total cost of inputs. Obviously, the higher the costs, the less are the profits. Hence, maximising profits implies that the firm must choose that input bundle which costs the least among those that can produce the output. We now go back to Figure 8.8(a) and ask which input bundle entails the minimum total costs, given that the firm is producing y_0. The answer is the tangency point *C*. Because at any other combination of inputs (such as *A* or *B*), the iso-cost line will be at a higher level compared to that at *C*. The input levels, L_C of labour and K_C of capital associated with the tangency point *C* (Figure 8.8(b))then constitute the cost-minimising choice of inputs.

In general, we say that the optimal or the cost-minimising choice of inputs is governed by the condition that the iso-cost line be tangential to the isoquant. Since the slope of the former is the input price ratio and that of the latter is the *MRTS* (equal to the ratio of *MPP*s), we can write the condition elaborately as,

Cost-minimisation Condition:

$$MRTS = \text{ratio of } MPP\text{s} = \text{factor price ratio.} \tag{8.1}$$

When, for example, the factors are denoted by *L* and *K*, with *w* and *r* as the respective factor prices, we can write this condition as:

$$\frac{MPP_L}{MPP_K} = \frac{w}{r}. \tag{8.2}$$

Figure 8.8 Cost Minimisation

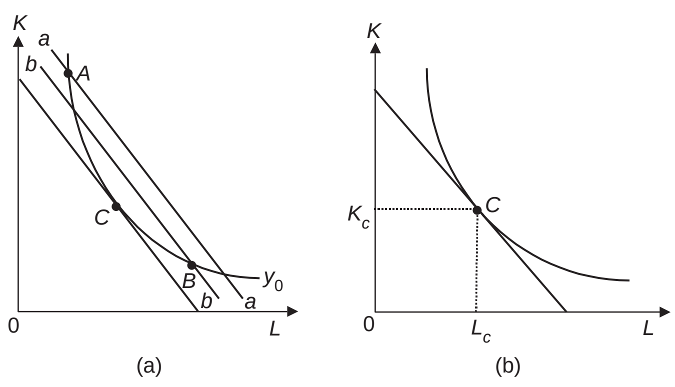
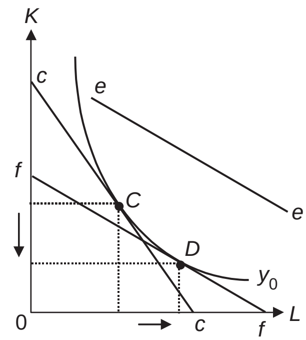


Figure 8.9 An Input Price Change



A Change in an Input Price

Having derived the cost-minimising condition, we can analyse how a change in a factor price or the output level can affect the optimal choice of inputs.

Suppose labour becomes cheaper.[8] This will make iso-cost line flatter (as its slope equals *w/r*). Turning to Figure 8.9, suppose *C* is the original point of choice and *cc* is the associated iso-cost line. With a lower value of *w*, the new iso-cost lines will be flatter than *cc* such as, *ee*, or any other line parallel to *ee*. The new cost-minimising input bundle is the tangency point between the isoquant y_0 and the iso-cost line *ff*, which is parallel to *ee*. This is the bundle *D*. Compared to *C*, the new input combination has more labour and less capital. This makes sense because labour being relatively cheaper now, the firm is hiring more labour and less capital—that is, it is employing a more labour-intensive technique.

A Change in Output

Suppose the firm decides to produce a higher quantity, y_1, than the original output y_0 (while input prices are unchanged). This means a higher isoquant. The effects are shown in Figure 8.10. In both panels, the point *C* is the original cost-minimising input bundle. In panel (a), the new bundle is indicated at *A*. Compared to the original input bundle, the firm hires more of both inputs. Panel (b) illustrates the case where the firm employs more of one input and less of the other. You can notice that the effect of an increase in output is similar to that of an income effect in demand theory. Like normal and inferior goods, we define normal and inferior inputs. An input is said to be **normal** or **inferior** as, at constant factor prices, the demand for it increases or decreases with output. In panel (a), both

[8]It could happen if, for example, workers from outside migrate into the area.

Figure 8.10 An Increase in Output

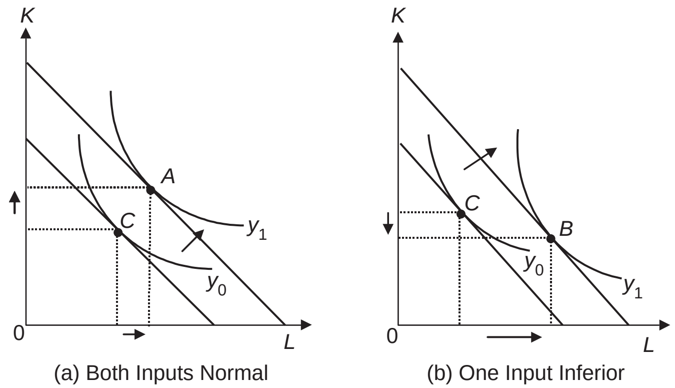
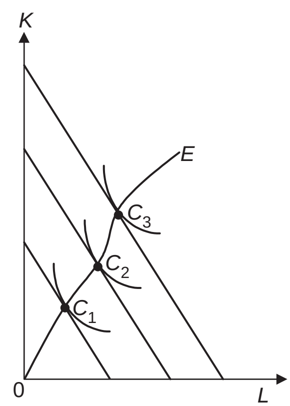


Figure 8.11 Expansion Path



inputs are normal whereas in panel (b), labour is the normal input while capital is the inferior one.

Note, however, that irrespective of whether some inputs are inferior, a higher output is always associated with a higher iso-cost line, that is, total cost always increases with output.

Suppose we extend this exercise and let output increase continuously. The effects are shown in Figure 8.11. The cost-minimising points are C_1, C_2, C_3 and so on. If we join these points, the line (0*E*) is called the **expansion path**, defined generally as the locus of cost-minimising input bundles at different levels of output, given the input prices. The expansion path indicates a firm's pattern of hiring inputs as it plans to expand its production.

Mathematically Speaking

Isoquants, Iso-cost Lines and Cost Minimisation

With two inputs 1 and 2, we can write the production function as $y = f(x_1, x_2)$. Totally differentiating this we have $dy = (\partial f/\partial x_1)dx_1 + (\partial f/\partial x_2)dx_2$. Along an isoquant, there is no change in output, that is, $dy = 0$. Hence $(\partial f/\partial x_1)dx_1 + (\partial f/\partial x_2)dx_2 = 0$. We can rewrite this as:

$$\frac{dx_2}{dx_1} = -\frac{\partial f/\partial x_1}{\partial f/\partial x_2} < 0. \tag{8.3}$$

This proves that an isoquant is downward sloping and its slope = *MRTS*, which is, in absolute value, equal to the ratio of marginal products.

Let w_1 and w_2 denote the respective input prices. The total cost has the expression $C = w_1x_2 + w_2x_2$. Totally differentiating it, $dC = w_1dx_1 + w_2dx_2$. On an iso-cost line $dC = 0$ since C is constant. Hence, $w_1dx_1 + w_2dx_2 = 0$, implying

$$\frac{dx_2}{dx_1} = -\frac{w_1}{w_2}.$$

This proves that the iso-cost line is downward sloping and the absolute value of its slope is equal to the factor price ratio.

Cost minimisation states that $C = w_1x_1 + w_2x_2$ is minimised along an isoquant. Keeping in view (8.3),

$$\begin{aligned} dC &= w_1dx_1 + w_2dx_2 = w_1dx_1 + w_2(dx_2/dx_1)dx_1 = [w_1 + w_2(dx_2/dx_1)]dx_1 \\ &= \left[w_1 + w_2\left(-\frac{\partial f/\partial x_1}{\partial f/\partial x_2}\right)\right]dx_1. \end{aligned}$$

If C is minimised, $dC/dx_1 = 0$. This implies that the term inside the brackets [] is zero, that is,

$$w_1 + w_2\left(-\frac{\partial f/\partial x_1}{\partial f/\partial x_2}\right) = 0,$$

$$\text{Or} \quad \frac{\partial f/\partial x_1}{\partial f/\partial x_2} = \frac{w_1}{w_2}.$$

The last equation is same as the cost-minimisation condition stated in (8.2).

* * * * * *

NUMERICAL EXAMPLE 8.3

Consider the input bundles along the isoquant U in Table 8.8. Suppose the price of capital services is r = Rs 60 and that of labour services (the wage rate) is w = Rs 5. Which of the input bundles among A to F minimises the total cost of producing the

Table 8.8 Cost Minimisation along the Isoquant U

Input Bundles	*K, L*	*Total Cost: r = 60; w = 5*	*Total Cost: r = 80; w = 5*
A	1, 100	560	580
B	2, 70	470	510
C	3, 45	405	465
D	4, 25	365	**425**
E	5, 10	**350**	450
F	6, 0	360	480

output associated with isoquant U? Suppose capital becomes more expensive—r increases from Rs 60 to Rs 80. Now, which input bundle minimises the total cost? Compare the new cost-minimising input bundle to the old cost-minimising input bundle.

This is straightforward. For each bundle, we know the level of each input and its price. Multiplying them gives the total cost of hiring that input. Summing it over the two inputs gives the total cost of the input bundle. For instance, consider input bundle B having 2 units of capital and 70 units of labour. With $r = 60$, 2 units of capital will cost 120 and with $w = 5$, 70 units of labour will cost 350. Thus the total cost of input bundle B is equal to $120 + 350 = 470$. Table 8.8 enumerates the total cost of all input bundles at the two input-price combinations. When $r = 60$ and $w = 5$, the total cost is the minimum at the input bundle E; the minimised total cost is equal to 350. When $r = 80$ and $w = 5$, the total cost is minimised at the input bundle D; the minimised total cost is equal to 425.

Compared to the input bundle E, the input bundle D has less capital and more labour because hiring capital is costlier. Put differently, capital is relatively more expensive and labour is relatively cheaper and hence the firm hires less capital and more labour. Furthermore, in the new situation, in absolute terms, one input price is the same while the other is higher. This is why the minimised cost in the new situation (425) is greater than that in the original situation (350).

Economic Facts and Insights

- Most businesses are individual proprietorships.
- In individual-proprietorship and partnership firms, each owner has unlimited liability, meaning that if the company is due for a payment and other owners are not in a position to pay (for example, someone dies or is currently

(continued)

residing in a foreign country), then a owner who is available is responsible for the entire amount due.

- Limited liability and separation of management from ownership are the salient features of a corporation.
- In predicting the behaviour of a firm, profit maximisation has worked as good as other alternative assumptions as the goal of a firm.
- In India, public sector firms are becoming more and more profit-oriented.
- Primary factors provide their services to a production activity whereas raw materials are totally used up in a production process.
- The contribution of an individual input to output is captured by the notions of total physical product, marginal physical product and the average physical product.
- The marginal physical product and the average physical product curves are inverse U-shaped curves because of the law of diminishing returns.
- The basis of the law of diminishing returns lies in that as more and more of a single factor is used in production, the proportion of this factor with respect to other factors becomes higher and, therefore, less suitable for production.
- A profit-maximising firm will not use any input to the level that its marginal product is negative.
- A profit-maximising firm will employ input combinations that minimise total cost at any given level of output.
- The cost-minimisation condition is one where the marginal rate of technical substitution, which is same as the ratio of marginal products of inputs, is equal to the ratio of input prices.
- Given an output, an increase in an input price would lead a firm to use less of that input. Thus an increase in the wage rate will lead a firm to use a less labour-intensive (that is, a more capital-intensive) technique. Similarly an increase in the rent will induce a firm to use a less capital-intensive, that is, a more labour-intensive technique.
- An increase in output may or may not lead to an increase in the employment of all inputs, that is, the employment of some inputs may decrease.

EXERCISES

8.1 What are the differences between a partnership form of business and an individual proprietorship?

8.2 What is the difference between a partnership business and a corporation?

8.3 'A private firm makes its decisions with the objective of maximising its profits.' Comment.

8.4 'In the real world, many firms go out of business. Thus, profit maximisation is a bad assumption.' Defend or refute.

8.5 Name four public sector firms in India that have been disinvested, that is, sold to the private sector.

8.6 Some years ago, a minister of a particular province made a comment to the effect that some private-sector firms incur losses in business and, hence, there is nothing wrong with public-sector firms making losses. Do you agree or disagree? Give reasons.

8.7 How defensible is the assumption of profit-maximisation as the main objective of a public sector firm today in India? Briefly discuss.

8.8 Many large, private firms contribute regularly to charity and other social causes. Is the behaviour consistent with the profit-maximisation hypothesis? Briefly discuss.

8.9 Briefly discuss the reason behind diminishing marginal returns to an input.

8.10 Why is the *MPP* curve of an input either downward sloping throughout or inverse-U shaped?

8.11 'A profit-maximising firm will never hire so much of an input, that its marginal product is negative.' Defend or refute.

8.12 Differentiate between diminishing returns (to an input) and diminishing returns to scale.

8.13 A 15 per cent increase in all inputs leads to only a 5 per cent increase in the output. Are the returns to scale increasing, constant or diminishing?

8.14 Some input combinations and the respective outputs associated with a given production function are given in the following table. What are the returns to scale (increasing, constant or diminishing) if the firm moves across the following input combinations:
(*a*) B to C, (*b*) C to D, (*c*) D to E, (*d*) E to F and (*e*) F to G.

Input Combination	*Labour*	*Capital*	*Output*
A	0	0	0
B	1	2	5
C	2	4	11
D	3	6	18
E	4	8	24
F	5	10	30
G	6	12	35

8.15 In terms of an isoquant, what is the economic region of production and why?

8.16 What assumption on the marginal product implies that an isoquant is downward sloping?

8.17 Prove that an isoquant to the right of a given isoquant implies higher output.

8.18 Prove that two isoquants cannot intersect.

8.19 Why is a typical isoquant convex to the origin?

8.20 Consider the isoquant V in the Numerical Example 8.2. Does it look convex to the origin? Why or why not?

8.21 What does Leontief technology mean? What shape of an isoquant does it imply?

8.22 'If the technology is Leontief, the returns to scale are constant.' Defend or refute.

8.23 What is an iso-cost line?

8.24 What is the slope of an iso-cost line equal to and why?

8.25 State the cost-minimisation problem facing a firm?

8.26 'By definition, cost is zero if a firm does not hire any input. Hence cost-minimisation essentially means shutting down the operation of the firm.' Comment.

8.27 Suppose a firm is committed to produce a given amount of output. There are two factors of production: capital and labour. Let rent and wages be their prices. Suppose this level of output is produced by an input combination such that the ratio of the marginal product of capital to that of labour exceeds the ratio of rent to wages. How should the firm change the input combination so as to lower the total cost of producing this level of output and why?

8.28 Assume that there are two variable factors of production (as in the isoquant analysis in the text). 'Given the level of output, if the price of an input increases, the firm may employ more of that input.' Defend or refute.

8.29 For a cost-minimising firm, how would an increase in output affect the total cost?

8.30 Briefly explain what is meant by an expansion path.

8.31 Briefly explain a normal input and an inferior input.

8.32 'If an input is inferior, its marginal product must be negative.' Comment.

8.33 'An increase in output would motivate a cost-minimising firm to employ more of all variable inputs.' Comment.

8.34 Derive the cost-minimising condition (8.2) by using the Lagrangean method outlined in the Appendix.

8.35 Consider the Cobb-Douglas production specified in the text. Suppose $\alpha + \beta > 1$. What does this imply towards the returns to scale? How would your answer change if $\alpha + \beta$ equals one or less than one?

9 Costs of Production and the Financing of a Firm

CONCEPTS

- Explicit Costs
- Implicit Costs
- Accounting Costs
- Economic Costs
- Short-run Cost Concepts
- Long-run Cost Concepts
- Fixed or Total Fixed Cost
- Overhead Costs
- Variable Cost or Total Variable Cost
- Total Cost
- Marginal Cost
- Average Fixed Cost
- Average Variable Cost
- Average Cost or Average Total Cost
- Plant Size
- Economies of Scale
- Division of Labour
- Diseconomies of Scale
- Social Cost
- Private Cost
- Externality
- Positive Externality
- Negative Externality
- Plough Back of Profits
- Retained Earnings
- Loans from Financial Institutions
- Mortgage
- Equity and Debt Instruments

Towards the end of the last chapter we saw that as output increases, the total cost rises. But there is much more to it than just that. In this chapter we study in detail various types of costs and their relation to output.

To begin with, there are explicit costs and implicit costs. **Explicit costs** are those, which are directly paid to other parties by an entrepreneur or a company running a business. They include, for example, the costs of labour, raw material, machinery purchased and so on. **Implicit costs** are those for which there is no direct payment but indirectly there is a cost involved.

Suppose you own a two-storey building. You live on the first floor and operate a small publishing company on the ground floor. Obviously, you do not have any rental cost of business operation. But there is an implicit cost. If you had otherwise rented out the ground floor to some other party, you would have obtained some rental income. By using it for your own business, you are effectively losing that income—and that is an implicit cost. Similarly, if you use your own savings in the business, the interest income foregone is an implicit cost.

Explicit costs are more commonly called **accounting costs**. Accounting costs plus implicit costs of the kind described above reflect the true cost of running a business and are called **economic costs**. In economic analysis, costs always refer to economic costs.

In this chapter, we will be concerned with various cost concepts based on the time horizon of an entrepreneur, namely, the short run and the long run. There is no particular calendar time like a month, quarter or year that distinguishes between the short run and the long run. Rather, as will be seen, the distinction is drawn from a production planning perspective.

SHORT-RUN COST CONCEPTS

If you think of a firm at a given point of time, like a snapshot, everything is fixed. The firm is producing a given amount by using a given amount of inputs and the inputs are paid their prices. All costs are given or fixed. But if we imagine the functioning of a firm over a relatively short length of time (like a movie), we can distinguish between costs that are fixed and those that are not. Suppose you run a clothing store. In a span of, say, one or two months, it is likely that the rental cost of the rooms you use are fixed in the sense that how much you pay as rent does not depend on how much you sell or produce. You may have signed a lease with the landlord for six months with a specified rent and your landlord (unless he is very kind) is going to charge you that rent—irrespective of how well you are doing in your business or even if you decide to produce nothing. That is why, generally, the rental cost is considered fixed in the short run. But typically labour costs are not fixed because workers can be hired and fired on a short notice by a firm. Costs of raw materials (for example, cloth bought in the wholesale market) are not fixed as you can buy more or less of them depending on the state of your business.

As another example, suppose Dr Juneja owns a diagnostic centre. It is located in a one-storey flat, inherited from his father. He employs about 30 people including nurses, technicians, management employees and manual workers. By taking loans from a bank he has bought several high-tech machines that do CAT scan, MRI, ultrasound tests and so on. Every month he repays the bank Rs 2 lakh in instalment towards his loan plus interest. This is an example of fixed cost because it is independent of how many patients come to Dr Juneja's centre for service. There is also an implicit rental cost of the flat, equal to the rent foregone by using the flat for the diagnostic centre. This is also a fixed cost. However, depending on how this venture is going, Dr Juneja can hire more or less of nurses or manual workers within a short notice. Hence, payment to nurses and other workers are not fixed.

Over a longer time horizon, however, an entrepreneur can think of explicitly or implicitly renting a different amount of space, a different plant size, different number of machines and so on. In other words, in the long run there are no fixed costs.

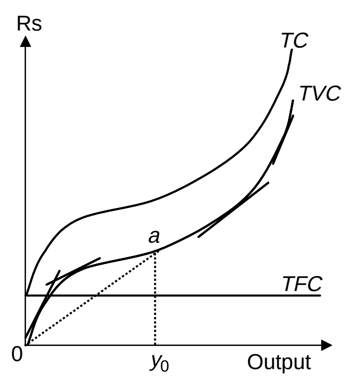
Returning to the short run, we then say that a firm has two types of costs, **fixed cost** and **variable cost**. Fixed costs are those costs that do not change with output. In common business terminology, these are called **overhead costs**. Variable costs refer to those that vary with the output. Typically, rental costs of land and capital are fixed and labour and raw material costs are variable.

More formally, these are respectively called **total fixed cost** (*TFC*) and **total variable cost** (*TVC*). The sum of the two costs is simply called the **total cost** (*TC*). That is,

$$TC = TFC + TVC. \tag{9.1}$$

These costs—*TFC*, *TVC* and *TC*—when graphed against the output, give us the *TFC*, *TVC* and *TC* curves respectively. See Figure 9.1 (but ignore for now the tangents, the point *a* and the dotted lines from this point). The *TFC* curve is a horizontal straight line (having zero slope) because *TFC* is independent of the output level. The *TVC* curve is upward sloping because producing more would cost more.

Figure 9.1 *TFC*, *TVC* and *TC* Curves



Since *TC* is the sum of *TFC* and *TVC* and costs are measured along the vertical axis, the *TC* curve is the vertical sum of the *TFC* and *TVC* curves. That is, at any output level, if we measure the *TFC* and *TVC* on the vertical axis and add them up, we get the corresponding point on the *TC* curve.

Like the marginal product and average product, we define marginal cost and average cost. **Marginal cost** (*MC*) is the addition to the total variable cost per one extra unit produced. It is also equal to the addition to the total cost per one extra unit produced since the difference between the *TVC* and *TC* is fixed. In the graph, *MC* is the slope of the *TVC* and the *TC* curves. You see in Figure 9.1 that as the output increases, the slope of *TVC* initially falls and then rises. This means that the *MC* curve (measuring *MC* against output) will be downward sloping first and then upward sloping. Put differently, it is U-shaped as shown in Figure 9.2(a). Also, since *MC* is the addition to the *TVC*, the area under the *MC* curve equals the *TVC*.[1]

Similarly, we can define

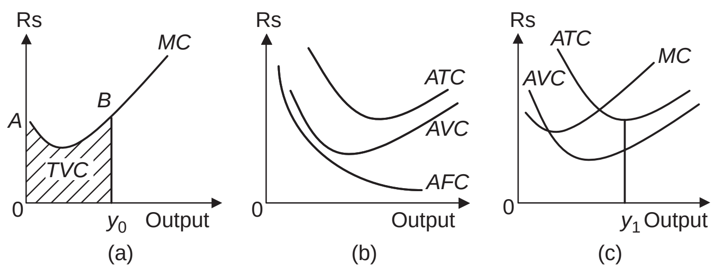
Average Fixed Cost (*AFC*) ≡ *TFC*/Output

Average Variable Cost (*AVC*) ≡ *TVC*/Output

Average Total Cost (*ATC*) ≡ *TC*/Output,

where the symbol '≡' means 'equal to by definition'. The *AFC*, *AVC* and *ATC* curves are drawn in Figure 9.2(b). The *AFC* is uniformly downward sloping by its definition—as output increases, its denominator increases while the numerator remains unchanged.[2] The shapes of the *AVC* and *ATC* curves depend on the shape of the *TVC* or the *TC* curve. Referring back to Figure 9.1 again, see that at output y_0, for instance, $TVC = y_0a$ and thus $AVC = y_0a/0y_0$, which is the slope of the ray $0a$. If we let the output gradually increase from zero, this slope increases up to a point and then decreases. This implies that the *AVC* curve will be U-shaped. The argument behind the *ATC* curve being U-shaped is similar.

Figure 9.2 *MC*, *AVC* and *ATC* Curves



[1]It cannot be equal to *TC* as it cannot account for the fixed cost.

[2]Indeed, the *AFC* curve is shaped like, what is called in geometry, a rectangular parabola, analogous to the unitarily elastic demand curve.

Finally, turn to Figure 9.2(c) (and ignore for now the output marked y_1). The mathematical relationship between the 'average' (A) and the 'marginal' (M) holds. Remember from the last chapter that $M < A$ if A is falling and $M > A$ if A is rising. Since *AVC* falls initially, $MC < AVC$; when *AVC* rises, $MC > AVC$. The implication is that the *MC* curve must cut the *AVC* curve at the latter's minimum point. The same relationship holds between the *MC* curve and the *ATC* curve.

NUMERICAL EXAMPLE 9.1

A firm's total cost of producing 4 units of output is Rs 70. Its marginal cost schedule is given in Table 9.1. What is the firm's total fixed cost? Derive the firm's *TC* schedule and *TVC* schedule.

The marginal cost of producing 4 units is Rs 6, which is the additional cost of producing the fourth unit. Since the total cost of producing 4 units is Rs 70, the total cost of producing 3 units must be Rs 70 – Rs 6 = Rs 64. Extending the same logic, the total costs of producing 2 units, 1 unit and 0 units are respectively equal to Rs 59, Rs 52 and Rs 42. The total variable cost of producing 0 units is 0 by definition. Hence Rs 42 being the total cost of producing 0 implies that it is equal to the total fixed cost. Using the relationship between the marginal cost and the total cost, and the marginal cost information for outputs 5, 6 and 7 units given in Table 9.1, we obtain the total cost at these output levels equal to Rs 78, Rs 89 and Rs 103. We now have

Table 9.1 *MC* Schedule (Numerical Example 9.1)

Output	*Total Cost (Rs)*
0	–
1	10
2	7
3	5
4	6
5	8
6	11
7	14

Table 9.2 *TC* and *TVC* Schedules (Numerical Example 9.1)

Output	*TC*	*TVC*
0	42	0
1	52	10
2	59	17
3	64	22
4	70	28
5	78	36
6	89	47
7	103	61

the entire total cost schedule. Deducing Rs 42 (the *TFC*) from it, we obtain the *TVC* schedule. These schedules are given in Table 9.2.

Why are *MC*, *AVC* and *ATC* Curves U-shaped?

The U-shape of the *MC*, *AVC* and *ATC* curves followed from the *TC* and *TVC* curves drawn in Figure 9.1. But this is not the basic economic reason behind why the *MC*, *AVC* and *ATC* curves are U-shaped—we could have drawn these curves first and then obtained the *TC* and the *TVC* curves from these. The underlying reason is the law of diminishing returns, studied in the last chapter.

Recall that this law states that initially the *MPP* of a factor may be increasing but after a certain point it must start to diminish. This means that if some other inputs are kept unchanged—there are fixed factors and hence there are fixed costs in the short run—initially when a factor's *MPP* is increasing, a gradual increase in output by a given amount would require a decrease in the rate of increase of the variable inputs and, therefore, a decrease in the rate of increase of the total cost of variable inputs. But the rate of increase of the total cost of variable inputs is equal to *MC* by definition. Thus, initially, *MC* decreases with output. After a certain point when diminishing returns set in, a gradual increase in output by a given amount would require an increase in the rate of increase of the total cost of variable inputs. This translates into *MC* increasing with output. In summary then, *MC* initially decreases with output and then increases with it after a certain point. That is, the *MC* curve is U-shaped.[3] The U-shape of the *MC* curve implies that the *AVC* and *ATC* curves are also U-shaped and that is why *TC* and *TVC* curves look like the way they do in Figure 9.1.[4]

Shift of the Short-run Cost Curves

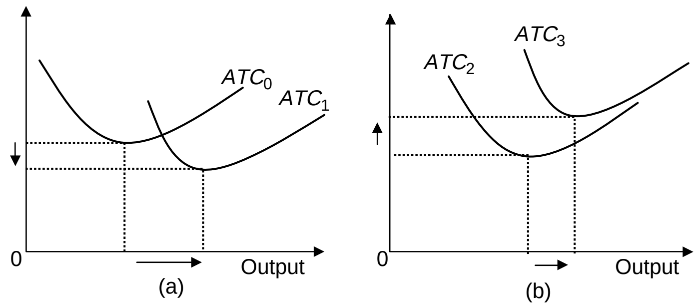
What are the factors that shift the short-run cost curves? These are technology, input prices, the level of fixed factors and business taxes. A technology improvement enables more output being generated from the same levels of inputs. It would lower costs and, in general, shift down *TC*, *TVC*, *AVC*, *ATC* and *MC* curves. An increase in the price of fixed factors would shift up the *TFC*, *AFC*, *ATC* and *TC* curves, while *TVC*, *AVC* and *MC* curves will remain unchanged. An increase in the price of variable factors would not shift *TFC* and *AFC* curves but would shift the other curves up.

The positions of the short-run variable and marginal cost curves depend crucially on the levels of fixed inputs. Since fixed inputs typically are heavy machinery, land and so on, the levels of fixed inputs define what is briefly called the **plant size**. Hence we can say that the plant size determines the positions of *TVC*, *AVC* and *MC* curves. In particular, the output at which the *ATC* attains its minimum can be

[3]In other words, the *MC* curve is the mirror image of the *MPP* curve.

[4]As a special case, if the *MPP* of each factor starts to decrease right from the beginning of its employment, then the *MC* curve will be upward sloping throughout and so will be the *AVC* and *ATC* curves.

Figure 9.3 An Increase in the Plant Size



interpreted as the most efficient output level corresponding to the plant size. For instance, if you go back to Figure 9.2(c), the most efficient output level relative to the underlying plant size is y_1. An output higher or less than y_1 means over-utilising or under-utilising the plant so that the unit cost is greater than the minimum *ATC*.

Now consider an increase in the plant size. With a bigger plant, the most efficient level of output will be higher. This is illustrated in both panels of Figure 9.3. However, will it mean a lower minimum *ATC* or a higher minimum *ATC*? That depends on the initial plant size. If it is relatively small, an increase in plant size would reduce the minimum *ATC* because of increasing returns to scale (to be discussed later). Otherwise, if the initial plant size is sufficiently large, a further increase in the plant size will raise the minimum *ATC* because of decreasing return to scale (to be discussed later). These possibilities are shown respectively in panels (a) and (b) of Figure 9.3.

As discussed in Chapter 3, businesses pay various taxes to the government like excise tax and VAT, because of their production activities. Such tax payments add to the direct production costs of firms. Thus an increase in the rate of business taxes will shift up the short-run *ATC, AVC* and *MC* curves.

LONG-RUN COST CONCEPTS

In the long run, a producer has more options compared to the short run. Contracts can be changed. Plant sizes can be increased or decreased. In fact, the quantity of any input can be varied. There are no fixed costs. All costs are variable. Thus we simply say 'total cost' instead of 'total variable cost.' The *TC* curve shapes up like the *TVC* curve, passing through the origin. This is shown in Figure 9.4(a) and marked as the *LTC* curve ('*L*' standing for 'long run').

We can link the long-run total cost with the isoquant analysis in the previous chapter. Each point on the long-run *TC* curve is the minimum cost associated with the corresponding isoquant. (The same holds for the total variable cost in the short run.) Consider, for instance, an output level of y_0. Panel (b) of Figure 9.4 shows its

Figure 9.4 Long-run Total Cost Curve

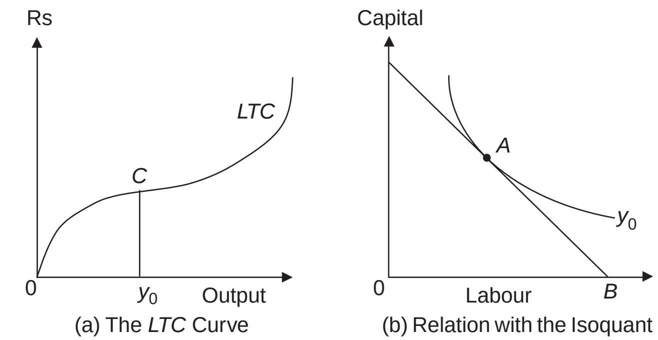
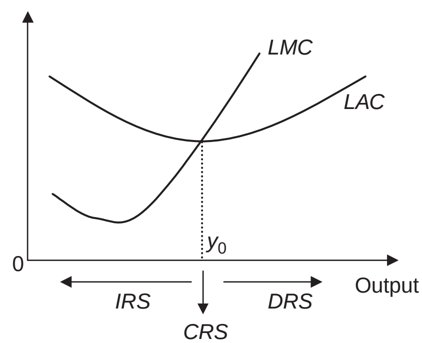


Figure 9.5 Long-run Average and Marginal Cost Curves



isoquant. Cost minimisation occurs with the input bundle A. The total cost of this bundle, in terms of say labour, is *B*. In terms of money this is equal to 0*B* × wage rate, which in turn equals the total cost y_0C in panel (a).

However, the concept of marginal cost remains the same. We call it the long-run marginal cost or *LMC*. There is no difference between *AVC* and *ATC*. Instead, we call it the long-run average cost or *LAC*.

In general, both *LMC* and *LAC* curves are U-shaped, similar to the short-run marginal and average cost curves, as shown in Figure 9.5. But the underlying reason is different and not the law of diminishing returns. This does not mean that for any factor, this law only holds in the short run but not the long run. The difference, however, is that since all factors are variable, it is the nature of returns to scale (studied in the last chapter) which determines how the long-run costs may change with output. For instance, if there are increasing returns to scale (*IRS*), a proportionate change in all inputs leads to a more than proportionate change in the output. This means, for example, that a 10 per cent increase in output can be obtained by a 7 per cent increase in all inputs. In terms of costs, it means that a 10 per cent increase in output will lead to a 7 per cent increase in the total cost and, thus, implies a decline in the average cost. Hence,

IRS ⇒ LAC falls with output.

Similar reasoning implies:

Decreasing Returns to Scale (DRS) ⇒ LAC rises with output
Constant Returns to Scale (CRS) ⇒ LAC is constant.

Therefore, a U-shaped *LAC* curve, as shown in Figure 9.5, which means that *LAC* first declines, then remains constant (at the minimum point) and finally increases, is built on the assumption that as a firm contemplates higher and higher output in the long run, it experiences *IRS* first, *CRS* next and *DRS* finally. The economic rationale behind this sequence is as follows.

Starting from a small scale of output or plant size, an expansion enables a firm to implement further **division of labour**. That is, the firm will be able to allocate its workers to the specialised jobs they are best at. For instance, if initially a firm has one secretary who does typing as well as answers customers, then, as business improves and the firm is in a position to hire, say, two secretaries, it can hire one who is really good at typing and another who is good at dealing with the public. Compared to the initial situation, twice the amount of original work is done now and that too more efficiently. Division of labour tends to reduce the long-run average cost. Another benefit from expansion of output could come from volume discounts in buying inputs. Higher output requires more inputs and in purchasing inputs in bulk, a producer may get discounts. This would also tend to lower the average cost. These are the reasons why the long-run average cost may fall with an increase in output. In other words, these are the sources of increasing returns to scale—also called **economies of scale**.

However, after a certain point when the output reaches a critical limit, inefficiency in management creeps in due to overcrowding and congestion of inputs. These are **diseconomies of scale**, leading to decreasing returns to scale. Any further expansion of output leads to an increase in the plant size accompanied by an increase in the long-run average cost.

In between *IRS* and *DRS*, there are neither economies nor diseconomies of scale; instead, the firm experiences *CRS*. If this happens over an interval of output, the *LAC* will have a flat segment. Otherwise, it will be just a point.

Our reasoning of why the *LAC* curve is U-shaped is complete. It is the U-shape of the *LAC* curve that explains the same shape of the *LMC* curve.[5]

A change in technology, a change in an input price or a change in the rate of business taxes shifts the long-run average and marginal cost curves too. As one would expect, a technology improvement shifts them downward, while an increase in an input price or business taxes shifts them upward.

[5]This is in contrast to the short run, where the shape of the marginal cost curve determines the shape of the average variable and average total cost curves.

RELATIONSHIPS BETWEEN SHORT-RUN AND LONG-RUN COST CURVES

The plant size is given in the short run whereas it can be varied in the long run (by definition). Hence, the long-run cost at any given level of output should take into account the firm's choice of the plant size. Figure 9.6 depicts a family of short-run *AC* and *MC* curves. As discussed earlier, a larger plant size is associated with a higher level of the most efficient output. But the minimum *AC* decreases or increases with plant size as the initial plant size is relatively small or large. Figure 9.6 exhibits this pattern.

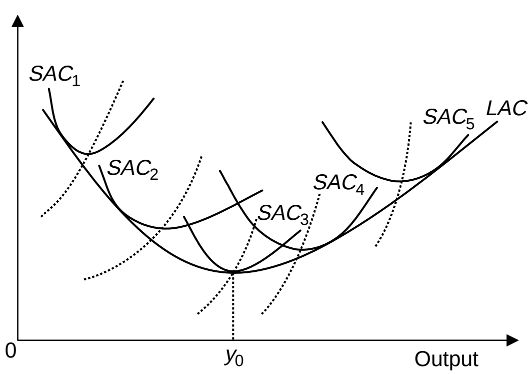
Two points are significant in this regard. (*a*) The long-run *AC* cannot be independent of the short-run *AC*. (*b*) At any given level of output, the long-run *AC* cannot exceed the short-run *AC*, because there is more flexibility in the long run. Put differently, the firm always has the choice, in the long run, to select the plant size chosen in the short run and hence the *LAC* cannot be higher. (*a*) and (*b*) imply that the *LAC* curve must be tangential to or an envelop of the family of short-run *AC*s.

Notice the output level y_0 at which the *LAC* as well as the associated short-run *AC* (*SAC*) is minimised. At any output level below (or above) y_0, the tangency point lies to the left (or right) of the minimum point on the corresponding *SAC*.

SOCIAL AND PRIVATE COSTS OF PRODUCTION

The cost curves we have studied so far relate to firms or an industry. Do they also reflect the society's cost of producing the same good? Of course, the firm or the industry in question is a part of the society and thus its costs must be included in the society's cost. But the industry's and the society's cost may not be the same

Figure 9.6 The Short-run and the Long-run Cost Curve



always because productive activity by one firm or industry may affect other firms, other industries or households.

Typically, the production of many industrial goods generates industrial waste, causing pollution. Many steel plants emit smoke, which has a very high carbon dioxide content. Wastes from chemical plants are highly polluting. Suppose a chemical plant does not have a proper disposal system and it simply channels its waste to a nearby river.[6] Many people (and animals) use the river water for a variety of purposes. Obviously, the polluted water causes health problems—it is an environmental concern.

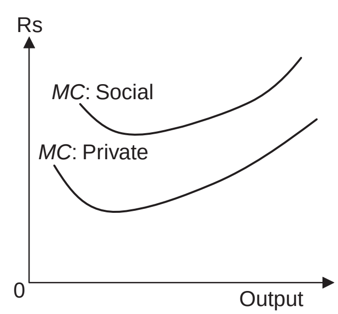
Are the chemical plant's marginal and average costs same as the society's marginal and average costs? The answer is no. The plant's own costs of producing output in terms of payments to labour, raw material, capital, land and so on are its private costs. But they do not include the environmental costs of the output to society at large. The social costs, which must include the cost to the environment, are higher.

In this example, we say that an increase in the firms' output causes **an externality** to society and it is a **negative externality** in that the firm's output, via creating or increasing pollution, adversely affects society.

We term the firm's own cost as **private cost**; **social cost** is defined as the sum of the private cost and the cost of the externality caused to society. Figure 9.7 illustrates these curves in terms of the marginal cost when a negative externality is present.

In some cases there may be a **positive externality**. Suppose a new nursery has come up close to your neighbourhood. You, as part of the general public, enjoy the smell of flowers and the greenery from plants and small trees. This is an example of positive externality—in this case, the social marginal cost will be less than the private marginal cost.

Figure 9.7 The Private and Social Costs



[6]In reality, many chemical plants do have proper disposal systems.

HOW DO FIRMS FINANCE THEIR COSTS?

Several concepts regarding costs have been discussed. But you may ask a basic but relevant question that we have not touched upon as yet. That is, how does a firm finance its costs? Put differently, what are the sources of funds from which a firm is able to pay the factors of production?

In general, there are four sources. One is the **plough back of profit** or what is called **retained earnings**, referring to a part of its own accumulated profit being used towards covering some costs. It simply means using the firm's own money. A small entrepreneur may use part of his profit income for own consumption and invest the rest in his firm. A corporation pays out a part of its profit earnings to shareholders, who are the owners of the firm. The rest are retained earnings, which are 'ploughed back' to the firm.

Businesses however, do not always, use their own money for covering costs. **Loans from financial institutions** are another source. A loan is a financial contract between the lender (the financial institution like a commercial bank) and the borrower (the firm), specifying the duration of loan and an interest rate. The lending institution typically reviews the project for which the loan is being requested. Further, as a security to itself, it holds some property of the firm (borrower) as the **mortgage**. If, for some reason, the firm is not able to pay back the loan, the lending institution has the right to sell off the mortgaged property and use the proceeds to recover its loan. If the sale of the mortgaged property does not fully recover the loan, the firm owners are personally liable to pay the rest.

New stocks and bonds are the remaining two ways of financing. Through issuing these, the firm raises funds from the general public. These are already described in Chapter 7. Basically, stocks offer an ownership of the firm whereas bonds do not. Stocks and bonds are respectively **equity and debt instruments** issued by a firm.

There are trade-offs between these alternative ways of raising funds. It is not in the interest of the company's management to always plough back a large fraction of profits because this would mean less dividend payments and a smaller return to stockholders. Hence, the public would not want to invest much in the company via buying its stocks. It will be harder for the company to raise money in the future through issuing stocks. On the other hand, it is not wise to always pay back most of the profit earnings to the shareholders because, all else the same, this would leave less amount for plough back and expansion of the company's operation. A balance has to be struck between dividend payments and retained earnings.

Raising funds through stocks, on one hand, and loans and bonds, on the other, have their advantages and disadvantages. It is interesting to observe that while for an investor stocks are riskier than bonds, for the firm it is the opposite. Since lenders to the firm and bondholders must be paid, even when the times are bad, it is risky for the firm to raise funds through loans and bonds. On the other hand, if profits are high, stockholders will be paid more in the long run (as they are the owners), while lenders and bondholders will be paid only to the extent of the interest rate inherent in a bond. Thus, bonds are 'cheaper' to the firm than stocks but riskier.

It is also interesting that from the viewpoint of a firm's management, plough back is the easiest choice because there is no outside scrutiny. If, instead, a company has to issue new bonds or stocks, or even apply for loans to a bank, it has to do a lot of paper work; submit it to the right authorities; and then the company's performance is evaluated. There is no such thing as perfect management. Hence, some weaknesses are bound to come out in the open. Yet, as we have just discussed, the easiest choice of plough back has its own limitations.

Economic Facts and Insights

- In economic analysis, costs always refer to economic costs—the sum of explicit (accounting) costs and implicit costs.
- Fixed costs are independent of the output, while variable costs are not.
- The total variable cost in the short run or the total cost in the long run is the minimised total cost in the cost minimisation problem.
- The short-run average variable and average total cost curves are U-shaped because the short-run marginal cost curve is U-shaped. The explanation of the latter lies in the law of diminishing returns.
- A technological improvement shifts the cost curves downward, while input price increases or increases in business taxes shift them upward.
- An increase in the plant size may increase or decrease the minimum average total cost when there are increasing or decreasing returns to scale respectively.
- The concepts of marginal product and the law of diminishing returns are valid even in the long run when all factors are variable.
- In the long run, all costs are variable.
- The long-run average cost is U-shaped because, as output expands, a firm typically experiences increasing returns to scale, followed by constant and decreasing returns to scale. The U-shape of the long-run average cost curve implies the U-shape of the long-run marginal cost curve.
- Division of labour and volume discounts on purchase of inputs are sources of economies of scale, whereas management inefficiency due to the large scale of operations, overcrowding and congestion of inputs cause diseconomies of scale.
- Negative (positive) externalities imply that social marginal cost is higher (less) than the private marginal cost.
- Pollution imposes a negative externality on society.
- Firms finance their costs through retained earnings, loans from financial institutions and through issuing new bonds and stocks.

(continued)

- As a means of financing its costs or operation, issuing bonds are cheaper but riskier than issuing stocks for a firm.
- Issuing new bonds and stocks or applying for sizeable loans from financial institutions are generally accompanied by an outside scrutiny of the concerned firm's performance.

EXERCISES

9.1 Give a hypothetical—yet related to the real world—example of an implicit cost.

9.2 'Economic cost is a part of accounting costs'. Defend or refute.

9.3 What is an overhead cost? Give two examples.

9.4 'Marginal cost is the addition to the total cost, not to the total variable cost, per one extra unit produced'. Defend or refute.

9.5 What is the area under the marginal cost curve equal to—total cost or total variable cost?

9.6 Explain why the *MC*, *AVC* and *ATC* curves are U-shaped?

9.7 Briefly state and explain the relationship between *MC*, *AVC* and *ATC* curves.

9.8 *AVC* is minimised at a level of output, which is _____ than that at which *ATC* attains its minimum value. Fill in the blank and give reasons.

9.9 What is the vertical difference between the *TC* and *TVC* equal to and why?

9.10 A firm's total cost schedule in the short run is the following. What is its total fixed cost? Derive the marginal cost schedule.

Output	*Total Cost (Rs)*
0	10
1	25
2	35
3	43
4	54
5	69
6	91
7	121
8	161

9.11 Briefly describe the relationship between the short-run average cost curve and the long-run average cost curve.

9.12 How does an increase in business taxes shift the marginal and average cost curves?

9.13 How does a technical progress shift the marginal and average cost curves?

9.14 How would an increase in input prices affect the marginal and average cost curves?

9.15 Briefly describe the link between the cost-minimisation problem studied in the previous chapter and the long-run cost curves.
9.16 Suppose the long-run marginal cost is constant (independent of the output). That is, *LAC* curve is a horizontal line. Draw the long-run marginal cost curve relative to this *LAC* curve. Give reasons.
9.17 Suppose a firm experiences increasing returns to scale initially over a range of output, constant returns to scale over another range and then decreasing returns to scale. Draw the long-run average cost curve.
9.18 Briefly explain the relationship between the returns to scale and the shape of the long-run average and marginal cost curves.
9.19 'Increasing and decreasing returns to scale respectively imply the downward portion and the upward portion of the long-run average cost'. Defend or refute.
9.20 What are the reasons behind a firm experiencing economies of scale first, followed by diseconomies of scale?
9.21 What do positive and negative externalities mean?
9.22 The production activity of a firm causes a negative externality. How will the social marginal cost be different from the private marginal cost?
9.23 A chemical plant owned by an Indian company operates in Kenya and it causes pollution in that country. Is there a difference between private cost and social cost in terms of this company's production of chemicals in this plant?
9.24 What are the different sources of funds available to the firm from which it can finance its costs?
9.25 What are retained earnings?
9.26 In raising funds from the public through issuing stocks and bonds, which is the riskier way for the firm? If it is riskier, then why would a firm choose that way to finance its expenses?

10 Profit-maximisation, Perfect Competition and the Supply Curve

CONCEPTS

- Accounting Profit
- Economic Profit/Abnormal Profit/Producer's Surplus Profit
- Perfect Competition or Perfectly Competitive Market
- Price Taker
- Price Line
- Total Revenue
- Marginal Revenue
- Shut-down Condition
- Break-even Price
- Market Period

Given our understanding of various concepts relating to production and costs in the last two chapters, we are ready to analyse the profit-maximising behaviour of a firm and this will lead us to the supply curve, which was introduced in Chapter 3.

ACCOUNTING PROFIT AND ECONOMIC PROFIT

Profits, by definition, are equal to the difference between the total revenue generated from selling the output and the total cost of output. Depending on which part of the cost is included in calculating profit, there are two notions of profit.

One is **accounting profit**, meaning total revenues minus total explicit costs. The other is **economic profit**—also called **abnormal profit** or **producer's surplus** or simply **profit**—which is equal to the total revenues minus the economic costs (the sum of explicit and implicit costs).

An example should make these differences clear.

NUMERICAL EXAMPLE 10.1

Dr Sahu owns the ground floor and the basement of a building in Bhubaneswar. He lives on the ground floor and rents the basement to a bank for Rs 2 lakh a year. He used to work for a local hospital, but last year he decided to take the basement off rent and opened a clinic of his own there. For one year, the annual 'turnover', a common business term meaning total revenues collected from services offered to patients, was Rs 20 lakh. The salary expenses paid to his employees amounted to Rs 4 lakh. The cost of electricity, phone and other utilities ran into Rs 3 lakh. The cost of medical supplies and other raw materials was Rs 4 lakh. These are all the expenses he had to incur for the business. What are his explicit costs, accounting profit and economic profit?

Salary expenses, utility costs, and costs of medical supplies and raw materials sum up to give explicit costs, equal to Rs (4 + 3 + 4) lakh = Rs 11 lakh. The foregone rental income is Rs 2 lakh, which is the implicit cost. The (total economic) costs, including explicit and implicit costs, are equal to Rs (11 + 2) = Rs 13 lakh. Hence the accounting profit = Revenues – Explicit costs = Rs (20 – 11) lakh = Rs 9 lakh. But the economic profit = Revenues – Total Costs = Rs (20 – 13) = Rs 7 lakh.

In what follows, we always refer to the economic profit.

REVENUES

How are revenues related to a firm's sale of output in the market? The answer depends on the **market structure**—meaning the number of firms operating in the market, the nature of competition between firms, the ease of entry into and exit from the industry by a firm and the nature of the product. In this chapter we focus on one particular market structure called perfect competition, which is of central importance in economics.

Perfect Competition

Perfect competition or **a perfectly competitive industry** is defined by the following three characteristics:

(*i*) All firms sell a very homogeneous (that is, an identical) product.
(*ii*) There are a large number of buyers and a large number of sellers (producers or firms) in the industry.
(*iii*) There is free entry and exit.

Consider, for instance, wheat or onions. Each of these is a very homogeneous product—same or identical everywhere (unlike cars that come in various designs, quality and so on). The same is true of leather footballs or cooking gas. In general, by a 'very homogeneous' or 'identical' product, we refer to a very standardised item for which inherent quality differences are minimal; which does not come in different designs; and on which consumers do not have different perceptions.

As for more examples of which goods are homogeneous and which goods are not, various computer-related items are fast becoming standardised items these days, which was not true 20 years ago. A CD R-W drive, for example, is almost the same everywhere in the world with minimal quality differences among different brands. In contrast, TVs are available in different sizes, features and looks. They are certainly not homogeneous. Different types of cotton shirts may have the same 'quality' but their styles may be different so that they are not necessarily homogeneous from the perspective of consumers. Sometimes, products like toothpaste or lipstick may be basically the same, but due to the effect of advertising, consumers see them differently—and their perception matters. Such products are not considered homogeneous from the viewpoint of economic analysis.

The important implication of a product being homogeneous is that all firms producing and selling it have to charge the same price. This is because, between two firms, if one charges a higher price than the other, no one will buy from the former.[1]

Given that a product is homogeneous, the existence of a large number of firms implies that each firm is very small compared to the whole market. Thus no single firm can influence the price. In other words, a firm in a perfectly competitive industry is a **price taker**—it takes the market price as given. We can say that the market price is *exogenous* to any individual firm. This is another way of saying that in a perfectly competitive market, no single firm possesses any market power.[2]

The last feature leads to a simple relationship between a firm's output and a firm's **total revenue** (*TR*), equal to total money generated from sales. By definition,

[1]This is, of course, barring transport costs and ignoring informational problems facing consumers regarding which price prevails where.

[2]While no single firm has any power to dictate or influence price, nonetheless, a price prevails in the market through the interaction between sellers and buyers. In the next chapter, we shall study how the price is determined in a market.

TR = price (p) × quantity sold or produced (y).[3] Since p is given to the firm, its total revenue must equal $p \times 1$ if one unit is sold, $p \times 2$ if two units are sold and so on. Plotting these points against the output we obtain the total revenue curve, as shown in Figure 10.1. It is a ray from the origin with a slope equal to p everywhere because p is given to the firm.[4] A decrease or an increase in p rotates this curve clockwise or anti-clockwise respectively.

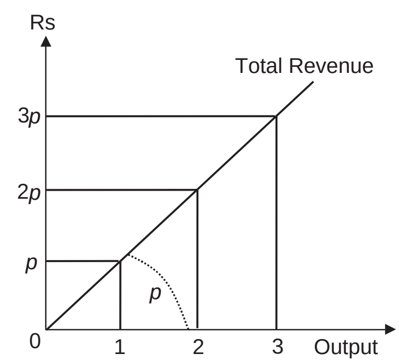
By now you might guess that with any 'total', there is always a 'marginal' and an 'average'. This is true here. We define **marginal revenue** (MR) as addition to the total revenue from one extra unit produced and sold. In Figure 10.1 notice by how much TR increases as one extra unit is produced. As output is increased from 0 to 1, $MR = p - 0 = p$. Similarly, if output increases from 1 unit to 2 units, $MR = 2p - p = p$. Thus, under perfect competition, $MR = p$. This is because p is exogenous to a firm.[5]

We now define average revenue (AR) as TR/output. Since $TR = py$, we have $AR = py/y = p$. Thus $AR = p$. Note that this has nothing to do with p being given to a firm. It holds in competitive and non-competitive markets.

As you will see shortly, it will be useful to have another graph called the **price line**, defined as a plot of price facing a competitive firm against its output. However, since p is exogenous to the firm it is simply a flat line as shown in Figure 10.2. From the price line we can measure the total revenue. If, for instance, the output = y_1, $TR = 0p_0 \times 0y_1$ = the area $0p_0A_1y_1$. Similarly at $y = y_2$, the total revenue equals the area $0p_0A_2y_2$ and so on. Hence, the (rectangular) area under the price line measures total revenue for a perfectly competitive firm.

Sometimes the price line is referred to as the 'demand curve facing a competitive firm' in the sense that irrespective of how much a firm wants to sell or equivalently how much the consumers are buying from the firm, the price they are

Figure 10.1 The Total Revenue Curve

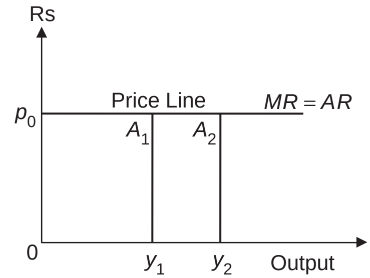


[3]It is presumed that whatever quantity is produced is sold. There are no inventories.

[4]If the firm has any influence on the price, the total revenue curve will not be a straight line.

[5]If the market structure is not competitive, *MR* will not be equal to *p*.

Figure 10.2 The Price Line



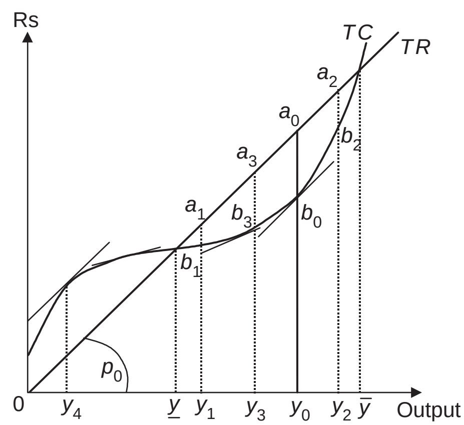
paying does not change.[6] We note that this demand curve facing a firm has price elasticity, equal to ∞.

PRODUCER'S EQUILIBRIUM IN THE SHORT RUN

We are ready to analyse the profit-maximising behaviour of a competitive firm. Since profits are the difference between total revenue and total cost, one way to understand profit-maximisation is to look at the curves representing these in a single diagram as in Figure 10.3.

We can 'read' profits at any level of output. Notice that at outputs $\underline{y}$ and $\overline{y}$, the two curves intersect, implying that $TR = TC$ and thus profits are zero. In other words,

Figure 10.3 Profit-maximisation by a Competitive Firm



[6]This is not, however, contrary to the law of demand, because here we are talking about the demand for the product by consumers from only one firm, not their total demand for the product.

these are the 'break-even' levels of output. If the output is less than $\underline{y}$ or greater than $\overline{y}$, $TR < TC$ and hence profits are negative, that is, there are (abnormal) losses.

However, between $\underline{y}$ and $\overline{y}$, $TR > TC$ and there are (abnormal) profits. Obviously, the firm would produce somewhere in this range. See Figure 10.3 and note that producing y_3 generates profit (equal to a_3b_3), which is higher than the profit (a_1b_1) associated with y_1 or the profit (a_2b_2) associated with output y_2. But what is the level of output at which the profit is the maximum? Realise that it has to be the output y_0 at which the slope of the TR curve (= the price p_0) is equal to the slope of the TC curve (= the marginal cost), that is, the tangent to the TC curve is parallel to the TR line. Any output higher or less than y_0 yields less profit compared to y_0. We can then say that profit is maximised at that level of output at which price equals marginal cost.

This condition does not, however, completely identify the profit-maximising output. Notice that, at the output y_4 also, the slope of the TR line is equal to the slope of the TC curve, that is, $p = MC$ holds. But, obviously, profit is not maximised at y_4. Instead, there is a loss at this output. Apart from profit and loss, there is another difference between y_4 and y_0—MC (the slope of the TC curve) decreases with output around y_4, while it increases with output around y_0. This is the basis on which we can differentiate between y_0 and y_4, and identify the profit-maximising output y_0. In general then, the profit-maximising output or the **optimal output** of a competitive firm is characterised by the following conditions:

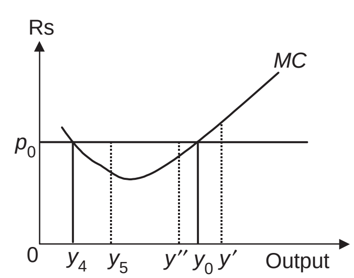
$$p = MC; MC \text{ is increasing in output.} \tag{8.1}$$

Often, the condition that MC is increasing is implicitly understood by economists and the profit-maximisation condition for a competitive firm is (loosely) stated as '$p = MC$'.

The economic reasoning behind the profit-maximising conditions (8.1) can be understood via Figure 10.4, which depicts the price line at $p = p_0$, and a firm's MC curve corresponding to the TR and TC lines in Figure 10.3.

We see that $p = MC$ holds at outputs y_0 and y_4, which correspond to the same points in Figure 10.3. Around y_4, MC is increasing, while around y_0, it is decreasing in output. Applying economic reasoning, we now argue that profit is

Figure 10.4 Profit-maximisation Once Again



maximised at y_0, not at y_4. Suppose that the firm is producing at y_0. Consider an alternative of producing a slightly higher amount, say y'. Since *MC* is increasing, we see that $MC > p_0$. By definition, *MC* is the cost of producing the additional amount. Similarly, $MR = p$ is the additional revenue. Thus, $MC > p_0$ means that the additional cost is greater than the additional revenue. It implies that producing an amount slightly higher than y_0 yields less profit. Similarly, if the firm produced a bit less than y_0, say y'', $MC < p_0$. This holds also because *MC* is increasing in output around y_0. That $MC < p_0$, as output is reduced from y_0, means that savings in cost are less than revenues foregone. It is also a worse situation compared to producing at y_0. This proves that around y_0, producing y_0 yields maximum profit.

On the other hand, suppose the firm is producing at y_4. Because *MC* is decreasing, if the firm produces a slightly higher amount (say y_5), the market price would exceed *MC*, that is, extra revenue to be earned would be higher than extra cost to be incurred. This will bring more profits. Therefore, producing at y_4 cannot be maximising profits. This completes the argument that out of the two output levels where $p = MC$, the profit-maximising one is where *MC* increases with output.

Mathematically Speaking

Profit-maximisation by a Competitive Firm

A competitive firm's total revenue has the expression, py. Let $C(y)$ denote the total variable cost function and F the total fixed cost. Note that the marginal cost is the first derivative of the total variable cost function, equal to $C'(y)$. Let $C''(y)$ denote the second derivative of the total variable cost function; if $C''(y)$ is positive (or negative), the *MC* increases (or decreases) with output. We can now express profits, π, as the following function of y:

$$\pi = py - C(y) - F.$$

The firm chooses y so as to maximise π. The first-order condition of this maximisation is:

$$\frac{d\pi}{dy} = p - C'(y) = 0. \tag{8.2}$$

This says that $p = MC$. The second-order condition is that $d^2\pi/dy^2 < 0$. (The first-order and second-order conditions of unconstrained maximisation such as this is outlined in the General Appendix.) Totally differentiating $d\pi/dy$ with respect to y,

$$\frac{d^2\pi}{dy^2} = -C''(y).$$

Thus, the second-order condition is met if and only

$$C''(y) > 0 \tag{8.3}$$

(8.2) and (8.3) are indeed the conditions of profit-maximisation summarised in (8.1).

* * * * * *

PROFITS, LOSSES AND THE SHUT-DOWN CONDITION

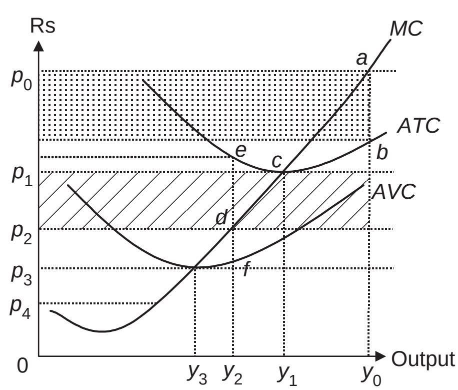
Profits and losses can be seen in a different diagram, which is quite useful. Turn to Figure 10.5, depicting various price lines and the firm's *MC*, *AVC* and *ATC* curves. At price p_0, the profit-maximising output is y_0, as the *MC* at this output is equal to p_0 and the *MC* curve is rising. The firm is making profits. The *ATC* is indicated as *b* and the profit per unit of output, $p - ATC$, is the line segment *ab*. Total profits are shown by the rectangular dotted area, equal to output times the profit per unit of output.

Consider next the price p_1. The price line goes through the minimum point on the *ATC* curve, that is, p_1 equals the minimum *ATC*. The optimal output is y_1. Notice that at this output $p = ATC$ and thus profits are zero; in the business language, the firm is 'breaking-even'. We can generally define **break-even price** as that price at which profits are zero. The break-even price for a competitive firm is equal to the minimum *ATC*.

It is clear that at any price less than p_1, the firm incurs losses. Consider p_2, for example. The optimal output is y_2. The *ATC* is indicated at the point *e* and *ed* measures losses per unit. Total losses are shown by the rectangular shaded area. Similarly, you can calculate losses at the price p_3. They will be greater. At a further lower price such as p_4, losses will be even greater than at p_3.

A natural question arises here—should the firm not shut down (that is, produce nothing) if losses are being incurred? The answer is—not necessarily. Why? It is because in the short run the firm has to pay the fixed cost even if it decides not to produce any amount. For example, it has to pay the rent according to the lease agreement, repay loan to the bank and so on. In other words, shutting down does not mean that the profits are zero. It also involves a loss, equal to *TFC*. It is then possible that by not shutting down, the firm's losses are less as compared to shutting down. If the firm shuts down, its losses are *TFC*. Instead, if it operates (and produces

Figure 10.5 Profits, Losses and the Shut-down Condition



where $p = MC$), the losses are $TC - TR = TFC - TVC - TR = TFC - AVC \cdot y - py = TFC - (AVC - p)y$. Therefore, the difference between the two levels of loss = $(AVC - p)y$. It then follows that the firm shuts down if $p < AVC$, that is, if the market price is less than the average variable cost at the profit-maximising output.[7]

It turns out that the above condition can be stated with more precision. Consider the particular price, which is equal to the minimum AVC. In Figure 10.5, this price is p_3. Notice that at this price if the firm follows the $p = MC$ rule, the optimal output is y_3. Since at this output $p = AVC$, the firm should be indifferent between shutting down and not shutting down. It then follows that at any market price less than p_3 (equal to the minimum AVC), we have $p < AVC$ and the firm should shut down. Therefore, we can state that at the **Shut-down Condition**,

$$p < \text{minimum } AVC. \tag{8.4}$$

NUMERICAL EXAMPLE 10.2

A competitive firm has a short-run total cost schedule as given in Table 10.1. What is the market price at which it will break even? What is the market price at which it will decide to shut down?

Dividing the total cost by output gives the *ATC*. See the level at which the output *ATC* is minimised. The minimised *ATC* itself is the break-even price. As we go down Table 10.1 from output equal to 1 to output equal to 8, the *ATC*s are respectively 40, 24, 18, 15, 13.6, 13, 13.85, 14.25 and 16. Hence the break-even price is Rs 13.

Next, note in particular that at zero output, the total cost is Rs 30. Hence, Rs 30 is the total fixed cost. If we deduct this from the total cost, we get the *TVC*. As we increase output, starting at 1, the *TVC*s are respectively equal to 10, 18, 24, 30, 38, 48, 65, 85 and 114. Dividing these by the respective outputs, *AVC*s are respectively equal to 10, 9, 8, 7.5, 7.6, 8, 9.27, 10.625 and 12.67. The minimum *AVC* is the price at which the firm will shut down and this is equal to Rs 7.5.

Table 10.1 Short-run Total Cost Schedule (Numerical Example 10.2)

Output	*Short-run Total Cost (Rs)*
0	30
1	40
2	48
3	54
4	60
5	68
6	78
7	95
8	115
9	144

[7] If $p = AVC$, the firm is indifferent between shutting down and not shutting down.

NUMERICAL EXAMPLE 10.3

Consider a competitive firm whose short-run total cost schedule is given in Table 10.1. The market price it faces is Rs 20. How much should it produce in order to maximise its profits? Is the firm making profits or incurring losses?

We first derive the marginal cost schedule from the total cost schedule. This is given in Table 10.2. The price = marginal cost condition is satisfied at output = 8. We also check that marginal cost is increasing around this level of output (it is 17 at output = 7, 20 at output = 8 and 20 at output = 9). Thus, profit is indeed maximised at output = 8. The maximised profit is equal to $TR - TC$ = Rs 20 × 8 – Rs 115 = Rs 160 – Rs 115 = Rs 45.[8]

SHORT-RUN SUPPLY CURVE

Once we understand profit-maximisation and the shut-down behaviour, it is easy to derive the supply curve, introduced in Chapter 3. Recall that a firm's supply curve is one showing the various quantities it is willing to produce and supply to the market at various prices, given technology and input prices. It is the counterpart of the demand curve in consumer theory. In consumer theory, the demand for a commodity was derived from the consumer's optimal choice of the consumption bundle. Likewise, the supply curve is derived from the producer's optimal choice of output.

Turn back to Figure 10.5. If the market price is p_0, we know that the profit-maximising output is y_0, where $p_0 = MC$. Thus, y_0 is what the firm would supply to the market at price p_0. The graph of this price-quantity pair is simply the point a. Similarly, at the price p_1, the firm would supply the quantity y_1 and the corre-

Table 10.2 Short-run Marginal Cost Schedule (Numerical Example 10.3)

Output	*Short-run Marginal Cost (Rs)*
0	–
1	10
2	8
3	6
4	6
5	8
6	10
7	17
8	20
9	29

[8]Another way to check whether the firm is making a profit or a loss is to compare the price with the break-even price. We already know from the Numerical Example 10.2 that the break-even price is Rs 13. Thus the market price is greater than the break-even price and, therefore, the firm must be making profits. Moreover, note that profit is maximised at output = 7 also. Two alternative answers arise because output is not measured on a continuous scale in this example—output only takes integer values, not decimals.

sponding point on the *MC* curve is *c*. The locus of points *a*, *c* and so on along the *MC* curve is then the supply curve. However, not all the points on the *MC* curve would lie on the supply curve. We have just seen that the firm does not operate below the minimum *AVC*, which is also a point on the *MC* curve. We can then say that a firm's short-run supply curve is its short-run marginal cost curve over and above the minimum *AVC* point. Since *MC* is rising, over and above the minimum *AVC* point on it, the supply curve is upward sloping. This is the economic foundation behind the supply curve.

What factors can shift a firm's supply curve? Since it is a part of the *MC* curve, whatever shifts the *MC* curve also shifts the supply curve. From what we have learnt in Chapter 9 regarding the shift of the short-run *MC* curve, it follows that a technological improvement shifts the supply curve to the right, whereas an increase in an input price or in the rate of a business tax shifts the supply curve to the left.[9]

Time Horizon and the Market Period

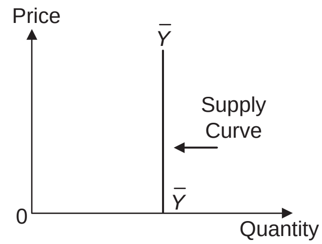
Remember that we are considering a short-run period, meaning an interval of time during which a firm can vary some, but not all inputs.

However, if the time period is very short, it is only natural that a firm cannot vary the employment of any input. Such a short period is called market period. How does the supply curve in a market period look like?

For instance, think of an agricultural good like paddy. It takes months between preparing the soil for the seeds and getting paddy ready for sale in the market during the spring season. How does the supply curve of paddy look like during the spring time? Given that it is already harvested and ready for sale, the quantity is fixed, irrespective of the price it is going to fetch. Accordingly, the supply curve for paddy during spring must look vertical, as shown in Figure 10.6, where $\bar{Y}\bar{Y}$ is the given quantity available for sale.

In sum, a period so short that no inputs and output can change is called **market period** and the corresponding supply curve in the market period is vertical.

Figure 10.6 Supply Curve in the Market Period



[9]From the individual, firm-level supply curves, the market supply curve is obtained by horizontally summing these. In Chapter 3, we have already discussed which factors shift the individual and market supply curves and how as well as the concept of price elasticity of supply.

Equilibrium in the Long Run

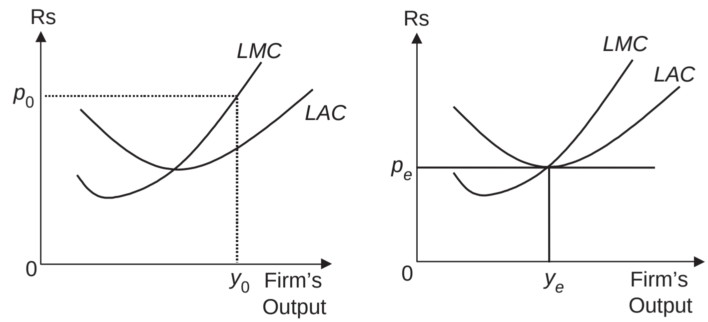
In the long run, when all—rather than some—inputs are variable, the profit-maximising behaviour of the firm remains qualitatively the same as in the short run. Profit is maximised at the output where the long-run marginal cost (*LMC*) is equal to the price and *LMC* is increasing in output. For instance, in Figure 10.7 that shows the *LMC* curve and the *LAC* (long-run average cost) curve, if market price is p_0, the firm produces y_0 and there are abnormal profits. At p_e, the firm produces y_e and the profits are zero. This is the break-even price, equal to the minimum *LAC*.

However, there is one important difference with the short run. Since there are no fixed costs, the firm would shut down—and leave the industry—as long as any loss is made. Why should a firm stay if it expects to lose in the long run? In view of Figure 10.7(b), we see that losses occur when price is less than p_e. Hence, output or supply is zero as long as the price is less than the minimum average cost. This means that the long-run supply curve of a firm is the long-run marginal cost curve over and above the minimum long-run average cost point.

Comparison with the Short-run Supply Curve

Although both short-run and long-run supply curves are upward sloping, there is a qualitative difference between them (apart from that in terms of which part of the marginal cost curve they belong to). Recall that in the market period, no inputs can be varied and the supply curve is vertical. Over the short run, the firm has some flexibility, that is, some input level can be changed and the supply curve is upward sloping rather than vertical. Thus, the short-run supply curve is more elastic than the market-period supply curve. In the long run, the firm has even more flexibility as it can change the level of any input. Therefore, it can adjust its output more with respect to a change in the product price. Accordingly, the long-run supply curve is

Figure 10.7 Long-run Competitive Equilibrium



(a) Positive Profits

(b) Zero Profits

more elastic than the short-run supply curve. In terms of graph, it means that the long-run supply curve is flatter than the short-run supply curve.

This is illustrated in Figure 10.8. Suppose that initially in a long-run equilibrium the firm was producing Y_0 at price p_0. If there is a price change from p_0 (upward or downward), the output or supply adjustment is greater along the long-run than along the short-run supply curve.

Entry and Exit

Notice that free entry and exit is a characteristic of a perfectly competitive market and we have not said anything about it as yet. But it is relevant because in the long run, firms would like to enter into the industry or exit from it depending on whether they are earning (abnormal) profits or incurring (abnormal) losses. See Clip 10.1 for a newspaper article on how taxes have forced entrepreneurs in the movie-theatre industry to quit a market.

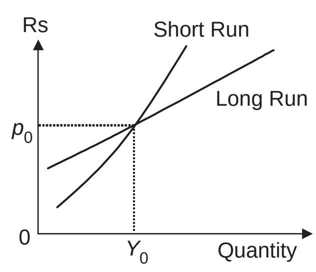
Industry equilibrium in terms of entry and exit is attained in the market when no firms have any incentive to enter or quit. This will happen when economic profits are zero. That is, entrepreneurs earn only their opportunity cost or what is called **normal profit**.

What is the zero-profit condition? We can write profit as (price – LAC) × output. Thus, profits are zero if and only if price = LAC. Combining this with the profit-maximising condition of a firm, the overall firm-cum-industry equilibrium conditions are then:

$$\text{Price} = LMC = LAC. \tag{8.5}$$

The firm-industry equilibrium is illustrated back in Figure 10.7, panel (b). The equilibrium price is p_e, equal to the long-run minimum average cost, and the equilibrium firm-output is y_e. We see that at this price-output combination, both the profit-maximising (price = MC) condition and the zero-profit (price = AC) condition are met.[10] Implicit here is that there is an equilibrium number of firms operating in the

Figure 10.8 Short-run versus Long-run Supply Curve



[10]An important property of this long-run competitive equilibrium is that if we know the long-run average cost curve and we know the long-run price of the product. In this sense, price is determined by technology only, not by demand—an observation made first by Alfred Marshall, who is regarded as the father of modern microeconomics.

Clip 10.1: 21 May 2005, Hindustan Times

Gurgaon cinemas give up, plan exit

S.K. Ahuja
Gurgaon, May 20

THE SHOW is over for Gurgaon multiplexes. Grappling with heavy losses, multiplex players have said that they will have to shut shop unless Haryana cuts its steep entertainment tax of 50 per cent.

DT Cinemas will start shutting its theatres one at a time from July and has shelved plans to open up more multiplexes in the state. Other multiplex chains like INNOX and FAME have also put off plans to enter the Haryana market for the same reason.

"We have suffered a loss of Rs 6 crore in two years. We will shut two of our theatres in July and all six later if the government does not exempt tax. We cannot sustain losses anymore," said Kajal Aijaz, CEO, DT Cinemas.

Aijaz said though DT Cinemas was willing to wait five years for its occupancy rate to pick up, the entertainment tax was crippling its functioning.

Ajay Bijli of PVR Cinemas, which has seven screens in Gurgaon, confirmed that PVR was losing Rs 2 crore a year and demanded that the state's entertainment tax be done away with. "Looking at the mounting losses, we will have no option but to shut down in due course," he told HT.

With cinemas being big crowd-pullers, this is not good news for the malls-cum-multiplexes that have seen a spate of closures by stores recently.

Admitting that the high tax was a big deterrent to entering the market, Shyam Shroff of FAME said: "We had plans to set up a multiplex in Gurgaon but we backed out because of the high tax structure."

DT and PVR have formed an association, The Multiplex Association of Haryana, to take up their cause with the government. But though Aijaz, its president, has already met Haryana Chief Minister B.S. Hooda, there hasn't been any progress on the tax front so far.

market. Relative to market demand, too many firms would mean abnormal losses and too few firms would imply abnormal profits. A diagram such as Figure 10.7(b) cannot, however, show the equilibrium number of firms.

Realise that in the presence of entry and exit, an individual firm's supply curve remains the same as before. But the nature of the industry supply curve—the horizontal sum of individual supply curves—is quite different because the number of firms in the industry changes in response to profits or losses being made. It is interesting that the long-run industry supply curve with free entry and exit of firms is a horizontal straight line (rather than an upward sloping curve). How is that?

Given the cost curves, we know the minimum *LAC*, say, equal to m.[11] Consider the following possibilities. (*i*) Suppose the long-run market price is less than m.

[11] In Figure 10.7(b), $m = p_e$.

Then no firm will be present in the industry in the long run; the industry output will be zero. (*ii*) Suppose the price is greater than *m*. New firms will keep entering and thus the industry output will be arbitrarily large. In the language of mathematics, it is infinity (∞). (*iii*) If the price is exactly equal to *m*, the industry output is indeterminate from the supply side. Of course, in the last case, the equilibrium quantity supplied to the market must equal the quantity demanded along the market demand at the price equal to *m*.

By combining the possibilities (*i*), (*ii*) and (*iii*), the industry supply curve is then a horizontal line whose height is equal to the minimum *LAC*, as shown in Figure 10.9.

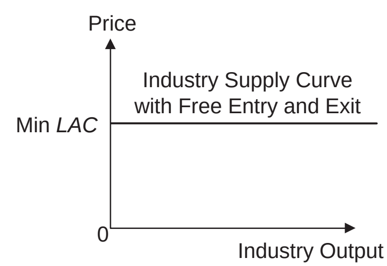
The practical implication of such long-run equilibrium with free entry and exit is that the price is driven down to its unit cost. Note that the unit cost of a product depends, in part, on technology. Thus, in the long run, the price of a product is essentially dictated by its technology—an important remark made, a long time ago, by Alfred Marshall.

A good example of competition driving down the price to the unit cost as firms enter and leave the industry is the personal computer market. The technology is so standard that assembling a computer takes no more than an hour and loading software is no more than a two–three-hour job. Entering this business and getting out of it are rather easy. As you can look around and see, there are many computer vendors in the market in most parts of India. You ask any vendor today about how much profit they make per unit sold. The business is so competitive that it is no more than Rs 1,000.

However, in some other businesses, the entry and exit processes are costly and time consuming. In the old era of licence raj, to start any major business one was supposed to obtain a licence and a number of clearances from the government. Any experienced, aged entrepreneur will tell you how painful it was to obtain these permissions in terms of crossing numerous layers of bureaucracy, corruption and delay. However, in the post-liberalisation era, it is a much simpler process (yet, in comparison to many other countries, still very costly).

Of course, getting a licence may be just one deterrent for entry into a business. Another may be the initial capital needed to set up a plant and getting production

Figure 10.9 Long-run Market Supply Curve with Entry and Exit



going. When the technology is such that the initial capital requirement is huge, it works as a natural deterrent to entry. Typically, in these industries those who are able to enter and survive earn abnormal profits over a long period of time (as it is hard for others to enter).

NUMERICAL EXAMPLE 10.4

A competitive industry has many identical firms. Each firm's long-run marginal cost schedule is given in Table 10.3. The market demand schedule for the product is given in Table 10.4. In the long-run equilibrium, how many firms would operate in this industry?

From the long-run marginal cost schedule, we can derive the long-run total cost schedule by adding up marginal costs. From the latter, we can derive the long-run average schedule through dividing the total cost by output. The *LAC* schedule is given in Table 10.5. We notice that this is minimised at output = 4. Hence, in the long-run equilibrium, each firm produces 4 units. The (minimised) *LAC* = Rs 8.5; this is the long-run market price as in the long-run equilibrium, the market price is equal to the minimum *LAC*. We now utilise Table 10.4. At price = Rs 8.5, 40 units are demanded. In equilibrium, 440 units are produced in the industry. Since each firm produces 4 units, the number of firms in operation must be equal to 440/4 = 110.

Table 10.3 Long-run Marginal Cost Schedule (Numerical Example 10.4)

Output	*LMC*
1	10
2	9
3	8
4	7
5	9
6	11
7	12

Table 10.4 Market Demand Schedule (Numerical Example 10.4)

Price Rs	*Quantity Demanded*
5.5	468
6	465
7.5	460
8	450
8.5	440
9	420
9.5	400

Table 10.5 Long-run Average Cost Schedule (Numerical Example 10.4)

Output	*LAC*
1	10
2	9.5
3	9
4	8.5
5	8.6
6	9
7	9.43

Economic Facts and Insights

- Economic analysis and decision-making take into account economic profit, not accounting profit.
- The product being homogeneous and the industry being populated with a large number of firms implies that an individual firm is a price taker in a competitive market. It means that no single firm has any market power.
- A perfectly competitive firm maximises profits by producing at a level of output such that the market price equals marginal cost and marginal cost is increasing in output.
- In the short run, a firm may be making profits or incurring losses.
- Even if a firm is making losses in the short run, it should not necessarily shut down (because the firm has to pay the fixed cost whether or not it shuts down).
- The supply curve of a competitive firm is a part of its marginal cost curve.
- In the market period, the supply of a product is fixed.
- The long-run supply curve is more price elastic than the short-run supply curve of a firm.
- In the long run, with free entry and exit, firms make zero or normal profit.
- The long-run firm-cum-industry equilibrium is characterised by the market price being equal to the long-run marginal cost as well as the long-run average cost.
- The long-run industry supply curve with free entry and exit is a horizontal straight line.
- In the long-run competitive equilibrium, the price of a product is essentially governed by its technology.

EXERCISES

10.1 What is the difference between accounting and normal profits? Illustrate it through an example.

10.2 Mr Guha owns a publishing company in the basement of the building that he owns. Within a quarter (three months), the company earns Rs 10 lakh revenue. The salary payments to employees are Rs 2 lakh. Raw material and machine maintenance costs are Rs 3 lakh. Machines were bought previously by taking loan from a bank; the principal and interest cost per quarter is Rs 4 lakh. If Mr Guha would have rented out his basement, he would have earned a rental income of Rs 2 lakh. What is his accounting profit and economic profit for the quarter?

10.3 What are the features characterising a perfectly competitive market?

10.4 Give three examples of products or services (different from those in the text), which are not homogeneous.

10.5 Give two examples of products or services (different from those in the text), which are homogeneous.

10.6 Which assumptions of perfect competition imply that individual firms in perfect competition are price takers.

10.7 What is the shape of the total revenue curve of a perfectly competitive firm? What is the reason behind it?

10.8 Why is $AR = MR$ for a competitive firm?

10.9 Assume that the marginal cost increases with output. Briefly explain why 'price = marginal cost' is the profit-maximising condition?

10.10 At a given level of output the, 'price = marginal cost' condition is met but the marginal cost is decreasing in output. Why is not profit maximised at this level of output?

10.11 Define break-even price and what is this price equal to (in principle) for a perfectly competitive firm in the short run? What is this price equal to in the long run?

10.12 'In the short run, a competitive firm always makes abnormal profits.' Defend or refute.

10.13 What is the shut-down condition? How is this condition derived?

10.14 A competitive firm's total variable cost schedule is given as follows. What is the shut-down price in this example?

Output	*Total Variable Cost (Rs)*
0	0
1	5
2	9
3	12
4	18
5	25
6	33
7	42
8	54

10.15 'At the shut-down price, the firm breaks even'. Defend or refute.

10.16 'If a firm shuts down, its short-run profits are zero'. Defend or refute.

10.17 'A firm's short-run supply curve is its marginal cost curve over and above the minimum *AVC* point'. Defend or refute.

10.18 What is 'market period'? What is the elasticity of the supply curve in the market period?

10.19 'A firm's, long-run supply curve is more inelastic than its short-run supply curve'. Defend or refute.

10.20 What conditions characterise the long-run equilibrium in a competitive industry (with entry and exit) and what is the rationale behind them?

10.21 A firm's *LAC* curve is minimised at output equal to 20 and the minimum *LAC* is equal to Rs 50. The market demand schedule is given as follows.

Price (Rs)	*Quantity Demanded*
10	1,950
20	1,700
30	1,400
40	1,200
50	1,000
60	900
70	800

In the long-run equilibrium with free entry and exit, how much is sold in the market and at what price? How many firms operate in the market? (It is implicit here and in the text that all firms have the same cost function.)

10.22 Taking into account the entry and exit of firms in the long run, how does the long-run supply curve of a competitive industry look like and why?

10.23 Explain why Marshall observed that the long-run price in a competitive market is determined mainly by technology, not demand.

10.24 Discuss briefly some of the deterrents of entry into a market.

10.25 A competitive firm's total cost is given by the mathematical relation $y^2 - 10y + 28$, where y denotes the output. Considering its profit-maximisation in the short run, if the market price is 20, how much will this firm produce?

11 Demand, Supply and Market Equilibrium

CONCEPTS

- Excess Demand
- Excess Supply
- Market Equilibrium
- Market Clearing
- Equilibrium Price
- Walrasian Stability
- Marshallian Stability
- Non-sustainable Industry
- Increase in Competition
- Consumer Price
- Producer Price
- Exchange Rate
- Currency Appreciation
- Currency Depreciation
- FAD theory of Famine
- Sen's Distribution Theory of Famine
- Price Mechanism
- Invisible Hand
- Indirect Intervention
- Direct Intervention
- Control Price
- Rationing
- Shortage
- Support Price or Minimum Support Price
- Public Distribution System
- Targeted Public Distribution System
- Buffer Stock

We have studied consumer behaviour and producer behaviour separately. In a market, consumers and producers interact and, therefore, in some sense, their behaviour must be consistent with each other.

More specifically, consider the market for, say, chewing gum. From its demand curve we know how much of it will be demanded at various prices. From its supply curve we can tell how much of it will be supplied at different prices. But what will be the actual price of chewing gum prevailing in the market?

In the very long run, when firms can enter into or leave an industry depending on whether profits are made or losses incurred, price equals the minimum long-run average cost (as we saw in Chapter 10). But if we consider a time horizon not so long, the price must be influenced by the forces of both demand and supply. That price will prevail in the market at which the quantity demanded of chewing gum is equal to the quantity supplied of it. This is a notion of matching the behaviour of producers of chewing gum with that of the consumers. It is called market equilibrium (to be defined more precisely later), which is the focus of the current chapter. In the process, we will get to learn the fundamental mechanisms behind the functioning of a market economy.

Demand and supply are like the twin blades of a pair of scissors. Just as the twin blades work in coordination with each other to produce results, so do the demand and supply forces in a market economy.

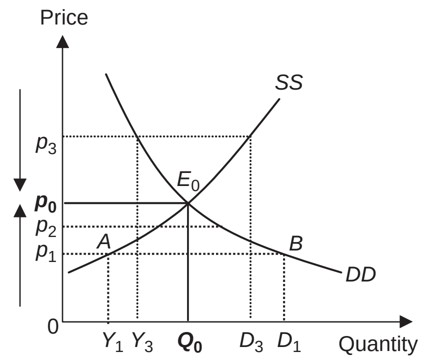
MARKET EQUILIBRIUM, DETERMINATION OF PRICE AND QUANTITY, AND STABILITY

Consider the market for, say, ruled paper that you buy in stationery shops. Suppose these are available in one particular quality and you buy them by weight. Let the demand curve and the supply curve of this paper be *DD* and *SS*, as shown in Figure 11.1. The question is—what will be the price of this paper in the market and how much of it will be purchased and sold in the market?

Suppose that initially the price of this paper is p_1. Notice that at this price the consumers are willing to buy the amount D_1, while the producers are willing to supply the amount Y_1. There is a mismatch. Consumers are demanding more than what the producers are supplying. The difference *AB* or Y_1D_1 is the **excess demand**, defined generally as the excess of the quantity demanded over the quantity supplied.[1] Will the price of paper remain stable at p_1? No. Excess demand will create competition among buyers to get hold of the limited amount available and this will raise the price. Suppose the price rises to p_2. Will it be stable at p_2? The answer is again no. The magnitude of excess demand is less than before, but it exists nevertheless, which will further push up the price. Indeed the

[1]This is somewhat different from the concept of 'excess demand' in macroeconomics where it refers to the excess of total planned expenditure on newly produced goods over the total income.

Figure 11.1 Market Equilibrium



price will keep increasing as long as excess demand exists. The upward pointing arrow shows this. It is easy to see that this process of price adjustment will settle at p_0, at which there is no excess demand.

Instead, if initially the price is above p_0, say at p_3, the producers are supplying more than what the consumers are willing to buy. There is **excess supply**, defined generally as the excess of quantity supplied over quantity demanded. It is equal to Y_3D_3. Excess supply generates competition among the sellers to sell the limited amount that the consumers are willing to purchase. This causes the price to fall. The process will continue until the price converges to p_0 at which there is no excess supply.

At p_0 then, there is **market equilibrium**, defined as the situation where excess demand or excess supply is zero. Put differently, market equilibrium occurs when quantity demanded and quantity supplied are equal to each other, or equivalently the demand and the supply curves intersect (at E_0 in the example). It is also called **market clearing**. The corresponding price is called the **equilibrium price** (p_0 in the example). The corresponding quantity demanded and supplied (Q_0 in the example) is the equilibrium quantity exchanged.

Furthermore, the fact that the price, irrespective of where it is initially, eventually settles at p_0 implies that the market equilibrium is *stable*.

Walrasian and Marshallian Stability

Note that the above argument of stability in analysing Figure 11.1 is based on the adjustment process that the price of a product adjusts upward or downward as there is excess demand or excess supply respectively. Stability of a market based on such a price adjustment process is called **Walrasian stability**, named after a 19th century French economist Leon Walras, who was the first to rigorously examine the issue of stability of market equilibrium via price adjustment.

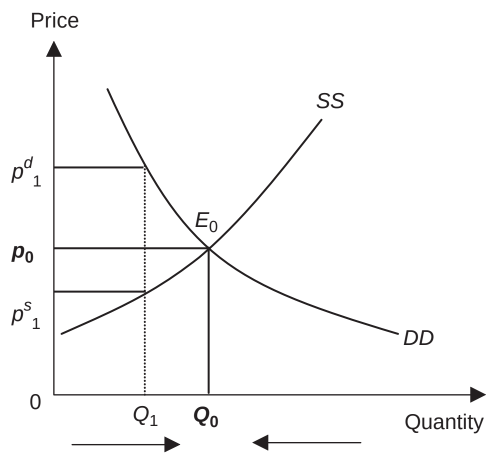
Alfred Marshall proposed a mechanism of market stability that is based on quantity adjustment, which goes by the name of **Marshallian stability** and is illustrated in Figure 11.2.

It has the same demand and supply curves as in Figure 11.1 but the stability argument proceeds in a different way. Suppose that at a given point in time, the quantity Q_1 is available in the market for exchange. The corresponding price on the demand curve is p_1^d, which can be interpreted as the 'demand price' in the sense that in order to purchase or demand Q_1, the consumers are willing to pay p_1^d. Similarly, the corresponding price on the supply curve is p_1^s, the 'supply price'—meaning that in order to supply the quantity Q_1, the firms are willing to accept the price p_1^s. The Marshallian adjustment process states that the quantity will increase or decrease as the demand price exceeds or falls short respectively of the supply price. The underlying logic is that if, for example, consumers are willing to pay more (or less) than what the producers are willing to accept, more (or less) of the commodity will be available to the market by the suppliers.

At the quantity Q_1, the demand price = p_1^d is greater than the supply price = p_1^s. Thus, according to the Marshallian adjustment process, the quantity in the market will increase from Q_1. The rightward pointing arrow indicates it. Where does this process lead to? The answer is Q_0. The corresponding price, of course, is p_0. The opposite happens but the quantity again converges to Q_0, if initially the quantity in the market were greater than Q_0. Therefore, the market equilibrium is stable under the Marshallian quantity adjustment too.

At a theoretical level, both adjustment processes seem reasonable. But there is not enough statistical evidence till today to tell us which one is more appropriate to assume. However, both processes lead to the same outcome in the case of a single market of the kind we are studying here. So, for us, it does not really matter.

Figure 11.2 Marshallian Stability



But, from now on, we will base our arguments on the Walrasian price mechanism and stability, which is more commonly used by economists and is somewhat more intuitive than the quantity adjustment mechanism.

Sustainable versus Non-sustainable Industry

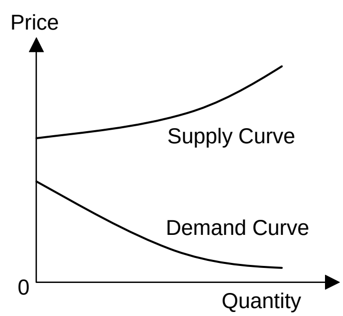
Figures 11.1 and 11.2 assume that the demand and supply curves intersect in the positive quadrant. Suppose that they do not, as in Figure 11.3. It does not mean that this can never happen or there is no equilibrium. It only means that the industry is not sustainable and the good is not produced, that is, the equilibrium production and consumption are zero because the cost is too high relative to what the consumers are willing to pay.[2]

However, it does not necessarily mean that the consumers cannot buy this product at all. If imports are allowed and the price of imports is low enough, then consumers can have it, although the domestic industry for the product is not sustainable and hence the domestic supply is zero. That is, the good is totally imported. For instance, at this point of writing, products like laser jet printers, computer memory chips, commercial jet planes and certain kinds of drugs are totally imported. Note the following.

(*i*) An industry may not be sustainable in one country, but sustainable in another.

(*ii*) Over time, a product, not sustainable earlier, can become sustainable in a country. For instance, colour TVs were not produced in India till the end of the seventies. In the eighties, the technology of producing colour TV became available in India and the market was large enough to start and sustain the colour TV industry.

Figure 11.3 A Non-sustainable Industry



[2]Put differently, the demand price is less than the supply price at any positive quantity sold.

(*iii*) If a product becomes sustainable, it does not imply that we should not import any of it at all. Domestic production and imports can co-exist and benefit us.

Mathematically Speaking

Solving the Market Equilibrium

Suppose the demand and the supply functions are given respectively as $q_D = a - bp$ and $q_S = c + gp$, where $a, b, g > 0$ and $a > c$. Because b and g are positive, the demand function is downward sloping and the supply curve is upward sloping. The restriction $a > c$ ensures that both curves intersect in the positive quadrant.

Market equilibrium is defined as $q_D = q_S$, that is, quantity demanded equals the quantity supplied. Given the expressions of q_D and q_S, we then have

$$q_D = q_S \Leftrightarrow a - bp = c + gp.$$

This is one equation with one variable, p. We can solve it and let the solution be denoted as p_0. From the above equation,

$$p_0 = \frac{a-c}{b+g}.$$

This is the equilibrium price. Notice that if a were less than or equal to c, p_0 would have been negative or zero.

Given p_0, we solve the equilibrium quantity, say Q_0, by substituting the solution expression of p_0 into either the demand function or the supply function. Substituting into the demand function,

$$a - bp_0 = a - \frac{b(a-c)}{b+g} = \frac{ag+bc}{b+g} = Q_0.$$

We would obtain exactly the same expression of Q_0 if we substituted the expression of p_0 into the supply function, $c + gp$.

* * * * * *

DEMAND AND SUPPLY SHIFTS

We now return to the situation where there is an equilibrium with a positive quantity exchanged. Our demand-supply 'model' can now be put to work in explaining various phenomena.

In reality, prices and quantities of products and services vary over time and space. A given piece of land that used to sell for Rs 1 lakh in Gurgaon district of Haryana 20 years ago may be selling now for Rs 50 lakh or even more. The South Indian delight *dosa*, of a particular kind and quantity may be selling for Rs 40 in Delhi, whereas in Chennai the same may be costing only Rs 5. Our demand-supply analysis helps us understand such phenomena among others.

Methodologically, either the demand curve or the supply curve (or both) can shift and as a result price and quantities change. Hence, we need to first understand how such shifts affect price and quantity.

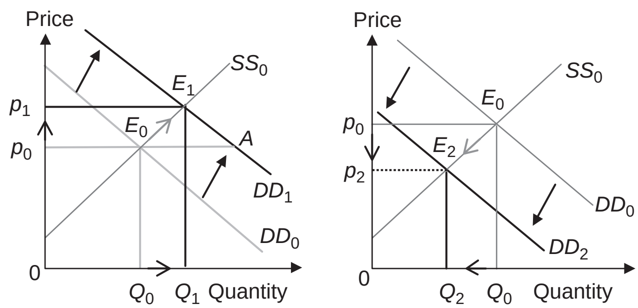
Demand Shifts

Turn to Figure 11.4(a). Initially, the demand and the supply curves are DD_0 and SS_0 respectively. They are drawn as straight lines just for graphical simplicity. The equilibrium point is E_0, p_0 is the equilibrium price and Q_0 is the equilibrium quantity transacted. Suppose there is an increase in demand. Let DD_1 denote the new demand curve. We see that E_1 is the new market equilibrium point, p_1 is the new equilibrium price and Q_1 the new equilibrium quantity. Hence an increase in demand raises price and quantity.

The 'economic process' behind these changes is the following. As demand increases (from DD_0 to DD_1), at the original price (p_0), there is excess demand (E_0A). This tends to push up the market price. As a result, the new equilibrium price is higher. There is no change in supply or the supply curve and, because of a higher price, the producers move up along the (unchanged) supply curve and supply more to the market. Thus a higher quantity is sold in the new equilibrium.

Panel (b) of Figure 11.4 exhibits a decrease in demand and its effects. As the demand curve shifts to the left from DD_0 to DD_2, at the original price, there is excess supply (equal to BE_0). This pushes the price down. The producers move down along the supply curve. Consequently, the equilibrium price and the equilibrium quantity

Figure 11.4 Effects of Demand Shifts



(a) Demand Increase (b) Demand Decrease

transacted both fall. The price and quantity at the new equilibrium are p_2 and Q_2 respectively.

Supply Shifts

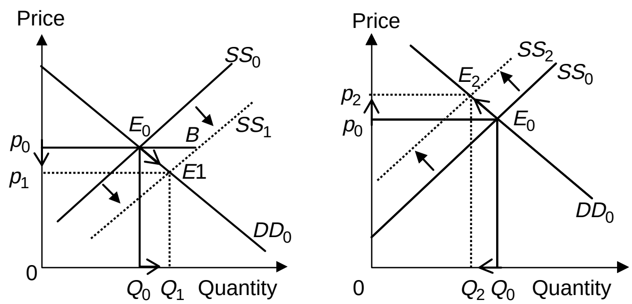
These are shown in Figure 11.5. In panel (a), there is an increase in supply (from SS_0 to SS_1). At the original price, this creates an excess supply (E_0B). The price is pushed down. There is no change in the demand curve. As the price falls, the consumers move down and to the right on their demand curve. In summary, the price falls (from p_0 to p_1) and the quantity increases (from Q_0 to Q_1).

Panel (b) of Figure 11.5 demonstrates a decrease in supply. The supply curve shifts to the left from SS_0 to SS_2. Initially a decrease in supply causes an excess demand. The price rises. Consumers move up on the demand curve and buy less. These implications are the opposite of those of an increase in supply.

Simultaneous Shifts

It is possible that both curves shift at the same time. The net impact is then the sum of the two individual impacts. For instance, in both panels of Figure 11.6, the demand increases and the supply decreases. Individually, both these shifts tend to increase the price and, therefore, the market price increases unambiguously. However, an increase in demand tends to increase the equilibrium quantity, while a decrease in supply tends to decrease the equilibrium quantity. Thus the net impact of these two shifts on the quantity is ambiguous, depending on the relative magnitude of the two shifts. In panel (a), the demand-shift effect is less and the equilibrium quantity is less. In panel (b), the demand-shift effect is more than the supply-shift effect and, thus, there is an increase in the equilibrium quantity available in the market.

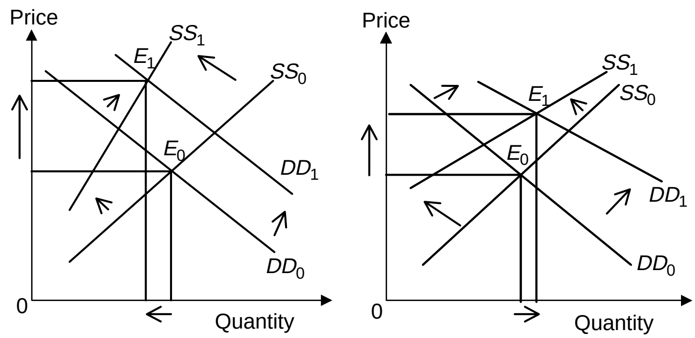
Figure 11.5 Effects of Supply Shifts



(a) Supply Increase

(b) Supply Decrease

Figure 11.6 Effects of a Simultaneous Increase in Demand and Decrease in Supply



(a) Quantity Decreases (b) Quantity Increases

Table 11.1 Effects of Demand and Supply Shifts

Source(s) of Shift	*Effect on Price*	*Effect on Quantity*
(A) Demand ↑	↑	↑
(B) Demand ↓	↓	↓
(C) Supply ↑	↓	↑
(D) Supply ↓	↑	↓
(E) Demand ↑ & Supply ↑	↑↓	↑
(F) Demand ↑ & Supply ↓	↑	↑↓
(G) Demand ↓ & Supply ↑	↓	↑↓
(H) Demand ↓ & Supply ↓	↑↓	↓

You can similarly work out the effects of other joint shifts. Various outcomes are summarised in Table 11.1.

We now examine the effects of various sources of demand and supply shifts by considering one at a time.

Sources of Demand Shift

In Chapter 2, we learnt that the demand curve shifts if, for instance, income and prices of related goods change. We will consider some of these shifts.

Suppose the total income of consumers in the market increases. If the good is normal, the market demand curve will shift to the right. In this case, row (A) in Table 11.1 applies. The equilibrium price will increase and more quantity will be available in the market. If the good is inferior, row (B) applies. The effects are opposite—price and quantity will both decrease.

As a real-life example, in the nineties, the prices of flats and land in many cities in Kerala sky-rocketed. Why? Because, many Keralites went to the Middle East for work and sent back their money to buy real estate. There was a major increase in the income of people in that region. Real estate is obviously not an inferior good. As our demand-supply 'model' predicts, its price increased substantially.

Similarly if the price of a substitute good increases, the demand curve of the good in question will shift to the right. In this case, row (A) in Table 11.1 applies. Price and quantity will increase. For instance, if coffee becomes more expensive for some reason, people will switch to tea. The demand curve for tea will shift outward. The price of tea will increase and more of tea will be sold in the market. In fact, this happened in the world market of tea in the late nineties. Brazil is the biggest producer of coffee in the world market. In the mid-nineties, the coffee production in Brazil was severely curtailed by bad weather. The coffee prices in the world market increased substantially and remained higher than before for a few years. It was observed that during the same years, the tea prices also went up in the world market without any major problems in the production of tea as such.

On the other hand, if the price of a complementary good rises, the demand curve shifts to the left; (B) in Table 11.1 applies here and both price and quantity fall.

If peoples' taste for a product improves (or deteriorates), the demand curve shifts to the right (or left). As a result, the price and the quantity both increase (or decrease) in the market. As a hypothetical example, suppose that medical research shows that eating 200 gm of *moong dal* a day significantly improves memory. This will surely increase the demand for *moong dal* by many consumers. You will soon observe that *moong dal* is costlier and more of it is sold in the market. As a real-life example, the demand for air travel plummeted right after 11 September 2001, which was because of the fear of terrorism and can be interpreted as a change in taste for air travel. As a consequence, the price of airline tickets within the US went down considerably and there were much fewer scheduled flights.

Even if the individual demand curves remain the same, the market demand curve can shift if the number of consuming households changes. An increase (or a decrease) in the number of consumers in the market will shift the market demand curve to the right (or left). As a result, price and quantity both will increase (or decrease). As an example, in the eighties, Delhi witnessed a surge in the price of land. It occurred because many people migrated from Punjab to Delhi, following disturbances in Punjab, and this shifted the market demand curve for real estate in Delhi.

Sources of Supply Shift

In Chapter 3, we learnt that the market supply curve can shift because of changes in technology, input prices, business taxes and the number of firms in the market.

As a technology improvement shifts the supply curve to the right, row (C) in Table 11.1 applies. The price falls and the quantity increases. Computer chips of any given speed are a prime example. Because of rapid technological progress in producing micro-chips, the supply curve of chips of a given speed has been shifting to the right and thus the price of the same micro-chip has been falling over time. A particular chip becomes obsolete after a few years. Another example is photograph printing. By the end of 2002, in Delhi, an instant photograph (by Polaroid technology) cost around Rs 120 for 8 copies. By the end of 2004, due to digital photography and the facility of printing from computer files on glossy paper, it cost Rs 100 for 25 copies.

In Chapter 3, we also learnt that increases in input prices shift the supply curve to the left. In such a situation, as shown in row (D) of Table 11.1, the price rises and the quantity sold falls. The opposite happens if input prices fall. Desktop computers are a good example. They have inputs like memory chips and motherboards and the prices of these inputs are falling over time. Thus, a computer with same configurations and quality is becoming cheaper over time.

As another example, between 2004 and 2005, the crude oil price in the world market soared. In the US, it led the average retail petrol price to increase from $1.87 per gallon in September 2004 to $2.90 in September 2005, a 55 per cent increase. Petrol is used as an input in many businesses, which typically involve transport. In October 2005, FedEx, the premier courier service company announced that it would increase its air shipment rates by 3.5 per cent.

An increase or a decrease in business taxes works in a similiar manner as an increase or a decrease in an input price. For an application of how the price of a service is affected by input prices and taxes, see Clip 11.1.

Clip 11.1: 'Unfair Air Fare'

On 13 October 2004, *Hindustan Times* ran an article by the title: 'Unfair Air Fare'. The main point was that air fares in India are very high compared to rates outside—in some cases about three times as high. For example, based on 21 days advance ticket booking, a one-way flight from Delhi to Trivandrum cost Rs 13,120 and the flying time was two hours and 45 minutes. In comparison, a Delhi to Singapore flight cost Rs 15,750 one-way and the flying time was five hours and 30 minutes.

While the airlines blamed the high fares on fuel prices, the article argued that flawed policy and 'back-breaking' taxes were also significant factors. For instance, because of government policy on fuel prices, the Air Turbine Fuel (ATF) rates in India were about two and a half times more than elsewhere. In Singapore, one kilolitre of ATF cost Rs 15,100, whereas in India it cost Rs 30,000. There was also 8 per cent excise tax and sales taxes ranged from 4 to 29 per cent in different states. Furthermore, navigation charges and landing fees constituted about 30 per cent of operating costs whereas in other countries they were about 15 per cent of operating costs.

This is an illustration of how input prices and business taxes affect the market price of a service.

We also learnt in Chapter 3 that an increase or a decrease in the expected future price of a product may motivate sellers to hoard and flood the market with their goods in the current period. In other words, as the price of a product in the future is expected to decrease or increase, the current supply curve will shift to the right or to the left with the result that the current price will decrease or increase.

As a prime example, think of stocks of various companies, which are sold in the share market—pretty much like other commodities. Investors' expectations play a big role in determining a stock price. If an investor speculates that the price of a stock that he is holding is to going to double or halve in the next few weeks, he is going to buy more of this stock or sell off some of it respectively. If many investors think the same way, it will impact the stock market today—irrespective of whether their expectations turn out to be correct in the future. As an example, on Monday, 17 May 2004, the 'Sensex', an index of selected stocks in the Indian stock market, fell precipitously by 565 points (11 per cent). This was completely unprecedented in the history of the Bombay Stock Exchange. Why? It was because the UPA registered an unexpected win in the general elections at that time. This led many important investors to think that the country may roll back the reforms, which would adversely affect business, profits and the stock prices in the future. Thus there was a decrease in the expected future prices of stocks. The investors sold their stocks in huge volumes, which resulted in a huge drop in the stock prices across the board. Some people have called this day the 'Black Monday' of the Indian stock market.

A change in the number of firms affects the market supply curve too. An increase in the number of firms, for example, would shift the market supply curve to the right. As a result, the price will fall and a higher quantity will be sold. In common language this is the case of an **increase in competition**. The telecommunication market is a good example. In the early nineties, MTNL (Mahanagar Telephone Nigam Limited) was the sole provider of the ISD facility in Delhi. An ISD call from India to America cost over Rs 50 per minute. Between 2004 and 2005, there were other land-line providers like Tata-Indicom, Reliance and Touchtel, which charged less than Rs 10 per minute.

ANALYTICAL APPLICATIONS

The demand-supply analysis is not only helpful in understanding what happens in some isolated markets but also has broader applications. We can use it to analyse how economic policies affect markets; how the exchange rate between different currencies may be determined; and even how catastrophes like famines occur. In what follows, we consider three applications in different areas of economics.

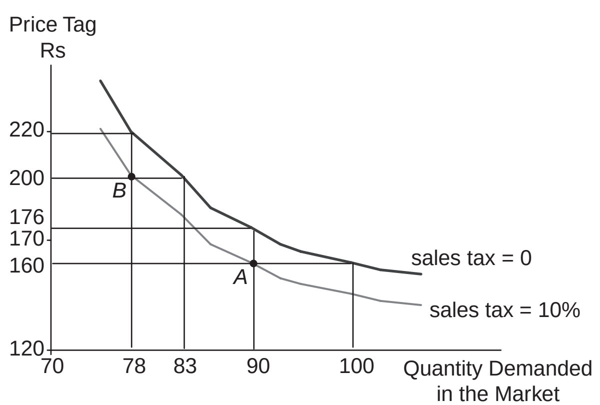
Application A: Burden of a Sales or a Service Tax

Imagine that, after arguing with your parents, you got Rs 3,000 from them to buy a leather jacket that you badly wanted. You are excited. You go to a pricy department store, which has a big collection of leather jackets of very good quality. You are a person who cannot compromise on quality. Although you know that there are other shops selling leather jackets, you are unsure of their quality. So this is the only store for you. Imagine that you do find a jacket that you really want to have and it has a price tag of Rs 2,699. Not only is it within your budget but you will also have Rs 301 left in your pocket to spend on other enjoyments. You think of spending this left-over money on food at your favourite restaurant with your friends in the evening. With your jacket in hand, you walk to the check-out counter in the store and the sales person tells you (with a smiling face) that the bill comes to Rs 2,968 and 90 paise. You are jolted but are politely told that there is a sales tax of 10 per cent, which you have to pay as the consumer and so the total comes to this amount. You have no choice and you buy the jacket. You are happy about the jacket, of course, but your mood is spoiled. You wonder (and despair) why you, the consumer, have to pay this tax instead of the seller.

Take it easy. Indirectly, the producer is also sharing the tax burden partly and this is what this application is all about. Consider Figure 11.7. The heavy line is the market demand curve for a product when the sales tax is zero. As an illustration, four price-quantity combinations are shown. Note that, since there is no sales tax, whatever price is written on the sales tag is the price the consumers pay.

Now suppose that the state government levies a sales tax of 10 per cent. Will the demand curve remain the same? No. For instance, previously, at price Rs 160, the consumers were buying 100 units. How many will they buy now? Note that, with a 10 per cent sales tax, a price tag of Rs 160 actually means Rs 176 for the consumer (Rs 160 plus 10 per cent of it). So, at the price tag of Rs 160, the consumers will buy whatever they would have bought at Rs 176 without the sales tax. The answer from the graph is 90 units. Thus, the corresponding point *A* (with the price tag equal to Rs 160 and the quantity demanded equal to 90) is on the new demand curve.

Figure 11.7 Sales Tax and the Shift of the Demand Curve



Similarly, at a price tag of Rs 200, the consumers were previously buying 83 units. Now in the presence of a 10 per cent sales tax, the same price tag means Rs 220 for a consumer. Hence, the quantity demanded will be equal to whatever the consumers would have bought at Rs 220 without any tax, that is, 78 units. Point *B* is on the new demand curve. If we join points like *A* and *B*, we get the new demand curve, which is shown as the thin line.

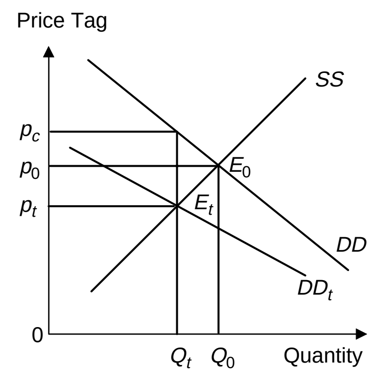
Two important points need to be noted. (*i*) The vertical distance between the old demand curve and the new one is the sales tax paid. For instance, at the quantity purchased equal to 90, the difference is Rs 16, which is 10 per cent of Rs 160. Similarly, at the quantity = 78, the difference is Rs 20, which is 20 per cent of Rs 200. (*ii*) At each quantity sold, the corresponding point on the old demand curve measures the **consumer price—**the price that the consumers actually pay, inclusive of the tax. The corresponding point on the new demand is the **producer price—**the price that the producers receive. The difference is the tax per unit which goes to the government.

Once these points are understood, the rest of the analysis is straightforward. Turn to Figure 11.8 depicting the market for any particular good. The supply curve is denoted by *SS*. The demand curve without any sales tax is *DD*. The point E_0 marks the market equilibrium; p_0 is the equilibrium price, same as the consumer and the producer price. This is the starting situation.

Suppose the government imposes a sales tax. This vertically shifts the demand curve down. Let DD_t be the new demand curve, intersecting the supply curve at E_t. The new equilibrium price is p_t and the new amount sold is Q_t, less than what was produced and sold before. From point (*ii*) above, p_t (on the new demand curve) is the producer price and p_c (on the old demand curve) is the consumer price. The distance p_tp_c is the sales tax per unit, paid to the government.

Compare this with the old equilibrium price p_0. How much more are the consumers paying? It is p_cp_0. This is the burden of the sales tax on them. The not-so-obvious point is that the producers are taking a price cut, p_0p_t, as they were

Figure 11.8 Sharing of a Sales Tax Burden



receiving the price p_0 before and are now receiving p_t. Hence, p_0p_t is the tax burden on the producers. The sum of the tax burdens is equal to $p_cp_0 + p_0p_t = p_cp_t$, the tax per unit accruing to the government. In other words, not just the consumers but also the producers share the sales tax burden, although the tax is levied on the consumers. Of the total burden, the shares are p_cp_0/p_cp_t and p_0p_t/p_cp_t for the consumers and the producers respectively.

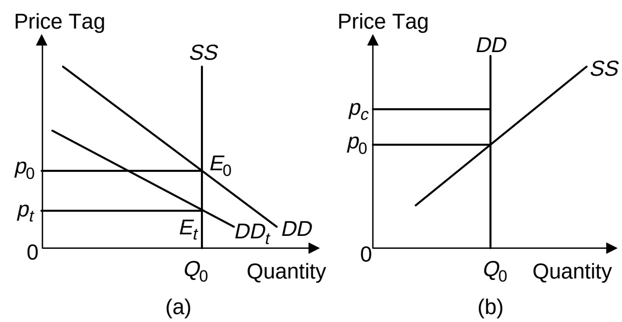
The economics behind both parties sharing the tax burden is as follows. A sales tax, by lowering the demand curve, lowers the base price (or the price tag) of the product. This means that, for the consumers, the new price is less than the old price plus the tax, that is, they do not bear the full burden of the tax. The very fact that the base price falls means that the producers are also sharing the burden.

What factors determine who shares how much. It is the elasticity of demand and supply. To see this, turn to Figure 11.9. The two panels show two extreme situations. In panel (a), the vertical supply curve has zero price elasticity. In this situation, as the sales tax is levied, p_t is the producer price but the consumer price is same as the original price p_0. Hence, there is no price increase for the consumers. It is thus the producers who share the full burden of the tax. The economic intuition is that the producers cannot adjust at all and are hence forced by the market to take the full burden of the tax, although legally or technically the tax is paid by the consumers (as it appears on the receipt).

On the other hand, panel (b) depicts a vertical demand curve having zero price elasticity of demand. In this case, there is no shift of the demand curve, since the consumers have *to* purchase Q_0. Of course, the supply curve does not shift. This implies that there is no change in the price tag or the producer price. Thus, producers do not share any burden of the tax. The consumers pay p_0 plus the tax, marked by p_c. They share the full burden. Why? Because they cannot adjust their demand for the product at all. It is the opposite of the situation illustrated in panel (a).

Combining the reasoning in the two extreme cases, it follows that if the supply curve is normal and upward-sloping and the demand curve is normal and

Figure 11.9 Elasticity and Sales Tax Burden



downward sloping, then both parties must share the burden, as shown earlier in Figure 11.8.

Moreover, the pattern of burden-sharing in the extreme cases illustrates that who shares how much depends on the elasticities: the higher the elasticity of demand (or supply) relative to supply (or demand), the higher is the share of the burden by producers (or consumers).

Note that the service tax works the same way. It is a tax to be paid by the consumer for a service. Yet the producer will also indirectly bear a part of this tax burden unless his supply curve is perfectly elastic.

Application B: Exchange Rate Determination

International commerce involves the exchange of currencies. Different exchange rates between different currencies prevail in the currency market. For example, on 30 June 2004, the exchange rate between US dollar and Indian rupees was Rs 45.99 per dollar. On the same day, the exchange rate was $1.81 = £1, between US dollar and British pound. The exchange rate between a given pair of currencies changes over time too. Compared to the dollar-rupee exchange rate on 30 June 2004, on 31 March 2005, it was $1 = Rs 43.62.[3]

Variations in the exchange rate can be understood by the demand-supply analysis also—once you realise that the exchange rate is simply the price of one currency in terms of another. When we say, for example, that $1 = Rs 46, it means that if you sell or buy one dollar, you get or have to give Rs 46, that is, the price of one dollar in terms of rupees is Rs 46.[4] Equivalently, the price of a rupee in terms of dollars is $ 1/46. Note that it does not matter whether we talk about the price of dollars in terms of rupees or the price of rupees in terms of dollars because, by definition, one is the inverse of the other. So let us stick to the term 'price of dollars in terms of rupees' as the exchange rate between the two currencies. The question is—how is this price determined in the market for currencies?

Just like the price of a banana is determined by the demand and supply of bananas, in the currency market the price of dollar is determined by the demand and supply of dollars (in exchange for rupees). In the context of commerce between India and the US, who demands dollars? It is the Indian importers of American goods and services. For instance, you want to do a master's in the US, for which there are application fees charged by American universities. They would want their fees in dollars, so you would have to demand dollars and pay them in dollars. How does it work? You go to a bank, apply for dollars, deposit your rupees and the bank gives you a draft in dollars in the names of the respective universities. Indirectly, the bank is placing your request (demand) for dollars

[3]See the web publication by the Federal Reserve Bank of the United States, http://www.federalreserve.gov/releases/H10/hist/

[4]This statement ignores the difference between the buying and the selling rate, which is called the 'spread'.

through its correspondent bank, which is authorised to deal with foreign currency directly.[5]

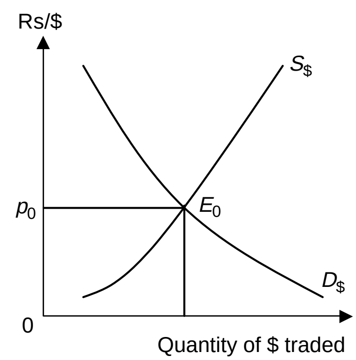
The point is that, given the price of dollar, Indians (through their representative banks) demand a certain amount of dollars in the currency market—this demand originates from our desire to import goods, services and assets from the US. For instance, if an Indian company wants to acquire a company in the US, it has to demand dollars because the owners of the selling company would want their payment in dollars and not in rupees.

This implies that if the price of dollar changes (that is, if there is a change in the exchange rate), our demand for dollars will change. To continue with the example of applying to US universities, if, hypothetically, the exchange rate falls from $1 = Rs 46 to $1 = Rs 16 only, it is almost a certainty that the applications from India to do a master's will increase. This means that a decrease in the price of dollars will increase the demand for dollars. Hence, there is a normal, downward sloping demand curve for dollars. This is shown in Figure 11.10 by the $D_\$$ curve.

Consider who would supply dollars into the currency market. It is the American importers of Indian goods, services and assets. If Bill Gates wants to buy a flat in Bangalore he would, through his bank, have to supply dollars to the currency market, get rupees in exchange and pay the Indian owner of the flat in rupees.[6] In general, how would an increase in the price of dollar affect the 'quantity supplied of dollars' (in exchange for rupees)?

This is a bit tricky. If the American import demand for Indian goods is absolutely inelastic, they would have to pay or supply less dollars because one dollar is now fetching more rupees and hence for the same good priced in rupees,

Figure 11.10 The Currency Market



[5]In the process, of course, the bank is charging you a service fee. But we can ignore it for analytical convenience.

[6]If the flat is priced at, say, 1 crore and the current exchange rate is, say, $1= Rs 50, he would have to supply $1 crore/50 = $2,00,000 to the currency market.

they would have to pay less in dollars. Then, the quantity supplied of dollars will decrease in the currency market. On the other hand, if their demand for Indian goods is highly elastic, cheaper Indian goods will generate a huge increase in their import demand for Indian goods and as a result they will end up supplying more dollars to the currency market. Thus, whether the total supply of dollars will increase or decrease depends on the price elasticity of import demand for Indian goods by the Americans. If we apply the relationship between a price change and the total expenditure on a product as studied in Chapter 2, it follows that the total supply of dollars (foreign currency) will increase or decrease as the price of dollar increases according as the foreign import demand for domestic goods is greater or less than one respectively. For expositional simplicity, we will assume here that this elasticity is greater than one. This then implies that the total supply of dollars will increase with the price of dollar. Figure 11.10 shows an upward sloping supply curve of dollar, $S_\$$.

The intersection of the $D_\$$ curve and the $S_\$$ curve determines the price of dollar or the exchange rate. In Figure 11.10, the equilibrium in the foreign currency market is at E_0. The equilibrium price is p_0. If p_0 happens to be 45, then the market exchange rate is \$1 = Rs 45; if $p_0 = 50$, then the market exchange rate is \$1 = Rs 50; and so on. This is how the exchange rate between two currencies is determined by the forces of demand and supply.

The demand-supply framework can be applied to understand what factors affect the exchange rate and how. Let us take an example. Suppose, within a given period, the Indian economy experiences an inflation rate of only 5 per cent, while in the US economy it is 12 per cent. All other things remaining unchanged, this would mean that American goods will be costlier relative to Indian goods. As a result, we would like to demand less of American goods and Americans would like to demand more of Indian goods. In turn, this will mean that we would like to demand less of dollars in the currency market and Americans would like to supply more of dollars. The $D_\$$ curve will shift to the left and the $S_\$$ curve to the right. The price of dollar will fall—one dollar will command less rupees or one rupee will command more dollars.

Hence, as a result of US inflation being higher than inflation in India, the rupee will be stronger, that is, it will **appreciate** and the dollar will be weaker, that is, it will **depreciate** in the currency market.

Application C: Explaining Famines

Famines have terrible consequences—millions of people die of starvation and diseases. Although these are less frequent now than they were, say, hundred years ago, they nevertheless occur in modern times, the Sudanese famine in 1998 being the most recent.[7] In what follows, we will illustrate the research (work) on famine by Amartya Sen, a Nobel laureate in economics.

[7]According to BBC, somewhere between 60,000 to 1 lakh people died in Sudan that year because of the famine.

The standard notion is that a famine occurs due to a drastic shortfall of food grain production and supply for some reason, for example, a severe drought or a massive flood. Sen calls this the **FAD theory of famine**, where FAD stands for 'food availability decline'. His 'thesis' is that not all famines are due to FAD. A famine may be 'man-made'. A large section of the population of the poor in a region can be deprived of food, even when there is no decline in the total supply of food.

We can understand famines by analysing demand and supply forces. We will first outline the FAD theory and then follow it by a discussion of Sen's theory of famines (1981).[8]

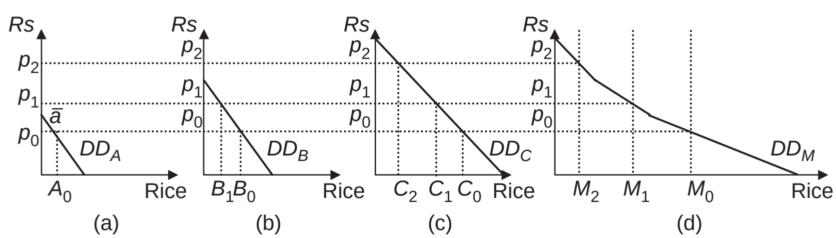
THE FAD THEORY

A famine occurs in the form of epidemics and deaths from pervasive starvation. In turn, starvation, that is, deprivation from basic food items is reflected mostly in terms of the staple food of the region. Hence, with regard to the question of what causes a famine, we can limit ourselves to examining how a large section of a region's population is not be able to obtain the minimum amount of staple food for survival. Let us call this staple food rice.

Now turn to Figure 11.11, which depicts the individual and market demand curves for rice as well as the market supply curve of rice. Let us say that there are three families, A, B and C, in the market. The panels *(a)*, *(b)* and *(c)* graph the demand curves of these families respectively. The B-type family's demand curve lies to the right of that of the A-type and the C-type family's demand curve lies to the right of that of the B-type. We can interpret the A-type as the poorest, the B-type as the next poorest and the C-type as the richest.

See that when the price of rice is p_1, the A-type family cannot afford to buy any rice at all, but the B-type or the C-type can. As shown, this price is above the point $\bar{a}$ at which the DD_A curve intersects the price axis. Hence, the A-type family's quantity demanded is zero. This is not true for the B-type or the C-type family. The former

Figure 11.11 FAD Theory of Famines



[8]There is no claim on the originality of the ideas, but the demand-supply exposition is a result of the author's own work. For Sen's theory of famines, see: Sen, Amartya. 1981. *Poverty and Famines: An Essay on Entitlement and Deprivation.*: Oxford University Press.

demands the amount B_1 and the latter the amount C_1. If the market price is p_2, the A and B-type families cannot buy rice but the C-type can; its quantity demanded is C_2.[9] The rightmost panel depicts the market demand curve, DD_M, as the horizontal summation of the three individual demand curves.

This graph assumes that there is one family of each type. But it is not restrictive. If there are more families of any given type, the market demand curve is obtained by horizontally summing up the demand curves of all the families. The resulting curve will look similar to DD_M.

From the supply side, let, initially, the total available amount be M_0. It is drawn vertically to represent that after the harvest, this is the total, potential amount available for consumption.[10] The equilibrium price is then p_0. At this price, all families are able to buy rice. The types A, B and C respectively buy A_0, B_0 and C_0. This situation can be interpreted as one in which there is no starvation or famine.

Now suppose that (for some reason) there is less total amount available, equal to M_1. The equilibrium price is higher, equal to p_1. Notice that at this price, the poorest cannot afford to buy any rice, but the other types can. We can think of this situation as 'starvation'—some people at the lower income levels just cannot afford to buy enough food for survival. If, instead, the total available amount were much less compared to the initial situation, for example, equal to M_2, the price would have risen to p_2 and both A-type and B-type families would have been driven out of the market. We can interpret this as a situation of famine or massive starvation. Whether exactly two types of families are deprived of the staple food or not is immaterial. This situation generally represents that a large number of people are under starvation. This is the FAD theory. In summary, it says that a drastic fall in the total availability of food can raise the price of the staple food to such an extent that many families at the lower end of income levels may no longer be able to afford to buy the minimum amount needed for a healthy survival. Consequently, massive starvation and famine may follow.

SEN'S THEORY

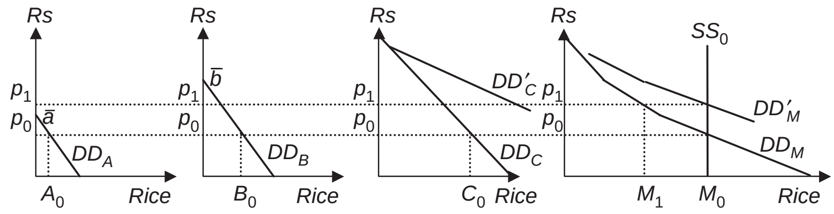
Sen argued and emphasised that a famine could be an **income distribution problem,** meaning that if the distribution of income becomes so unequal that a large section of a region's population becomes relatively much poorer than before, they can no longer afford to buy a minimal amount of food even when there is no decline in the total availability of food. Sen went on to argue convincingly that this was indeed the case in the Bengal famine of 1943–44, which claimed the lives of 3 lakh people.

How can it be that when there is no decrease in food availability, a large number of people cannot buy the minimum amount of staple food for survival?

[9] If the price of rice is p_3 or higher no one can buy rice; but such a price cannot prevail in equilibrium and hence is irrelevant.

[10] We can instead draw a standard upward sloping supply curve. But this will *not* change the analysis.

Figure 11.12 Sen's Distribution Theory of Famines



The key is that, the supply remaining unchanged, if the demand for the staple food by people who are not poor increases sufficiently, it can drive the poor families out of the market and force them to starve.

Turn to Figure 11.12, which is similar to Figure 11.11. Suppose that starting from the situation where p_0 is the equilibrium price and every family is able to buy some rice, the demand curve for rice by the C-type increases from DD_C to DD'_C. Note that the market demand curve shifts to DD'_M. As a result, the equilibrium price increases to p_1 and the A-type family cannot afford to buy anything. The same reasoning implies that if the C-type's demand curve had increased more, the price increase would have been greater and even the B-type could have been thrown out of the market. The whole point is that if the demand for the staple food by the non-poor increases sufficiently—because the non-poor become even richer—the price would rise sufficiently so that a large number of people in the poor classes are unable to buy the minimal amount of the staple food and hence a famine occurs.

Note that there is no decline in the total availability of food. But, because of demand pressures from one class of people, this class is able to claim a large share of food via the market mechanism at the expense of the remaining people. This is the gist of **Sen's distribution theory of famine**.[11]

PRICE MECHANISM—THE INVISIBLE HAND

The purpose of laying down the above applications is to drive home the point that the reach of the demand-supply analysis is long. Many important problems

[11] In establishing that this was the nature of the Bengal famine, Sen first provided data (statistics) to show that a decrease in total supply of rice was *not* the case of Bengal in 1943. This task is not as simple as it may seem. To go back to relevant archives scattered across libraries in India and verify against a dominant belief is a painstaking job. He documented that the supply of rice in Bengal in 1943 was only about 5 per cent lower than the average of the previous five years. It was, in fact, 13 per cent higher than that in 1941 and there was no famine in 1941 (Sen, 1981: 58). He then outlined the factors prevailing in Bengal during the war period that led to distributional shifts of the kind described above and other 'supply' factors like speculation and prohibition of inter-province export of cereals prevailing at the time that aggravated the situation so as to result in a famine.

in economics can indeed be meaningfully studied and understood using this analysis.

All through these examples, we have seen how demand and supply curves shift and how they affect prices and quantities in a market economy. Such shifts occur not just in a few selected markets but in almost all markets and almost always over time. The beauty is that there is no central or supreme planning authority that has to coordinate the actions of innumerable consumers and producers in the market. A change in the behaviour of one group—either producers or consumers—gets reflected in a price change, which acts a signal to the other party and both parties adjust in the end in a mutually consistent way. This is called **price mechanism**. Adam Smith, considered the father of economics, compared price mechanism to an 'invisible hand' and held that the behaviour of the consumers and the producers is coordinated in a market economy by **an invisible hand**.

A fundamental question is whether such free working of the price mechanism or the free operation of a market serves a society the best. The answer is 'not always.' As a glaring example, we saw that a catastrophe like a famine can occur in a market economy. There are other less extreme situations where a free market may not work in an ideal fashion. In such situations, there is a justified role for the government to intervene in the market in some desirable way. A systematic analysis of situations where a free market may fail to act in the society's best interest and of the best corrective forms of policy intervention will be taken up in Chapter 12.

In what follows, we take as granted that as a matter of practice, governments do intervene in the markets and we will study the implications of such interventions.

GOVERNMENT POLICY IN A MARKET

A government can interfere in a market in a variety of ways.

(*a*) It can levy taxes or grant subsidies to either consumers or producers or both.
(*b*) It can control prices in the form of ration and support prices.
(*c*) It can regulate entry into a market through an exclusive licensing system and other requirements (or what is called licence raj).
(*d*) It can regulate mergers and acquisitions of companies and the pricing policy of firms (different from controlling the price itself).
(*e*) It can choose to do the business itself as a public sector firm (for example in the steel and civil aviation sectors) or in services like electricity and telecommunication.

This is by no means an exhaustive list. For instance, through its credit policy, the government influences the loan market. In the following, we restrict ourselves to the consequences of (*a*) and (*b*) only.

Business Taxes and Subsidies

Levying taxes and granting subsidies on commodities and services are called **indirect interventions** in the sense that these policies are not meant to forcefully ensure a particular price or particular quantity that the government believes to be right for the society. Rather the end is achieved indirectly by influencing the price and that is why these are called indirect interventions. In many cases, indirect taxes are levied for revenue purposes only.

We have already discussed the effect of sales or service taxes—the consumer price increases, the producer price falls and, therefore, both parties share the burden of this tax, although it is the consumers who pay this tax directly to the government.

Exactly in the same way, taxes that are directly paid by the producers, for example, an excise duty or a VAT, are borne by producers and consumers both. Who shares how much is dependent on the price elasticity of demand and supply as in the case of sales tax.

While we have focussed on taxes that reduce the produced and consumed, in certain markets the government encourages production and consumption by granting subsidies. For instance, through nationalised banks, the government provides subsidised loans to small businesses in rural areas.

Control Price and Support Price

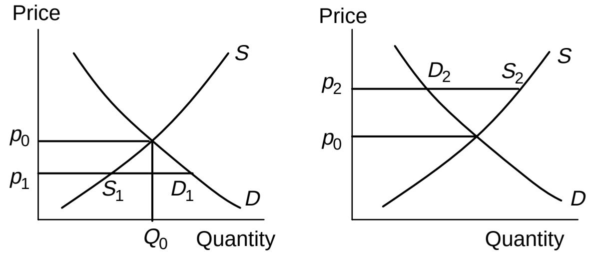
As we shall see, these are prices set directly by the government in certain markets and hence called **direct interventions**.

CONTROL PRICE

For necessary food items like rice and wheat, it is felt that if the prices are totally left to the free market, the poor and the needy may not be able to afford them. The government sets a low price and the products are sold in the ration or 'fair price' shops. This is illustrated in Figure 11.13(a). Suppose the demand and the supply curves in this figure refer to wheat. If there are no interventions, the free market price of wheat will be p_0 and the quantity Q_0 of wheat will be produced and sold. If it is presumed that the price p_0 is too high for the poor, the government would step in and set a price below p_0, say p_1. This is a situation of control price. What are the implications?

Notice that at the price p_1, which is below the equilibrium price, the consumers want to buy the quantity D_1, while the producers are willing to supply less, equal to S_1. Hence, there is an excess demand or a **shortage**. How is this 'excess demand' problem handled? By **rationing** the quantities, the consumers can buy only a limited amount so that their total rationed demand is equal to S_1. This is why the fair price shops are called ration shops. Indeed, any price control scheme must be accompanied by a rationing rule.

Figure 11.13 Control Price and Support Price



(a) Control Price (b) Support Price

There is a further implication. Because of the shortage, some people will always want to pay a higher price in order to obtain more than the 'rationed' amount. Thus the ration-shop licence holders have an incentive to illegally sell at a higher price though back door means. In other words, a price control programme invariably begets a black market, which otherwise would not exist.[12]

In India, there were about 4.77 lakh ration shops as of 2004. They come under what is called the 'public distribution system' or PDS, which has recently evolved into TPDS, 'targeted public distribution system'. PDS or TPDS is instituted by the central government. See Clip 11.2 for a brief account of the PDS/TPDS.

Clip 11.2: PDS/TPDS in India

The public distribution system (PDS) came into existence during World War II. Through this the government sells grains, mostly rice and wheat. (Other products like sugar and edible oil used to be included on a limited scale.) Through the Food Corporation of India (FCI), the government procures grain at a procurement price and stores them. In distributing the grains across the country for consumption, the central government first provides the states and union territories these commodities at 'central issue prices' (CIPs). In turn, the states and union territories fix the end retail price by taking into account the margin for wholesalers, transportation costs, local taxes and so on.

(continued)

[12]The same is true in currency markets. When currency rates are artificially set different from the equilibrium rates, *hawala* markets emerge.

Historically, the primary objectives of the PDS have been to ensure price stability of food grains and provide the poor an access to basic food at reasonable prices. Till 1997, the access to the PDS (ration shops) did not depend on the household income. Virtually anyone having a permanent address could obtain a ration card. In 1997, PDS became a targeted PDS (TPDS), meant to target the poor families. The TPDS was initially a two-tier subsidy system—one for the poor below the poverty line (as defined by the Planning Commission for different states in 1993–94) and the other for the poor above the poverty line. In 2000–2001, Antyodaya Anna Yojana was initiated to target the poorest among those who were below the poverty line. Hence, as of now, it is a three-tier subsidy system. Accordingly, there are CIPs for each grain (rice or wheat). For instance, in 2000–2001, for the poorest of the poor, the end-retail prices were Rs 2/kg for wheat and Rs 3/kg for rice. For the rest below the poverty line, these are Rs 4.15/kg for wheat and Rs 5.65/kg for rice. For those above the poverty line and yet relatively poor, these prices are Rs 6.10/kg for wheat and Rs 8.30/kg for rice (which are at par with the government's cost). The state governments shoulder the responsibility of identifying the poor.

How far has the PDS or the TPDS reached the poor? In 1999–2000, 36 per cent of the below-poverty-line households and 31 per cent of the poor but above the poverty line reported purchasing rice or wheat from the PDS. Various case-studies point to a number of reasons for the relatively low participation rate of the poor—like the difficulty in obtaining ration cards, uncertain ration supplies and inferior quality of PDS grains. Furthermore, there is also fraudulent or illegal diversion of grains from the PDS—consistent with the black-market implications of a ration system. According to Ramaswami (2007), the future of the food subsidy programme may not lie in a centralised PDS but in a regionally differentiated one. Another option could be food coupons or food stamps, where food is supplied by the private sector.

Reference

Ramaswami, B. 2007. 'The Public Distribution System', in Kaushik Basu (ed.), *Oxford Companion to Economics in India*. New Delhi: Oxford University Press.

Control price programmes do not just exist for food grains. Rental markets in metropolitan areas like Chennai, Delhi, Kolkata and Mumbai are other prime examples. There are rent control acts for each of these cities. Essentially, these acts have frozen the rent for a large number of urban properties over several decades. At current prices, the controlled rents are ridiculously below their market rates—in some cases even less than 1000th fraction of the market rates. An account of the implications of such acts is given in Clip 11.3.

Clip 11.3: Some Implications of Rent Control Acts

Not only are rents abysmally low but also it is very hard under the law to remove a tenant. For instance, in Delhi, to remove a tenant, the landlord has to prove, among other things, that he/she has no other place to live. But the landlord would be living somewhere else; perhaps he/she has some other property; hence it is difficult—almost impossible—to establish that one has no other place to live.

What are the implications? There is no incentive for property owners to maintain the rental property in terms of quality other than what is (legally) required for some bare minimum level (for example, providing electricity and water connections). The renters have absolutely no incentive to vacate. If they die, according to the law, the rental property is automatically inherited by their children. Who does not want some space in an urban area, which is almost free of cost? If the landlord really wants to get some tenant out, so that he/she can use the property, typically and illegally, the tenant is paid a huge sum of money. Consider the fairness of all this—to reclaim your own property, you have to pay somebody else! Another way to remove a tenant is to use threat and violence.

Interestingly, according to Suketu Mehta in his best seller, *Maximum City: Bombay Lost and Found*, a majority of police complaints and crimes (taking small and big occurrences together) in Mumbai owe their origin to the Bombay Rent Control Act! That is, violence and crime are by-products of control price regimes too.

Why do the law makers not remove the rent control acts? As Suketu Mehta aptly points out, it is because the tenants far outnumber the landlords and exert the political pressure to continue with the existing law.

SUPPORT PRICE

While some consumers of certain food grains are offered a price control programme, growers of various food grains are offered a price support programme by the government. That is, the government guarantees them a price *higher* than the market price. This is called a **support price** or a **minimum support price** (MSP). For instance, for common paddy, during the *kharif* market season (September–October) of the financial year 2003–2004, the MSP was Rs 550 per quintal. During the *rabi* market season of 2002–2003, it was Rs 620 per quintal for wheat.

What is the rationale for the support price? It is thought that since farmers are poor, price fluctuations resulting from the operation of free markets may hurt their well-being, especially if the prices become very low. Therefore, a minimum support price should be set. Panel (b) of Figure 11.13 illustrates this. While the free market price is p_0, the government support price is some price higher than p_0, like p_2.

What are the implications? At p_2 the amount D_2 is sold in the free market. The quantity produced however is Q_2. There is excess supply, equal to D_2Q_2. Where does this amount go? The answer is that it is the government who purchases this 'excess'. This is indeed necessary for the price support system to work. Typically, the government stores the unsold amount in warehouses. It is called a **buffer stock**. If in a particular year or season, the production is low (perhaps because of bad weather), the quantities can be released from the buffer stock to meet the demand at reasonable prices.

See Clips 11.4 and 11.5 about some particulars and assessments of the MSP program in India.

Clip 11.4: The MSP Programme in India

India has an elaborate MSP programme since the sixties. Currently the central government announces MSPs for 24 major crops. Typically, MSPs are announced each year by taking into consideration the recommendations by the CACP (Commission for Agricultural Costs and Prices). Several factors go into the MSP calculation by the CACP such as cost of production, international market price situation, trends in domestic market prices and demand and supply situation. But the most important factor is the cost of production. MSPs are uniform across the country. They are meant to serve as a long-term guaranteed minimum price by the government from the viewpoint of investment decision by the growers.

However, the general economic efficiency of this programme is being increasingly questioned. Note that it is essentially a subsidy programme—the excess over what is sold in the private market is paid for by the government. It is eventually the tax-payers, who pay for this subsidy programme. The greater the discrepancy between the MSP and the market price, the higher is the cost of this subsidy programme. There are also storage costs in keeping the excess in warehouses. Furthermore, over time, in particular, the MSPs of wheat and rice have increased substantially. It has been seen that this has resulted in crop diversion from cotton, oil seeds and coarse grains to wheat and rice, especially in more efficient states like Haryana and Punjab. Thus, crop growing has become somewhat less diversified. A related view is that the MSP programme has mainly benefited only a small number of rich farmers.

For a critical view on the current state of the MSP programme, see the editorial column of *The Economic Times*, dated 6 November 2004, which is reproduced in Clip 11.5.

Clip 11.5: Economic Times on the Current MSP Programme

THE ECONOMIC TIMES

SATURDAY 6 NOVEMBER 2004

Rethink food strategy

Time We Junked Traditional MSP

THE government has once again raised the support price for wheat and assorted other rabi crops. The saving grace is that the minimum support price of wheat has been raised by just Rs 10 per quintal, or less than 2%, significantly below the rate of inflation. The disgrace is that MSP has not been slashed, as it ought to have been. The new buzz on MSPs is that they ought to be linked to demand. This is the lesson drawn from the build-up of stocks prior to the drought of 2002-03, when the government kept on ractcheting up procurement prices of foodgrains and priced grain out of the reach of private trade. But policy has to go beyond the need to link up procurement with demand. The country has three functional commodity exchanges in place, which offer quotes on commodity futures and derivatives. Policy should integrate these markets with actual commodity trade, so that farmers have an assurance of price even without official intervention. Support prices should be low enough to not distort the working of these commodity markets, which should reflect demand and supply adequately. The government must remove impediments to efficient price discovery by the commodity markets. These include the absence of rural roads, primitive laws that restrict trade in agricultural produce, insufficient storage infrastructure and wholly inadequate alignment with global markets on account of high tariffs and inefficient ports.

Open-ended procurement at high prices from a small part of the country — principally the north-west — have combined to make a mockery of food subsidy. Food subsidy has become a means of transferring public resources to a small section of rich farmers and jacking up food prices to an extent where it becomes unaffordable by the poor. This strategy of ensuring food security resulted in the government having to subsidise grain exports to cattlefeed manufacturers in southeast Asia, so that soaring stocks would not bloat the food subsidy bill further. The drought saved the system from collapsing under the weight of its own contradictions. Rather than softpedal implementation of the flawed strategy, the government should alter it completely.

Economic Facts and Insights

- Market equilibrium is achieved at that price at which quantity demanded is equated with quantity supplied. In this sense, price mechanism coordinates the decision-making by consumers and producers.
- A non-sustainable domestic industry does not imply that the consumers are not able to consume the product at all, because they may be in a position to import the product.
- An increase (or a decrease) in demand causes the equilibrium price and the quantity transacted to increase (or decrease), whereas an increase (or a decrease) in supply causes the equilibrium price to fall (or rise) and the equilibrium quantity transacted to increase (or decrease). These implications form the basis of how various demand and supply shifts, individually or together, may affect the price and the quantity transacted.
- Although a sales tax or a service tax is paid by the consumers, the burden of such tax payments falls on both producers and consumers. The higher the price elasticity of demand relative to the price elasticity of supply, the greater is the share of burden borne by the producers.
- The exchange rate between two currencies can be interpreted as the price of one currency in terms of another. This is determined by the demand and supply of foreign currency. In turn, this is dependent on the domestic country's demand for foreign goods, services and assets and the demand for domestic goods, services and assets by the foreign countries.
- Famines in terms of many poor people not being able to buy the staple food can result from a drastic decrease in the supply of the staple food or if the distribution of income becomes very unequal.
- Free market or laissez faire may or may not work in the best interest of an economy.
- The government can influence a market through direct and indirect interventions.
- Control price regimes are seen in the markets for food grains and urban property rental markets. Support price systems exist in case of some major crops.
- The motivation behind a control price system lies in providing essential products at affordable price to the poor. Support price systems exist on the rationale that they protect farmers from price fluctuations, especially downward swings in agricultural prices.
- Control price systems necessarily imply shortages, leading to black marketing and sometimes crime and violence.
- A support price system leads to surplus production, which is stored as buffer stocks.

EXERCISES

11.1 According to Walras, the price adjusts in a competitive market to clear the market. Explain the process of Walrasian stability of equilibrium.

11.2 According to Marshall, the quantity adjusts in a competitive market to clear the market. Explain the process of Marshallian stability of equilibrium.

11.3 'Under any price below (or above) the market price, there is excess demand (or excess supply) in a market'. Defend or refute.

11.4 Illustrate the situation where an industry is not viable or sustainable. Does it mean that the consumers cannot consume a product if the domestic industry is not viable?

11.5 How will a simultaneous increase in demand and a decrease in supply affect the market price and quantity transacted?

11.6 How will a simultaneous decrease in demand and supply affect the market price and quantity transacted?

11.7 All else the same, how will an increase in the VAT affect the market price and quantity of a product?

11.8 When you go to see movies, the ticket price includes an entertainment tax. Who is likely to bear most of the burden of an entertainment tax? Discuss.

11.9 Suppose the government completely eliminates tariff on the import of steel. Assume that foreign steel and domestic steel are substitutes of each other. Through demand and supply curves, illustrate how this will affect the price and quantity produced of domestic steel. All else the same, how is this likely to affect the market price of bicycles produced in India?

11.10 'Since the consumers pay the sales tax or the service tax, the burden of this tax is borne by the consumers only'. Defend or refute.

11.11 If the demand curve is vertical, who shares the whole burden of a sales tax and how?

11.12 If the supply curve is vertical, who shares the whole burden of a service tax and how?

11.13 Suppose stock prices soar (that is, become 'bullish') in India, but they remain stagnant ('bearish') abroad? How may it affect the value of rupee in the currency market and why?

11.14 Suppose India continues to produce better quality products for exports. How would it affect the value of rupee in the currency market and why?

11.15 'When we import more of British goods, it increases our demand for British pound in the international market'. Agree or disagree. Give reasons.

11.16 If we import more of Chinese goods, our demand curve for yuan (the name of the Chinese currency) will shift to the ____ and it will ____ the value of rupee in terms of yuan. Fill in the blanks and explain.

11.17 Mrs Rao argues that economists say inconsistent things like as the price rises, demand falls but as demand falls, the price falls. Do you agree or disagree? Give reasons.

11.18 Describe the FAD theory of famines.

11.19 Explain Sen's distribution theory of famines.

11.20 How are decisions taken by numerous consumers and producers in a market economy coordinated?

11.21 What is meant by the 'invisible hand' in the working of a market economy?

11.22 What is price mechanism and why is it so fundamental to the working of a market economy?

11.23 Why is a system of control price typically associated with rationing of demand?

11.24 Explain how a regime of control price is likely to breed black marketing?

11.25 What is the support price? Which kinds of goods are given support prices in India?

11.26 'The support price is fixed below the equilibrium price'. Defend or refute.

11.27 Suppose that in a market the demand equation is given by $q_D = 10 - p$ and the supply equation by $q_S = -2 + 3p$, where q_D and q_S respectively stand for the quantity demanded and the quantity supplied and p is the price. Solve for the equilibrium price and quantity in this market.

11.28 Referring to the previous question, suppose that the supply equation were given by $q_S = -11 + p$. What is the equilibrium quantity transacted in this market?

12 Optimality of a Competitive Market Structure, Market Failure and Corrective Measures

CONCEPTS

- Social Welfare or Surplus
- Social Efficiency
- Market Failure
- Pigovian Tax
- Demerit Good
- Public Goods
- Non-rivalry
- Non-excludability
- Free Rider Problem
- Asymmetric Information
- First-best Policy Intervention
- Second-best Policy Intervention
- Government Failure
- Over-intervention

In Chapter 11, we talked about the smooth working of the price adjustment in solving the wants of consumers and producers. This is a great virtue of a perfectly competitive market structure. But is it the best in some sense, for society or the economy? The answer is yes—under certain conditions. This is a profound result as it means that in some situations, in a decentralised market economy in which no economic agent has any market power, there is no need for any policy intervention. A laissez faire or a completely free market approach is indeed the best policy.

Because this result holds in some conditions, it is quite useful since this implies that the rationale for an economic policy intervention arises when the actual conditions deviate from those 'ideal conditions'. The nature of the deviation then dictates the kind of policy intervention or the nature of the corrective measures. In this chapter, we will probe into normative economics and the foundations of economic policy.

SOCIAL WELFARE

Since we are talking about the society's interest, we must define some notion of social welfare. In the context of a market for a single good, in equilibrium a certain amount is produced, which, on one hand, costs the society in terms of resources used and, on the other, benefits the society in terms of providing consumption. Thus we may define it as

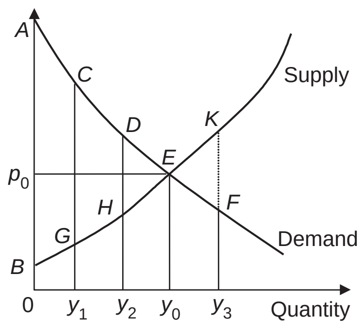
$$\begin{pmatrix}\text{Social Welfare} \\ \text{or Social} \\ \text{Surplus } (W)\end{pmatrix} = \begin{pmatrix}\text{Total social utility} \\ \text{obtained from} \\ \text{consuming a good}\end{pmatrix} - \begin{pmatrix}\text{the total social} \\ \text{cost of its} \\ \text{production.}\end{pmatrix} \tag{12.1}$$

It can be expressed in another way. Assume that the total social utility is same as the total utility to the consumers of the product. Similarly, assume that the total social cost of production is equal to the total cost of production faced by the producers. Under these assumptions, we can write W = total utility to the consumers – total cost to the producers. Next, recall from Chapter 5 that consumer surplus is defined as the total utility measured in money – the total expenditure on the product. Substituting this into the last expression of W, we then have social surplus = consumer surplus + {total expenditure on the product – the total cost to the producers}. Note that the total expenditure on the product is, by definition, the total revenues accruing to the firms. Hence, the term inside the curly brackets is simply the profits or producer surplus of the firms. Therefore, we can rewrite,

$$\begin{pmatrix}\text{Social Welfare or} \\ \text{Social Surplus } (W)\end{pmatrix} = \begin{pmatrix}\text{Consumer's} \\ \text{surplus}\end{pmatrix} + \begin{pmatrix}\text{Producers'} \\ \text{surplus.}\end{pmatrix} \tag{12.2}$$

This is quite intuitive. In the market for a good there are two types of economic agents—consumers and producers. Social welfare, as defined in (12.2), is simply the sum of the welfare of the two parties.

Figure 12.1 Measuring Social Welfare



Measuring Social Welfare

Figure 12.1 depicts the market demand curve and the market supply curve of a particular good. The competitive equilibrium is attained at the point E, with the price equal to p_0 and the quantity equal to y_0.

Focus now on the demand curve. Remember from Chapter 5 that the concept of consumer surplus presumes that the total utility from consuming a product, measured in money, which may be termed as the total gross benefit, is equal to the area under the demand curve. Thus if, for example, the amount y_1 is consumed, the total gross benefit = area $0ACy_1$. If y_2 is consumed, it is equal to $0ADy_2$ and so on.

We have already learnt that the supply curve represents the marginal costs at various levels of output and also that the area under the marginal cost curve is the total variable cost. Thus, the area under the supply curve equals the total variable cost. For instance, at the output y_1, it is equal to the area $0BGy_1$ and so on.

Hence, applying the definition (12.1), social welfare at any given quantity = Area under the demand curve–area under the supply curve–total fixed costs (*TFC*).

Being a given amount, *TFC* does not matter in comparing social surplus associated with different quantities. For convenience, we can ignore it (that is, assume *TFC* = 0) and say that the social welfare or the social surplus (W) = Area between the demand curve and the supply curve.

For example, at y_1, $W = ACGB$ and at y_2, $W = ADHB$. At any point to the right of y_0, the expression of W is a bit different. For instance, at y_3, $W = 0AFy_3 - 0BKy_3 = AEB - EFK$.

Socially Optimal Output and the Competitive Equilibrium

Now we can easily compare social surplus at various quantity levels. For instance, W at y_2 is greater than W at y_1. More generally, as we consider higher and higher levels of output to the left of y_0, the social surplus is greater. At y_0, it is equal to area AEB. Moreover, if we consider any level of output higher than y_0 (such as y_3),

W is less. It then follows that the social surplus is maximised at y_0. This is where price is equal to marginal cost, that is, $p = MC$ is the condition for social surplus maximisation or, what is called, **social efficiency**. The economic rationale behind it is that having an extra unit of a product provides utility to the society equal to the marginal utility, which, in turn, is equal to the price of the product, and the extra social cost of having it is its marginal cost. Therefore, the social surplus is maximised when the extra benefit or utility is equal to the extra cost.

Notice immediately that the condition $p = MC$ holds under perfect competition, that is, in terms of Figure 12.1, y_0 is the amount produced in competitive equilibrium. Thus, the social surplus is maximised, that is, social efficiency is attained at the competitive equilibrium. In other words, a competitive market structure serves society the best when social welfare is defined and measured in the way we have. It is indeed a central result in the whole of the science of economics.

And no one as yet has described the beauty of this result more tellingly than Adam Smith. According to him, in a perfectly competitive market, every one, a producer or a consumer, is maximising his/her self-interest (either utility or profit), being guided by the market price on which he/she has no control. No one is doing any one else any direct favour. Yet pursuing self-interest serves the society's interest. Put differently in a perfectly competitive market of the kind we have studied, there is no conflict between individual interest and social interest. We have 'proved' this geometrically, but it can be proven in a very general setting (which is beyond our scope).

MARKET FAILURES

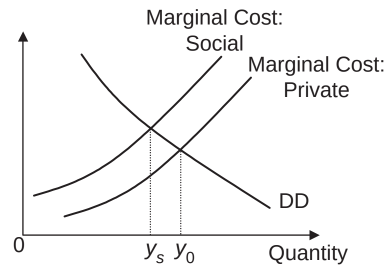
This central result then implies that a market fails to perform in the best interests of the society that is, there is a **market failure** if either the market structure is not perfectly competitive or there are conditions within a competitive market such that private interests do not match with the interests of society. Some of the non-competitive market structures and the policy implications in such markets will be discussed in the following chapters. Here, in the remainder of this chapter, we will identify some of the situations in which there is a market failure even when the market is perfectly competitive.

Externalities

In Chapter 9, a distinction was drawn between private cost and social cost based on the presence of some externality. If such a situation exists, the quantity at which social welfare is maximised differs from the competitive equilibrium and the 'market fails'.

Recall from Chapter 9 the example of a negative externality, where a set of factories are emitting pollution and causing health problems for the surrounding neighbourhoods. The social marginal cost (SMC) at each level of output is greater than the private marginal cost (PMC). This is illustrated in Figure 12.2. The private

Figure 12.2 A Negative Externality



marginal cost curve is the competitive supply curve as in Figure 12.1 and the competitive equilibrium holds at the quantity y_0. However, the total social cost is the area under the social marginal cost curve. Thus, the social surplus is maximised at the intersection of the demand curve and the social marginal cost curve. The corresponding quantity is y_s, which is the socially optimal level of output. Clearly, the market fails. More exactly, the market 'overproduces' compared to what is best for the society (by the amount $y_s y_0$ in Figure 12.2). The economic reason is clear. That is, private producers, in the pursuit of their self-interest, do not take into consideration the negative externality they are causing to society. That is why they are overproducing in comparison to what is most desirable for society.

What is then the corrective measure? It is to tax the producers such that their own (private) cost, equal to the marginal production cost plus the tax per unit of output, coincides with the social marginal cost. Such a tax shifts the equilibrium from y_0 to y_s. In other words, the optimal policy intervention is to impose a tax on output, exactly equal to the marginal externality cost (in the example, the health cost of pollution) being imposed on the society. Put simply, the optimal policy is to make the producers pay for the health cost of the society.

Mark, however, that if the government imposes a very heavy tax such that the tax-inclusive marginal private cost curve lies to the left of the social marginal cost curve, the social surplus is not maximised. It is because the product, though causing pollution, is nonetheless used for society's consumption and generates obvious benefits. Thus, taxing it too much leads to such a large reduction of this benefit that it outweighs the extra benefit from reduced pollution. Some well-meaning yet overenthusiastic environmentalists seem to sometimes overlook this and clamour for draconian measures to curtail production of goods that cause pollution. Our economic analysis says that in designing policy, we should, of course, take into consideration the pollution cost to the society but the benefit side of industrial goods should not be ignored. What is called for is a measurement of the (health) cost of pollution or the cost of a negative externality, and a policy which comensurates with the measured cost.[1]

[1]Of course, a long-term policy would be to encourage new environment-friendly technologies.

Figure 12.3 illustrates the optimal intervention issue in a different way. It measures the output tax along the x-axis. Increasing the tax rate, on one hand, has the benefit of reducing production and thereby reducing pollution or the negative externality. This defines the marginal benefit curve resulting from increasing the tax rate. On the other hand, less output means less consumption of the good, which is a cost to the society. In our example, the chemicals produced by the chemical factories could be used in various useful drugs, so that the less the output, the less is the availability of useful drugs for various ailments. This is captured by the marginal cost curve. The intersection of the two curves defines the magnitude of the optimal tax rate—and this is the tax rate at which the private producers' tax-inclusive marginal cost curve coincides with the social marginal cost curve in Figure 12.2. Such an optimal tax rate is called a **Pigovian tax**, named after a famous English economist Arthur Pigou, who was the first to separate the private cost from the social cost and analyse how the government, by mixing taxes and subsidies, can correct market failures.

There is still another way to look at a Pigovian tax, as in Figure 12.4. It measures the tax rate on the x-axis, while social welfare is measured on the y–axis. When the tax rate is zero, social welfare is measured at $0A$, which is the free market level of welfare. As the tax rate is gradually increased, the gap between the

Figure 12.3 Negative Externality and the Pigovian Tax

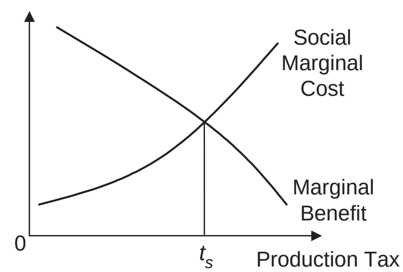
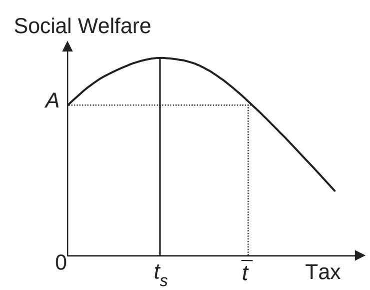


Figure 12.4 Pigovian Tax Once Again



private and the social cost gets narrowed and, as a result, social welfare gradually improves. It attains maximum level at the tax t_s; this is where the social marginal cost is equal to the private marginal cost. Beyond this, there is again a discrepancy between social and private costs and the level of social welfare falls with an increase in the tax. At any tax rate above $\bar{t}$, the associated level of welfare is even less than $0A$, the welfare level associated with no intervention. The point is that high-handed, heavy policy interventions have the danger of making the matter worse as compared to a free market approach, even if there exists a reason for policy intervention.

So far we have concentrated on negative externalities. But there can be positive externalities. Consider the nursery industry in a town. Suppose there are many nursery firms spread out within a city. While these sell plants to customers, their very presence cleans the air and offers greenery. In this case, the social marginal cost curve is below the private marginal cost, implying that the socially optimal output is greater than the free market level of competitive output. The market also fails in such a situation. What is the optimal policy intervention? It is a production subsidy, that is, a negative Pigovian tax.

Now consider a slightly different situation. Suppose a competitive industry has two groups of firms—one having technology that is environmental friendly, that is, it does not cause pollution problems and the other having polluting technology (as in our original example). Both groups of firms coexist. Without loss of generality, let there be one firm in each group. Let these firms be called firm 1 (the 'clean' firm) and firm 2 (the 'dirty' one).

Here also the social marginal cost exceeds the private marginal cost for the industry as a whole. Figure 12.2 applies. A free, competitive industry will produce more (at a point like y_0) than what is socially efficient (y_s). Some kind of a production tax is required to improve social efficiency.

Consider first a uniform production tax on both firms such that their marginal cost curves shift upward and the amount y_s is produced in equilibrium by both firms together. That is, this tax policy leads the industry to produce a total amount that is consistent with social efficiency. But is this policy intervention the most efficient or simply the efficient? The answer is 'no', because in this case efficiency is governed not just by the total amount produced but also by an allocation of this total output.

More precisely, the overall social efficiency requires the condition $p = SMC_1 = SMC_2$, where SMC stands for social marginal cost. Since firm 1 does not pollute, $SMC_1 = PMC_1$, where PMC stands for private marginal cost. But for firm 2, $SMC_2 = PMC_2 + MC_x$, where MC_x is the marginal cost of pollution. Thus the efficiency conditions can be stated as:

$$p = PMC_1 = PMC_2 + MC_x.$$

That p must equal PMC_1 implies that firm 1 should not be taxed at all, whereas $p = PMC_2 + MC_x$ means that firm 2 should be taxed, as in our original example, with the marginal tax being equal to MC_x.

Hence the best policy intervention is to impose an appropriate Pigovian tax only on firm 2, which pollutes and thus causes negative externality. There is no need to tax firm 1, which is not causing any harm to others.

We have looked at externalities caused by some production activities and how these are imposed on the consumers or households. There are examples of externalities caused by one production activity towards another, or by one type of consumption activity towards others. A prime example of the latter is smoking. Smoking by smokers causes health problems for others too who inhale it. Therefore, a free market equilibrium in terms of producing and consuming cigarettes is not socially optimal. There is over-consumption of this **demerit good**. What is the nature of optimal policy? It is a tax on the consumption of cigarettes. The sales tax imposed on cigarettes serves this purpose. Also, various public-area restrictions for smokers can be viewed as taxing cigarette consumption (insofar as it increases the inconvenience of smoking in public). By the way, in 2007 Chandigarh became the first smoke-free city in India.

Public Goods

Governments intervene in different markets not just by taxing or subsidising a product or service but also by becoming the supplier. We are referring here to not just running a business in the so-called public sector but also providing basic services like national defence, administrative services and so on.

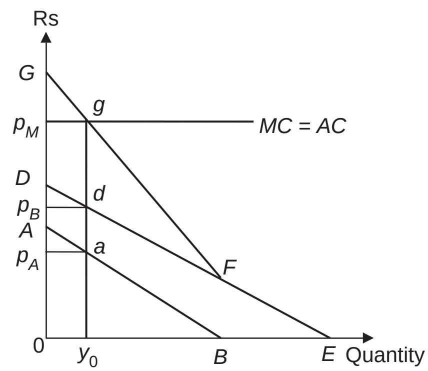
The latter are **public goods**. By definition, a pure public good is one having two characteristics—non-rivalry and non-excludability. **Non-rivalry** means that, if a fixed amount of the good is available for consumption, one person's consumption of it does not reduce the amount available for others. Think of national defence. In a sense we 'consume' national defence in that it provides us a service of safety. A fixed magnitude of national defence is enjoyed by all of us. The availability of national defence to you does not reduce the availability of it to me. In other words, we are not 'rivals' in consuming the national defence service. As another example, think of plantation of trees along national highways to prevent soil erosion as well as provide greenery and a clean environment. This service is available to all. My enjoyment of looking at the greenery does not reduce anything of what is available to you. Formally, if you and I each demand a public good equal to, say, y our total demand for this good is equal to y, not $2y$.

Non-excludability means that if it is available to one, it is available to all. No one can be excluded. It is not same as non-rivalry. Consider, for example, a fenced enjoyment park, which is only available to those who pay an entrance fee. Non-excludability does not hold here because those who do not pay can be excluded. But it is characterised by non-rivalry. If you and I both pay and enter the park, your consumption of the services of the park does not reduce mine (unless of course it is too crowded, so that the presence of other people hinders my enjoyment).

As we will now see, both these characteristics together create a situation where a public good cannot be provided by a private entrepreneur.

Consider non-excludability first. Since no one can be excluded, everyone will expect others to pay for it and, once the service is available, expect to consume it

Figure 12.5 Non-rivalry of Public Goods and Market Failure



free. That is, everyone will have the incentive to free ride on others. This is what is called a **free rider problem,** meaning that every one recognises that if others pay for a product or service, she/he does not have to pay and therefore no one has an incentive to pay. It is like saying that 'I prefer more police protection in trains. But, I do not want to pay for it because I know that if others somehow pay for it, I will enjoy the benefit too'. The end result is that only a few end up paying for the service. Thus, it will not be profitable for a private entrepreneur to supply the service. The market fails.

The problem caused by the non-rivalry characteristic of the public good can be understood via Figure 12.5. Suppose there are two individuals, *A* and *B*. Their demand curves for the public good are given respectively by the lines *AB* and *DE*. How does the aggregate demand curve look like? Unlike that of a private good, the aggregate demand for a public good is the vertical, not the horizontal, summation of the individual demand curves. In Figure 12.5, the lines *AB* and *DE* are added vertically to obtain the line *GFE,* which is the aggregate demand curve. How?

Suppose that the total available quantity is y_0. Because it is a public good, non-rivalry implies that y_0 is available to *A* and *B* each. From *A*'s demand curve, his total utility is equal to $0Aay_0$, the area under his demand curve. Similarly, the total utility to *B* is equal to the area $0Ddy_0$. The total utility to the two individuals together is equal to the area $0Ggy_0$, under the line *GFE*. The same holds at any other level of total quantity available for consumption—the total utility of the individuals together equals the corresponding area under the line *GFE*. It follows that the height of *GFE* at any given level of consumption, for example, gy_0, measures the society's marginal willingness to pay, which is the price. Thus *GFE* is the aggregate or the market demand curve.

Assume that there is no difference between the private cost and the social cost (that is, there are no externalities of the kind discussed earlier). Let *MC* be constant, which is shown by the horizontal line. (If *MC* is constant and there are no fixed costs, it must be equal to the average cost, that is, *AC* = *MC*.)

Where is social surplus maximised? It occurs at point g, with the total quantity equal to y_0. The question is—will this quantity be provided by a competitive, private market?

The answer is no. The problem is that a private producer cannot charge the 'right' price to each consumer because at the level of the common quantity consumed (as it is a public good), each consumer has a different marginal valuation—it is p_A for consumer A and p_B for consumer B. If the price p_A is charged to both consumers, it is too low for consumer B and hence social surplus is not maximised; if p_B is charged, it is too high for consumer A and thus at this price also, the social surplus is not maximised. Any other price is not right for any individual, and hence not right for society.

How about consumer A being charged p_A and consumer B being charged p_B? In principle, this will maximise social welfare. But there are two problems. First, almost surely, the private producer will not have information about the true valuation by each and every individual because every one has an incentive to reveal a near-zero valuation for the product in order to free ride. Second, even if it is possible to know their true valuation, it will be extremely costly to exercise such price discrimination across so many consumers. The end result is that the market fails.

This is why the government steps in such situations and provides public goods by financing it through (compulsory) taxes.

Minimum Consumption of Food by the Poor

The total social utility measured by the area under the market demand curve drawn in Figure 12.1 neither tells us or 'cares', so to speak, who consumes how much. Like in our analysis of famine in Chapter 11, if Figure 12.l depicts the total market for an essential product like food, it is possible that out of the total quantity consumed y_0, the poor are able to consume—rather afford—only a small portion of it, less than what is needed for normal nourishment. In other words, under nourishment may be present in a competitive equilibrium as depicted in Figure 12.1.

Therefore, any reasonable notion of social welfare should not ignore it. It must place a higher weight on a poor person's consumption of a necessity like food than that on a relatively rich person's consumption of it. Then, clearly, the competitive equilibrium does not maximise social welfare, that is, the market fails. The best corrective measure must involve subsidising the food consumption of the relatively poor such that no one is under nourished. Put differently, a targeted food subsidy programme is the best policy option. This is indeed the rationale behind the PDS/TPDS system in India.

However, how should such a programme be operationalised? This is where the problem of our PDS/TPDS programme lies. If the government itself takes upon the task of procuring and distributing food for the poor (as it has been thus far), inefficiencies are bound to creep in. Instead, some sort of a food stamp programme is likely to be a more efficient way of executing the scheme because such a programme can be implemented through the private market system and thus would not require resources needed for the tasks of separate procurement and separate distribution via fair price shops.

Redistribution

The formula (12.2) and the corresponding Figure 12.1 presume that the marginal valuation of one rupee is same for everyone. But in a society where there is a lot of economic inequality, this is not a good assumption. A very poor man's valuation of Rs 100 is very different from that of a millionaire. Thus, some consideration of distribution of income and wealth is desirable in any concept of social or collective well-being. A free competitive market, resulting in a very unequal society, with career opportunities not affordable to many cannot be construed as working in the best interests of the society.

If redistribution is an objective, what is then the right policy intervention? The answer is a progressive income tax system, meaning a relatively high rate for the rich and a relatively low rate for the poor. It is good to know that most countries do follow a progressive income tax system.[2]

Other Reasons for Market Failure

There are other sources of market failure, such as various forms of **asymmetric information**. For instance, think of an attempt by an Indian firm to introduce a product in the international market. The firm knows that the quality of its product is very good. But suppose that the customers in the international market do not know the quality of the Indian product (as it is new in the market and Indian products are yet to make a reputation). Hence, they will not be willing to pay a price commensurate with the cost of producing such a high-quality product. It is a situation of asymmetric information—product quality is known to the producers but not to the consumers.

In this situation, it is quite possible that in a free market, the Indian firm will incur a loss and, therefore, may not be able to enter the international market. This is clearly not an optimal situation. Subsidising such a firm till it is able to establish the reputation of its quality will improve social welfare.

Asymmetric information is typically present in credit markets. A lender may not know the credit record of a potential borrower. That is, a borrower knows his/her 'type' but the lender does not. Thus a lender may not be able to distinguish a good borrower from a bad borrower. The result is that all borrowers, good and bad alike, will be rationed in some ways by the creditors. This is not socially optimal. What is the right policy intervention? It is to develop and maintain a nation-wide computerised database that keeps tab of a person's past credit record.[3]

[2]If a tax system is too progressive, it creates negative work incentives and also encourages tax-evasive behaviour.

[3]India is currently in the process of developing such a database.

FIRST-BEST AND SECOND-BEST

If you go back to our earlier example of two groups of firms, one of which causes a negative externality to households, we saw that a uniform tax on both groups of firms reduces pollution, no doubt, but taxing only the group of firms that cause pollution was the best policy intervention. The latter is an example of a **first-best policy intervention**—generally defined as that policy intervention, which addresses the source of market failure in the most efficient manner. In comparison, a general production tax on both firms, which brings about the correct amount of total output but not the optimal allocation of it among the firms, is welfare-improving but not the best; it is an example of what is called a **second-best intervention**.

The distinction between the first-best and second-best policies is important since it clearly brings out what is the best among many policy options that work towards a particular goal (for example, reducing pollution). Indeed, in reality, governments do adopt, in many situations, second-best, not first-best policy measures.

For example, tax revenues are important for any government to function and to provide public goods like national defence, national highways, an administrative system and so on. Recall the discussion in Chapter 5 on direct versus indirect tax. Our conclusion was that income tax is a better method to raise revenue than an indirect tax. If raising revenue is the objective, then taxing income is the first-best option. Yet, for decades, India partly justified its use of heavy import-tariff regime on the grounds that tariffs are necessary for revenue purposes. An import tariff is an indirect tax as it changes the price of the imported goods for the consumers and producers. Thus tariffs are, at best, a second-best solution to raise revenues.

If we go back to our discussion of the PDS system, a blanket policy under which almost every household is entitled to some amount of grain at low prices is only a second-best policy, whereas a targeted PDS system (TPDS) is, at least in principle, the first-best.[4]

Consider a macro example. In the days of a pegged exchange rate system, India used various policies to help out its export sector. One of these was to use a favourable exchange rate for an export-oriented production unit. But manipulating the exchange rate has implications towards a country's balance of payments. Thus, it is only a second-best policy, whereas the most direct and the first-best policy option is to provide a direct production subsidy—if promoting this sector is the objective.

In general, the first-best intervention is one that addresses the 'problem' most directly and specifically without affecting other sectors or units in the economy.

A SUM-UP: CONSERVATISM VERSUS LIBERALISM

We started out by saying that a freely functioning, competitive market structure serves a society's interest the best. But we also saw how competitive markets may

[4]As we discussed earlier, there are costs involved and other difficulties in running a TPDS system.

'fail' and a role for government intervention arises. What are we advocating then? Should the government, as a matter of practice, regularly interfere with the market or not? Those who say 'yes' are labelled as liberals and those who say 'no' are the conservatives. The former charge that the latter's argument for free operation of a market is justified only in case of a competitive economy without problems like that of externalities, public goods and so on. But actual economies do have problems that markets cannot correct by themselves. Therefore, it is imperative that governments intervene in the markets on a regular basis.

To their benefit, the conservatives are generally aware of the concern of the liberals. But their worry is that in the name of intervention, the government may make things even worse, that is, there could be the problem of **government failure** also. It can arise due to lack of information and political imperatives.

Turn to Figure 12.4. Suppose the government does not have good information on the marginal social benefit and marginal social cost functions. Then it is quite possible that the government may 'overshoot'. It may select a tax (as in the example) higher than what is truly optimal, which may even exceed $\bar{t}$, with the result that intervention is worse than what the free market delivers despite its problems. This is a generic problem and not specific to the tax example. Another source of lack of information is that the government may not have the administrative set-up to correctly identify the beneficiaries of a programme or those individuals firms that cause externalities. The outcome could very well be, at best, a second-best situation with the net benefit being less than the cost of administering a programme or regulation.

Even if the government may have fairly precise information, political imperatives may force policy makers to overshoot. If, for example, a liberal government is coming to power after a gap, it may wish to please its constituency by going overboard—by going beyond what may be the optimal degree of intervention or by deliberately choosing a second-best option. As an extreme example, in the name of redistribution of income, it may impose such a heavy tax on the relatively rich that it effectively kills any incentive to be an entrepreneur and produce something for the society; in such a society, equality will be achieved, no doubt, but everyone will be poor.[5] To summarise, lack of information and political imperatives may lead to **over-intervention** or a second-best intervention.

In sum, there is no simplistic answer to whether conservatism or liberalism is the way to follow in advocating economic policy. A careful statement will be something like the following. Yes, there is a role for the government to play but only in a very selective manner. However, if there is not enough information about the parameters of a market, the benefit of doubt should go to the free market not to policy activism, unless the problem is visibly acute.

This sounds a bit conservative, but many economists are likely to agree with this assessment.

[5]Put differently, a sufficiently heavy intervention to redistribute a pie will reduce the size of the pie itself.

Economic Facts and Insights

- In measuring social welfare or surplus, consumers' benefit is represented by consumer surplus and producers' benefit by the producers' surplus or profits.
- Social surplus in case of a single good is measured by the area between its demand curve and supply curve. It is maximised when price equals the social marginal cost.
- The social surplus is also maximised at the competitive equilibrium since the price = marginal cost condition is met.
- Market failure occurs when a market is not perfectly competitive.
- Even when markets are perfectly competitive, market failure may occur because of externalities, public goods, minimum consumption requirements by the poor, asymmetric information and so on.
- A Pigovian tax/subsidy corrects the market failure due to the presence of negative/positive externalities.
- In case of a public good, non-excludability leads to a free rider problem.
- The market demand curve for a public good is the vertical summation of individual demand curves.
- High-handed policy interventions have the danger of making the matter worse, as compared to a free market approach, even if there is a valid reason for policy intervention. This is due to lack of information and political imperatives.
- Second-best interventions also emerge because of political compulsions.

EXERCISES

12.1 How is social welfare/surplus defined in relation to production and consumption of a particular good and how does it relate to consumer surplus and profits?

12.2 What is the condition for maximising social welfare and what is the logic behind it?

12.3 'Equilibrium in a competitive market attains a social optimum or efficiency.' Defend or refute.

12.4 'In a market economy, all are 'selfish' in the sense that they wish to maximise their own benefit without being directly concerned about the benefit of others. Yet, their selfish behaviour may lead to a socially desirable outcome'. Agree or disagree? Give reasons.

12.5 What does the term 'market failure' mean?

12.6 Diagrammatically explain how the market fails in the presence of a positive externality.
12.7 What is a Pigovian tax? Explain the concept through an example (other than the one given in the text).
12.8 The production process of an industry causes a negative externality to the society. Then the socially optimal output is ____ than the output in the unregulated market equilibrium. Fill in the blank and give reasons.
12.9 'If an industry is causing pollution, in the interest of social well-being, it should be taxed as heavily as possible'. Comment.
12.10 What are the two characteristics of a public good? Briefly explain these.
12.11 'The market demand curve for a public good is obtained by horizontally adding the individual demand curves for such a good'. Agree or disagree. Give reasons.
12.12 What problem does the non-excludability characteristic of a public good cause towards paying for the good?
12.13 What problem does the non-rivalry characteristic of a public good cause towards paying for the good?
12.14 What is meant by a first-best policy intervention? Explain it through an example.
12.15 What is meant by a second-best policy intervention? Explain it via an example.
12.16 What is meant by government failure in terms of policy?
12.17 Give your opinion on whether the following policies are first-best or second-best.

(*a*) A steel industry in a country suffers in terms of employment and output because of cheap import of steel from abroad. On the grounds of protecting national interest, the government bans the import of steel into the country.
(*b*) A fair amount of cheating was detected in a city in Gujarat during the entrance exams to civil services. As a result, the government plans to ban all civil service entrance exams in Gujarat. If anyone from Gujarat wants to appear in these exams, they have to travel to Maharashtra.
(*c*) Some criminals in a country were able to escape from the country because they had passports. To tackle this problem, the government decides to ban the issuance of passports to all its citizens for the next two years.
(*d*) Some private hospitals in a large city have failed to comply with some government regulations for a number of years. To respond to such problems, the government bans the establishment of new private hospitals for the next three years.

13 Game Theory and Economic Applications

CONCEPTS	
• Nash Equilibrium	• Non-cooperative Game
• Payoff Matrix	• Game
• Zero-sum Game	• Non-zero-sum Game
• Cooperative Game	

Game theory is not about how a particular game is played or won. It deals with strategic behaviour in a situation of interdependence among individuals, species or organisations, that is, in a situation where the behaviour of one party affects the well-being of another. Technically speaking, it is a branch of mathematics, which has wide applications in various subjects. Economics analyses the behaviour of people and organisations vis-à-vis other people and organisations in a strategic environment. This is how game theory is relevant for economics. In the last few decades, game theory has become an integral part of economics and is considered essential for any higher study in the subject. Since 1995, seven scholars have been awarded Nobel prizes in economics for their contributions to game theory.

Many economic problems can indeed be meaningfully posed in game theory. For example, when firms compete in advertising, one firm's advertising strategy depends on its rival firm's choice of an advertising strategy. One country's economic policy may affect another's welfare and, therefore, the latter's choice of an appropriate policy.

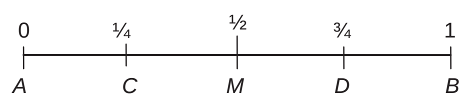
This chapter provides an introduction to game theory via illustrative examples.

ICE-CREAM ON THE BEACH

Think of a sea beach as a straight line from 0 to 1, shown in Figure 13.1. On a sunny, warm day it is crowded. Assume that the crowd is evenly spread out and everyone wants to consume one ice-cream. Assume further that there are exactly two ice-cream vendors. Both sell one type of ice-cream only and there is no quality difference between the ice-creams sold by them. This implies that there is no preference as such for ice-cream from any particular vendor over that from the other. Hence the price of ice-cream bought from either vendor will be the same. The central question is—where are the vendors going to situate themselves on the beach?

It is a 'two-person' or a 'two-player game', a game between the two vendors. There are many, indeed infinitely many, possible strategies available for a vendor—he could choose any point on the real-line interval [0, 1]. Are both going to select the same spot, that is, stand at the same location on the beach like at *A*, *C*, *M*, *D*, *B* or any other point? Or, are they going to locate themselves at different points like one at *A*, the other at *B;* or one at *C;* the other at *D;* or at any other pair of points? There is an important but natural assumption to be used. That is, people on the beach incur a discomfort to walk to and from the ice-cream vendor on the hot sand

Figure 13.1 Location of Ice-cream Vendors



and the longer they have to travel to get their ice-cream, the greater is their total discomfort. This can be viewed as an additional cost of having an ice-cream.

Let us try to figure out the answer. One possible answer is that one vendor stands at *C* (marked ¼) and the other at *D* (marked ¾). It is natural to think that this is the right answer because with these locational choices people on the left-half of the beach go to the vendor at *C* and those on the right-half of the beach go to the vendor at *D*. No one has to travel more than half of the beach length. Each vendor serves half of the market.

However, we will now argue that the strategies (*C*, *D*) do not constitute the answer. Starting with (*C*, *D*), suppose the vendor at *D* moves a little to the left. He will then have all his previous customers (from *M* to *B*) plus a few more on the immediate left of the mid-point *M*, to whom this vendor will now be nearer. As a result, the vendor on the right-half captures more than 50 per cent share of the market. Will the vendor on the left-half continue to stay at the point *C*? Unless he is a fool, he will move rightwards and recapture or perhaps extend his market share.

Where will this process end? The answer is the point *M*. That is, both vendors will locate themselves in the middle. This is the answer. It is an *equilibrium* in the sense that if both vendors are at *M*, neither has any incentive to change its position. Indeed, it is the Nash equilibrium of this game, named after a Nobel laureate in economics, John Nash. In general, a collection of strategies by the players constitutes **Nash equilibrium** if no single player has any incentive to change his strategy.

This example explains something, which we observe in the market place. In providing a service or a product, the providers (shops) have an incentive to cluster geographically rather than spread themselves out. For example, you must have somewhere noticed two petrol pumps side by side on the road and wondered why they did not locate themselves separately. You may also find two similar eating joints side by side at the same location.

Our example illustrates a very general and important point as well. Notice that in our example the socially most desirable solution is that the vendors locate themselves at points *C* and *D* because this will minimise the total discomfort from travelling for all people on the beach, while each firm keeps its natural market share of one-half. But the strategic equilibrium is not consistent with this, that is, unlike in a competitive market, the social welfare is not maximised at the Nash equilibrium. You can connect this with our analysis in the last chapter by noting that here there are only two players, not infinitely many, and hence each player has some power. This is why the Nash equilibrium entails market failure. Pursuing self-interest is not in the best interests of the society.

To further understand this point, note also that, in our example, the vendors play what is called a **non-cooperative game**, that is, the players do not select their strategies jointly in order to maximise their joint profits. But if they would have acted cooperatively, they would have selected the points *C* and *D*, which would have served the interest of the society the best and they would have been as well-off as at the point *M* (each serving 50 per cent of the market). This illustrates that cooperation is valuable for a society.

PRISONERS' DILEMMA

Two guys, *A* and *B*, have committed a murder. The police know it but only have circumstantial evidence and want confessions. Without the confessions the police expect to get, from the court of law, only a mild, two-year prison term for each.

In attempting to make them confess, the police keep *A* and *B* in two separate cells—so that they cannot communicate (there are no cell phones) or cooperate—and offers them a deal. Each one is told separately, 'If you do not confess, but 'your friend' does, you get 20 years in prison and he gets only seven years. If you both confess, each of you will get a 12-year prison sentence'.

In this game, note that there are two strategies available to each player—'confess' (C) or 'hang tough' (T). The question is, what would the prisoners choose? To arrive at an answer let us consider the utility to a player from various combinations of strategies. In game theory a player's utility is typically called his payoff. For instance, if *A* confesses (chooses C), while *B* hangs tough (chooses T), *A* gets a seven-year sentence (his payoff is –7) and *B* gets 20 years (his payoff is –20). Various payoffs associated with various strategy combinations in the game are listed in Table 13.1. For each combination of strategies there are two numbers, showing the players' payoffs. The first number is *A*'s payoff and the second is that of *B*. Table 13.1 is a **payoff matrix**.

In order to determine which of the four strategy combinations will be chosen, we must use a criterion. The standard assumption in game theory is the notion of Nash equilibrium, that is, we seek that combination of strategies such that no player has an incentive to deviate from it.

What about (T, T)? It is not a Nash equilibrium, since either player can deviate (that is, move to C) and be better off (prison sentence will be reduced by 13 years). What about (C, T), that is, *A* plays C and *B* plays T? From this combination, *B* can deviate and benefit by choosing C instead; he can reduce his prison sentence by 8 years. Indeed you can check that the Nash equilibrium is (C, C), that is, both players confess. This is the solution to the Prisoners' dilemma game.

The point of this example is that, from the players' perspective, the best strategy combination is of course (T, T). But they are unable to attain it because they are denied any opportunity to communicate and cooperate. Put differently, a non-cooperative game does not achieve the best for the players.

Table 13.1 Payoffs in the Prisoners' Dilemma Game

Prisoner A	*Prisoner B*	
	C	T
C	(–12, –12)	(–7, –20)
T	(–20, –7)	(–2, –2)

ADVERTISING GAME

Suppose there are two firms, *A* and *B*, in an industry. Each produces a product, which is a direct substitute of the other. There are two strategies available to each firm—advertise and do not advertise. Assume that advertising costs Rs 10 lakh. When neither firm advertises, each firm earns a profit of Rs 50 lakh. If one firm advertises and the other does not, some customers switch from the latter to the former. The former advertising firm's (net) profit is Rs 60 lakh, while the latter earns a profit of Rs 30 lakh. If both firms advertise, then their profits are Rs 40 lakh each.

The payoff matrix is given in Table 13.2. The left-hand figure is the payoff of firm *A* and the right-hand figure is that of firm *B*.

What is the Nash equilibrium of this game? Applying its definition, you can deduce easily that (advertise, advertise) is the Nash equilibrium. Notice that this game is analogous to the Prisoner's dilemma. Compared to the Nash equilibrium, both players can be better off—and the society will benefit—if they choose (no ad, no ad) instead. Yet, they cannot do it because they are 'behaving' (choosing their strategies) in a non-cooperative manner.

WORLD TRADE

Almost all countries in the world engage in international trade. They export certain items and import some. Divide all trading countries of the world into two types—developed countries and developing countries and think of them as two countries. Table 13.3 shows their payoffs from various combinations of trade policy.

If both countries impose trade restrictions against each other, their benefits are (60, 60), measured in some units, where the first entry is the benefit to the developing countries and the second entry is that to the developed countries. On the

Table 13.2 Advertising Game

Firm A	*Firm B*	
	Ad	No Ad
Ad	(40, 40)	(60, 30)
No Ad	(30, 60)	(50, 50)

Table 13.3 Payoffs in the Trade Policy Game

Developing Countries	*Developed Countries*	
	Free Trade	Trade Restrictions
Free Trade	(100, 100)	(30, 160)
Trade Restrictions	(160, 30)	(60, 60)

other hand, suppose they cooperate and eliminate trade restrictions against each other. Both would benefit from specialising and exporting products that they are relatively better at producing. The benefits associated with free trade (zero trade restrictions) are denoted as (100, 100). Note that the same numbers for both developing and developed countries do not mean that both gain equally. For instance, for one country the gains may be measured in terms of dollars and for the other it may be in terms of some other currency. Indeed, a fundamental lesson learnt from international economics is that there are many sources of benefits from free trade to a nation as a whole. While this is not our subject matter here, simply note that, compared to both countries choosing trade restrictions, each country's payoff is higher under free trade permitted by each other. However, if one country follows trade restriction, while the other does not, the former gains at the expense of the latter. For instance, if developing countries play 'trade restrictions', while the developed countries play 'free trade', the former is better off than under free trade (that is, 160 versus 100) and the latter is worse off.

What is the equilibrium in this game? You can check that it is 'trade restrictions' by each country. This is again analogous to the Prisoners' Dilemma game. The implications for cooperation are also similar. If both are able to cooperate, they will choose free trade and both will be better off.[1]

FORMAL NATURE OF A GAME, ZERO-SUM AND NON-ZERO-SUM GAMES, AND NON-COOPERATIVE AND COOPERATIVE GAMES

Game theory goes far beyond the two-player or two-strategy examples we have studied. In general, a **game** (as in game theory) is defined by:

(*a*) a set of players;
(*b*) a set of strategies available to each player; and
(*c*) a list of payoffs for each player from each combination of strategies by all players.

A game is 'solved' by using some notion of an equilibrium, for example, the Nash equilibrium.

Apart from the examples already given and the ones that will be given in the exercises, there are numerous economic problems where the tools of the game theory are applied. For instance, the interaction between labour unions and firms can be thought of as a game where the union sets the wage rate and the management of the firm decides the level of employment. In a political economy context, two parties, as players, choose the economic policies so as to please their own party

[1] In the best of spirits, WTO can be seen as an organisation that attempts to induce cooperation and discipline across countries so as to ensure global free trade.

supporters and win votes from rival party sympathisers. In Chapter 15, we will use game theory concepts to study the behaviour in an oligopoly market.

In general, there are two kinds of games: zero-sum games and non-zero-sum games**. Zero-sum games** are those in which the sum of the payoffs to players in each combination of strategies is zero. This is equivalent to saying that in a two-player or a two-team game, one wins and the other loses; there are no other possibilities. Any game, which is not a zero-sum game, is a **non-zero-sum game**, defined by the property that the sum of payoffs do not always add up to zero.

None of the examples given so far are zero-sum games. Indeed, most economic situations in which strategies are important are not zero-sum in nature. This is important to know because there are still some people who erroneously think that international trade is a zero-sum game, for example, if there are two countries and one country benefits from free trade, the other must lose by definition. Instead, in most trading situations *all* trading partners benefit, although they may not equally benefit. Similarly, some think that allowing multinational companies to make profits in a country is a zero-sum game—since these companies are making profits, the host country must be losing. This is also a mistaken notion. While the multinationals may be making profit, they may be creating employment or at least enhancing the salary of some workers. Is this not good for the workers? Also, they may be making and selling a product, which otherwise would not be available at a reasonable price in the market. Is this not good for the consumers? While this is not a forum to formally discuss the merits and demerits of international trade and investment, the general point is that such economic activity is generally non-zero-sum in nature.

In all the examples given above, there is a unique Nash equilibrium. But this need not be the case; there may be two or more equilibria (see Exercise 13.3). In some games, a Nash equilibrium may not even exist (see Exercise 13.4); this does not mean, however, that game theory cannot be used. There are many other equilibrium notions and different ways of using strategies. Indeed, game theory is vast, rich and fascinating.

We will end our introduction to game theory by noting a different classification of 'game situations'. There are **non-cooperative games** (all the above examples and those in the exercises are of this kind) and **cooperative games**. Non-cooperative games are those in which players act, that is, choose their strategies, on their own without any direct understanding or communication with other players. In cooperative games, the players maximise their joint payoff or surplus and distribute it in a way that benefits them all (though not necessarily equally). Thus, in a cooperative game equilibrium, players are better off than in the corresponding non-cooperative game equilibrium.

This raises a question of why the players or economic agents do not always play a cooperative game. There are two problems. In some situations cooperation is not permitted. Think of the prisoner's dilemma game. Will the police allow the murder suspects to cooperate in admitting or not admitting to the crime? As we shall see in Chapter 15, firms or companies are not allowed to cooperate or collude in deciding the prices of their products (because it would undermine competition). In other situations cooperation is permissible. For instance, two neighbouring countries may be engaged in a situation of potential conflict with each other, each spending enormous amounts on defence. They always have the option of cooperatively reducing their conflict and improving their welfare. But the availability of the option of cooperation does not mean that the players will necessarily choose it over a non-cooperative game (the preceding example should make it clear). It is because there may be issues of reputation and commitment—one player may not expect to obtain a credible commitment towards cooperation from another. This is where social or political institutions come in. Our aim here is not to get deeper into these aspects, but just note that roles of reputation and institutions are studied in the framework of game theory.

Economic Facts and Insights

- Game theory deals with strategic behaviour in a situation of interdependence among individuals, species or organisations, that is, in a situation where the behaviour of one party affects the well-being of another.
- Social welfare is not maximised at the Nash non-cooperative equilibrium.
- A non-cooperative game does not achieve the best for the players.
- Game theory illustrates that cooperation is valuable to the society.
- Most economic situations in which strategies are important are not zero-sum in nature.
- In a cooperative game equilibrium, players are better off than in the corresponding non-cooperative game equilibrium.

EXERCISES

13.1 In the major cities in India, getting water supply by illegally using booster pumps attached to the city water line is not an uncommon practice. Consider two neighbouring households *A* and *B*. If neither switches on the pump, each gets 10 units of supply. If both turn on the pump, each gets 7 units (implicit here is that the electricity cost of using the pump in terms of buying water is 3 units). But if one switches on while the other does not, the former gets 15 units and the latter gets 5 units. Construct the payoff matrix. Using Nash equilibrium, what does the game theory predict in terms of behaviour towards switching on the booster pump?

13.2 There are two players in the game *P1* and *P2*. *P1* has three strategies: (a_1, a_2, a_3). *P2* also has three strategies: (b_1, b_2, b_3). Their payoffs are given in the following table. Is this a zero-sum or a non-zero-sum game? Which strategy combination is a Nash equilibrium?

P2	P1		
	a_1	a_2	a_3
b_1	(–2, 2)	(–2, 2)	(–2, 2)
b_2	(2, –2)	(1, –1)	(2, –2)
b_3	(–2, 2)	(–2, 2)	(–2, 2)

13.3 There is a game, called the *Battle of Sexes*. The story of this game is that a man and his wife are deciding how to spend the evening. There are two options: go to a movie or go to an Odissi dance programme. For the husband, the movie is preferable to the dance programme, whereas the wife prefers the dance programme to the movie. However, both prefer to be together, irrespective of whether they go the movie or to the dance programme. The payoffs are as follows:

Wife	Husband	
	Movie	Dance
Movie	(5, 3)	(0, 0)
Dance	(0, 0)	(3, 5)

Show that in this game there are two Nash equilibria. Which are they?

13.4 Yusuf and you play the following game. Each one of you hide a one-rupee coin in your palm. Both of you choose whether you want head or tail up inside your palm. Then you open your palms at the same time. If there is a matching of head-head or tail-tail, you win, meaning that Yusuf must give his one rupee to you. If there is no matching (that is, you have head up, while Yusuf has tail up or vice versa), then Yusuf wins (you give one rupee to Yusuf). Construct the payoff matrix and show that there is no Nash equilibrium in this game.[2]

13.5 Suppose there are two multinational firms A and B, contemplating to outsource a component of their production process to a small developing economy: Sunderland. If both enter, they recognise that competition between them in a relatively small market would raise local wages to the extent that outsourcing will not be profitable and their global profit will fall by 2 units. If one enters and the other does not, then the global profit of the former increases by one unit and that of the latter falls by one unit. If none enter, their profits are unchanged. The following summarises and illustrates the payoff matrix.

Firm B	Firm A	
	Enter	Don't enter
Enter	(–2, –2)	(1, –1)
Do not enter	(–1, 1)	(0, 0)

[2]Such a game is called a penny-matching game.

In game theory, it is an example of what is called a *Hawk-Dove game*. The Hawk and the Dove respectively stand for an aggressive and a passive strategy. In our example, 'enter' is the Hawk strategy and 'do not enter' is the Dove strategy. The Hawk–Dove game has the property that if both players play Hawk, then both are damaged. If both play Dove, then both are better off. If one plays Hawk and the other plays Dove, then the Hawk benefits and the other loses, compared to both playing Dove. Is there any Nash equilibrium in this game? If yes, how many and what are they?

13.6 Construct two hypothetical examples (not any extensions of examples given in the text or exercises) relating to real-word situations in which the number of players is greater than two.

14 Monopoly

CONCEPTS

- Imperfectly Competitive Market or Imperfectly Competitive Market Structure
- Monopoly
- Patent or Patent Rights
- Patent Life
- Natural Monopoly
- Cartel
- Price Maker
- Synergy
- Price Discrimination
- Arbitrage
- Marginal Cost Pricing
- Two-part Tariff
- Average Cost Pricing
- A Fair Rate of Return Method

In Chapter 10, we studied a perfectly competitive firm and a perfectly competitive market. The central feature of the latter is that no single firm has any market power. All other forms of market structure in which some firms or economic agents have some market power go under the name of **an imperfectly competitive market** or **an imperfectly competitive market structure**. Depending on the specific features, there are different types of imperfectly competitive markets such as monopoly, monopolistic competition and oligopoly. In this chapter we will study monopoly.

'Mono' means 'one' and 'poly' means 'seller'. Thus 'monopoly' means one seller. By definition, there is little competition in a monopoly market. It is the opposite of a perfectly competitive market in which there are infinitely many sellers. It is implicit that entry into a monopoly market is very difficult or very costly (otherwise, monopoly cannot last for long in the first place).

The term 'monopoly' should be understood in the context of a given geographical location and time. In India, before liberalisation in the power sector got underway in the nineties, the generation, transmission and distribution of electricity were in the hands of State Electricity Boards (SEBs). These were monopolies in the respective states. In providing air travel inside the country, Indian Airlines had a monopoly for a long time. Both are examples of 'state monopolies'. BCCI (Board of Control for Cricket in India) is a monopoly in providing international-level cricket games involving the Indian team. It generates revenues from various sources—the public who go to see these matches by purchasing tickets, the sale of TV rights and commercial sponsorships—and guarantees money for the Indian team's participation abroad. In 1998–1999 its profits were about Rs 8.37 crore.

As another instance, you hear people use the term 'xeroxing' to mean using a photocopying machine. Xerox Corporation is an American company, which discovered the plain-paper photocopying machine in 1959. In the sixties, it was the only company that manufactured and sold plain-paper photocopying machines.[1] This is an example of a private monopoly.

HOW MONOPOLY ARISES

There are several alternative reasons behind an emergence of monopoly in a market.

(*a*) The government gives licence to only one company for providing a product or service in a given locality or space. For instance, till 2002, VSNL (Videsh Sanchar Nigam Limited) had monopoly in India in providing international telephony service.

[1] The plain-paper photocopy machine has been considered as, by far, the most successful commercial product in history. Today there are many well-known companies, besides Xerox, in the world market that produce photocopying machines, for example, Canon, Mita, Panasonic, Ricoh, Royal, Sharp and Toshiba. Many fax machines also have copying capabilities.

(*b*) Big private companies—typically in developed countries—engage in research and come up with new products or new technology in producing an existing product. As a reward for their risk of investing in research, they apply to their government for a patent, which is an official recognition that they are the originators of the new product or technology. If a firm is granted the patent, no one else can use their technology without obtaining licence from them. Hence the firm can exercise monopoly power. Xerox was an example in the sixties.

A patent or what is called a **patent right** is not granted forever. It is valid only for a certain number of years (after which other firms can freely copy the technology if they can). This period is called **patent life**. In most developed countries, the patent life varies between 15 to 20 years. In Australia it is 20 years; in the US it is currently 17 years. In India patent laws were relatively weak. But recently, because of our commitment to WTO, they have become stricter. See Clip 14.1 on patent laws in relation to the world economy and India.[2]

Clip 14.1: Patent Laws

Most developed countries have comprehensive patent laws. During the patent life the patent holder can sell licence to other firms for using its technology (legally). Typically, the licence is sold to firms who operate in markets other than where the patent holder operates, for example, in a different country. The enforcement of patent law is also strict in developed countries. A patent holder can take to court some other firm, who may be using its technology without a license, and get a fairly quick decision.

In India, the patent law and its enforcement, till very recently, were rather passive. This is because research and development, discoveries and inventions were not in the focus of Indian firms. Barring a few exceptions, we generally imported technology from abroad.

The most important patent legislation in India is the Indian Patent Act of 1970. It provided that any invention of a new product or a process of production, which is useful and not obvious, is patentable. But it explicitly did not allow product patents in the drug and food sector.

Recently, however, as an obligation of being a member of WTO (World Trade Organization), India and other countries had to revamp their patent laws. In India, a major amendment to the Patent Act of 1970 was done in 1999,

(continued)

[2]As a specific example, an international drug company called Eli Lilly had a patent on a very widely used anti-depressant called *Prozac*. This patent expired in 2003.

by which both product and process patents were allowed in the food and drugs sector. The amended law has become operational since 2005.

Before the law came into effect and even now, there was/is a fear in India that, because of our being a member of WTO, we are forced to honour patent protection. As a result, particularly in the drug sector, Indian companies will no longer be able to sell many essential drugs at affordable prices. Once multinational companies start to sell them, the drug prices are going to skyrocket, and many poor people will be denied access to these drugs.

However, (see Clip 14.2, which reproduces the editorial column of *The Times of India*, 26 March 2005) it is unlikely that drug prices will soar up. First, the costs of drugs for minor ailments like headaches and stomachaches are unlikely to change. Second, patent right protection will, to some extent, reduce the profits made by Indian producers, who now have to pay royalties to multinationals, rather than increase the drug prices. Finally, and perhaps the most significant, it will punish the 'copycats' (as it should) and encourage domestic research.

(*c*) An expiry of a patent right does not mean that other firms possess the technology automatically. Sometimes, a firm is able to maintain its technical and organisational superiority, which is hard to be copied by others. Thus technical superiority may be a source of monopoly power. For example, IBM (International Business Machines, an American firm) had virtual monopoly in the computer business for a fairly long time.

(*d*) The production processes of certain products have large economies of scale relative to demand, meaning that the average cost curve is downward sloping up to a very high level of output relative to market demand. In such a case, producing a small amount entails high average cost, whereas producing a large amount offers a fairly low average cost. If many firms enter the industry, each one will have to produce a relatively small amount, incurring a high average cost above the market price and, therefore, will be unable to sustain itself in the market. As firms exit from the market, competition among the surviving firms will continue to eliminate the relatively smaller size firms. In the end one firm survives. Once the only firm is established in the market enjoying a low average cost, it is fairly immune from entry by others as potential new entrants are typically small in size. In other words, in the presence of strong economies of scale, monopoly emerges 'naturally'. Such a monopoly is called a **natural monopoly**. Providers of utilities like electricity and water are typical examples. For instance, running two sets of electric poles, wires and cables for a given area will be much costlier than running one.[3]

[3]Until a few decades ago, natural monopolies were present in providing telephone services. But this has changed because of new technologies.

Clip 14.2: How will Our Compliance with the New Patent Law under the WTO Regime Affect Us?

16 Saturday, March 26, 2005

EDIT PAGE

A THOUGHT FOR TODAY

It's a major achievement.

— Rupchand Pal

Seize the Day

New patents law an opportunity, not a threat

When the WTO was created 10 years ago, members like India decided to protect intellectual property rights (IPRs). Stripped of jargon, this meant that inventors would get some monopoly over the fruits of their creativity: Copies would carry royalty price tags, and each monopoly would last 20 years. That seemed to make sense. If copycats stole every new idea, no inventor would ever make profits, so they wouldn't work. That's why capitalism, which thrives on competition, grudgingly grants patents and copyrights — a limited kind of monopoly — for original stuff. But monopolies have problems: For example, who's to judge what a new medicine or technology is worth? Or, if a poor nation has an epidemic, is it OK to make cheap copies of costly medicines? Ten years later, India has taken its call on these issues. The result is a new law on patents, which makes big changes to the 1970 law. The older legislation, which allowed copying if the manufacturing process was different from the original, boosted copycat drugs, IT and manufacturing businesses. This helps to explain how India became the fourth largest manufacturer of medicines, the leading exporter of generics, and how it cut prices of AIDS drug-cocktails by 98%. The new regime will make life tougher for copycats, but amendments inspired by the Left parties soften the blow.

Most of these amendments relate to drugs, the main area of concern. Taken together, they mean that patients don't need to worry about soaring medical bills for a few years. In any case, costs of most drugs used for minor illnesses or headaches are unlikely to change. Some copycat drugmakers are unhappy about the prospect of having to pay royalties to multinationals, but experience shows that these payments are negotiable: GlaxoSmithKline once demanded 40% licence fee for an AIDS cocktail from South Africa, but finally settled at 5%. If there's a health emergency, New Delhi has powers to scrap patents or ask foreign drugmakers to sell stuff cheap. However, if you look beyond the scare-stories, prospects begin to look very good. With intellectual property protection, India becomes a very attractive destination for research. A 'mail box' containing patents applications over the last 10 years has about 8,500 requests, and 1,500 of those are from desi companies. Sceptics say because India's research budgets are a small fraction of American and European ones, we'll never catch up. That is wrong, simply because local costs too are a fraction of western ones. India's patents law is now one of the best in developing countries. It's an opportunity, not a threat.

Clip 14.3: OPEC and the World Oil Market

Organization of Petroleum Exporting Countries (OPEC) was constituted in 1960 by Iran, Iraq, Kuwait, Saudi Arabia and Venezuela. Qatar joined it in 1961, followed by Indonesia and Libya in 1962, United Arab Emirates in 1967, Algeria in 1969 and Nigeria in 1971. Currently there are two other countries in OPEC, namely, Ecuador and Gabon. The aim of the OPEC countries is to set production quotas so as to manipulate the price of petrol in the world market.

Besides the OPEC, there are other major producers of oil. For example, America was and still is a big producer of oil. But its consumption exceeds its production and it imports oil. India also produces oil and is an importer.

The oil shortage of the seventies motivated many other countries to explore oil. By mid-eighties there were other countries, who used to be importers of oil earlier but had become exporters, for example, Mexico, The Netherlands and Russia.

(*e*) Sometimes, firms retain their individual identity but they coordinate their outputs and pricing policy so as to act as a monopoly. This is called a **cartel**. The OPEC (Organization of Petroleum Exporting Countries) in the seventies is a prime example of a cartel that led to virtual monopoly in the world market for oil.[4] See Clip 14.3 for how the constitution of OPEC changed over time.[5]

PROFIT-MAXIMISATION

For any firm, profits are equal to the difference between total revenues and total costs. The cost structure facing a monopolist is similar to that of a competitive firm because like a competitive firm, it has to pay for the use of many inputs, for example, labour and raw material, and assuming that these inputs are used in other sectors of the economy as well, a monopoly has to buy these inputs competitively.[6] Thus the concepts like total cost, average cost and marginal cost, and their general shapes apply to a monopoly firm as well. Moreover, in a monopoly industry there is no issue of entry and exit (as long as the monopoly power continues). Hence, except that there are differences between the short-run and

[4]There are also other reasons for monopoly or near-monopoly, for example, mergers and acquisitions. In the early nineties, in the tea industry, Brooke Bond and Lipton merged and subsequently they merged with Hindustan Lever. This left out Tata, another large tea firm.

[5]It is not an exhaustive list of factors leading to monopoly power. For instance, some firm may be the sole possessor of a scarce input and, thereby, enjoy monopoly power in the product market. The diamond syndicate of South Africa is an example.

[6]That is, a monopoly will have market power in the product market but not necessarily in the input markets.

long-run cost curves, there is no essential behavioural difference of a monopoly industry between the short run and the long run.[7] Thus, typically, in the monopoly analysis it is not necessary to be particular about differentiating the two time periods. We just say average, marginal and total costs without referring them to the short run or long run.

The revenue structure facing the monopolist is, however, quite different from that of a competitive firm. A perfectly competitive firm is very small compared to the market and thus a price-taker. This is obviously not true for a monopoly. Being the only firm serving the market, it has market power and it is a **price maker** so to speak. This important difference implies that the total revenue changes with output in a manner which is different from what happens for a perfectly competitive firm. For the latter, as output increases, the price remains unchanged. But a monopoly firm faces the entire market and, therefore, as it increases or decreases its output, it cannot expect that the market price will remain unchanged—price will change according to what consumers are willing to pay along the demand curve. The monopolist has to take this into account. Hence, the market demand curve is a constraint facing a monopoly firm.

An example of a market demand schedule is given in Table 14.1. As the monopolist faces this demand schedule by itself, it means that if, for example, she wants to sell 5 units, she (the monopolist) must charge a price equal to Rs 9. The reason is the following. At any price higher than this, she will be able to sell only less than 5 units. Further, she can always sell 5 units by charging Rs 9 each since 5 units are demanded at the price = Rs 9. Thus there is no reason to sell 5 units for any price less than Rs 9. By similar argument, if the monopolist wants to produce and sell 6 units, the price charged will be Rs 6, and so on.[8]

Table 14.1 A Demand Schedule

Price (in Rs)	*Quantity Demanded (units)*
3	7
6	6
9	5
12	4
15	3
18	2
21	1
24	0

[7]This does not imply that monopoly power continues indefinitely. For example, if the monopoly power is obtained by virtue of a patent, it expires with expiration of the patent. The point here is that as long as there is monopoly, there is little analytical difference between short run and long run in terms of output and price determination.

[8]Thus it is not true that the monopolist can charge any price for its product at its own sweet will—unlike what many believe. It could have, only if the demand curve was totally vertical, that is, there were absolutely no substitutes available. But for most products there is almost always some available substitute.

Table 14.2 *TR*, *AR* and *MR* under Monopoly

Output (1)	*Price (Rs) (2)*	*TR (Rs) (3)*	*AR (Rs) (4)*	*MR (Rs) (5)*
0	24	0	–	–
1	21	21	21	21
2	18	36	18	15
3	15	45	15	9
4	12	48	12	3
5	9	45	9	–3
6	6	36	6	–9
7	3	21	2	–15

We now write 'Output' in place of 'Quantity Demanded' and present the same demand schedule with the output listed in increasing order from 0 to 7 in the first column of Table 14.2, while the second column lists the price. These columns also represent the demand schedule given in Table 14.1.

Next we multiply output by price and get the total revenue (*TR*). This is given in column (3). Dividing *TR* by output gives average revenue, *AR*, since, by definition, *AR* = *TR*/output. This is shown in column (4). *TR* being equal to price × output, we have *AR* = price × output/output = price, that is, *AR* is always equal to price.[9] Thus the entries in column (4) are same as those in column (2). Remember from Chapter 10 that marginal revenue (*MR*) is defined as the addition to the total revenue from one extra unit sold. The last column gives the *MR* schedule.

We note the following properties of the three revenue concepts:

(*a*) *MR* decreases with the output. Initially it is positive and after a certain level of output it turns negative.
(*b*) *TR* increases or decreases with output, as *MR* is positive or negative.
(*c*) (*a*) and (*b*) imply that *TR* first increases and then it decreases with output. Thus, the graph of *TR* against output, the *TR* curve, rises initially and then falls. It means that, if output is measured on a continuous scale, *TR* reaches maximum when *MR* = 0.
(*d*) Since *AR* = price, the *AR* curve is always the same as the demand curve facing the firm.
(*e*) Except for the first unit, at all other levels of output, *MR* < *AR*. This follows from the relationship between the 'average' and the 'marginal', that is, if the 'average' is falling (or rising), the 'marginal' is less (or greater) than the 'average'.
(*f*) Consider the range of output over which *MR* > 0. Since the discrepancy between *MR* and *AR* arises because the average revenue is falling with quantity, the rate of decrease in the *AR* determines the proportional difference

[9]This is true except when output is zero. At zero output, *TR*/output = 0/0, therefore, AR is not defined.

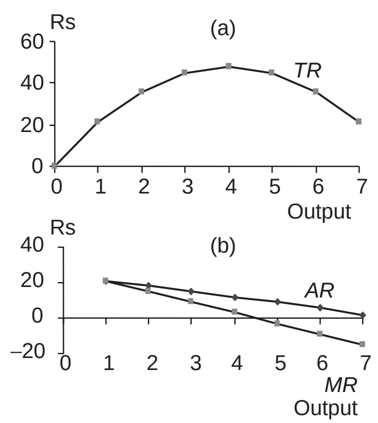
between *MR* and *AR*. In turn, the rate of decrease in *AR*, meaning the rate of decrease in price as quantity demanded increases, depends on the price elasticity of demand. Hence, the higher the price elasticity of demand, the flatter is the demand curve and thus the less is the decrease in the price or *AR*. It then follows that the higher the elasticity, the smaller is the ratio, *AR*/*MR*.[10]

Panels (a) and (b) of Figure 14.1 respectively graph the *TR*, *AR* and *MR* curves corresponding to Table 14.2. The *TR* curve is inverse U-shaped as *TR* initially increases and then decreases with output.

Figure 14.2 depicts a smooth, hypothetical *TR* curve and the associated *AR* and *MR* curves. As you notice, *TR* reaches its maximum when $MR = 0$.

There is a relationship between the sign of the *MR* and the magnitude of the price elasticity of demand. In the range of output where $MR > 0$, an increase in output or quantity sold results in an increase in the total revenue. This is same as saying that a decrease in price (as quantity sold increases) is associated with an increase in the total expenditure (= total revenue). Now, go back to Chapter 2 and check the relationship between a price change, elasticity and the change in total expenditure—a decrease in price is associated with an increase in total expenditure if and only if the price elasticity is greater than unity. Thus, $MR > 0$ means that the price elasticity is greater than one. Similarly, $MR < 0$ means that the price elasticity is less than one.

Figure 14.1 *TR*, *AR* and *MR* Corresponding to Table 14.2



[10]Let e_p denote the price elasticity of demand. By definition, $e_p = -\frac{p}{y}\frac{dy}{dp}$, where y is the quantity sold. Thus $1/e_p = -\frac{y}{p}\frac{dp}{dy}$. For the monopoly firm, $MR = d(TR)/dy = d(py)/dy = p + y\frac{dp}{dy} = p\left(1 + \frac{y}{p}\frac{dp}{dy}\right) = AR\left(1 - \frac{1}{e_p}\right)$. Hence, $AR/MR = \frac{e_p}{e_p - 1}$. As a special case, in the case of a perfectly competitive firm, $e_p = \infty$. Thus $AR/MR = \lim_{e_p \to \infty} \frac{e_p}{e_p - 1} = 1$, that is, $AR = MR$ or $p = MR$.

Figure 14.2 General *TR*, *AR* and *MR* Curves

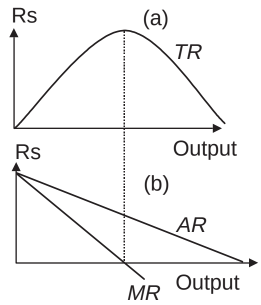
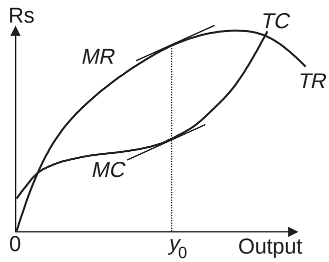


Figure 14.3 Profit-maximisation by a Monopoly Firm



Profit-maximising Rule

Once we know the shapes of revenue and cost curves, we are in a position to derive the profit-maximising condition. Figure 14.3 depicts hypothetical *TR* and *TC* curves. The vertical difference between the two curves measures profit. Where is the profit maximised? Just as for a competitive firm, analysed in terms of Figure 10.3, the profit-maximising level of output is y_0, at which the slope of the *TR* curve equals the slope of the *TC* curves. These slopes are respectively equal to the *MR* and the *MC*. Thus we have:

Profit-maximisation Condition:

$$MR = MC.$$

This is indeed a very general condition of profit-maximisation by any firm.[11]

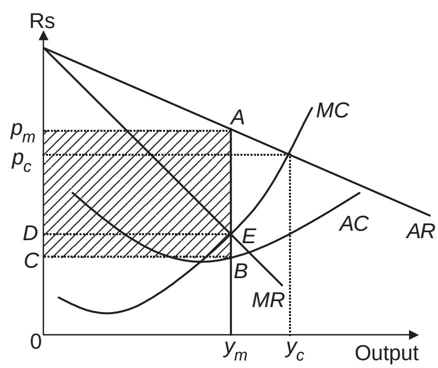
[11] Recall that for a competitive firm $MR = p$ and thus the condition $p = MC$ is a special case of the general condition $MR = MC$.

The $MR = MC$ condition is quite intuitive. At very low level of output, $MR > MC$. Thus adding output by one unit generates additional revenues greater than the additional cost. Thus the firm will obtain more profits if it increases its output. On the other hand, at a very large level of output, $MR < MC$. This means that if the firm reduces output, the savings in cost will be greater than the revenues lost and hence profits will be higher. Thus, profit reaches its maximum at the level of output where $MR = MC$.[12]

Note that at the profit-maximising output, $MR = MC$, which together with $MC > 0$ implies that $MR > 0$. We have discussed earlier that $MR > 0$ if and only if the price elasticity of demand is greater than one. Hence, at the monopoly equilibrium the quantity and price are chosen such that the price elasticity of demand exceeds unity.

A monopolist's profit-maximisation is typically illustrated in a diagram like Figure 14.4. It has AR and MR curves drawn as straight lines for simplicity, but they need not be. It also has the AC and MC curves. The profit-maximisation condition $MR = MC$ is met at the point E, where the MR and the MC curves intersect. The corresponding level of output, y_m, is the profit-maximising or the optimal output. The firm also chooses the price. The monopoly price, say p_m, is the one at which the quantity demanded is equal to y_m. If we extend the vertical line passing through E, its intersection with the AR curve (which is the demand curve) marks p_m. The profit per unit of output is the difference between AR and AC at the optimal output. This is equal to the vertical distance AB. Thus, total profit = $AB \times$ quantity produced = $AB \times 0y_m$ = the area $ABCp_m$. This is shown by the shaded area. (Ignore for now the output level marked y_c and the price p_c.)

Figure 14.4 Profit-maximisation by a Monopolist Once Again



[12]In analysing profit-maximisation by a competitive firm, we also had a condition that MC must be increasing in output. The equivalent of this condition here is that the 'MR curve must be falling faster than the MC curve'.

Sometimes the monopoly equilibrium is shown in terms of a simpler diagram like Figure 14.5. Suppose that the firm's average cost is constant (independent of the output). This implies that the marginal cost is constant and equal to the average cost, that is, *AC* = *MC*. The monopoly output is shown at the intersection of the *MR* curve and the horizontal *MC* curve. The monopoly price is read from the *AR* curve and the profit is indicated by the rectangle $CDFp_m$.

Can a monopolist incur a loss? Yes, but this is unlikely since it is the only firm supplying the product for which, by definition, there are no close substitutes available.

NUMERICAL EXAMPLE 14.1

A monopoly firm faces the demand schedule given in Table 14.3 and the total cost schedule given in Table 14.4. What is its profit-maximising output and price?

The demand schedule is the *AR* schedule. Multiplying the quantity demanded with price gives the *TR* schedule. The *MR*s are derived as additions to the *TR*. The *AR*, *TR* and *MR* schedules are given in Table 14.5. From the total cost schedule, the *MC* schedule can be derived; this is given in Table 14.6.

Figure 14.5 Monopoly Equilibrium when *AC* = *MC*

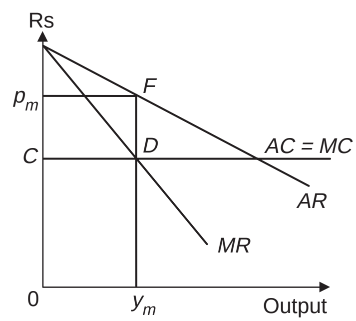


Table 14.3 Demand Schedule (Numerical Example 14.1)

Price (in '000 Rs)	*Quantity Demanded*
5	8
6	7
7	6
8	5
9	4
10	3
11	2
12	1

Table 14.4 Total Cost Schedule (Numerical Example 14.1)

Output	*Total Cost (in '000 Rs)*
1	3
2	7.5
3	12.5
4	18.5
5	25
6	32
7	39.5
8	47.5

Table 14.5 *AR*, *TR* and *MR* Schedules (Numerical Example 14.1)

Output	*AR*	*TR*	*MR*
1	12	12	12
2	11	22	10
3	10	30	8
4	9	36	6
5	8	40	4
6	7	42	2
7	6	42	0
8	5	40	–2

Table 14.6 *MC* Schedule (Numerical Example 14.1)

Output	*MC*
1	3
2	4.5
3	5
4	6
5	6.5
6	7
7	7.5
8	8

We see that the condition $MR = MC$ is met at output = 4. Thus it is the profit-maximising output. Since this amount is sold in the market, from Table 14.3, the corresponding market price (that is, the monopoly price) is Rs 9,000.

Mathematically Speaking

Deriving a Monopolist's Profit-maximising Condition

Let the industry demand function be given by $y = f(p)$, where y is the quantity sold and p is the price charged. We can 'invert' this function and write $p = g(y)$. We can also interpret y as the quantity produced. The function $g(y)$ is called the inverse demand function. For example, if $y = a - bp$ is the demand function, the inverse demand function is given by

$$p = g(y) \equiv \frac{a-y}{b}.$$

The total revenue = quantity produced × price = $yg(y) \equiv R(y)$. This is the total revenue function, and $MR = R'(y)$ is the derivative of the function $R(y)$. Similarly, if we let the total cost function be denoted by $C(y)$, we have $MC = C'(y)$.

The monopolist's profit = $TR - TC = R(y) - C(y) \equiv \pi(y)$. This is maximised with respect to y. The first-order condition is

$$\pi'(y) = R'(y) - C'(y) = 0, \text{ or}$$
$$R'(y) = C'(y), \text{ i.e., } MR = MC.$$

MONOPOLY VERSUS PERFECT COMPETITION

In Chapter 12, we learnt that at the equilibrium under a perfectly competitive market structure, social surplus is maximised (under certain conditions). It is, therefore, imperative to know how other market structures compare with perfect competition. The following are the general features of monopoly in comparison to perfect competition.

(*a*) In perfect competition, profit-maximisation leads to a supply curve that tells us how much a firm produces at different market prices that are exogenous to the firm. In monopoly, however, the firm decides the output and the price. There is no question of the optimal level of monopoly output at different prices. Hence, there is no supply curve as such under monopoly.

This does not mean, however, that demand and supply forces do not interact. Shifts in the demand (AR) curve or in the MC curve will affect a monopolist's choice of output and price.

(*b*) For a monopoly, $p > MR$, and it selects an output level where $MR = MC$. These two relations imply that $p > MC$, which is verified in Figure 14.4. At the profit-maximising output y_m, the price charged = $0p_m$, which is greater than the marginal cost = $0D$. It is the market power possessed by the monopolist, which enables him to charge a price higher than the marginal cost. In comparison, recall that the price is equal to the marginal cost in perfect competition. Thus, we can say that a monopoly charges 'too high a price' for its product. Moreover, the monopoly price being higher than the competitive price, it follows that along a given demand curve, less is sold and, therefore, less is produced under monopoly than under perfect competition. In summary, we can then say that the monopolist produces less and charges a higher price as compared to that in perfect competition. This is the fundamental behavioural difference between monopoly and perfect competition.

(*c*) Recall from Chapter 12 that the social surplus is maximised at that level of quantity where price equals marginal cost and this condition holds at the competitive equilibrium. The monopoly price being different from the marginal cost implies that the social welfare or surplus is less under monopoly than under perfect competition.

NUMERICAL EXAMPLE 14.2

Let Table 14.3 describe the total demand schedule facing a competitive industry, and let Table 14.6 be the supply schedule of this industry (given price = marginal cost). How much will be produced and what will be the price of the product in equilibrium? How does this equilibrium compare with the corresponding monopoly equilibrium worked out in Numerical Example 14.1?

Plotting the demand and supply schedules yields that the quantity demanded is equal to the quantity supplied at price = Rs 7,000. The equilibrium quantity produced is 6 units. In the Numerical Example 14.1, the monopoly price is Rs 9,000 and the monopoly output is 4 units. Hence, in the monopoly equilibrium the price is higher and the quantity sold is less.

MERITS OF MONOPOLY

The points (*b*) and (*c*) in the earlier section summarise what is wrong with a monopoly market and form a rationale for restricting monopoly power in practice. Indeed these features are sometimes played up sufficiently high so as to generate a strong negative public sentiment against a monopolist. Some people even think that a monopolist is simply an evil that exploits the public.

But before rushing to such a conclusion some positive aspects of a monopoly should be noted too.

Suppose that initially there are two firms in an industry and both are somewhat inefficient. Their *MC* curves are at a high level and consequently they charge a higher price and produce less than what they would if the *MC* curves were at a lower level. They realise, however, that if they merge with each other—and thereby become a monopoly—they can reduce their costs. For instance, one firm may have excellent technical manpower but may not have good marketing skills, whereas the other may not have good technical manpower but possesses superior marketing knowledge. By merging, the resulting monopoly firm's *MC* curve will be at a lower level and thus it will be more efficient. In the business language it is called **synergy**. This, by itself, will induce the monopoly to charge a price, which is less, and produce a quantity, which is greater than what it would be if both firms were competing with each other.

Another major benefit from encouraging monopoly to a certain extent is that monopoly power and profits provide incentives for inventions and innovations. In reality, these activities are very risky propositions. Often they materialise from individual efforts and persistence. Why would someone invent a product if he/she is not allowed to enjoy monopoly profits for a few years? This is indeed the reason behind granting patents.

Finally, recall the natural monopoly case. If there are huge economies of scale, the presence of a single (monopoly) firm will imply lower unit costs as compared to the situation where there are many small firms. Hence, there is a natural benefit arising from a natural monopoly.

PRICE DISCRIMINATION

In the real world, there are examples where, for the same product or service, a producer charges different prices from different consumers. For the same movie, a cinema hall owner prices the afternoon show lower than the evening show. Utility companies like telephone and electricity fix different rates for household and industrial (business) customers. In the international markets, the same company, for the same product, may charge different prices from consumers in different countries. For instance, famous Japanese electronics companies typically price their products lower in the American market than in the Japanese market itself.

This practice is called **price discrimination**, meaning different prices being charged in different markets for essentially the same product. Why does a firm do it? It is because the demand conditions may vary across markets. Typically, the demand for an evening cinema show is higher than that for an afternoon show. Therefore, the movie hall charges a higher price for an evening show. In what follows, we make this observation more precise by (*a*) identifying the conditions under which a firm can price discriminate, (*b*) deriving the profit-maximising rule under price discrimination and (*c*) deducing from this rule the exact nature of the difference between markets that tell us why in one market the price charged may be higher than in another.

Suppose there are two markets a monopolist is serving, say market A and market B. Let p_A and p_B denote the prices charged in these two markets respectively. If $p_A = p_B$, there is no price discrimination. Suppose there is price discrimination and $p_A < p_B$. Then the consumers in market A have an incentive to buy the product in market A at a lower price, sell it themselves in market B for a higher price and make a profit. In economics, this is an example of what is called an **arbitrage** behaviour—meaning the act of weighing the cost or price of a product/service in two different markets, buying or selling accordingly and making a profit. If the consumers in market A can actually do it, that is, if reselling from the low-price to the high-price market is feasible, the price of the product in market A will keep rising and that in market B will keep falling till $p_A = p_B$. Therefore, price discrimination is feasible only if there is no opportunity to resell the product from a low-price market to a high-price market. We assume this.

In reality, the consumers are often not allowed to resell. In the movie ticket example, an afternoon-show ticket is no good for an evening show. In case of differential electricity consumption rates for household and business purposes, it is illegal for businesses to buy electricity from households and vice versa. Also, the physical distance between two markets may be very long so that the transportation cost of goods may be high enough to deter any profitable arbitrage by consumers.

Profit-Maximising Condition

We now analyse how a monopoly firm maximises profits given that price discrimination is feasible. Denote the total revenues generated from the two markets by TR_A and TR_B. The grand total of revenues from the two markets equal $TR_A + TR_B$, and the firm's profit can be expressed as $TR_A + TR_B - TC$. We can indeed 'break up' the firm's profit-maximisation problem into two sub-problems: (*i*) how should the total amount produced be allocated between the two markets so as to maximise the grand total revenue and (*ii*) what is the optimal level of the (total) output? Price discrimination essentially relates to problem (*i*) as it pertains to selling the product in the two markets. So the question before us is—how should a monopoly firm allocate its sales between the two markets, given that the total amount of the quantities sold in the two markets (= output) is fixed?

The answer to this question is simple—allocate the sales such that $MR_A = MR_B$. That is, the total revenues across more than one market are maximised if the marginal revenues are equalised across the markets.

The logic behind this principle is straightforward. Suppose the sales in markets *A* and *B* are such that $MR_A > MR_B$. Then let the firm move one unit of the product from market *B* to market *A*. By doing this, it will lose revenues = MR_B in market *B* and gain revenues = MR_A in market *A*. Since $MR_A > MR_B$, it follows that the total revenues will be higher by this reallocation. Moreover, we know that the *MR* in any market falls with the quantity sold in that market; thus if the firm keeps shifting sales from market *B* to market *A*, MR_B will keep rising and MR_A will keep falling. But as long as $MR_A > MR_B$, this sales-reallocation process will continue to increase the total revenues. Hence, total revenues are maximised at that sales allocation where $MR_A = MR_B$.

Once we understand the answer to the problem (*i*), the answer to (*ii*) is an extension of the $MR = MC$ principle, namely,

$$MR_A = MR_B = MC.$$

This is the overall profit-maximising condition for a price-discriminating monopolist.[13]

Ranking of Prices

The $MR_A = MR_B$ condition can be used to derive the ranking of prices across the markets. Suppose the demand for the product is more elastic in market *A* than in

[13]If there are more markets like *C and D*, the profit-maximising rule is that $MR_A = MR_B = MR_C = \ldots = MC$.

market B. Recalling the relationship between AR/MR and the price elasticity of demand discussed earlier, it follows that $AR_A/MR_A < AR_B/MR_B$. Thus, $AR_A/AR_B < MR_A/MR_B$. By definition, $AR_A = p_A$ and $AR_B = p_B$. Moreover, $MR_A = MR_B$ means $MR_A/MR_B = 1$. Hence, the last inequality implies that $p_A/p_B < 1$, that is, the price charged in market A is less than that charged in market B. Remember that the demand in market B is less elastic. Thus, under price discrimination a monopolist charges a higher price in a market in which the demand is less elastic. Put differently, the ranking of prices across the markets is the inverse of the ranking of price elasticities across the markets.

We can now explain various price-discrimination phenomena in the real world. Afternoon movie tickets are cheaper because the demand for movies is more elastic in the afternoon than in the evening. Business customers are charged more for power than households because the demand for power by the former is less elastic.[14]

NUMERICAL EXAMPLE 14.3

A monopolist has already produced 9 units of output, which will be sold in two markets, A and B, between which price discrimination is feasible. The marginal revenue schedules in the two markets are given in Table 14.7. Suppose the monopolist allots 4 units to be sold in market A and 5 units in market B. Show that this does not maximise the monopolist's total revenue from the two markets combined. Should the monopolist sell more in market A and less in market B, or vice versa? What is the sales allocation, which will maximise its total revenues and what will be the prices charged in the two markets?

If 4 units are sold in market A and 5 units in market B, the marginal revenues are 15 and 7 respectively. These are unequal and thus cannot maximise total revenues. Moreover, since the marginal revenue is greater in market A than in

Table 14.7 ***MR* Schedules (Numerical Example 14.3)**

Output	MR_A	MR_B
0	–	–
1	24	11
2	21	10
3	18	9
4	15	8
5	12	7
6	9	6
7	6	4
8	3	2
9	0	0

[14] For various reasons, the Japanese market is more closed to foreign competition, compared to the American market. Thus, the Japanese firms face more competition in the latter market, which makes the demand for their product more elastic in the American market than in the Japanese market. Therefore, they benefit from charging a lower price in the American market than in the Japanese market.

market B, total revenues are higher if more (or less) is allocated to market A (or B). When 6 units are sold in market A and 3 in market B, the marginal revenues are equalised and all 9 units are sold. Hence, total revenues must be maximised at this allocation of sales. (Note that marginal revenues are also equal when 7 units are sold in market A and 6 in market B. But this cannot be the answer since the monopolist has only 9 units to sell in the two markets.) In order to determine the prices charged, we need to first derive the total revenue schedules in the two markets (by adding the respective marginal revenues). Given the total revenue in market A at 6 units and that in market B at 3 units, we then divide them by 6 and 3 respectively to get the average revenues; these are the respective prices. Computing these you can get $p_A = 16.5$ and $p_B = 10$.

REGULATING A PRIVATE MONOPOLY

As a monopolist charges a price not equal to the marginal cost, the monopoly output is not socially optimal, that is, a monopoly market fails to perform in the best interests of the society. Hence, there arises a case for regulating a monopolist.

More specifically, since the monopoly price exceeds the marginal cost and as a result it produces a quantity less than what is socially optimal, any policy or regulation that induces the monopolist to produce more is social-welfare improving.

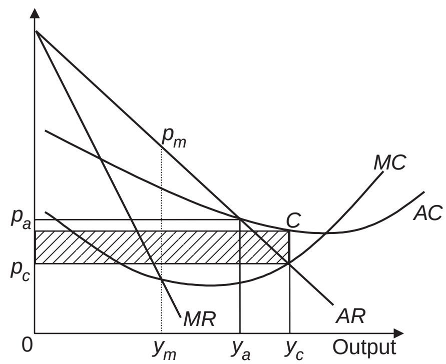
One idea is to control the price to be charged by a monopolist. In Figure 14.4, the competitive, social surplus-maximising price is p_c, equal to the marginal cost. Hence, if the regulatory body sets the price equal to p_c, the monopoly problem, from the perspective of social welfare, is solved. This is called a policy of **marginal cost pricing**. Of course, the monopolist is not happy with such a price control, as its profits are less compared to the situation where it can set price such that $MR = MC$.

However, the matter is not this simple. Suppose the monopoly market in question prevails because of a patent right. If, as a regular practice, the price is controlled after a new product or process is introduced into the market, it would act as a disincentive towards R&D (research and development) and innovation. Monopoly profit can be seen as returns to investment in R&D and innovation. Why would an entrepreneur take the risk of such investment if the returns are low? On the other hand, innovations improve product quality, bring new products and, therefore, constitute a fundamental source of enhancement of the standard and quality of life. Therefore, while $p = MC$ principle may be the best policy at a given point of time, after the product is available, from a long-term perspective such a policy of controlling monopoly price may not work in the best interests of the society.

Now consider an industry in which a monopoly does not arise out not due to R&D. A 'natural' candidate is a natural monopoly, as defined earlier, which is characterised by a decreasing average cost for a sufficiently long range of output. Figure 14.6 exhibits this case. Without any regulation, the natural monopolist produces y_m, at which the condition $MR = MC$ is met.

Suppose the regulator sets the price p_c, forcing the monopolist to produce y_c, such that price equals MC. However, there is a problem. That is, at this output, the

Figure 14.6 Regulating a Natural Monopolist



monopolist is incurring a loss as the average cost (= Cy_c) exceeds the price p_c. This is indicated by the shaded area. If such a situation persists, the monopolist will leave the industry. The product will not be available at all for consumption. This is bad news for social welfare. Put differently, regulating price equal to the marginal cost is not sustainable.

There, however, is an alternative. This is to set price equal to the marginal cost and provide a lump-sum subsidy to the monopolist, covering its loss. One way to make it operational to some extent is to charge a fixed fee from the consumers and a relatively low rate for actual usage. This is essentially what is called a **two-part tariff**. However, there is a difficulty here too. A relatively large amount of a fixed fee effectively discriminates against small users, who are typically the relatively poor.

The next best alternative is to set price equal to the average cost, shown as p_a in Figure 14.6. This is called **average cost pricing**. Compared to the monopoly equilibrium, the monopolist produces a higher quantity (equal to y_a) and the social surplus is greater. And the monopolist earns no more than its normal profits.

Indeed, such a regulation is widespread in many countries, especially for public utilities like electricity and water. In practice the price is not set exactly as is depicted in Figure 14.6. There are variations. Historically, a **fair rate of return method** has been used extensively. Essentially there is a regulatory body that assesses the unit cost of production by the firm. A 'fair' rate of return for investment by the monopolist is added to this price. The resulting price is the regulated price that the firm is entitled to charge.[15]

There is a problem with average-cost regulation also. That is, since the regulated price is made dependent on the cost of production, the monopolist has an incentive to inflate its cost in the eyes of the regulator. The cost shown to the regulator may very well exceed the true cost. On the other hand, it is very difficult for the regulatory body itself to have information on all transactions made by the firm and,

[15]In India, various state electricity regulatory commissions are in charge of regulating electricity tariffs.

therefore, on the precise nature of its costs of production. By jacking up the costs shown to the regulator, the firm is able to ensure a price higher than its true unit cost and, thereby, earn an abnormal profit. Social welfare is less than otherwise.

There is still another option—let the government do the business itself. A prime example in almost any country is postal service, which is typically subsidised. In India, rail-transport is another example of a subsidised government-run business, controlled by the Ministry of Railways. Interestingly there have been suggestions for establishing a tariff commission for railways, but the railways ministry, so far, has been opposed to it. See Clip 14.4 for a short review of the state of affairs of Indian railways.

A general point of this discussion is that regulating a monopoly because it charges a price higher than the marginal cost may not be desirable. Even if it is desirable, executing the regulation effectively is not easy.

Clip 14.4: The State of Indian Railways

Almost all of us have grown with Indian Railways (IR). We can easily remember our childhood excitement of looking at a train, counting its bogeys and some of our very first train journeys. Reflecting back over time, we can observe that the number of trains has increased substantially; services may have somewhat improved. But what is the IR's overall record in terms of meeting the need in the modern age of growing business and competing modes of transport? What is its financial state? Interestingly, until very recently, the picture was indeed saddening, but dramatic turnarounds seem to be happening now.

In 1999–2000, the losses over providing passenger service were estimated near Rs 4,100 crore. The cost of hauling a short-distance passenger train for 250 km was about Rs 75,485 whereas with 100 per cent occupancy of passengers, the revenues generated was about Rs 22,000; this amounts to a minimum loss of Rs 53,485 per 250 km. Of course, there has been a cross-subsidisation within the railway services—'profits' from providing freight services have partially made up for losses in the passenger segment.

How about the overall financial position? In 1998–99, the total working expense was Rs 27,835 crore, while the revenues were Rs 29,619 crore. The ratio of the two, called the 'operating ratio,' was 93 per cent. Of course, the higher the operating ratio, the greater is the degree of inefficiency. In 2001–02, this ratio rose to 96 per cent.

On the face of it, an operating ratio less than 100 per cent may not appear bad. But note that these calculations do not involve all normal operating costs. For instance, the cost of overdue equipment and railway tracks (they wear out over time) is not included. If these are included, it is a 'heavy loss making entity in normal commercial accounting terms.' Furthermore, IR reaches out to the central government and other sources for financing its investment and in 2001–02, it had to defer its interest payments to the central government.

(continued)

It is interesting to compare IR with China Railways. Over the period 1992–2002, China's railways network grew by 24 per cent (measured in kilometres of train tracks), whereas IR's network grew by 1 per cent. The Chinese investment in railways during the same period amounted to 85 million US dollars, compared to only 17 billion US dollars for India.

The concern over the future of IR led to an expert group study chaired by Rakesh Mohan. The Mohan Report was submitted in 2001. It noted that IR's freight market share has been consistently falling due to overpricing and poor quality of services. Further, IR is facing a financial crisis that needs to be addressed sooner than later. If 'business as usual' inertia continues, IR 'will be led to a fatal bankruptcy.' The Report recommended, among other things, a restructuring of the organisation, separation of railways from the government and at least 25 per cent reduction in manpower (as there is too much over-employment).

It is heartening, however, that a dramatic turnaround has occurred lately under the leadership of the current Railways Minister Mr Lalu Prasad Yadav. While detailed analysis of sources of this turnaround and their specific effects was not available at the time of writing (except that positive results have been achieved through better leverage of existing assets and considerable reduction of unit cost of operation), it is a fact that IR's market share of freight traffic, after a long-term decline, has improved. In 2005–06, the operating ratio was less than 85 per cent and that too without any hike in fares or freight rates.

References

1. Harral, Clell and Jit Sondhi. 2005. 'Comparative Evaluation of Highway and Railway Development in China and India 1992–2002,' Report submitted to the South Asia Energy and Infrastructure Unit, International Bank of Reconstruction and Development, Washington D.C.
2. Poddar, Saroj K. 2006. Welcome Address to the Conference on Private Partnership in Indian Railways, organised by FICCI, 16 October, 2006, New Delhi.

Economic Facts and Insights

- Board of Control for Cricket in India is a monopoly of its kind as there are no other rival institutions in India that provide competition in organising international level cricket games involving the Indian players.

(continued)

- In India, patent laws were relatively weak. But recently, because of our commitment to WTO, these have become stricter.
- Monopoly in a market can arise due to various reasons such as exclusive government licence, patents, technological superiority, and economies of scale.
- It is not true that a monopolist can choose any price arbitrarily because it is the sole producer of the product or service. The market demand curve acts as a constraint facing a monopoly firm.
- $MR = MC$ is a general condition of profit-maximisation of a firm.
- The monopolist charges a price higher than the marginal cost. This is because of the market power possessed by the monopolist.
- Compared to perfect competition, a monopolist produces less and charges a higher price.
- Since the monopoly price exceeds the marginal cost and as a result it produces a quantity less than what is socially optimal, any policy or regulation that induces the monopolist to produce more is social-welfare improving.
- Merits of monopoly lie in synergies exploited from joint activity, R&D and economies of scale.
- Under price discrimination, a monopolist charges a higher price in a market in which the demand is less elastic. Put differently, the ranking of prices across the markets is the inverse of the ranking of price elasticities across the markets.
- In case of a natural monopoly, marginal cost pricing would imply losses.
- Utilities are typically subject to a fair rate of return regulation by the government.
- Regulating a monopolist is not easy. There are theoretical and practical problems associated with any form of regulation.

EXERCISES

14.1 What are the different types of imperfectly competitive market structures?

14.2 What are patent rights?

14.3 Collect information on BCCI from the internet and other sources and write a short essay on how far BCCI is a private monopolist.

14.4 What is a natural monopoly?

14.5 Why is a monopoly firm a price maker?

14.6 'Because there is no competition, a monopoly firm can charge any arbitrarily high price for its product'. Defend or refute.

14.7 Since, by definition, a monopoly firm is the only seller of a product, if it wants to sell more, compared to a given quantity being sold, why would it have to charge a lower price to be able to do so?

14.8 The *MR* schedule of a monopoly firm is given below. Derive the *TR* and *AR* schedules.

Output	*MR (Rs)*
1	26
2	20
3	15
4	11
5	6
6	0
7	–5
8	–7

14.9 'For a monopoly firm, the *MR* may exceed the price'. Defend or refute.
14.10 'Average revenue is less than the price for a monopoly firm'. Defend or refute.
14.11 The higher the price elasticity, the ____ is the ratio of *AR* to *MR*. Fill in the blank and explain.
14.12 What is the condition of profit-maximisation by a monopoly firm? What is the logic behind it?
14.13 Why is it that at the monopoly equilibrium, the price elasticity of demand exceeds unity?
14.14 Compare price and industry output under perfect competition and monopoly.
14.15 Compared to a perfectly competitive industry, a monopoly produces ___ and charges a ____ higher price. Fill in the blanks and give reasons.
14.16 'At the monopoly equilibrium, social welfare is not maximised'. Defend or refute.
14.17 What is meant by synergy from a merger of two firms?
14.18 In the text we have outlined the benefit of synergy from a merger or an acquisition (briefly called M&A in the business language). Can you think of a demerit of M&A in terms of management and efficiency of a company?
14.19 What are the merits of monopoly?
14.20 What is meant by price discrimination?
14.21 What is the condition under which price discrimination is feasible?
14.22 What is the condition of revenue maximisation when a monopoly can price-discriminate between different markets? Give reasons.
14.23 Under price-discriminating monopoly, how are prices across markets correlated with the price elasticities of demand? Give reasons.
14.24 What is marginal cost pricing?
14.25 Explain why a marginal-cost pricing regulation may lead a monopoly firm to incur losses.
14.26 Under marginal cost pricing, can a monopoly earn any abnormal profits? Give reasons.
14.27 What are the problems associated with regulating a natural monopolist?
14.28 What is a two-part tariff?
14.29 What is a fair rate of return regulation?

15 Monopolistic Competition and Oligopoly

CONCEPTS

- Monopolistically Competitive Market
- Product Differentiation
- Excess Capacity Theorem
- Advertising or Selling Costs
- Persuasive Advertisement
- Informative Advertisement
- Oligopoly
- Duopoly
- Kinked Demand Curve Model
- Cournot Model
- Best Response Curve
- Bertrand Model
- Anti-trust Regulations

Perfect competition and monopoly are the two extreme forms of market structure. The former has infinitely many firms producing a homogeneous product with the implication that any particular firm has no market power and behaves as a price taker. The latter has only one firm.

In this chapter we study monopolistic competition and oligopoly, which are intermediate between perfect competition and monopoly.

MONOPOLISTIC COMPETITION

This market structure has the following features: (*a*) There are a large number of sellers and buyers. (*b*) There is free entry and exit in the long run. (*c*) Moreover, there is **product differentiation**. That is, each firm produces a brand or variety (of the same product) that is unique, that is, different from what any other firm produces. The varieties produced are very close substitutes of one another. Products like toothpaste, soap and lipstick are good examples.[1] Service providers like tailors who make suits and fancy hair-cut salons are also good examples; each of these basically provides the same service, yet each one is slightly different from the other in terms of style, feel and look.

The features (*a*) and (*b*) respectively imply that each firm is small relative to the market and earns zero abnormal profits in the long run. Thus (*a*) and (*b*) are competitive features. However, (*c*) is a monopoly feature in the sense that any given brand is produced only by one firm; hence each firm has some monopoly power. Put differently, product differentiation confers some market power to a monopolistically competitive firm, even though it is small compared to the entire market. Monopolistic competition is thus a market form that has monopolistic and competitive features.

Equilibrium in the Short Run

The monopoly feature implies that a monopolistically competitive firm also faces AR and MR curves for its brand and it maximises profits at the level of output, where $MR = MC$. It charges a price, which exceeds marginal cost.

Analytically, all these are analogous to monopoly, except for one qualitative difference. That is, since there are close substitutes available for any particular brand, the demand curve facing a monopolistically competitive firm (unlike that facing a monopoly firm) is very elastic, implying that the AR curve must be quite flat. Figure 15.1 draws a short-run situation of a representative firm in monopolistic competition. The AR curve is relatively flat. The MR curve corresponds to the AR curve. The firm maximises where MR equals MC, producing the amount y_0 and charging price p_0.

[1]There are at least 7 brands of lipstick available in the Indian market: *Avon*, *Elle*, *Lakme*, *Loreal*, *Maybelline*, *Revlon* and *Tips & Toes*. There are many more brands of toothpaste, for example, *Acquafresh*, *Anchor*, *Amar*, *Babool*, *Cibaca*, *Close-Up*, *Colgate*, *Forhans*, *Meswak*, *Neem*, *Pepsodent*, *Promise* and *Vicco Bajradanti*.

Figure 15.1 Monopolistic Competition in the Short Run

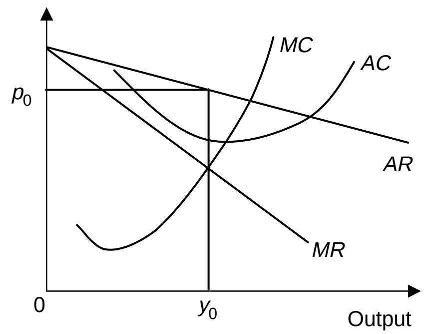
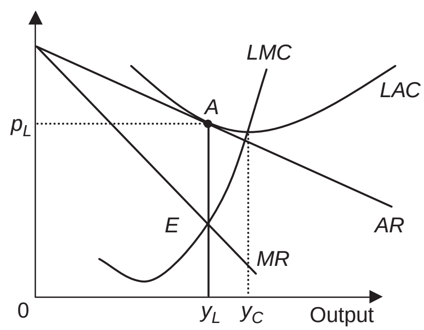


Figure 15.2 A Monopolistically Competitive Industry in the Long Run



Long-run Equilibrium

There is, however, a major difference between monopolistic competition and monopoly in the long run. Unlike in monopoly, there is free entry and exit, implying that abnormal profit is driven to zero. This is equivalent to $P = LAC$, where the letter '*L*' refers to the long run. Together with the profit-maximising condition $MR = LMC$, we can then compactly write the long-run equilibrium conditions in monopolistic competition as:

$$MR = LMC;$$
$$P = LAC.$$

Figure 15.2 illustrates the long-run equilibrium. Mark that at the point where $MR = LMC$, AR is also equal to LAC; therefore profits are zero. The representative firm produces output y_L and charges price p_L.

Even though the long-run profits are zero, the situation is not exactly analogous to the long-run, competitive equilibrium. Notice in Figure 15.2 that whereas under perfect competition, a firm would have produced at the minimum point of the long-run average cost (that is, at y_C), in monopolistic competition, the equilibrium output lies to the left of where the *LAC* is minimised—such as at y_L. That is, under monopolistic competition, in the long run there are increasing returns to scale and unit cost is not minimised. In microeconomic theory this result is known as the **excess capacity theorem** in the sense that in equilibrium, the production is not optimised from the viewpoint of using up all economies of scale without stepping into the range of diseconomies of scale.

Social Welfare Implications

The excess capacity theorem means that each firm in the industry is so small that it is not able to fully exploit scale economies till they are exhausted. In other words, compared to the social optimum (which takes into account production efficiency), there are 'too many' firms, implying that each firm is 'too small'. While this is a source of welfare 'loss', compared to the social optimum, it provides a welfare gain of a kind we have not discussed so far.

Suppose society positively values the variety available to a consumer. Compare a situation where you walk into a garment shop and find men's shirts of three colours to another where you find men's shirts of many different colours and combinations. All else the same, you are better off in the latter situation. If so, a large number of firms imply greater variety, which is a source of welfare gain.

Therefore, we cannot generally say whether, compared to the social optimum, the monopolistically competitive equilibrium entails too many or too few firms and varieties. In any event, to the extent that price is not equal to marginal cost implies some welfare loss compared to what is the first-best for the society.

Advertising

Although the features of monopolistic competition are a combination of perfect competition and monopoly, in terms of decision-making, there is one aspect of it, which is different from both perfect competition and monopoly. That is, monopolistically competitive firms typically engage in advertising, that is, they incur **advertising costs** or what are also called **selling costs**. (Remember the concept of advertising elasticity in Chapter 2.) It is because of the need to maintain a perception in the mind of the potential consumers that their respective brands are different (and more tasteful or classy), compared to other brands. This is called **persuasive advertisement** and its purpose is to lure away consumers from other brands to your brand. (In perfect competition, the product is perfectly homogeneous and hence there is no scope to engage in persuasive advertisement. In monopoly, since there is no competition, there is no need to engage in persuasive advertisement.)

Realise that persuasive advertisements do not benefit the consumers as a group; they only switch consumers from one brand to another. But they involve resources, which can be potentially used for production; advertising itself is a big industry. Therefore, advertising costs are 'wasteful' from the viewpoint of the society.

However, not all advertising costs are wasteful. On many occasions, there is **informative advertisement** (for example, about health), which is useful for the consumers. Further, even if it is persuasive advertising, an opportunity to advertise may increase competition, reduce prices and thereby improve social welfare. Indeed, a classic study by Lee Benham found that in America the price of eye-glasses was cheaper on the average by $20 (at 1963 prices) in the states that allowed advertising than in the states that did not allow advertising.

OLIGOPOLY

Features

An **oligopoly** market is one that is inhabited by a limited number of sellers or producers. A special case where there are only two firms is called a **duopoly** market. In oligopoly the product may be homogeneous or differentiated. The number of firms being limited means that entry and exit processes are limited. This could be due to government regulations with respect to entry and exit, or the technology, which may involve huge fixed costs, deterring small firms from entering the industry or surviving in the market for long.

There are numerous examples of an oligopoly market such as passenger aircrafts, air-conditioners, cars and cellular telephone services.

The most crucial behavioural difference of oligopoly with other market structures studied so far is that each firm in the industry is big enough to hold market power to the extent that a change in the price, advertising or output decision by any one firm significantly affects the profit of any other firm. Put differently, each firm is 'important' for any other firm, implying that there is a game-theoretic, strategic interaction among firms. In deciding price, quantity and so on, each firm must take into consideration the effects of its own action on the actions to be chosen by others and vice versa. Whether the product is homogenous or differentiated does not have any bearing on whether the market is an oligopoly or not. One can talk about a homogeneous-product oligopoly or an oligopoly with product differentiation.

As a result, the analysis of oligopoly is relatively complicated. Unlike perfect competition, monopoly or monopolistic competition, there is no single unified 'model' of oligopoly. There are several 'models' of oligopoly instead. In what follows, we outline a few of them.

Kinked Demand Curve Model

It is observed that prices in industries with a small number of firms (that is, oligopoly), fluctuate less than prices in industries with a large number of sellers and buyers. Compare for instance the prices of products like television, telephone services, washing machines on the one hand, and those of stocks of different companies and different metals on the other. The latter vary from day to day, even from one minute to the next, whereas the former are relatively stable. The price stability in an oligopoly market is explained by the kinked demand curve analysis.

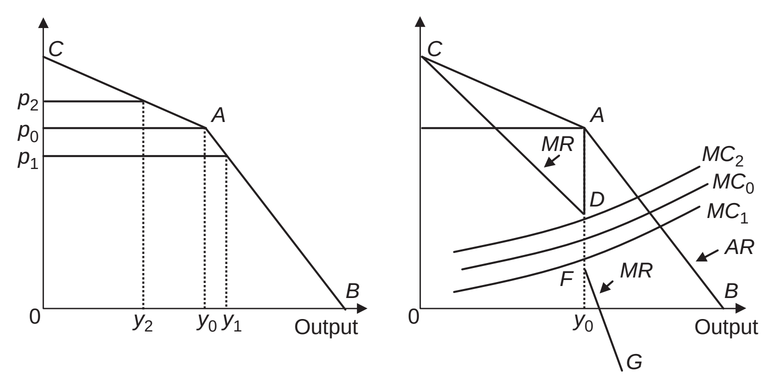
Consider an oligopoly industry in which the firms produce a differentiated product. Think of a particular firm, say firm *X*, which is initially producing y_0 and charging the price p_0. This quantity-price combination is shown at the point *A* in the left panel of Figure 15.3. Starting from this initial situation, suppose that the firm cuts its price, say to p_1. Will the other firms 'keep quiet' and not change their prices? No, because if they do, they will lose customers to firm *X*, reducing their profits. Their optimal response will be to lower prices too. Thus, the firm initiating the price cut (firm *X*) will gain only a modest increase in sales (to the extent that there is a decline in the price of *all* brands, which encourages the consumers to buy more of the product across the board). At p_1, the quantity sold by firm *X* is marked by y_1. The key point is that below the price p_0, the demand curve is relatively inelastic and steep—reflecting a price-cut response by other firms. This is illustrated by the relatively steep line segment *AB*.

What happens if firm *X*, instead of lowering its price, increases it, say to p_2? The other firms will not mind this at all as they stand to gain. Since the brand produced by firm *X* is costlier, it is likely to lose a lot of sales to other firms. That is, the quantity sold by firm *X* will decrease by a relatively large amount—from y_0 to y_2. Hence, above the price p_0, the demand curve facing firm *X* will be very elastic (flat). This is indicated by the segment *AC*.

Combining what happens with a price cut and a price hike, *CAB* is the entire demand curve facing a particular firm *X*. (The two segments are drawn as straight lines for simplicity only, but they need not be so.) Note the 'kink' (sharp edge) at the point *A*. What is the economic rationale behind such a kink? It is the big difference in the price elasticity along the demand curve on the two sides of the price p_0, which arises from the asymmetry in the response of other firms with regard to a price cut and a price rise initiated by a particular firm.

As the demand curve facing a firm is its *AR* curve, the line *CAB* in the left panel of Figure 15.3 is marked as the *AR* curve in the right panel. What is the

Figure 15.3 The Kinked Demand Curve Model



corresponding *MR* curve? In the output range from 0 to y_0, the *MR* curve will have a segment like *CD* which corresponds to the segment *CA* on the *AR* curve. In the output range beyond y_0, the *MR* curve has the segment *FG,* corresponding to the segment *AB* on the *AR* curve. Hence, the *MR* curve is discontinuous.

Now let the curve MC_0 denote firm *X*'s marginal cost curve, passing through the discontinuous part *DF*. At what level of output is the firm's profit maximised? The answer is y_0 and the reason is as follows. At any level of output below y_0, $MR > MC$, so that profit can be increased by increasing the output. Similarly, at any output above y_0, $MR < MC$ and, therefore, profit will increase by decreasing the output. These two statements imply that the profit is the maximum at the output level corresponding to the kink.

Given this equilibrium, suppose there is some variation in the cost conditions facing the firm such that the *MC* curve shifts to say MC_1 or MC_2. Notice that this has no impact on the firm's profit-maximising output. It continues to produce at the kink and charge the price p_0. In other words, the price chosen by the firm is insensitive to a change in cost conditions—at least to some extent (that is, as long as the *MC* curve passes between the points *D* and *F*). This is how the kinked demand curve model explains why price variations are not likely to be large in oligopoly markets.

While this model is useful, it has two major deficiencies. First, the model does not explain the position of the kink on the demand curve. It starts with an arbitrarily given output-price combination and then argues that it is the profit-maximisation combination. Second, although it appeals to the notion of strategic interaction (that is, how one firm responds to a change in strategy by another), it is still not grounded in equilibrium concepts in game theory, for example, the Nash equilibrium as discussed in Chapter 13.

Quantity Competition: The Cournot Model

Cournot, a French economist and mathematician in the 19th century, formulated a model of duopoly competition in quantities produced and sold, which could be founded on the notion of Nash equilibrium and, therefore, is consistent with game theory.[2]

Let the two firms be called *A* and *B*. Assume that the product is homogeneous. Figure 15.4 depicts the industry demand curve *DD*. The questions are: (*i*) what will each firm produce and (*ii*) what will be the market price? Of course, if we have the answer to question (*i*), the answer to question (*ii*) is immediate—add up the two firms' outputs to get the industry-level output and the equilibrium price is the one at which the total quantity demanded is equal to the industry output. Let y_A and y_B denote the output produced and sold by firm *A* and firm *B* respectively. For instance, if $y_A + y_B = Y_0$, the market price, as shown in Figure 15.4, is p_0.[3]

[2] The duopoly model is capable of being easily generalised to the case of more than two firms.

[3] We are not assuming that the market price is p_0. Individual quantities and the market price are determined within the model.

Figure 15.4 Market Price in the Cournot Model

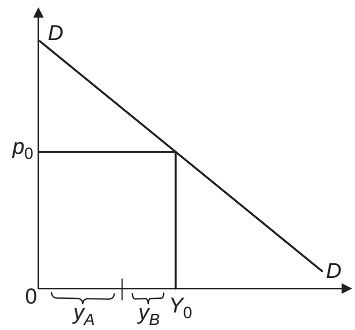
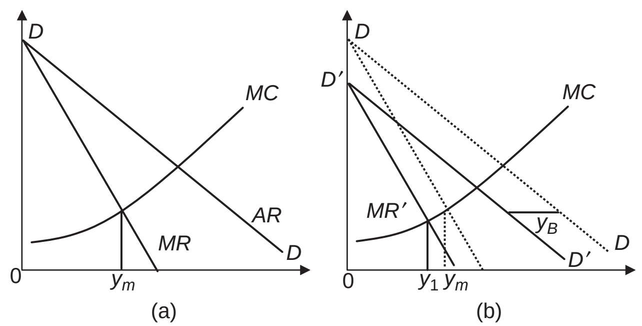


Figure 15.5 Output Choice in Cournot Duopoly



Recall that Nash equilibrium is defined by a set of strategies chosen by the players such that any single player will not benefit from changing his strategy. In the present context, the players are firms *A* and *B*. The strategies are (the individual) outputs chosen by the firms. A particular combination of outputs, say (y_A^0, y_B^0), will qualify as a Nash equilibrium if, given y_B^0, firm *A* has no incentive to choose any output other than y_A^0, that is, the y_A^0 is the profit-maximising output of firm *A* when $y_B = y_B^0$, and likewise y_B^0 is the profit-maximising output of firm *B* when $y_A = y_A^0$. How are y_A^0 and y_B^0 determined?

Consider first the profit-maximising behaviour of firm *A* at various possible values of y_B. Suppose that $y_B = 0$. Then firm *A* has the entire market, that is, it is a monopoly. The market demand curve is its *AR* curve. Panel (a) of Figure 15.5 illustrates this. The *AR* curve drawn is the same as the demand curve in Figure 15.4. Figure 15.5(a) also depicts the marginal cost curve of firm *A*. Its profit-maximising output is y_m, at which $MR = MC$. Put differently, at $y_B = 0$, the best response of firm *A* is to produce output equal to y_m.

Now suppose firm B produces a small quantity. Then, compared to $y_B = 0$, at any given quantity sold by firm A, the total quantity sold in the market will be higher and hence the market price will be less. That is, the average revenue will be smaller. Thus the AR curve facing firm A will lie to the left of its AR curve when $y_B = 0$. This is shown in the panel (b) of Figure 15.5, which depicts the same market demand curve as in panel (a). The AR curve of firm A is $D'D'$, such that the horizontal difference between DD and $D'D'$ is the amount produced by firm B. The corresponding MR curve is marked by MR'. The associated profit-maximising output by firm A is equal to y_1. This is the best response of firm A.

Note that y_1 is less than y_m, that is, a higher output by firm B implies a smaller profit-maximising output by firm A. It is because if one firm is able to sell a higher quantity than before, it is equivalent to a decrease in the demand curve facing the other firm. Hence, the other firm produces less in equilibrium. Such reasoning implies that the equilibrium output of firm A decreases with the output of firm B. This relationship is graphed in Figure 15.6(a). The curve is marked R_A, which we can call the best response curve of firm A. In general, the **best response curve** shows a player's best (optimal) strategies corresponding to different strategies chosen by other player(s).

Similarly we can analyse the profit-maximising behaviour of firm B at various levels of output by firm A. By the same logic, the equilibrium output of firm B decreases with the output of firm A. This relationship is shown by the R_B curve in Figure 15.6(b). This is the best response curve of firm B.

Figure 15.7 graphs the best response curves of the two firms together. Mark the point of intersection, N. This is the Nash-equilibrium point, solving the Nash-equilibrium outputs, y_A^0 and y_B^0 respectively for firm A and firm B. Why does the intersection point constitute Nash equilibrium? The point N being on R_A means that when $y_B = y_B^0$, firm A maximises its profit by producing y_A^0 and thus it has no incentive to choose any output other than y_A^0. The point N also lies on R_B, meaning that when $y_A = y_A^0$, y_B^0 is the profit-maximising output of firm B and thus firm B has no incentive to choose any output other than y_B^0.

Figure 15.6 Best Response Curves

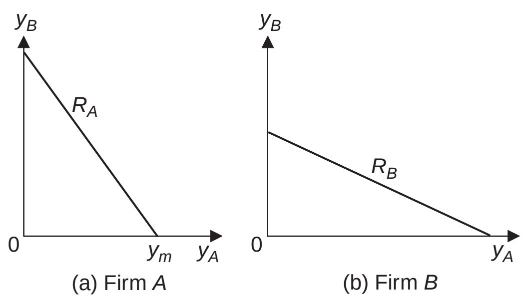
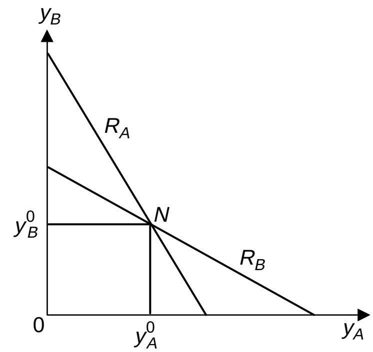


Figure 15.7 Nash Equilibrium in Quantity Competition



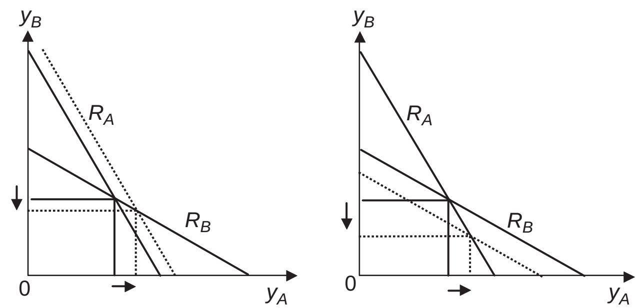
We can now use this quantity-competition framework to determine the effects of changes in 'parameters' facing the industry. The term 'parameter' here means a basic condition or state of the industry, which is not influenced by the choices made by producers or consumers, but which can change due to extraneous factors.

For example, suppose firm *A* is able to discover a new way of organising workers in the production activity such that its *MC* curve shifts down. How will it affect the equilibrium quantities produced by both firms? Turn first to Figure 15.5. If the *MC* curve shifts down, at any given level of output by firm *B* (which fixes the position of the *MR* curve facing firm *A*), the intersection of the *MR* curve and the new *MC* curve will occur at a higher level of output by firm *A*. In other words, the profit-maximising output of firm *A* is higher than before at any given level of y_B. This implies that the R_A curve in Figure 15.6 will shift to the right (but there will be no shift of the R_B curve).

This shift is shown in Figure 15.8(a). In the new (Nash) equilibrium, firm *A* produces more and firm *B* produces less. These are reasonable outcomes. A cost reduction method discovered and executed by a firm will enable it to produce more in the market and force its rival to produce less.

Consider another application. In India, steel is sold by Indian and foreign firms. Among Indian firms, SAIL (Steel Authority of India) is the largest, having plants at Rourkela, Durgapur, Bhilai and Bokaro. Tata Iron and Steel Company is also a big manufacturer of steel. However, for analytical purposes, assume that there is one domestic producer of steel. India also imports steel from many countries, on which it imposes an import duty. For the sake of argument, assume that there is only one foreign firm who sells steel in India. Hence, ignoring quality difference between domestic and foreign steel, we can depict the Indian steel market as a duopoly, having a domestic firm and a foreign firm. Note that paying import duty adds to the marginal cost of selling steel in India by the foreign firm. Go back to Figure 15.6 and interpret R_A and R_B as the best response curves of the domestic

Figure 15.8 Changes in Parameters



(a) Cost Cutting Innovations (b) Tariff Protection

firm and the foreign firm respectively, keeping in mind that R_B takes into account the tariff cost facing the foreign firm. Using Figure 15.7, in equilibrium, the former sells y_A^0 of steel and the latter sells y_B^0.

Suppose that, starting from a situation of 10 per cent import duty on steel, the Indian government slaps a higher tariff (duty) at 15 per cent.[4] How would it affect the firms? The tariff hike would shift the foreign marginal cost curve up. As a result, the foreign firm, firm B, will produce and sell less at any given level of output by the domestic firm. Hence, the R_B curve will shift inward, as shown in Figure 15.8(b). As we see, at the new equilibrium, the domestic firm produces more, while the foreign firm produces and sells less. Thus, the domestic firm benefits from import protection, while the foreign firm is worse off. This does not imply that import tariffs are good for India's economy. The welfare of consumers must be taken into consideration.

NUMERICAL EXAMPLE 15.1

Two firms, Assam-Chicken (firm A) and Bengal-Chicken (firm B) supply broiler chicken to the city of Guwahati. They compete in quantities, that is, play a Cournot game. Chickens sold by both firms are of the same size and quality and quantity is measured by the number of chickens produced and supplied to the market. Suppose the best response function of Assam-Chicken is given by $y_A = 5{,}000 - 3y_B$, that is, if $y_B = 500$, Assam-Chicken produces $5{,}000 - 1{,}500 = 3{,}500$ chickens. Similarly, let the best response function of Bengal-Chicken be specified by $y_B = 2{,}000 - (0.5)y_A$. How many chickens will be supplied to the market by each firm and both firms together in equilibrium? Suppose $Q = 3{,}150 - p$ spells the market demand curve for chicken in Guwahati. What will then be the market price of broiler chicken?

Rewrite the best response functions as:

$$y_A + 3y_B = 5{,}000$$
$$y_A + 2y_B = 4{,}000.$$

[4]Under the WTO system, a country can unilaterally increase import tariffs on industrial goods only on a temporary basis.

These are two equations in two variables: y_A and y_B. Subtracting the second equation from the first, we get the solution of the output supplied by Bengal-Chicken: $y_B = 5{,}000 - 4{,}000 = 1{,}000$. Substituting this value in either equation gives the solution of output by Assam-Chicken: $y_A = 2{,}000$. Thus both firms together sell:

$$y_A + y_B = 2{,}000 + 1{,}000 = 3{,}000.$$

This is the total quantity sold in the market, that is, $Q = 3{,}000$. Substituting this into the demand function, $p = 3{,}150 - Q = 3{,}150 - 3{,}000 = 150$, that is, a broiler chicken will be sold in Guwahati for Rs 150.

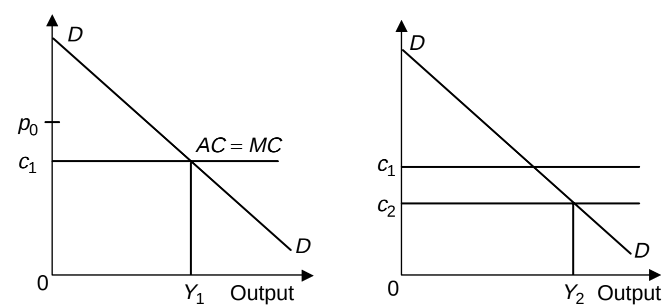
Price Competition: The Bertrand Model

The Cournot quantity competition model assumes that firms strategically choose their quantities of production, while the price adjusts to clear the market (as shown in Figure 15.4), after quantities are chosen. It is easily arguable, however, that firms strategically choose prices instead. This view was advanced forcefully in the late 19th century by a French mathematician, named Joseph Bertrand. It turns out that different outcomes follow under price competition, as compared to quantity competition.

Suppose again that there are two firms and both sell a homogeneous product. Let p_A and p_B denote the prices charged by firm *A* and firm *B* respectively. Realise that, because the product is homogeneous, if $p_A < p_B$, for example, no one would buy the product from firm *B*—why would anyone pay a higher price for exactly the same product? Thus firm *A* would capture the entire market. Similarly, if $p_A > p_B$, firm *B* will acquire the whole market. Once we understand this, finding Nash equilibrium is a simple task.

Look at Figure 15.9. The curve *DD* depicts the market demand. For simplicity it is assumed that the marginal cost and the average cost are constant (independent of the output), that is, $MC = AC$. Panel (a) assumes that the average cost is same

Figure 15.9 Bertrand Price Competition



(a) Symmetric Costs (b) Asymmetric Costs for both firms, equal to c_1. This is the symmetric cost case. Panel (b) depicts the asymmetric-cost case, where the average cost of firm B (= c_2) is less than that of firm A (= c_1).

Let us begin with the symmetric-cost case. Suppose that any of the two firms, say firm A, charges a price higher than c_1, say equal to p_0. What is the optimal pricing strategy of firm B? There is no need to 'calculate' profits by weighing marginal revenue and marginal cost. The answer is 'charge a price slightly less than p_0'. Why? It is because by doing so the firm B will capture the entire market. Of course, firm A can recapture the entire market by charging a price slightly less than what firm B is asking. This reasoning implies that the firms will continue to cut the price till no firm makes any abnormal profit. Thus, $p_A = p_B = c_1$ constitute the Nash equilibrium. Given $p_A = c_1$, firm B has no incentive to charge a higher price (thereby losing the entire market) or a lower price (thereby incurring losses). Similarly, given $p_B = c_1$, firm A has no incentive to change its price from $p_A = c_1$. In summary, when the product is homogeneous and average costs are the same across the firms, in Bertrand price competition, price = average cost is the Nash equilibrium. This implies that no firm makes any abnormal profits, that is, the profits are zero. Which firm sells how much? This is indeterminate. A natural assumption is that since the firms are symmetric, they will have equal shares in the market—each firm will have 50 per cent of the market, that is, sell $0Y_1/2$ in Figure 15.9(a). However, irrespective of who sells how much, the total quantity sold is determined from the market demand curve at the price equal to the average cost—$0Y_1$ in Figure 15.9(a).

It is noteworthy that under the same market and technology parameters, if the firms were competing in quantities, they would have made abnormal profits as in Cournot competition, the price is not driven down to the average cost. This yields a fundamental insight that price competition is more severe on the profits of firms in oligopoly than competition in quantity.

Consider now the asymmetric cost case. By the same logic, no price above c_1 can be in Nash equilibrium. But unlike the symmetric-cost case, $p_A = p_B = c_1$ do not constitute the Nash equilibrium. It is because firm B can still charge a price lower than c_1, drive firm A out of the market and yet make a profit higher than that at $p_B = c_1$. Hence, in Nash equilibrium, only the more efficient firm survives, and the market price is the one, which is marginally lower than the average cost of the less efficient firm.

Indeed we can generalise the above conclusion to the case of many firms. Rank the firms starting from the most inefficient to the most efficient. Under price competition, only the most efficient firm survives because it can 'under-cut' in terms of price any other less efficient firm, thereby, driving all other firms out and capture the entire market.

It must be noted that the result that only the most efficient firm survives is dependent on the product being homogeneous. If, instead, the product were differentiated, then each firm would have some market power (as in monopolistic competition), and unless the cost differences are very large, the relatively more efficient firms and the relatively less efficient firms (at least some of them) would co-exist in the market.

NUMERICAL EXAMPLE 15.2

There are two companies, *Gangajaal* and *Mahajaal*, which supply bottled water to the twin cities of Bhubaneswar and Cuttack. In procuring, processing and packaging water bottles, each firm incurs a constant average cost of Rs 8 per bottle. The demand curve of bottled water in Bhubaneswar and Cuttack is given by the equation $Q = 10{,}000 - 250p$. If the two firms play the Bertrand price competition game, what will be market price and quantity sold?

This is a case of symmetric Bertrand competition. Hence the equilibrium price will be driven down the (common) average cost, that is, p = Rs 8. Substituting this into the market demand function, the total quantity sold = $Q = 10{,}000 - 250 \times 8 = 10{,}000 - 2{,}000 = 8{,}000$.

Whether Cournot or Bertrand?

Because quantity and price competition yield very different outcomes, one may ask which one is more appropriate. There is no single answer to this question. It depends on whether in the industry concerned, the firms commit to quantities they produce and let prices clear the market or commit to prices and let quantities clear the market. In producing steel, for instance, it is relatively time-consuming to change production levels (within say month or a quarter). Thus, steel firms have to commit to a production level and it is reasonable to assume quantity competition. But, in providing telephone service, for example, the prices (tariffs) are quoted in catalogues and on websites, which are not changed frequently; thus it is reasonable to assume price competition (that is, treat price as the strategic variable).

Cartel Issues

The Nash equilibrium in quantity or price competition is comparable to the Nash equilibrium in the Prisoner's Dilemma game discussed in Chapter 13. Hence, essentially, oligopoly competition is a Prisoner's Dilemma situation. It is because in oligopoly, the firms compete with one another rather than collude.

Colluding or collusion means jointly deciding the total output for the industry and the market price, and formulating an allocation rule as to who will produce how much. If the firms do collude, collectively they behave like a monopoly or a cartel. A cartel essentially raises the price of the product by restricting output. Under the cartel, the total (monopoly) profits are higher than the sum of the profits in oligopoly competition (irrespective of the form of oligopoly competition). The quota system dictating who produces how much is devised such that every member (firm) of the cartel is better off (that is, enjoys higher profit), compared to oligopoly competition.

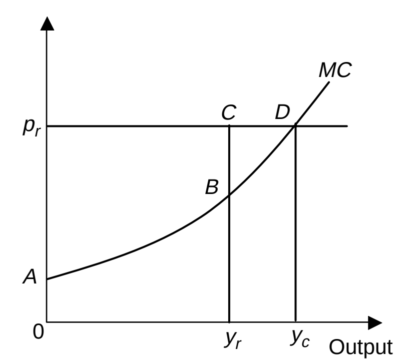
If so, in the real world, why do not firms always collude or form a cartel on a regular basis? One obvious answer is that colluding or forming a business cartel is illegal in many countries because of its anti-competitive (high-price) effects, which hurt the consumers. However, for products that are primarily exported by a country, there is no such concern because higher prices would hit the consumers in the foreign countries, not the domestic consumers. This is why, over time, many cartels have come up typically in international markets. OPEC is, of course, the best example. But cartels have been formed in many products including steel and vitamin.

However, there is an inherent problem with any cartel, even if it may be legal—a member of a cartel has an incentive to 'cheat' and secretly break the already agreed upon cartel rules. This can be understood by using Figure 15.10. The *MC* curve denotes the marginal cost curve of a cartel member. Suppose the cartel has fixed the price of the product at p_r, which is presumably higher than the price that would prevail under oligopoly competition. Such a high price is made sustainable by following a production (an export) quota rule in which each member is asked to produce a limited quantity (less than what it would have produced in oligopoly competition). Let this quantity or quota for the member in question be y_r. In the cartel regime, this member's total revenues are thus equal to the area $0p_rCy_r$. Its total costs are measured by the area under the *MC* curve, equal to the area $0ABy_r$. Hence, profits are equal to the area $ABCp_r$.

Suppose this is the original situation—there is a cartel, in which the member firms are charging the price p_r, and each member is producing a limited quantity, with y_r being the output of a particular member firm. Now observe that as long as the market price is given at p_r, the particular firm we are looking at can actually increase its output and sale up to y_C and make higher profits because if the market price is given, it is equal to *MR* and is over the output range 0 to y_C, with $MR > MC$. Indeed, at the output level y_C, the firm's profits will be equal to ADp_r, which is greater than $ABCp_r$. The implication is that this member-firm will have an incentive to break the cartel rules, over-produce and make greater profits.

Of course, once the 'cheater' sells more, there will be downward pressure on the market price. This will be 'felt' by the rest of the cartel members (who are 'honest' to the cartel) because they will face difficulty in selling their quota output at the cartel price p_r (since there is more quantity in the market than before—thanks to

Figure 15.10 Cartel Problems



the cheater). Whether the cheater is identified or not, others will come to infer that someone among them has broken the rules. What happens next? A strong possibility is that the cartel is abandoned and the industry reverts to oligopoly competition.

In summary, insofar as the cheater is concerned, it gains in the short run and loses in the long run. If the firms value the future or trust each other very much, then no firm may decide to cheat in the first place. If not, cheating is an inherent problem, making the cartel inherently unstable. This is why, in the real world, it is hard to find an example of a cartel that has lasted successfully for a long stretch of time. Even OPEC has had its problems over years. The average cartel life (before it busts itself due to cheating and enforcement problems or gets busted by authorities) is around four to six years.

ANTI-TRUST REGULATIONS

A common feature of almost any form of an imperfectly competitive market is that price exceeds marginal cost and, therefore, such a market entails a loss of social surplus. This is of concern especially from the viewpoint of consumers' welfare. After all, the social purpose of businesses is to create/produce goods and services so as to meet our consumption needs and enhance our standard of living.

Thus, increasing concentration of firms (meaning smaller number of large size firms) in an industry is worrisome insofar as it results in higher prices for the consumers. This is indeed the basis of monopoly regulation, which we discussed in Chapter 14. It is also the rationale behind the so-called **anti-trust regulations** with respect to oligopoly markets.

The origin of anti-trust regulations lies in the economic history of the US. During decades following the civil war in America, a selected group of aggressive entrepreneurs became extremely powerful, for example, Andrew Carnegie, Jay Gould, John Rockefeller and others. These men created huge empires of business for themselves and were accused of exploiting labour and using unfair labour practices. They were portrayed as ruthless businessmen who could do anything to increase their wealth and later came to be known as 'robber barons'. Intense public sentiments led to the first major anti-trust law in the form of the Sherman Act of 1890. This marked the beginning of anti-trust regulation in the US and in the rest of the world. The initial aim of these regulations was to contain market concentration on the presumption that this is bad for the consumers for the reasons described above.

However, recall from Chapter 14 that there are a few merits of monopoly too and these apply to an oligopoly market as well. Hence, the modern philosophy behind anti-trust laws and their enforcement is that being big in business may not be necessarily bad by itself. The key issue that an anti-trust body looks at is whether merger and acquisitions or product/pricing strategies by big firms that tend to eliminate competition, ultimately serve the interests of consumers. See Clip 15.1 for an account of anti-trust problems faced by Microsoft.

Clip 15.1: Microsoft's Anti-trust Woes

The software giant Microsoft had to face serious antitrust lawsuits for nearly a decade. Its rivals have claimed and brought attention to the antitrust authorities in the US and Europe that the firm was essentially either tying its products with its operating system or making it relatively difficult for products from other companies to run efficiently on the Windows platform. For instance, in the early- and mid-nineties Microsoft would 'force' computer makers to use Internet Explorer (IE) as a part of licensing Windows. The company however argued that IE was an 'integrated part' of its operating system. But in 1997, it was asked by the antitrust authorities to un-bundle IE from its operating system. Later on, the European Commission asked Microsoft to un-bundle Windows Media Player. Moreover, Sun Microsystems, which developed Java technology, alleged that Microsoft was breaking its contractual obligation by not using Java in IE. Novell Corporation, which developed the server software 'NetWare' (as well as the Word Perfect package), argued that the Windows server operating system did not allow NetWare to run efficiently. In 2004 and 2005, Microsoft reached out-of-court settlements with Sun, Novell as well as others (for example, Gateway) by paying them compensations running into hundreds of millions of dollar. Microsoft has spent well over 3 billion dollars in settling anti-trust litigations. From 2005, Microsoft's antitrust problems seem to be on their way out.

Anti-trust Practice in India

In India, the Monopoly and Restrictive Trade Practices (MRTP) Act of 1969 was the initiating legislation, whose aim was to restrict monopoly power. However, the MRTP interventions were nothing close to what is outlined above as the objective of a standard anti-trust legislation in a market oriented economy. They were mostly in the form of controls, licensing and permits. It was also largely ineffective, as both public-sector and private-sector monopolies thrived under it. Why? Because in terms of having real punitive powers, it was nearly toothless. To enforce its orders it had to depend on courts entirely, which would take years to reach any final decision. It did not have power to impose, on prima facie grounds, any initial penalties for breach of its directives, before the matter could be moved to the courts. It did not even have powers to enforce the attendance of a witness. The personnel at the disposal of MRTP did not have the right skills to deal with investigations of economic transactions.

Only very recently, India has passed the Competition Act in 2002, which has many innovative features. In effect, the MRTP Act has been repealed and the MRTP Commission dissolved. Following the Competition Act, a Competition Commission was set up in 2003. Its aim is to prevent 'anti-competitive practices, promote and sustain competition, protect the interests of the consumers and ensure freedom of trade' as the website of the Competition Commission states. Both goods and services sectors as well as private and public sector firms come under its purview. The Commission has the mandate to impose penalties initially on erring parties and provide relief to firms who are adversely affected by the anti-competitive strategies adopted by other firms. At the point of writing this book the Commission is in its formative stage, yet to initiate an anti-trust case.

Regulating the Telecom Sector

It is interesting that the telecom sector in India is oligopolistic, yet it is not subject to the 'standard' anti-trust regulations. Instead, there is a separate body called the Telecom Regulatory Authority of India, or briefly TRAI, established in 1997. Why is there a separate regulatory body for this industry only? It is because (*i*) this sector provides a 'basic' utility, like telephony, which is unlike automobile or an internet browser and (*ii*) there are a few special features of this sector.

Because the industry offers a basic utility, TRAI regulates telecom prices, similar to electricity regulatory commission fixing electricity tariffs. But there is much more. That is, the government recognises the vital role of the telecom sector in the growth and development of almost any other sector of the economy and, therefore, its objective is to ensure a telecom sector that is world-class.

The technology of this sector is changing almost 'day by day', which is posing new problems and challenges. For instance, internet services are facilitating telephony. Telephone and broadcasting industries are stepping into each other's markets. The working difference between wireless lines and wired lines is getting narrowed. There is also an issue of 'spectrum management' or 'frequency allocation'. As the demand for frequency is increasing dramatically, it must be used in an efficient manner, without interfering with those used by the national defence sector. Dealing with these complex issues and ensuring healthy competition in a vital sector like telecom are the reasons behind a separate regulatory body like TRAI.

Economic Facts and Insights

- In monopolistic competition, an individual firm, even though very small compared to the entire market, has some market power because its product is differentiated. This implies that the firm is able to set a price above the marginal cost.
- Even though monopolistically competitive firms have market power, they earn normal profits in the long run.
- Besides setting output and prices, firms in monopolistic competition and oligopoly engage in advertising.
- There is no unified theory of oligopoly.
- An oligopoly market is characterised by game-theoretic, strategic interaction among firms.
- Essentially, oligopoly competition yields a Prisoner's Dilemma situation.

(continued)

- The kinked demand curve model attempts to explain why in oligopoly markets, prices may be relatively stable.
- Under quantity competition, an increase in the costs faced by one firm leads, in equilibrium, to less output by this firm and more output by the rival firm.
- If the product is homogeneous and firms compete in prices, then the market sustains only the most efficient firm(s).
- Price competition is more severe on the profits of firms in oligopoly than competition in quantity.
- In a cartel where quantities to be produced by firms are restricted, that is, each firm has a quote, an individual firm has an incentive to cheat in terms of producing and selling more than its quota. This implies that cartels are inherently unstable.
- The average life of a cartel in the international market is between four to six years.
- The modern philosophy behind anti-trust laws and their enforcement is that being big in business may not be necessarily bad. Central is the issue of whether merger and acquisitions or product/pricing strategies by big firms tend to serve the interests of consumers.
- In India the recently constituted Competition Commission is in charge of anti-trust issues. The telecom sector, however, has a separate regulatory body, TRAI (Telecom Regulatory Authority of India).

EXERCISES

15.1 Briefly explain the excess capacity theorem.

15.2 Suppose at a given point of time a monopolistically competitive firm faces the *AR*, *MR* and the cost curves as shown in Figure 15.1. (a) Is this a situation of long-run equilibrium? (b) If not, will some new firms enter into or some existing firm exit from the industry?

15.3 The *LAC* schedule of a representative firm under monopolistic competition, is given as follows. In the long-run industry equilibrium, this firm produces an output, which is ____ than ____. Fill in the blanks and give reasons.

Output	*LAC*
0	–
1	20
2	16
3	13
4	11
5	14
6	18
7	24

15.4 'In a monopolistically competitive industry, there are no abnormal profits earned in the long run equilibrium. Therefore, the price must be equal to the marginal cost.' Defend or refute.

15.5 Explain why in the kinked demand curve model, the demand curve facing a firm is very elastic (or inelastic) at prices below (or above) the kink.

15.6 'The kinked demand curve model is used to explain price instability in some oligopoly markets.' Defend or refute.

15.7 Why are the best response curves downward sloping under quantity competition?

15.8 Suppose that in Kolkata, fish is supplied by two firms only—a local firm and a firm from Andhra Pradesh. While the local firm incurs the costs of production only, the outside firm incurs the costs of production as well as the costs of transportation by trucks from Andhra Pradesh to Kolkata. Assume that the two firms engage themselves in Cournot competition. (They both sell one type of fish and there is no quality difference.) Starting from an equilibrium situation, suppose that there is a diesel price hike. Using a diagram, demonstrate how this will affect the quantities sold by the two firms.

15.9 Consider the numerical example 15.2 in the text. Suppose that *Mahajaal* is able to cut its average cost from Rs 8 to Rs 7.50 per bottle, while the average cost of production by *Gangajaal* remains unchanged. What will be the market price and quantity sold in the new Bertrand equilibrium?

15.10 In a small town, there are two taxi stands serving the whole town. One has a fleet of cars all running on diesel and the other has cars all running on CNG. Think of this as a duopoly market, each 'producing' kilometres of travel. Starting from a given situation, suppose CNG becomes more expensive, while the cost of diesel remains unchanged. Assuming that the two taxi stands play a Cournot duopoly game, how will this affect the best response curves and the amount produced by the two taxi stands?

15.11 Suppose that in a cartel, the price of a product is fixed at p_r = Rs 150. A cartel member's quota of production is fixed at $y = 35$. This member's marginal cost function is given by $MC = 50 + 2y$. If this firm decides to cheat, by how much will it like to increase its output (from 35)?

15.12 Consider the symmetric Bertrand competition case, in which the average cost is constant. Is the social surplus maximised at the equilibrium?

15.13 'Firms making abnormal profits in an oligopoly market is the main concern of modern anti-trust bodies.' Comment.

16 Factor Markets

CONCEPTS

- Factor Price Line
- Value of the Marginal Product
- Marginal Productivity Theory of Distribution
- Labour Economics
- User Value
- Economic Rent
- Total Value Product
- Derived Demand
- Marginal Revenue Product
- Trade Union or Labour Union
- Asset Value

In all the previous chapters, product market was our focus—which goods or services are produced and consumed, and, if so, how much and at what price. In other words, we dealt with the central problem of 'what' facing an economy. Households are the demanders and firms are the suppliers in product markets.

In this chapter, we probe factor or input markets, that is, markets for different types of labour or skill, capital (that is, machinery and equipment), land and so on. In factor markets, the roles of firms and households are reversed—firms are the demanders and households the suppliers.

There are dissimilarities and similarities between product and factor markets. Dissimilarities stem from the demanders and the suppliers in a factor market being the opposite of who they are in product markets. Also, instead of the economy's central problem of 'what', the factor market analysis sheds light on the 'for whom' problem. In the labour market, for example, the price of the labour service is the wage rate. How are wages of different types of labour determined in a market economy? In general, wage incomes across individuals, incomes from land, incomes from money lent and so on, determine income distribution in an economy, which, in turn, determines the differences in the purchasing power over goods and services among individuals or households. This is how the factor market implications are linked to the central problem of 'for whom'.

The similarity between factor and product markets is that there is a demand side and there is a supply side of a factor. The equality between demand and supply of a factor determines the respective factor price.

FACTOR DEMAND

A Firm's Problem

At a given point of time, a firm faces different prices for different factors. Think of a private management institute. It has to employ instructors and the average salary for one may be Rs 15,000 per month. The institute may be situated in a rental property. It has to decide how many rooms or floors it should rent in a high-rise building. The market rent may be Rs 50,000 per month per 1000 sq. ft area. The question is, given factor prices, how much of different factors a profit-maximising firm should hire?

On the one hand, hiring more of factors will produce more output, generating more revenues (as long as the marginal revenue is positive). On the other hand, hiring more factors will cost more. Hence, there is a trade-off. As we shall see, a marginal condition characterises the profit-maximising principle.

Throughout the chapter, it will be assumed that the firm under consideration is perfectly competitive in the factor markets, that is, it is a small buyer of any factor such that it cannot by itself influence its price.

One Variable Factor

Suppose the employment levels of all factors, except one, are fixed, that is, there is only one variable input and the rest are fixed. Let this variable factor be called labour, measured in hours of work. (If all workers must work a given number of hours per day, we can measure labour as the number of workers hired.) In other words, we are not differentiating between different types of workers at the moment. The question is, how many labour hours (denoted by L) should a firm employ?

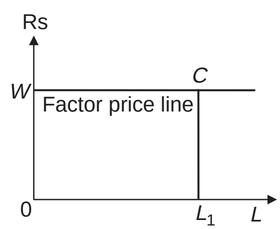
The total cost of fixed factors is fixed by definition. The total cost of the variable factor is easy to compute. Suppose that the wage rate is Rs 15 per hour. If the firm hires 4 hours of labour, the total cost of labour is Rs 15×4 = Rs 60. If 6 hours of labour are hired, the total cost of labour or the total wage bill is Rs 15×6 = Rs 90 and so on.

Suppose that, in general, the factor L costs W per unit, that is, the hourly wage rate is Rs W. Figure 16.1 draws the **factor price line** or the 'wage line' in this case. It is a horizontal line since the wage rate is unaffected by how many labour hours our small firm hires in the market. (The horizontal factor price line is the counterpart of the price line for a competitive firm in the product market.) The total factor cost or payment to the factor is equal to the area under the factor price line. For instance, if the firm hires L_1 labour hours, its total wage bill is the area $0WCL_1$.

On the revenue side, the way the firm's total revenue changes with different levels of employment of a factor is that: (*i*) a change in the employment of a factor affects output and (*ii*) a change in the output affects the total revenue. We have already studied (*i*) in Chapter 8. We also have analysed (*ii*) for a competitive firm in Chapter 11 and for a monopoly firm in Chapter 14. What remains to be done is to combine (*i*) and (*ii*). For now, assume that the firm is a perfectly competitive firm *in the product market*.[1]

From Chapter 8, recall in particular the definitions of total physical product (*TPP*) and the marginal physical product (*MPP*). The former refers to different levels of total output at different levels of employment of a factor, when the

Figure 16.1 Factor Price Line



[1] That is, the firm under consideration is small in the factor and product markets.

employment of other factors is unchanged. The latter is the increase in total output per unit increase in the employment of a factor, when the employment of all other factors is held constant. We also know the shapes of the *TPP* and *MPP* curves. The inverse U-shape of the *MPP* curve follows from the law of diminishing returns.

We develop two more concepts here. The first is the **total value product** (*TVP*), defined as $P \times TPP$, where P is the product price. This is indeed same as total revenue. The second one is the **value of the marginal product** (denoted by *VMP*), defined as $P \times MPP$. Equivalently, *VMP* is equal to the increase in *TVP* per unit increase in the employment of the factor. It is because an extra unit employed of a factor generates extra output equal to *MPP*, which will fetch extra revenues equal to the value of this extra output.

A particular *TPP* schedule is given in the second column of Table 16.1. Its associated *MPP* schedule is shown in the third column. To calculate the *TVP* and the *VMP*, we need to know the product price. Suppose that P = Rs 10. Table 16.1 gives the associated *TVP* and *VMP* schedules in the last two columns.

Particularly relevant for us will be the *VMP* schedule and its properties.

(*a*) The *VMP* schedule is proportional to the *MPP* schedule as it is obtained by multiplying the *MPP* schedule by price, which is constant. This implies that the law of diminishing returns also determines the nature of the *VMP* schedule; *VMP* initially increases with factor employment and then diminishes.

(*b*) The *TVP* of a particular level of factor employment is the sum of *MVP*s up to that level of employment. For instance, at $L = 3$, $TVP = 640$. This is equal to the sum of *MVP*s at $L = 1$ (200), at $L = 2$ (230) and at $L = 3$ (210).

Property (*a*) implies that the *VMP* curve, the graphical representation of a *VMP* schedule, will be inversely U-shaped, just as the *MPP* curve. This is illustrated in

Table 16.1 *TPP*, *MPP*, *TVP* and *VMP* Schedules

Labour Hours (L)	*TPP*	*MPP*	*TVP = P.TPP (Rs)*	*VMP = P.MPP (Rs)*
0	0	–	0	–
1	20	20	200	200
2	43	23	430	230
3	64	21	640	210
4	82	18	820	180
5	94	12	940	120
6	100	6	1000	60
7	100	0	1000	0
8	96	–4	960	–40
9	89	–7	890	–70

Product Price = Rs 10

Figure 16.2 The *VMP* Curve

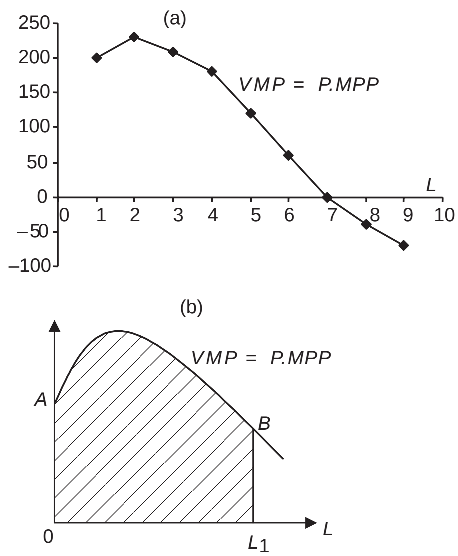


Figure 16.3 Factor Employment Decision

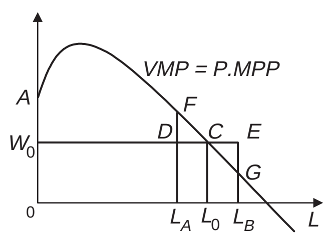


Figure 16.2(a). Property (*b*) implies that, if we draw a smooth *VMP* curve, the area under it will be equal to the *TVP* (that is, the total revenue). A general, smooth *VMP* curve is shown in Figure 16.2(b). For instance, at $L = L_1$, the *TVP* is equal to the shaded area $0ABL_1$.

We are ready to derive the principle that governs how many labour hours a profit-maximising firm should hire. Turn to Figure 16.3 that combines Figures 16.1 and 16.2(b). Let the factor price facing the firm (wage rate) be W_0. The answer is that the firm should hire up to that level, where the factor price line intersects the *VMP* curve, that is, it should hire L_0 labour hours. In other words, the general, profit-maximising principle of hiring a factor is that

$$VMP \text{ of a Factor} = \text{Its Price}. \tag{16.1}$$

This condition is parallel to the profit-maximising condition for selecting output, namely, $p = MC$. Indeed the conditions (16.1) and $p = MC$ are two sides of the same coin. The rationale behind condition (16.1) is analogous to that behind $P = MC$. At the level of employment L_0, *TVP* or total revenue = the area $0ACL_0$. The total factor cost = the area $0W_0CL_0$. Thus, the gross profit, the difference between *TVP* and total factor cost, is equal to the area W_0AC.[2] Consider any employment level less (such as L_A) or more (such as L_B) than L_0. We can compute that the profit is less than W_0AC. For instance, at $L = L_A$, it is equal to W_0AFD, which is equal to $W_0AC - CDF$. At $L = L_B$, it is equal to $W_0AC - CEG$. This proves that profit is maximised at $L = L_0$.

The law of diminishing returns is the key behind the principle (16.1). Starting from where the *VMP* of a factor = its price and the *MPP* is diminishing, if the firm hires one extra unit of the factor, the *VMP* will be less than the factor price, that is, the additional revenue generated is less than the additional cost incurred. This implies less profit than before. Similarly, if the firm hires one less unit than where *VMP* is equal to the factor price, the *VMP* will be higher than the factor price, that is, the revenue sacrificed by hiring one unit less will be more than the savings on the total factor cost (equal to the factor price). Thus, profit will be less. This proves why profit is maximised when condition (16.1) is met.

NUMERICAL EXAMPLE 16.1

A small garbage-collecting company named Clean-City collects garbage from homes in a city at Rs 4 per one home per day. The total physical product (*TPP*) schedule of labour employed per day by the company is given in Table 16.2. The *TPP* is defined in terms of the number of households served. All workers are equally efficient and they all work the same number of hours per day. The labour cost is Rs 28 per worker per day. How many workers will be employed by Clean-City?

Table 16.2 *TPP* Schedule (Numerical Example 16.1)

Number of Workers	*TPP*
0	0
1	15
2	31
3	45
4	56
5	65
6	72
7	77

[2] The adjective 'gross' is attached since fixed costs are not deducted. By definition, profit = gross profit − fixed cost. However, since the fixed costs are given, gross profit is maximised when profit is maximised and vice versa.

Table 16.3 *VMP* Schedule (Numerical Example 16.1)

Number of Workers	*VMP*
0	–
1	60
2	64
3	56
4	44
5	36
6	28
7	20

We first derive the *MPP* schedule and then multiply it with the market price, Rs 4, to obtain the *VMP* schedule. The *VMP* schedule is given in Table 16.3. The condition (16.1) is met at the level of employment equal to 6. Hence, 6 workers will be employed by Clean-City.

Factor Demand Curve

Note from Figure 16.3 that at $W = W_0$, the firm chooses the corresponding point on the *VMP* curve. Similarly at any other input price, the firm would choose the corresponding point on the downward portion of the *VMP* curve. It then follows that the downward portion of the *VMP* curve is the firm's demand curve for the factor.[3] It also means that a firm's demand curve for a factor is downward sloping.

FACTOR DEMAND CURVE SHIFTS

Since the factor demand curve is a part of the *VMP* curve, anything that shifts the *VMP* curve shifts the factor demand curve. We will consider the following sources of change.

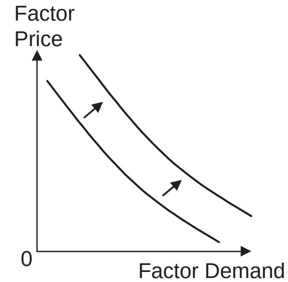
A Change in Product Price

As *VMP* = *P.MPP,* an increase in *P* increases the *VMP* at any given level of factor employment. As a consequence, the factor demand curve shifts to the right. This is illustrated in Figure 16.4. In general then, an increase (or a decrease) in the product price shifts the factor demand curve to the right (or left).

From this result, we can see a link between product and factor markets. For instance, consider the industry of a particular handicraft, whose product has a worldwide market. On the supply side, there are artisans who, with the help of raw materials and equipment, make the handicraft. Suppose that in an international exhibition this handicraft attracts a lot of attention. Many people and organisations

[3]This is parallel to the supply curve of a firm being the same as the upward sloping portion of the marginal cost curve. For simplicity, we ignore the 'shut-down' possibility here.

Figure 16.4 Product Price Increase and Factor Demand



around the world come to know about it, like it and place a purchase order for it. Consequently, there is an increase in demand for this handicraft. From the demand-supply analysis in Chapter 12, we know the effect—the price of this handicraft increases.

Now consider the (factor) market for artisans. The increase in the price of the handicraft will shift their *VMP* curve and hence the demand curve for artisans to the right.

The general point is that factor demand is, in a sense, derived from product demand. This is why factor demand is called a **derived demand**.

Technological Change

A change in technology can alter the *MPP* of a factor and thereby its demand curve, even when the product price is unchanged. If it is such that the *MPP* of a factor increases, then the demand curve for that factor shifts to the right. Figure 16.5(a) shows this effect. Otherwise, if the *MPP* of a factor decreases due to a technological change, then its demand curve shifts to the left. Figure 16.5(b) illustrates this.

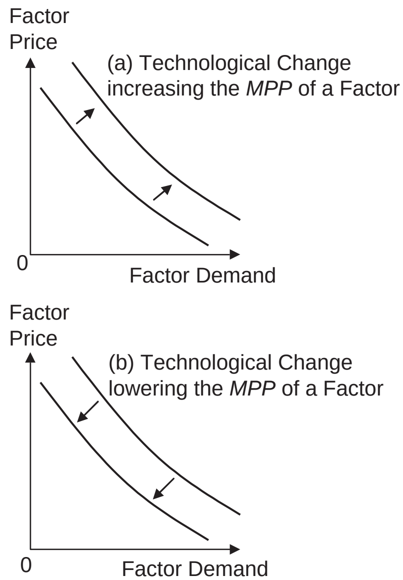
For example, it is widely believed by economists that in the recent two–three decades, the whole world economy has experienced technological progress that has increased the *MPP* of skilled labour, while its impact on the *MPP* of unskilled labour is not clear.[4]

MANY FACTORS AND THE MARGINAL PRODUCTIVITY THEORY OF DISTRIBUTION

We have assumed that the firm employs only one variable factor of production. In reality, firms employ many factors, for example, different types of labour, raw materials, power, various kinds of machines, land and so on. What are the

[4]Another possible source of a shift in the factor demand curve, which we have not discussed and which is something to be done in a more rigorous course in microeconomics, is the change in the employment of other factors.

Figure 16.5 Technological Change and Factor Demand



profit-maximising principles that govern the *simultaneous* demand/employment of more than one factor?

They are simply the extensions of the condition (16.1). If, for example, there are two factors, say X and Y, their respective prices are W_X and W_Y, and their respective marginal products are MPP_X and MPP_Y, the profit-maximising principles are then:

$$VMP_X = P.MPP_X = W_X,$$
$$VMP_Y = P.MPP_Y = W_Y. \tag{16.2}$$

That is, profit is maximised when the VMP of each factor is equal to its price. Note that even when there are two or more variable factors, the definition of MPP remains valid however.

Recall, from Chapter 1, the central problem of 'for whom' facing an economy, which concerns who earns how much. The conditions (16.2) imply a theory of this, that is, each factor earns the value of its marginal physical product. It is called the **marginal productivity theory of distribution**.

Of course, we have not yet discussed the supply side of a factor and hence our analysis has not 'determined' a factor price; we have only examined how much of a factor a firm would demand at a predetermined factor price. Yet, at the equilibrium quantity supplied, the VMP of a factor must be equal to its equilibrium price. In this sense, the marginal productivity theory holds even when the supply side is taken into consideration.

Mathematically Speaking

The Profit-maximising Conditions with Respect to Hiring Factors

Suppose there are n factors, 1, 2, 3, …, n, the quantities hired of which are x_1, x_2, x_3, …, x_n and whose prices are w_1, w_2, w_3, …, w_n. Let $y = f(x_1, x_2, x_3, \ldots, x_n)$ be the production function and let P denote the product price. By definition, the *VMP* of input 1 = $P\dfrac{\partial f}{\partial x_1}$. Similarly, the *VMP* of input 2 = $P\dfrac{\partial f}{\partial x_2}$ and so on.

We have total revenue = $Py = Pf(x_1, x_2, x_3, \ldots, x_n)$. Similarly, the total cost of n inputs = $w_1x_1 + w_2x_2 + \ldots + w_nx_n$. Thus, profits ($\pi$) equal

$$Pf(x_1, x_2, x_3, \ldots, x_n) - w_1x_1 - w_2x_2 - \ldots - w_nx_n = \pi\,(x_1, x_2, x_3, \ldots, x_n).$$

This is maximised with respect to x_1, x_2, x_3, …, x_n. The first-order conditions are:

$$\frac{\partial \pi}{\partial x_1} = P\frac{\partial f}{\partial x_1} - w_1 = 0$$

$$\frac{\partial \pi}{\partial x_2} = P\frac{\partial f}{\partial x_2} - w_2 = 0$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\frac{\partial \pi}{\partial x_n} = P\frac{\partial f}{\partial x_n} - w_n = 0.$$

These conditions are generalisations of (16.2). Each first-order condition states that the respective input is paid its *VMP*.

* * * * * *

MONOPOLY IN THE PRODUCT MARKET

We have derived the rule for hiring a factor, (16.1) or (16.2), under the assumption that the hiring firm is perfectly competitive in the product market. What is the rule if the firm is imperfectly competitive? There is a minor difference. Remember that a non-competitive firm (for example, a monopoly) is not a price taker. Hence, the extra revenue generated by selling one additional unit is the marginal revenue (*MR*), not equal to the price. Thus the extra revenue by hiring an additional unit of an input is equal to the *MPP* of that input × *MR*, defined as the **marginal revenue product (*MRP*)**. Accordingly, the hiring or the profit-maximising principle with respect to a particular input is that

$$MRP = MPP \times MR = \text{Input price.} \tag{16.3}$$

This is a more general condition than (16.1), stating that an input is paid its marginal revenue product. (As a special case, if the firm is competitive, then *MR* = Price and thus the *MRP* reduces to the *VMP* of an input.) We can draw an *MRP* curve, similar to the *VMP* curve. Remember that the marginal revenue decreases with output. Thus, as long as *MPP* falls with the employment of an input (due to diminishing returns), *MRP* will also fall, implying that the *MRP* curve will be downward sloping. Just as the downward portion of the *VMP* curve is the demand curve of an input for a competitive firm, the downward portion of the *MRP* curve is the demand curve of an input for a non-competitive firm in the product market.

Note that an increase in the demand for the product would shift the *MR* curve to the right. This would shift the *MRP* curve and the demand curve of an input to the right. Thus, the notion of derived demand is the same whether the firm under consideration is competitive or non-competitive in the product market.

So far we have looked at the demand side of any arbitrary input. Remember from Chapter 8 that there are three primary factors of production, namely, labour, capital and land. Each one of these has its own typical features. Next, we analyse these primary factors individually including their supply side and equilibrium in the respective market.

LABOUR

There are several kinds of labour. For qualitative analysis, however, we can classify them into two broad categories—skilled labour and unskilled labour. Typically, those who are in the labour market and have not gone to college or any professional school are termed as unskilled labour, while the rest are called skilled labour. Within skilled labour there are various types. Someone is a chemical engineer, someone else is a professional dress designer and so on.

In principle, if we horizontally add up the demand curve of a particular kind of labour by various firms in the economy, we obtain the total demand curve of that kind of labour. In Figure 16.6 it is shown as the downward sloping line *DD*. (Ignore the dotted line for now.)

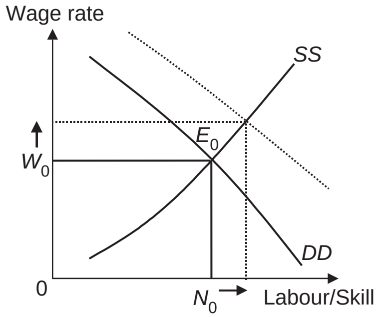
An increase in the wage rate of a particular skill would increase the number of individuals acquiring that skill. (For example, in the nineties, the lucrative salary of IT professionals led to a huge increase in the number of students specialising in IT related subjects.) Thus, the supply curve of a particular skill is upward sloping as shown by the *SS* curve Figure 16.6.

The intersection of the *DD* and the *SS* curves defines the equilibrium with W_0 as the equilibrium wage and N_0 as the equilibrium level of employment.

We can now determine how an increase in the product demand or product price affects the wage rate and employment. An increase in the product price would increase the derived demand for labour. It would shift the *DD* curve to the right. Let the dotted downward sloping line in Figure 16.6 represent the new demand curve. As a consequence, both the wage rate and the employment level increase.

In reality, over time the demand for the same kind of skill may very well change, causing a change in its reward. For instance, in the nineties, there was an

Figure 16.6 Demand, Supply and Market Equilibrium for a Particular Skill



incredible increase in the demand for computer software personnel. As a consequence, the wages in this sector saw a phenomenal increase. More recently, however, the demand for software professionals has somewhat slackened, which has depressed wages and employment for such personnel.

The demand-supply approach can also be used to understand why, at a given point of time, one type of occupation or skill commands a higher wage than some other type of skill. For instance, we consider skilled labour versus unskilled labour. First of all, skilled labour is able to perform relatively important tasks in the production process that unskilled labour cannot. Thus, the *MPP* of skilled labour is a higher than that of unskilled labour, that is, the derived demand curve of skilled labour will lie to the right of that of unskilled labour. Second, in a populous country like India, it is likely that a relatively smaller fraction of population gets college or professional education, that is, the supply of skilled workers is less than that of unskilled workers. Both these factors imply that skilled labour earns a higher wage.

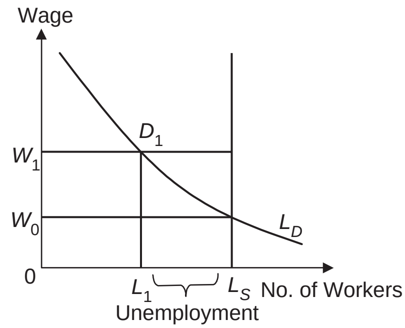
Trade Unions and Unemployment

There are hosts of other issues relating to labour. Indeed, **labour economics** is a field in its own right within the science of economics. In our limited treatment here, we consider one more feature of the labour market, namely, **trade unions** or **labour unions.**

These are workers' organisations in various sectors of the economy. They voice grievances of workers in a collective way. Sometimes they organise strikes and boycott work for days and weeks. They also try to bargain for higher wages than what the employers are willing to offer. The last aspect can be interpreted as wage-fixing or wage-setting by labour unions above the equilibrium (market) rate.

What effect does such wage-setting by trade unions have on the labour market? Turn to Figure 16.7, where L_S denote the total number of workers. Assuming for simplicity that L_S is given, the vertical line at L_S is the supply curve of labour. The

Figure 16.7 Labour Unions and Unemployment



demand curve for labour is denoted by L_D. If there were no wage-setting by trade unions, the intersection of the labour supply and labour demand curves would have determined the market-clearing wage at W_0.

Instead, suppose that the trade union bargains for the wage at W_1, higher than the market wage W_0. As a result, the firms will demand less labour. This is indicated at the point D_1 on the labour demand curve, or equivalently, at the point L_1 on the horizontal axis. We see now that there is *unemployment* of labour; L_1L_S measures the number of workers who are unemployed.

Thus, the wage-setting by the trade unions, on its own, contributes to the incidence of unemployment in an economy.[5]

CAPITAL AND LAND

A basic distinction between a factor like labour and other non-human factors like capital and land is that the latter have a user value and an asset value.

User value refers to the rental earning of a piece of equipment or land. It is the reward for the service rendered. If you already own a truck and you are getting Rs 70 per kilometre when you rent it out, this is the rental income or the user value of the truck. On the other hand, one can own a truck as an asset, which has a buying/selling price. If you own a trucking company and buy a second-hand truck at Rs 1 lakh, then this is the asset price. As another example, a 1500 sq. ft flat can be rented out in a central location in Delhi for Rs 40,000 a month, which is its rental income or user value; whereas if the buying/selling price of that flat is Rs 1.5 crore, then this is its asset value. Thus, **asset value of capital or land** refers to the buying/selling price of capital or land as an asset. As we noted in Chapter 7, an asset essentially generates a stream of income over time, which in the present context is the user value over time. In the absence of speculations, the asset value should be equal to the discounted sum of user values over time.

[5]The European economy is a prominent example of this.

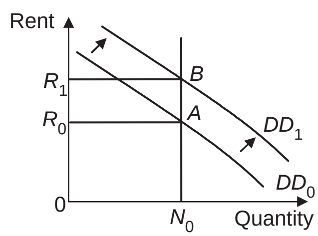
How is the user value or rent of capital/land determined? The answer is that it is similar to that of wage in a labour market. There is a demand curve for capital/land along its *VMP* curve; there is a supply curve. The intersection of the two determines the equilibrium rent. There is, however, a difference between capital and land. Capital is reproducible in the sense that its supply can constantly change. (Trucks are manufactured just as TVs or toys.) Thus, it is reasonable to suppose an upward sloping supply curve of capital. Equilibrium in the market for capital will then look like Figure 16.6.

But this is not the case with land. Its supply is naturally given.[6] Because the supply of land is fixed, the land rent is entirely governed by the demand for land. In Figure 16.8, the fixed supply of land is indicated by the vertical line at N_0. Let DD_0 denote the original demand curve for the use of land in production. The equilibrium rent is R_0. Suppose there is an increase in the demand for the use of land to DD_1. The new equilibrium rent is now R_1. The point to note is that there is no adjustment in payment to land because of any supply changes.

Note that the word 'rent' derives from a more general term, **economic rent**, defined as the excess of payment to a factor over the minimum amount, necessary to induce its supply. Put differently, economic rent is the difference between the payment to a factor and its opportunity cost. In the case of land, the supply is unchanged irrespective of whether the demand for land is high or low. Thus, the entire reward to land (except for the expenses required for the maintenance of it) is normally considered as its economic rent. In Figure 16.8, if the demand curve is given by DD_0, the whole area $0R_0AN_0$ is the economic rent. Similarly, if the demand curve is DD_1, the economic rent is equal to the area $0R_1BN_0$.

It may be remarked that when land is able to command a huge rent, it attracts a negative attention. It is thought that, unlike labour or capital, no effort or sacrifices are made to create its supply and yet someone is earning a lot because he or she happens to possess some land.[7]

Figure 16.8 Land Rent



[6]Of course, the supply of land for residential and business purposes in urban areas grows as rural and farm lands are gradually converted to these purposes. In some countries (like Hong Kong and Japan), some land is 'reclaimed' from the sea. But these processes are rather slow. Unless we are talking about what happens over a fairly long period of time, the supply of land can be treated as fixed.

[7]In the late 19th century, Henry George almost won the election for the mayor of New York on the platform that the government should be financed by a single tax, namely, tax on landlords only.

Economic Facts and Insights

- Profit-maximisation implies that inputs or factors are paid their marginal revenue product. If the firms are competitive, then the inputs are paid the value of their marginal product (since marginal revenue equals price).
- Because product market conditions affect the demand for factors, factor demand is called derived demand.
- A firm may be small (competitive) or large (for example, a monopoly) in the product market but small in an input market as a buyer.
- Compared to unskilled labour, skilled labour is generally paid more because its marginal productivity is greater (and, in a populous and relatively poor country like India, the supply of skilled labour is less than that of unskilled labour).
- In the last two–three decades, the whole world economy has experienced technological progress that has increased the marginal product of skilled labour, while its impact on that of unskilled labour is not clear.
- Wage-setting behaviour of labour unions contributes to the unemployment problem.
- Non-human factors like capital and land have asset value and user value (rent).
- Generally capital is reproducible, while land is not. Land, therefore, earns pure economic rent.

EXERCISES

16.1 What are the dissimilarities between product and factor markets?
16.2 Why is demand for a factor called 'derived demand'?
16.3 The *TPP* schedule of a factor is given in the following table.

Factor Employment	*TPP (units)*
0	0
1	8
2	18
3	26
4	35
5	41
6	45
7	48
8	50

Derive the *MPP* schedule.

16.4 Referring to the previous question, if the product price is Rs 10, what is the *VMP* when employment is increased from 5 units to 6 units?
16.5 Why is the *VMP* curve inverse-U shaped?
16.6 For a competitive firm in the product market, what is the profit-maximising rule for hiring an input? Give economic reasons.
16.7 What is the relationship between the *VMP* curve and the factor demand curve? What is the underlying reason?
16.8 'A change in the factor price and a change in technology both shift a factor demand curve'. Defend or refute.
16.9 How will a technology change that enhances the *MPP* of an input shift its demand curve?
16.10 What is the marginal productivity theory of distribution?
16.11 What is marginal revenue product?
16.12 How do the rules for hiring an input differ between a monopoly firm and a perfectly competitive firm in the product market?
16.13 Why are skilled workers paid more than unskilled workers?
16.14 What do trade unions do?
16.15 Show how wage-setting by trade union creates some unemployment whether the labour supply curve is vertical or upward sloping.
16.16 What is meant by user value of a factor like capital and land? What is meant by their asset value? What is the difference between the two?
16.17 In the modern age, unfortunately, some kinds of labour are bought and sold as capital and land. Think of bonded labour or illegal trade in human beings, including children. What is the nature of user value in this context? Suppose the government takes active steps to curb such activity. How would it affect the user value and asset value of such labour?
16.18 What does economic rent mean?

General Appendix

In economics we consider a consumer as someone who wants to maximise her satisfaction or utility, subject to her budget constraint (as in Chapter 4). We consider a producer as someone who wants to minimise the total cost of producing a given amount and maximise profits. Mathematically, these examples refer to maximising or minimising a function.

SINGLE-VARIABLE CASE

Consider the simplest case of y being a continuous and twice-differentiable function of a single variable x, that is, $y = f(x)$, where f is a continuous function and the derivatives f' and f'' exist. Suppose that $x \in [a, b]$. Figure A.1(a) illustrates the case where a function f is maximised in the interior of the interval of (a, b), that is, in the open interval (a, b). Notice that at the point where $f(x)$ is maximised, that is, at x_0, the slope of the function is zero and around x_0, the slope declines gradually (from being positive to zero and from zero to negative), as x increases. Thus, the conditions for the maximisation of $f(x)$ are that:

$$f'(x) = 0 \tag{A.1}$$

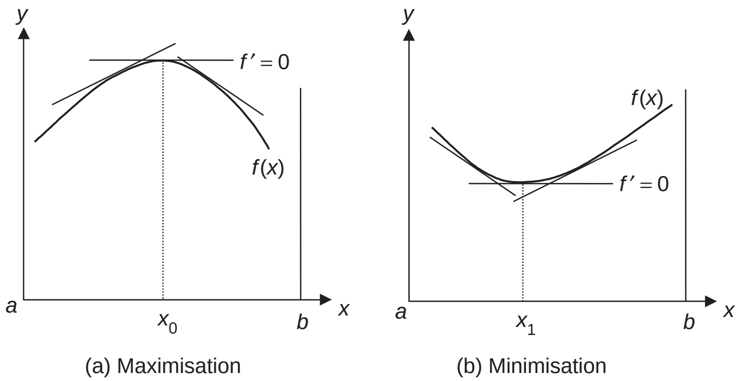
$$f''(x) < 0. \tag{A.2}$$

Figure A.1(b) illustrates the case where a function $f(x)$ is minimised at x_1. Notice that at this point the slope of the function is zero and, around x_1, the slope increases (from being negative to zero and from zero to positive), as x increases. Thus, the minimisation conditions are that:

$$f'(x) = 0 \tag{A.3}$$

$$f''(x) > 0. \tag{A.4}$$

Figure A.1 Maximisation and Minimisation of a Function



Equations such as (A.1) and (A.3) are called the **first-order conditions**. Thus, we see that the first-order condition of maximisation or minimisation of a function is the same, namely, $f'(x) = 0$. Inequalities like (A.2) and (A.4) are called the **second-order conditions**, which, of course, differ between maximisation and minimisation of a function.

MANY-VARIABLES CASE

We consider here the first-order condition only.

Suppose y is a function of many variables: $x_1, x_2, x_3, \ldots, x_n$. It is maximised with respect to all n variables and the maximum occurs in the interior of the intervals to which $x_1, x_2, x_3, \ldots, x_n$ respectively belong. Then the first-order conditions are generalisations of (0.1), that is, each partial is equal to zero:

$$\frac{\partial f}{\partial x_1} = 0, \frac{\partial f}{\partial x_2} = 0, \ldots, \frac{\partial f}{\partial x_n} = 0 \tag{A.5}$$

These are also the first-order conditions of a minimisation problem, that is, if the function $y = f(x_1, \ldots, x_n)$ is minimised with respect to $x_1, x_2, x_3, \ldots, x_n$.

CONSTRAINED OPTIMISATION

In microeconomics, we consider situations where a function $y = f(x_1, \ldots, x_n)$ is maximised or minimised, subject to a constraint: $g(x_1, \ldots, x_n) = 0$. What are the first-order conditions? We introduce a function called the Lagrangian, defined as

$$L = f(x_1, \ldots, x_n) + \lambda g(x_1, \ldots, x_n),$$

where λ is an 'extra' variable, called the Lagrangian multiplier. We have $L = L(x_1, ..., x_n, \lambda)$. Assuming interior solution once again, the first-order conditions are that:

$$\frac{\partial L}{\partial x_1} = \frac{\partial f}{\partial x_1} + \lambda \frac{\partial g}{\partial x_1} = 0, \text{ or } \frac{\partial f / \partial x_1}{\partial g / \partial x_1} = -\lambda \tag{A.6}$$

$$\frac{\partial L}{\partial x_n} = \frac{\partial f}{\partial x_n} + \lambda \frac{\partial g}{\partial x_n} = 0, \text{ or } \frac{\partial f / \partial x_n}{\partial g / \partial x_n} = \lambda \tag{A.7}$$

$$\frac{\partial L}{\partial \lambda} = g(x_1, x_2, ..., x_n). \tag{A.8}$$

We can eliminate λ from the above equations and obtain

$$\frac{\partial f / \partial x_1}{\partial g / \partial x_1} = \frac{\partial f / \partial x_2}{\partial g / \partial x_2} \cdots = \frac{\partial f / \partial x_n}{\partial g / \partial x_n}. \tag{A.9}$$

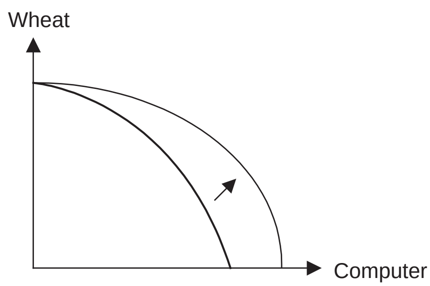
Equations (A.8) and (A.9) are also the first-order conditions without the Lagrangian multiplier.

Partial Answers to Selected Questions

CHAPTER 1

1.3 It is a 'positive' question because it is not concerned with whether associating with the WTO is good or bad for the agriculture sector.

1.5



(You need to argue why such a shift occurs.)

1.8 No; you refute.

1.10 It will shift out (to the right).

1.13 (a) Bribe Giving: The major benefit is the (value of) time and trouble saved. The cost is the amount of money paid as bribe.

1.16 Providing jobs means that the government must pay in terms of salaries and such like. From where does the government get the money? It has to be collected from the tax-payers. Collected revenues have alternative uses like maintaining national defence and building infrastructure. Thus, the need for tax revenues for other purposes is the constraint.

CHAPTER 2

2.2 Both goods are not normal. Seeing movies is a normal good, while eating out in a *dhaba* is an inferior good.

2.6 The demand for land will shift to the right. Income and market size as determinants of market demand are at work here.

2.11 No. The opposite is true.

2.13 Gupta family's demand for petrol is elastic (that is, the price elasticity is greater than one).

2.17 Since total spending is the product of price and quantity demanded, percentage increase in total spending = percentage increase in price + percentage increase in the quantity demanded. In this example, percentage increase in total spending = 15 and percentage increase in price = 20. Hence percentage increase in quantity demanded = 15 – 20 = –5 percentage, that is, there is a 5 percentage decrease in quantity demanded. Applying the formula, price elasticity = 5/20 = 1/4.

CHAPTER 3

3.2 It will imply a change in supply. The supply curve will shift to the right because the cost of one input used by the manufacturer has decreased.

3.4 The supply curve will shift to the right.

3.6 The individual supply curves will remain unaffected (although, because of foreign entry, after price adjustment, they would move down along their individual supply curves). The market supply curve will shift to the right.

3.8 No; you refute.

3.10 If both curves are rays from the origin, both have the same price elasticity, equal to one. If they intersect, at the point of intersection, the steeper curve is less price elastic.

3.12 The supply curve will shift to the right.

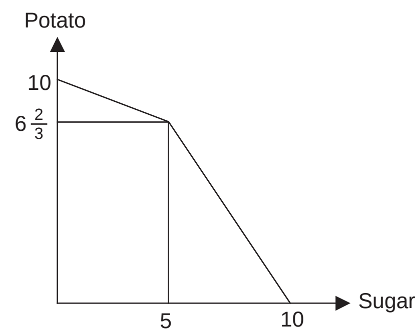
CHAPTER 4

4.2 The marginal utility schedule in terms of rupees or money is the following.

Minutes of Call	*MU in Rupees*
0	–
1	50/3 = 16 2/3
2	45/3 = 15
3	40/3 = 13 1/3
4	35/3 = 11 2/3
5	30/3 = 10
6	25/3 = 8 1/3
7	20/3 = 6 2/3
8	15/3 = 5

Since the price is Rs 8, the answer is 6 minutes.

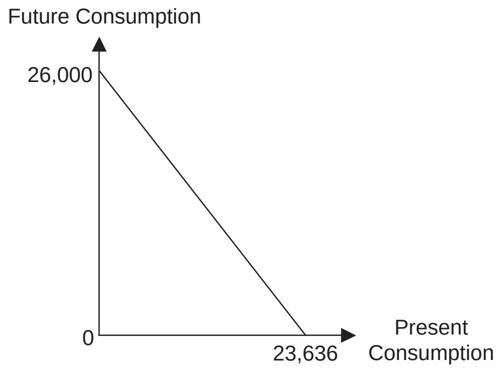
4.3 This family should consume more carrot *halwa* because the marginal utility (from consuming 3.5 kg of this good) in terms of rupees = 100/2 = Rs 50, which is greater than the price (Rs 40).
4.6 5 chicken rolls and 4 ice-creams. You have to work out the details.
4.9 Non-satiation.
4.10 Diminishing marginal rate of substitution.
4.12 Marginal rate of substitution is the ratio of marginal utilities.
4.14 It will look downward sloping.
4.17 It is because the prices are fixed irrespective of the quantity of a good purchased.
4.21 The budget line will not shift in any way.
4.23



4.29 By the substitution effect the quantity demanded for it will increase, while by the income effect the demand for it will decrease.
4.31 No; you refute.
4.36 It does not consider the income effect of a price change.
4.39 No; you refute.
4.40 At original prices, the consumer buys 12 units of good B (along with 5 units of good A). At new prices, this bundle will cost $5 \times 5 + 12 \times 5$ = Rs 85. Thus, the consumer's income has to be decreased by Rs 100 – 85 = Rs 15, so that he/she can buy the old bundle.

CHAPTER 5

5.2 She should prefer cash (according to the theory).
5.3 Both lead to parallel shifts; the former shifts the budget line to the right, whereas the latter shifts it to the left.
5.7 The consumer will prefer the indirect tax programme.
5.13 It will look as follows. The slope = 1.1.



5.16 $\left(\frac{224}{200}-1\right)\times 100 = 12\%.$

CHAPTER 6

6.1 It will be negatively sloped.
6.3 No. It can be a Giffen good.
6.5 It is the effect of (other) factors not included in the regression.
6.7 One intercept coefficient and several slope coefficients.
6.8 A positive sign.
6.14 In every year the sales peak around October (a combination of September–October or October–November) because of the concentration of holidays and festivals.

CHAPTER 7

7.1 (*a*) initial (cost of) investment, (*b*) its yield and (*c*) future sale value of the asset.
7.4 Here the figure \$2,000 is immaterial since prices are given. Let ρ denote the rate of return, which is like an internal compound interest rate. Thus, $520\,(1+\rho)^2 = 566$. Solving this, $\rho = 0.043$ or 4.3%.
7.5 The expected return on Stock A = $0.05 \times 0.6 + 0.09 \times 0.4 = 0.03 + 0.036 = 0.066$, that is, 6.6%. The expected return on Stock B = $0.04 \times 0.3 + 0.10 \times 0.7 = 0.012 + 0.07 = 0.082$, that is, 8.2%.
7.7 We have

$$\begin{aligned}\text{PDV} &= \frac{4}{1.1}+\frac{4}{(1.1)^2}+\frac{4}{(1.1)^3}+\frac{4}{(1.1)^4} = 3.636+3.305+3.005+2.732\\ &= 12.678 \text{ lakh, that is, Rs } 12{,}67{,}800.\end{aligned}$$

CHAPTER 8

8.10 It is because of diminishing marginal returns.
8.11 Yes; you defend.
8.13 It is a case of diminishing or decreasing returns to scale.
8.14 (*a*) and (*b*): increasing returns to scale; (*c*) and (*d*): constant returns to scale; (*e*): decreasing returns to scale.
8.16 The marginal product is positive.
8.19 It is because of diminishing marginal rate of technical substitution.
8.22 Yes; you defend.
8.24 The slope is equal to the ratio of input prices.
8.27 The firm should hire more of capital and less of labour.
8.28 No; you refute.
8.29 The total cost will increase.
8.32 No; you refute.
8.35 It is increasing, constant or decreasing returns to scale if $\alpha + \beta > 1$, $\alpha + \beta = 1$ or $\alpha + \beta < 1$ respectively.

CHAPTER 9

9.2 No; you refute.
9.4 No; you refute.
9.5 Total variable cost.
9.8 Less.
9.10 *TFC* = 10. The *MC* schedule is as follows:

Output	*MC (Rs)*
0	–
1	15
2	10
3	8
4	11
5	15
6	22
7	30
8	40

9.12 It would shift these cost curves up.
9.13 It shifts these cost curves down.
9.16 The *LMC* curve is horizontal and same as the *LAC* curve.
9.19 Yes; you defend.
9.21 Positive (respectively negative) externalities mean that the activity of one economic agent directly improves (respectively hurts) the interest/welfare of another economic agent.

9.23 If we think of India's welfare only, there is no difference between private cost and social cost because the pollution being caused affects Kenya, not India.

9.26 Issuing bonds is riskier because the firm has to pay the bondholders (who are the creditors) even when it is making losses. But, on the other hand, if the business is good and there are profits, servicing of bonds is less costly. There is no pressure to raise funds by equity (stocks), which amounts to stockholders taking additional risk.

CHAPTER 10

10.2 Revenues = Rs 10 lakh
Accounting costs = 2 + 3 + 4 = Rs 9 lakh
Accounting profit = 10 – 9 = Rs 1 lakh
Implicit cost = 2 lakh
Economic profit = 10 – 9 – 2 = – 1 lakh, that is, it is a loss of Rs 1 lakh.

10.6 (*i*) There are a large number of firms and (*ii*) the product is homogeneous.

10.7 It is a straight line through the origin because the market price is the same, irrespective of how much a firm produces and sells in the market. Put differently, it is a straight line because the firm is a price taker.

10.8 Because the firm is a price taker.

10.10 Because a marginal increase in output from the point where $p = MC$ holds and MC decreasing in output would fetch higher profit.

10.12 No; you refute.

10.15 No; you refute.

10.21 In the long run, each firm produces 20 units and the market price = minimum LAC = 50. From the demand schedule, at this price the quantity demanded = 1,000 units. Thus 1,000 units are sold at the price Rs 50. Since each firm produces 20 units, the number of firms must equal 1,000/20 = 50.

10.25 The total cost function $y^2 - 10y + 28$ implies the marginal cost function

$$MC = \frac{d}{dy}\left(y^2 - 10y + 28\right) = 2y - 10.$$

If the market price is 20, then $p = MC$ condition is: $20 = 2y - 10$. This gives $y = 15$.

CHAPTER 11

11.3 Yes; you defend.

11.5 The market price will increase, while the quantity transacted may increase, remain unchanged or decrease.

11.7 The price will increase and the quantity transacted will fall.

11.9 It will lower the price and quantity produced of domestic steel. Steel being an input in producing bicycles, a decline in the price of steel means a decline in the price of an input used in producing bicycle. This will shift the supply curve to the right. As a result, the price of bicycles will fall.
11.13 The rupee will appreciate (as foreigners would like to invest more in India and, in the process, would demand more rupees in the currency market).
11.14 Rupee will appreciate.
11.16 Right, depreciate.
11.23 It is because there is excess demand at any below market-clearing price.
11.26 No; you refute.
11.27 $p = 3$ and $q = 7$.
11.28 0.

CHAPTER 12

12.2 $p = MC$.
12.3 Yes; you defend. There are, however, conditions like no externalities or public goods.
12.4 Agree; you have to write the reasons.
12.8 Less.
12.9 No; you have to articulate the reason.
12.11 Disagree; you have to give reasons.
12.12 Free-rider problem.
12.15 Second-best policies refer to those, which achieve their intended objectives but at a higher social cost compared to some other policy that also achieves the same objective. You have to provide an example.
12.16 It refers to the situation where a particular type and magnitude of government policy serves the society the best, but either the government chooses a less efficient policy or the correct type of policy with too less or too high degree of intervention.

CHAPTER 13

13.1 Both will switch on their booster pumps.
13.2 Zero-sum game. (a_2, b_2).
13.4

Yusuf	*You*	
	Head	Tail
Head	(–1, 1)	(1, –1)
Tail	(1, –1)	(–1, 1)

13.5 Yes; 2. You have to identify them.

CHAPTER 14

14.1 Monopoly, monopolistic competition and oligopoly. Apart from these, there are others too like duopoly, duopsony and oligopsony.)
14.4 A firm whose long-run average cost curve is downward sloping (relative to demand).
14.6 No; you refute.
14.9 No; you refute.
14.11 Smaller.
14.13 As long as $MC > 0$, the condition $MR = MC$ implies that $MR > 0$. That is, at the monopoly equilibrium, an increase in the quantity sold brings more total revenues. From the consumers' perspective, this is same as saying that lowering price (and buying more) implies a higher total expenditure. From the relationship between price change and total expenditure discussed in Chapter 2, this can happen if and only if the price elasticity of demand is greater than one.
14.16 Yes; you defend.
14.20 Price discrimination means that consumers in different markets are being charged different prices for the same good.
14.21 There is no scope for the consumers in the low-price market to resell the product in the high-price market.
14.23 The greater the price elasticity in a market, the less is the price.
14.25 If it is natural monopoly, we have the average cost curve falling with output. This implies $MC < AC$. Hence, if the regulated price, p, is equal to MC, we have $p < AC$, which implies that the firm would incur a loss (at such a regulated price).
14.26 Yes, it will happen if the average cost increases with output. In this case, we have $MC > AC$. Hence, if $p = MC$, then $p > AC$; this means abnormal profit.

CHAPTER 15

15.2 (*a*) No. (*b*) New firms will enter.
15.3 less, 4 units.
15.4 No; you refute.
15.6 No; you refute.
15.8 Quantity sold by the firm from AP will decrease and that by the local firm will increase.
15.9 Market price = Rs 7.50 and quantity sold = 10,000 − 250 × 7.50 = 8,125.
15.10 The best response function of the CNG-run taxi stand will shift to the left. The amount produced by this taxi-stand will decrease and that by the other taxi-stand will increase.
15.11 15.

CHAPTER 16

16.2 It is because the demand for a factor depends on the demand for the product, which the factor (among other factors) helps to produce.

16.4 Rs 40.

16.5 It is because of diminishing marginal returns.

16.8 No; you refute.

16.9 It will shift to the right.

16.13 It is because the marginal product of skilled workers is higher and their supply is less than that of unskilled workers.

16.16 User value is simply the rental value. Asset value refers to the purchase or sale value of capital or land.

16.17 User value in this context is the value of labour services obtained. Suppose bonded labour is used for household work. Then it is the value of these services. If instead, it is used in a factory or some other production activity, then the user value is the *VMP* or the marginal revenue product. If the government cracks down on such activity at the purchase-sale points as well as where such labour is used, the demand for such illegal labour will fall and hence the asset value and user value will both decrease.

Index

About the Author

Satya P. Das is a Professor of Economics at the Statistical Institute, New Delhi. Previously, he was a Professor of Economics at Indiana University and University of Wisconsin-Milwaukee, USA, and has held visiting positions at various universities in different countries. He has authored the Class XII NCERT Textbook *Introductory Microeconomics* (2003) and the monograph *New Perspectives on Business Cycles: An Analysis of Inequality and Heterogeneity* (1993), and is a regular contributor to journals in economics. His current research interests include Economic Modelling of Terrorism, Impact of International Trade on Product Quality as well as various issues relating to Growth, Development and Distribution. He is the founding editor of *Indian Growth and Development Review*.