The McGraw-Hill Companies

# MANAGERIAL ECONOMICS

The McGraw-Hill Companies

# MANAGERIAL ECONOMICS

## SECOND EDITION

**G S Gupta**
*Former Professor*
*Indian Institute of Management*
*Ahmedabad*

**Tata McGraw Hill Education Private Limited**
NEW DELHI

*McGraw-Hill Offices*
**New Delhi** New York St Louis San Francisco Auckland Bogotá Caracas
Kuala Lumpur Lisbon London Madrid Mexico City Milan Montreal
San Juan Santiago Singapore Sydney Tokyo Toronto

The McGraw·Hill Companies

Tata McGraw-Hill

Published by the Tata McGraw Hill Education Private Limited,
7 West Patel Nagar, New Delhi 110 008

**Managerial Economics, 2/e**

Copyright © 2011, by Tata McGraw Hill Education Private Limited. No part of this publication can be reproduced or distributed in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise or stored in a database or retrieval system without the prior written permission of the publishers. The program listings (if any) may be entered, stored and executed in a computer system, but they may not be reproduced for publication.

This edition can be exported from India only by the publishers,
Tata McGraw Hill Education Private Limited

ISBN-13 : 978-0-07-106786-7
ISBN-10 : 0-07-106786-8

Vice President and Managing Director—McGraw-Hill Education: Asia Pacific Region: *Ajay Shukla*
Head—Higher Education Publishing and Marketing: *Vibha Mahajan*
Publishing Manager—B&E/HSSL: *Tapas K Maji*
Deputy Manager (Sponsoring): *Surabhi Khare*
Senior Development Editor: *Shalini Negi*
Executive (Editorial Services): *Yogesh Kumar*
Senior Production Manager: *Manohar Lal*
Senior Production Executive: *Atul Gupta*
Marketing Manager: *Vijay S Jagannathan*
Assistant Product Manager: *Daisy Sachdeva*
General Manager—Production: *Rajender P Ghansela*
Assistant General Manager—Production: *B L Dogra*

Information contained in this work has been obtained by Tata McGraw-Hill, from sources believed to be reliable. However, neither Tata McGraw-Hill nor its authors guarantee the accuracy or completeness of any information published herein, and neither Tata McGraw-Hill nor its authors shall be responsible for any errors, omissions, or damages arising out of use of this information. This work is published with the understanding that Tata McGraw-Hill and its authors are supplying information but are not attempting to render engineering or other professional services. If such services are required, the assistance of an appropriate professional should be sought.

Typeset at Print-O-World, 2579, Mandir Lane, Shadipur, New Delhi 110 008, and printed at Magic International Pvt. Ltd., Plot No. 26E, Sector-31 (Industrial), Site IV, Greater Noida 20306

Cover Design: Meenu Raghav, Graphic Designer, TMH

Cover Printer : Magic International

RAXCRRCCRQQAL

The McGraw·Hill Companies

*To the memory of my*

*Parents, Saduram and Narmada Gupta*
*and*
*Parents-in-law, Rikhabdas and Rukmini Bansal*

# PREFACE TO THE SECOND EDITION

The present edition of *Managerial Economics* is overdue. Since its first edition, published in 1990, significant developments have come up in the corporate world, such as cost minimization through mergers and takeovers, downsizing, globalization, etc., innovative methods of product pricing, including dynamic pricing, and the emergence of a variety of new products. These changes led me to realize the need for revising the first edition. Excellent feedback both from the students and faculty and persuasion from the publishers have given me enough dividends to undertake this endeavour, and it is a great pleasure to present this much-awaited second edition. This edition offers a thorough revision of each chapter, removes all ambiguities, incorporates the feedback received, and adds many new concepts, theories and applications.

## NEW TO THIS EDITION

The erstwhile two chapters on product pricing have been expanded to four chapters. In particular, measures of industry concentration, monopoly power, entry-exit barriers, pricing under natural monopoly, contestable markets, game theory, market power (dynamic pricing), asymmetric information, externalities, risk and uncertainty have been included.

An entirely new chapter on Theory of Consumer Behaviour (Chapter 3) has been added. This covers the ordinal or indifference curve approach to consumer behaviour, which has become so crucial not only to explain the choice of consumption basket but also to demonstrate leisure-work choice and thereby the labour supply, consumption-saving trade-off, return-risk trade-off, among other such behaviours.

Many new concepts, jargons and fundamentals have been introduced. These include market failures, public goods, common goods, merit goods, bandwagon and snob effects, sunk cost, duality theorem in production and cost, economies of scope, learning curve, transaction cost economies, backward bending labour supply curve, Lerner index, horizontal and vertical integrations, reservation price, consumer, producer and economic surpluses, dead weight loss, bilateral monopoly, economic rent, transfer pricing, adverse selection, moral hazard, principal-agent problem, Coase theorem, etc.

Case studies have become an important pedagogy even in economics. Though no real-life cases have been added, real-life examples of dynamic pricing, industry competition, mergers and takeovers, etc. have been cited at appropriate places. Also, the chapter-end cases have been expanded to practice micro economics' fundamentals.

To help students test their understanding of the fundamentals of managerial economics as well as to provide ready material for short quizzes to instructors, multiple choice questions and fill-in-the-blanks exercises have been added.

## ACKNOWLEDGEMENTS

I am grateful to the students and faculty alike who have sustained the textbook for over two decades and provided heartening feedback and suggestions for revision. Further, since the release of the first edition, I have taught managerial economics at a number of institutions, including IIM Ahmedabad, Illinois State

University (USA), Universiti Sains Malaysia, Mudra Institute of Communication, Entrepreneurship Development Institute of India and IIM Raipur, besides delivering guest lectures on selected topics at a few other institutions. Interactions with the students and faculty at these institutions have improved my own understanding of the subject and I am thankful to all of them.

The entire revision work was carried out during my almost five-month stay (April 2010–August 2010) with my children and grandchildren in Dubai, Dallas and Houston. The quality time that I spent with the family was doubly rewarded as I could concentrate on this revision during 9–5 on week days and had the pleasure of living with them all the time. The amicable environment and the required infrastructure were fully available, and there were no interruptions which are so unavoidable at otherwise workplaces. Interactions and playing with kids in evenings and weekends were so refreshing that it is hard to describe. At times of need, help on computer matters was readily available from my well-versed children. My wife, Lalita, provided all logistic and moral support, and instead of complaining much on depriving her on quality time with me, she encouraged me to work on this long-pending project. My children and grandchildren took pleasure in canvassing with their friends as how I was working on such a noble project.

The publishers, Tata McGraw-Hill, deserve sincere thanks for pursuing me for almost a decade to work on this edition, assuring full cooperation. In particular, I appreciate the encouragement from Tapas K. Maji and Vibha Mahajan to work on this edition and the help that I have received from Tapas K. Maji, Manohar Lal, and others in converting the soft copies of chapters into this finished product.

I hope that the new edition is received well both by the students and faculty of managerial economics. Needless to say, I solicit feedback from all readers of this book.

**G.S. Gupta**

# PREFACE TO THE FIRST EDITION

Managerial economics is concerned with decision making at the managerial level. The book presents the alternative theories of firm behaviour, decision making problems and different approaches to arrive at the most appropriate answers to such problems. This is accomplished through a brief and logical discussion of the various relevant concepts and techniques and, through hypothetical examples, to illustrate the decision-making process. At the end of each chapter, small caseletes are incorporated for students to work on in order to ascertain their understanding of the material presented in the corresponding chapter. The approach adopted is a good mix of verbal language, geometry and algebra. While the emphasis is on theory and normative answers, numerous practical examples are cited and integrated in the material presented in the text.

Managerial economics draws heavily from micro-economics, econometrics and operations research. Accordingly, the book borrows the relevant concepts from all these disciplines and also some from macro-economics. The approach of the text is to first pose problems relating to the decision-making process and then seek answers. The book does not give a detailed exposition of micro-economics or of any other disciplines,rather it accords a synthesis of all the relevant aspects of the related subjects to arrive for decision-making problems faced by firms.

The book is written primarily for students pursuing post-graduate courses in management. It should also be useful to those who are doing M Com and CA. Since many instructors of economics emphasize applications, the book could serve as a good reference book for those students who are doing MA in Economics as well.

I have co-authored another book on the subject-Mote, Paul and Gupta, Managerial Economics: Concepts and cases, Tata McGraw-Hill, New Delhi. The earlier volume, written in 1976, is concise and includes case studies. The present text which borrows nothing from the earlier one, is uptodate, quite elaborate, presents all the aspects of various topics concerning managerial economics and provides caselets on all decision-making aspects as practice exercises for students. Readers interested in real-life case studies and/or in concise material should still find the old text useful.

In the writing of this text, I have benefitted from the comments and suggestions which I have been receiving from my students at the Indian Institute of Management, Ahmedabad (1970-till date), Illinosis University, USA(1982-83,1985-86), and at several other educational institutions where I have delivered guest lectures. I express my thanks to all of them. Owing to my desire to bring out the present volume early, I had planned to provide caselets either in a separate hand-out or in the subsequent edition of this text. I am grateful to the referee for advising me to include the caselets here itself. A special word of thanks is due to B Sreekumar, K Ram Babu and T Sridham for efficiently typing the manuscript and to GD Patel for his able assistance in drawing graphs. Lastly, I appreciate the assistance of my wife, Lalita and the patience and cooperation of my children, Indu Jaya and Manish.

**G.S. Gupta**

The McGraw-Hill Companies

# CONTENTS

*Preface to the Second Edition* *vii*
*Preface to the First Edition* *ix*

**1. Introduction** **3**
1.1 Economics *3*
1.2 Managerial Economics *6*
1.3 The Firm: Objectives and Constraints *6*
1.4 Decision Process *14*
1.5 Basic Principles *15*
1.6 Chapter Plan *19*
*References* *19*

**2. Demand and Revenue Analysis** **23**
2.1 Meaning of Demand *23*
2.2 Types of Demand *25*
2.3 Determinants of Demand *30*
2.4 Demand Function *36*
2.5 Demand Elasticities *39*
2.6 Demand-Revenue Relationships *46*
*References* *50*
*Caselets* *51*

**3. Theory of Consumer Behaviour** **55**
3.1 Consumers' Preferences *55*
3.2 Consumers' Budget Constraints *61*
3.3 Consumers' Choice *63*
*References* *68*
*Caselets* *68*

**4. Demand Estimation, Analysis and Forecasting** **71**
4.1 Demand Estimation *72*
4.2 Analysis of Estimated Demand Function *80*
4.3 Demand Forecasting *84*
*References* *107*
*Caselets* *107*

The McGraw-Hill Companies



**5. Production Analysis and Inputs' Demand** **113**
5.1 Meaning of Production *113*
5.2 Production Function *113*
5.3 Production Analysis: Long-Run *120*
5.4 Production Analysis: Short-Run *131*
5.5 Production Analysis: Long-Run versus Short-Run *137*
5.6 Elasticity of Factor Substitution *138*
5.7 Production Analysis and Input Demand: A Generalization *139*
*References* *141*
*Caselets* *141*

**6. Cost and Supply Analysis** **147**
6.1 Cost Concepts *147*
6.2 Cost Function *153*
6.3 Cost-Output Relationship: Long-Run *155*
6.4 Cost-Output Relationship: Short-Run *161*
6.5 Cost Output Relationship: Long vs Short-Run *165*
6.6 Economies of Big Businesses *166*
6.7 Estimation of Cost Function *170*
6.8 Managerial Uses of Estimated Cost Functions *172*
6.9 Supply Function *176*
*References* *179*
*Caselets* *179*

**7. Preliminaries on Price and Output Determination** **183**
7.1 Price Concepts *183*
7.2 Price Determinants *185*
7.3 Conditions for Profit Maximisation *193*
7.4 Profit and Break-Even Analysis *194*
7.5 Pricing under Different Objectives *200*
7.6 Market Structure *201*
*References* *208*
*Caselets* *209*

**8. Price-output Determination–I (Pricing under Perfect Competition, Monopoly and Monopolistic competition)** **213**
8.1 Pricing under Perfect Competition *213*
8.2 Pricing under Monopoly *228*
8.3 Contestable Markets *237*
8.4 Pricing in Multi-Plant Firms *239*
8.4 Pricing in Multiple Products Firms *241*
8.5 Monopolistic Competition *242*
*References* *247*
*Caselets* *247*

The McGraw-Hill Companies



**9. Price–output Determination-II (Pricing under Oligopoly) 253**
9.1 Price-Output Determination Models in Oligopoly Markets *254*
9.2 Perfect (Explicit) Collusion (Cartel) Model *256*
9.3 Competition Oriented Models *260*
9.4 Tacit (Implicit) Collusion Models *263*
9.5 Game Theory *270*
9.6 Non-optimizing Models *277*
9.7 Price Rigidity *282*
*References* *285*
*Caselets* *285*

**10. Price–output Determination-III**
**(Pricing under Market Power, Asymmetric Information, Externalities and Risk) 289**
10.1 Market Power *289*
10.2 Price Discrimination *291*
10.3 Price Discrimination by Market Segments *293*
10.4 Inter-Temporal Price Discrimination *302*
10.5 Peak-Load Pricing *304*
10.6 Block Pricing *306*
10.7 Two-Part Pricing *308*
10.8 Commodity Bundling *309*
10.9 Transfer Pricing *310*
10.10 Pricing under Asymmetric Information *313*
10.11 Pricing under Externalities *317*
10.12 Pricing under Risk and Uncertainty *324*
*References* *326*
*Caselets* *326*

**11. Investment Analysis 331**
11.1 Meaning and Significance *331*
11.2 Time Value of Money *334*
11.3 Cash Flows and Measures of Investment Worth *338*
11.4 Investment Analysis *354*
11.5 Concluding Remarks *364*
*References* *365*
*Caselets* *366*

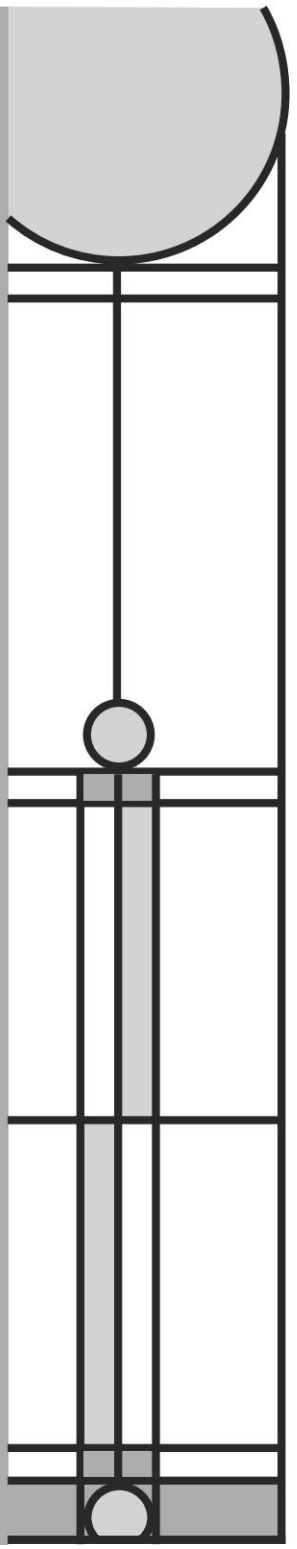
**Appendix A : Economic Optimization 369**
**Appendix B : Tables for Investment Analysis 387**
**Appendix C : Multiple Choice Questions 399**
**Appendix D : Fill in Blanks Questions 419**

**Index 425**

The McGraw-Hill Companies

# MANAGERIAL ECONOMICS

The McGraw-Hill Companies



CHAPTER 1

# INTRODUCTION

Profitability and productivity are the two important yardsticks that are applied to measure the health of an organization/firm. The significance of achieving optimum values of these is now widely recognized. The present volume concentrates on that part of economics which is directly relevant for the management of a firm.

## 1.1 ECONOMICS

The origin of the term 'economics' lies in the Greek words *oikon* and *nomos*, which mean 'laws of households'. This marks the significance of economics to households. To understand its relevance to managerial decision-making, a deeper understanding of its meaning is imperative.

Adam Smith, the father of economics, published his famous book "An Enquiry into the Nature and Causes of the Wealth of Nations" in 1776. To him, economics is concerned with material wealth. The said wealth is contained in the quantity and quality of goods and services. Thus, economics was considered as the discipline dealing with the production, exchange and consumption of goods and services. Alfred Marshall, in his book "The Principles of Economics", published in 1891, defined economics as the study of human-kind in the ordinary business of life. Lionel Robbins defined economics as a socia1 science concerned with the optimum allocation of scarce resources among competing ends. On combining all these facets, the subject has been defined as "a social science which covers the actions of individuals and groups of individuals in the processes of producing, exchanging and consuming of goods and services" (Henderson and Quandt, 1980). In brief, economics is the science of optimization under scarcity in the process of production, exchange and consumption of goods and services. As such, it deals with the estimation and comparison of benefits and costs associated with all economic decisions.

In its modern meaning, economics renders help in various matters of decision-making like:

(a) *Production decisions:* What to produce, how much of a good to produce and how to produce a given quantity of the chosen good?

(b) *Exchange decisions:* Who is the target group of buyers for each product and at what prices?

(c) *Consumption decisions:* What and how much to (buy and) consume?

On closer scrutiny, all these questions involve the problem of allocation of scarce resources among alternative ends. For example, the answer to (i) what to produce involves allocation of limited funds (expected equity and debt) among alternative investment projects, (ii) how much of a good to produce requires forecasting of demand, (iii) how to produce a good involves choice of input mix so as to maximize profit subject to given technology (production function) and prices of inputs, (iv) how much to consume of various goods requires the allocation of a given income (consumption expenditure budget) among those goods so as to maximize human satisfaction (utility), and (v) whom to sell how much involves price settings so as to maximize profit subject to given market conditions. Accordingly, economics is known as the **science of choice** under scarcity. In view of this broad definition, in addition to the above questions, economics also encompasses subjects like economic development and growth, business cycles, public finance, unemployment, poverty, inflation, money and banking, international trade and investment, and so on.

Ragnar Frisch has classified economics into two broad categories: **Microeconomics** and **Macroeconomics**. Once again the term 'micro' and 'macro' have been derived from the Greek words *Mikros* and *Makros*, which mean 'small' and 'large' respectively. Microeconomics deals with the behaviour of individual economic units and the markets where they interact. For example, it studies the decision making by a producer (firm) and a consumer, and the individual markets for each of the products they produce and buys. Thus, microeconomics talks about the behaviours of buyers and sellers of cars, quantity and price of cars, and the same for all other goods and services in the economy. In contrast, macroeconomics is concerned with the behaviour of the economy at large (like India and each of the other countries) and national aggregates, such as national income, general price index, total employment and unemployment, wage rate, money supply, interest rate, national saving and investment, fiscal deficit, exports, imports, balance of trade and foreign exchange rate, business cycles and growth rate of an economy, etc.

**Scarcity** and **uncertainty** are the two foundation stones of economics. Economics deals with the allocation of resources, which are scarce and versatile. Scarce, for they are inadequate to satisfy human needs which are seemingly unlimited. Versatile, as they could be used to satisfy alternative needs. Needs, though unlimited, differ in terms of their intensity, and thus could be ranked in the order of priority. Choice making is thus essential, which is attempted through the process of optimization.

Scarcity does not mean shortage; rather it means the demand for that resource/item at zero price exceeds its supply. In other words, anything which commands a price is a scarce item. Only scarce items are economic goods, and the rest are free goods. A commodity which is a free good today in a particular society might become an economic good tomorrow in the same society or might even be an economic good today in some other society. For example, water which was a free good all over in primitive society, has a price tag now in many cities but is still a free good in most rural areas. However, scarcity, though a necessary condition for choice, must accompany other alternatives for resource uses (versatility) to cause the problem of choice. For example, if I could do nothing else other than teach managerial economics at my present institution, then there is no problem with regard to the allocation of work to me. But the fact is, I can teach several other courses among other non-teaching activities, and at many other organizations. This complicates the situation—the problem of how to use time. In other words, because I am capable of doing alternative jobs, that is, I am versatile or I have an opportunity cost, besides I being a scarce factor, there is a problem of determining the best use of my time. Scarcity and versatility of resources lead to trade-offs, which binds all decisions-makers. For example, a consumer

has scarce and versatile income or consumption budget. If he/she spends more on one good, less is left for other goods. Similarly, an investor has scarce funds and time. The more these are spent on any project, the less is left for undertaking other projects. Also, a worker has scarce time, the longer he/she works (earns income), the less is left for leisure. Thus, all **people face trade-offs**.

Most decisions that we take have a time dimension. For example, an appropriate answer to the question 'what to produce' would require an understanding of the market for various commodities, among many other things, in future, which is nothing but uncertain. The decision makers have no alternative but to assess as best as they can this uncertain market and act accordingly. If their assessment turns out to be wide off the mark, their decisions would be far from the right. The methods of handling such uncertainties would be dealt with later.

**Markets and Governments.** All economies follow a mixed economic system today, where both the market (capitalism) and governments (communism) work hand in hand. While many goods are traded almost in a free market, many goods in markets are regulated by governments, and some goods are purely under governments monopoly. The mixed system exists because there are pros and cons of each, and different countries choose a different mix depending on its choice. In a pure market economy (capitalism), buyers and sellers are free to decide on all economic issues (like what to produce/consume) while in a pure command economy (communism), the governments' (planners) take all such decisions. Such polar systems are prone to market or government failures, and result in poor outcomes. Market failures arise due to (i) externalities (called third party effects) (ii) market power (imperfect competition: monopoly and oligopoly) (iii) ignorance and uncertainty (iv) immobility of some factors and time lags, and (v) presence of public goods (like national defence services and light houses in seas). Details on these concepts are discussed later. Suffice to mention here, externalities lead to inefficient decisions as the market based decisions ignore the costs/benefits accruing to those who are neither the producers nor the buyers of the good in question. Public goods (the ones which are non-rival and non-excludable from consumption), which are useful for the society, have to be available to everyone at no cost and thus would attract no one to produce in a pure market economy. Government failures result due to political interventions, corruption, delays, poor motivations towards performance, and so on. Due to these, different countries operate under varying mix of the market and government. While markets provide the forum for exchange between sellers and buyers, governments regulate those exchanges through taxes, subsidies, wage-price controls, government monopolies, pubic sector enterprises and so on. Accordingly, all economic units (firms, workers, consumers and investors) are required to follow the rules of both the market (demand and supply) and government while making choices.

Principles of economics have been designed by economists on the basis of reasoning and thus follow the **cause-effect** (known as **inductive**) **approach**. No two individuals are alike and thus generalization are hard to make. However, theories are developed on the basis of some reasonable assumptions about the behaviours of decision makers. In particular, economic theories have been derived on the basic assumption that all decision makers act **rationally**. Thus, these assume that as a consumer, one maximizes satisfaction from the bundle of goods and services one consumes, and as a worker, one maximizes the satisfaction from work (income) and leisure it gets. Note that rationality here means the bounded rationality which means everyone does what he/she considers the best for him/her and not what others think what is the best for him/her. For example, as a utility maximizing consumer, one may prefer giving a charity over, say, enjoying a personal car, if the former yields him/her more utility than the latter. Similarly, as a utility maximizing worker, he/she could choose a lot of leisure with little income, or as an organization, one may opt for more social service over high profits, and so on.

## 1.2 MANAGERIAL ECONOMICS

Managerial economics is considered to be applied microeconomics. That is, it is concerned with the applications of microeconomic principles to decision making by firms. Thus, a text in managerial economics ought to cover:

(a) Decision variables for firms
(b) Microeconomic principles
(c) Techniques from decision sciences

The decision variables for firms have been briefly discussed in the previous section. Incidentally, it must be noted here that 'firm' here includes both business and non-business ones. This is because the decision problems are more or less the same for both the groups of firms, though, as we shall see later, the two groups may pursue different objectives. For example, like a business firm has to choose one or more production lines from amongst many alternative lines available to it, even a non-business firm like a university has to decide what kind of disciplines (faculties) it should open and to what extent each discipline should be operated, from amongst a host of disciplines available in the field of education today. Also, managerial economics includes decision-making by all other groupings of firms, small, medium and large; proprietorship, partnership and corporation, and public sector and private sector firms.

The relevant microeconomic principles for the firms' decision making include those found in the demand theory, production and cost theory, pricing theory, and the theory of investment decisions. This book will cover each of these subjects in sufficient detail. Macroeconomics is not altogether irrelevant for decision making at the level of the firm. This is because the macroeconomic environment, which includes the behaviour of national aggregates (such as, income, price, unemployment, poverty) and, macroeconomic policy-making aspects (such as industrial policy, import quota, export promotion and tariffs, administered prices and controls, fiscal and monetary policies) affect firms' decisions. Nevertheless, because, there is little that an individual firm can do to affect the aggregate economy, the discussion of macroeconomic principles is not covered in a managerial economics course. Managers of firms are assumed to take the economic environment as given.

Any meaningful application of microeconomic principles necessitates use of some quantitative techniques. Some of the important techniques include methods of estimation, optimization, and discounted cash flow techniques. These techniques have come from the fields of statistics, operations research and finance, and thus a good manager of a firm needs to have a good knowledge not only of economics but also of these disciplines. Accordingly, all such techniques are briefly covered in this text.

## 1.3 THE FIRM: OBJECTIVES AND CONSTRAINTS

This section is presented under three subsections: firm, inputs and outputs; firms' objectives, and firms' constraints. A discussion of each of these follows.

## Firm, Inputs and Outputs

A firm is understood as an organization which converts inputs, which it hires, into outputs, which it sells. The primary inputs, called the factors of production (FOP), are classified as follows:

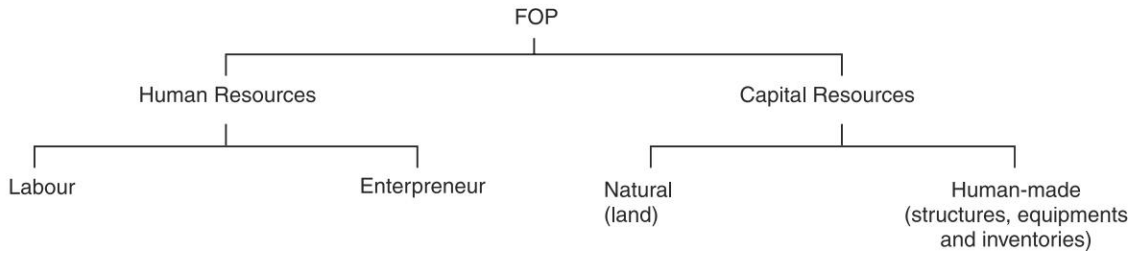



**Chart 1.1**

Under human resources, labour input includes both physical and mental labour i.e., both unskilled (blue collar) and skilled (white collar) labour and it is that part of human effort in an organization which is paid wages and salaries as its remuneration. The other kind of human resource, is entrepreneurial resource. An entrepreneur takes the initiative, coordinates, innovates and takes risk, and receives as compensation profits or loss. Under land resource, land resource has a rather broad meaning in economics—it includes all the resources created (gifted) by God. Thus, it consists of the barren land, minerals, forests, rivers, sea, mountains, etc., as initially discovered by mankind. Any development work which humankind has carried over all these is part of the human-made capital. It includes all construction on land, like roads, bridges and buildings (residential as well as commercial), all the equipment, such as plant, machines, and tools, and inventories, which consist of unsold finished and semi-finished goods, and raw-materials. In economic terms, the four factors of production are referred to as land, labour, capital (human-made), and entrepreneur (organization), and the remuneration they receive as rent, wage, interest (capital rental) and profit, respectively. In addition to these primary inputs, there are inputs such as raw-materials, supplies and intermediate goods which get consumed in the process of production.

The output of firms consists of goods and services they produce. It includes production in various sectors such as agriculture, industry, trade, transport, banking and communication; consumer goods as well as producer goods; and perishable as well as durable goods.

Today, production of almost any good (tangible or non-tangible) requires services of almost all the four factors of production. However, there are more than one input combinations to produce a given quantity of any good. Thus, the firm has to decide the best input-mix from all the alternatives it has. In other words, there is a problem of choice of technique of production. Also, the firm must decide what goods to produce. Such decisions depend heavily on the objectives the firm decides to pursue.

The firms are also classified into categories like private sector firms, government firms (owned entirely by the government and run through government budgets, like railways in India), public sector firms (government owns more than 50% of the share capital), joint sector firms and not-for-profit (non-business) firms. They are also classified according to the number of owners of a firm. On this basis, firms are known as proprietorship, partnership and corporations. The meaning of each of these types of firms, barring perhaps not-for-profit firms, is self explanatory. The somewhat less well understood group of firms include universities, public libraries, hospitals, performing arts groups, museums, churches, voluntary organizations, cooperatives, credit unions, labour unions, professional societies, foundations

and fraternal organizations. Some of these organizations provide services to a group of clients, such as the patients of a hospital. Others provide services primarily to members, such as the members of a cooperative or a club. The classification of firms into these categories is significant to appreciate and understand the objectives of firms.

## Firm's Objectives

There are a number of theories about the objectives of a firm. The important ones are the following:

(a) Profit maximization
(b) Firm's (economic) value maximization
(c) Sales (revenue) maximization subject to some predetermined level of profit
(d) Size maximization
(e) Long-run survival
(f) Management utility maximization
(g) Satisfying
(h) Other (non-profit) objectives

## Profit Maximization

The traditional theory of firm's behaviour assumes that the objective of firm owners is to maximize the amount of short-run profits. Before we dwell on the pros and cons of this theory, it is imperative that we understand the meaning of profit. Profit is defined differently in business and economics. The public and business community define profit as an accounting concept, where it is the difference between total receipts and the explicit (accounting) costs of carrying out the business; explicit cost is the payments made to the hired factors of production. This profit concept is gross of the implicit cost, which stands for the imputed cost of the self-owned factors of production employed in the business. The economic profit is the residual after both the explicit and implicit costs are deducted from the total receipts. To illustrate this important distinction, let us consider an example.

A carpenter makes 100 chairs per month and sells them at Rs. 450 per piece. His expenses on rent of the shop, cost of wood and other material are worth Rs. 15000 per month. He employs two workers whose monthly wage bills stand at Rs. 7200 and pays electricity bill of about Rs. 1500 per month. He has invested Rs. 150,000 in the form of machines, tool and inventories in the business, of which Rs. 75,000 is from his own fund and the remaining Rs. 75,000 is a loan from a bank at the interest rate of 18% per annum. Further, assuming imputed costs of his own time, his wife's time and his own savings of Rs. 75,000 for the month are Rs. 9000, Rs. 3000 and Rs. 750, respectively. The various calculations would then be:

| | |
|---|---|
| Total receipts | = Rs. 450 × 100 = Rs 45000 |
| Total explicit costs | = Rs. 15000 + 7200 + 1500 + 75000(1/12)(0.18) |
| | = Rs. 24,825 |
| Total implicit costs | = Rs. 9000 + 3000 + 750 |
| | = Rs. 12,750 |

Business (Accounting) profit = Rs. 45,000 – 24,825
= Rs. 20,175
Economic profit = Rs. 45,000 – 24,825 – 12,750 = Rs. 7,425

Thus, business and economic concepts of profit are different. In this book we will refer to profits only in its economic sense.

Economic profits are a powerful guiding force in the free enterprise system, particularly for a proprietorship firm. However, the present day world has both the private and public sector firms operating simultaneously, and most firms are either on a partnership basis or are corporations. The public sector firms are known to pursue social objectives, such as factor productivity and the supply of essential goods at reasonable prices. The partnership firms and corporations on the other hand care for non-profit criteria as well. Further, a firm is expected to continue for a number of years and it would be unwise for it to care for today's profit only, particularly if it impinges on future profits. For example, if short-run profits were the only criterion, there would be no expenditure on research and development, and no one would care for creating goodwill through good customer service and products quality. Nevertheless, firms are designed to make profits and profit is at least one of the factors on the basis of which the performance of firms is evaluated. There are various theories to explain profit-making by firms, the important ones are presented in what follows.

**Innovation Theory** Firms make innovations in new products, new production techniques, new marketing strategies, etc. These innovations are costly and must be rewarding for them to flow continuously. For this reason, innovating firms are sometimes awarded patent rights for a specific period of time, during which time no other firm is permitted to copy the product and/or technology. Profits are thus considered partly a reward for innovation.

**Risk-Bearing Theory** Firms invest large sums in the production system, expecting to produce goods and make profits on it. However, the production may run into difficulties, be delayed and there may not be an adequate market when production is ready. In consequence, rewards for entrepreneurship are highly uncertain. The firms take these risks and must be adequately rewarded.

**Monopoly Theory** Some firms are able to enjoy certain monopoly powers in view of being in possession of a huge capital, economies of scale, patent protection or socio-political powers. As a result, there is a lack of perfect competition and such firms are able to reap economic profits.

**Friction Theory** According to this theory, there is a long-run equilibrium of economic profit, which is zero (adjusted for risk). However, markets are seldom in equilibrium and that gives rise to economic profits or losses. For example, if winter is too severe or too prolonged, firms dealing in woolen garments would reap large economic profits while those dealing in items like ice cream or fans may run into losses. Thus, frictions in the economy also give rise to profits and losses.

**Managerial Efficiency Theory** This theory argues that economic profits can arise because of exceptional managerial skills of well-managed firms. For example, if firms that operate at an average level of efficiency can avoid losses, then those which operate at above that level must reap economic profits. Thus, existence of profit is essential to ensure good performance.

Each of these five theories has an element of truth and one or more of these may be valid in an individual case. An automobile firm might make profit for all the above reasons operating simultaneously while an ice cream vendor might make profit on a rainy day just because there is a heavy rush of tourists in his town on that particular day (i.e., Frictional theory).

While firms must or could make some profit, they may not aim to earn maximum profit in the short-run. This has been explained earlier and would become clearer through an understanding of alternative objective of firms.

## Firm's (Economic) Value Maximization

Since most firms are expected to operate for a long period, they are postulated to aim for maximum long-term profits instead of maximum short-term profit. Thus, if $\pi_1$ denotes expected profit in period 1, $\pi_2$ expected profit in period 2, and so on, what the firm aims at is not to seek the maximum value of any one of these profits ($\pi$'s) but the maximum value of their sum, adjusted properly for the time value of money. Thus,

$$\text{Value of the firm } (V) = \frac{\pi_1}{(1+i)} + \frac{\pi_2}{(1+i)^2} + \ldots \frac{\pi_n}{(1+i)^n}$$

or,
$$V = \sum_{t=1}^{n} \frac{\pi_t}{(1+i)^t} \tag{1.1}$$

The parameter $i$ denotes the appropriate interest rate, and $n$ the number of years the firm is expected to last. If $\pi$'s are interpreted as dividends per share, $\pi_n$ is inclusive of capital gains, if any, and $i$ includes the cost of equity capital, then the value of the firm equals the present value of a share of the firm. The above equation assumes that the reader is familiar with the concepts of 'discounting' and 'present value' which is discussed later in this chapter.

Equation 1.1 can be simplified further if we assume that profit grows at a uniform rate of $g$, and that the firm lasts for infinite period. Under these assumptions, $\pi_2 = \pi_1 (1 + g)$, $\pi_3 = \pi_1 (1 + g)^2$,

and so on. Substitution of these values in Eq. (1.1) gives

$$V = \frac{\pi_1}{(1+i)} + \frac{\pi_1(1+g)}{(1+i)^2} + \frac{\pi_1(1+g)^2}{(1+i)^3} + \ldots\ldots$$

The above infinite series is in geometric progression. Using the formula for an infinite geometric series summation, the equation becomes

$$V = \frac{\frac{\pi_1}{1+i}}{1 - \frac{1+g}{1+i}}$$ , which on simplification yields

$$V = \pi_1 \left[\frac{1}{i-g}\right] . \tag{1.2}$$

Equation (1.2) signifies that the firm value depends on the profit in the first year and two parameters, viz. interest rate and the growth rate in profit. Thus, the short-run profits are crucial even under the value

maximization theory. The assumption of a constant growth rate may be unreasonable but the essence of the theory is fairly valid.

The goal of value or wealth maximization is recognized today as the primary objective of a business firm. However, most firms have multiple objectives in the modern world and business firms are no exception to this rule. Further, non-business firms do pursue non-value maximization objectives.

### Sales Maximization Subject to Some Pre-Determined Profit

William J. Baumol has advanced a theory of firm behaviour in which he argues that a firm seeks a certain level of profit and within that constraint aims at maximum sales. The 'certain level of profit' presumably means the level of profit considered satisfactory by the shareholders. The variable for constrained variable for maximization, viz., sales, is in terms of revenue (rupees) and not in terms of physical units of goods and services. This is because, many firms are engaged in multiple products, and these products may not be additive in physical terms or/and may have different values per unit. For example, Godrej manufactures refrigerators and cupboards of different sizes, among other things, and it is impossible to add all these products in physical units. Also, just as in the short-term profit maximization and long-run profit (or value) maximization theories, one could postulate the constrained short-run sales or constrained long-run sales maximization theories, and choose the long-run alternatives only.

The constrained sales revenue maximization theory rests on the premise that a dichotomy exists between owners and the management. In the corporate world, firm is owned by numerous small share holders, who hardly have any say in the day to day management of the firm. They might attend annual general body meetings and are content if the decisions on dividend are fair in relation to dividends declared by similar enterprises. On the other hand, the firm is managed by salary earning professional managers who take decisions which serve their interests best while ensuring 'no serious objection' from the owners at the annual meeting. In general, paid managers' interests rest in their salary packages and perquisites which they enjoy. Quite often the salary and perquisites of decision-makers are linked directly with sales volume or market shares which their respective organizations enjoy and there are cases where manager get some small percentages of sales in the form of commissions as well. Further, promotions to higher positions within an organization or in some other firms also depend on the turnover a manager is handling. For all these reasons, Baumol has advanced the hypothesis that firms seek maximum sales subject to a profit constraint, which is satisfactory to the shareholders.

### Size Maximization

Some experts have suggested growth or size maximization as an alternative goal for firms. By growth they mean, an increase in sales, assets and/or the number of employees. Edith Penrose argues that managers have a vital interest in growth because "individuals gain prestige, personal satisfaction in the successful growth of the firm with which they are connected, more responsible and better paid positions, and wider scope for their ambitions and abilities" (Penrose, 1959).

### Long-Run Survival

K.W. Rothschild (1947) has suggested yet another alternative goal for the firm to pursue, that is, of assuring long-run survival for the firm. Under this objective, the firm seeks to maximize the probability of

its survival into the future. Such an objective would commensurate with the interests of the share-holders and the management. Through this objective, the owners of today would be able to provide security and business to their next generations. Likewise, management would be happy with this objective, for their present and future compensation depends on the firm's continued existence. A short career in a bankrupt firm would hardly provide a strong basis for a successful job mobility or job security.

Unlike other objectives of the firm, the objective of long-run survival is hard to measure and difficult to practice or achieve.

### Management Utility Maximization

O.E. Williamson's model of firm behaviour (1963) "focuses on the self-interest-seeking behaviour of corporate managers". The theory basically ignores the owner's interest whenever there is a dichotomy between owners and managers. To this extent it goes even beyond Baumol's hypothesis, where managers atleast ensure some minimum profit for the owners.

There are many variables in an organization which affect the management utility. Among these, the prominent ones are the salary including bonus, if any, perquisites, number of subordinates and the management's role in investment decisions. Again, the theory is somewhat vague since the numerous dimensions of management's utility may not always be in harmony and there is no perfect method of developing a combined yardstick which could merge all these into a single variable.

### Satisfying

Herbert Simon, a noble prize winner, had proposed an alternate theory of firm behaviour. According to his theory (Simon, 1956), firms do not aim at maximizing anything (profits, sales, etc.) due to imperfections in data and incompatibility of interests of various constituent of an organization. Instead, they set up for themselves some minimum standards of achievement, which they hope will assure the firm's viability over a long period of time. This would require satisfying all the constituents of the firm, including the stock holders, management, employees, customers, suppliers and government.

The satisfying objective, in fact, is a multiple goal and it is very difficult to practice and achieve. This is because, human-beings by nature want satisfaction not only in an absolute sense but in a relative sense as well. In other words, stockholders may be satisfied by, say, a dividend rate of 20% if the top management's salary, including perquisites is no more than rupees one lakh a year, but if the latter stand at rupees two lakhs a year, even a dividend rate of 40 per cent may be unsatisfactory. Similarly, employees may be content with a bonus of 8 per cent if the dividend rate is, say, 10 per cent and the profit rate on capital employed is around 10 per cent.

### Other (Non-Profit) Objectives

The non-profit objectives include goals such as, maximization of quantity and quality of output subject to a break-even budget constraint, administrators' utility maximization, maximization of factor productivity, and maximization of cash flows. Also, Herbert Simon's satisfying objective and Rothschild's maximization of long-run survival are, in fact, components of this broad group of objectives.

The rationale for these objectives are inherent in the nature of the public sector firms and not-for-profit organization. These units are engaged in the production of essential goods (such as gas, electricity, transport and electricity) and public goods (such as national parks, museums, national defence and flood

control), and render services to a group of clients (such as patients of a hospital) and to their members/contributors (such as the members of a trade union, of a cooperative firm or of a country club). The funds for such organizations come from general government funds, donations and members' contributions. Since there is no unambiguous objective function for these managers to maximize, they look for a level of service that satisfies those paying for it or those who deserve them most in the public eye, and thereby protect the job of the management.

The foregoing discussion would indicate that there is no unique theory of the firm's behaviour, that a firm might pursue more than one objective at the same time, which may not always be compatible, and that different firms may pursue different objectives at the same time. Also, it would be clear that the objective of the firm would depend upon whether it is a business firm on a public or private sector firm, or a proprietorship partnership or a corporate firm, or a profit-seeking or not-for-profit firm. However, it can be concluded that the more closer a firm is to the proprietorship and private sector unit, the more inclined it would to aim for the maximum long-term profit or the maximum value of the firm. Conversely, the more closer a firm is towards a big corporation and public sector unit, the more inclined it would be towards looking for maximum benefits to the management and factor productivity, which go generally hand-in-hand with maximum sales. Not-for-profit firms look for product quality and service to its members/clients. Notwithstanding this, profits are the most important yardstick for judging the success of a firm, and firms (particularly business firms) are designed to make profits and no firm can afford to go without any profits in the long-run.

## Firm's Constraints

Decision-making by firms is often subject to certain restrictions or constraints, which may be internal or external. The internal constraints refer to the ones imposed by the organization itself. For example, while deciding on what to produce, a firm might not like to explore each and every alternative good or service it could produce. If the promoter, for example, is a fresh engineer with a little money of his own and is a risk-averter, he might consider only a handful of small manufacturing firms to choose just one amongst them. Instead, if the promoter is a rich person, owning a vast business enterprise, and has the objective of creating employment for unskilled workers, he might decide to establish a handloom unit. Similarly, entrepreneurs with socialist leanings, may produce the essential good and sell them at no-profit-no-loss basis. These are some examples of internal constraints which have a significant effect on management decisions.

Among the external constraints, the important ones are resource constraints, output quantity or quality constraints, legal constraints, and environmental constraints. A brief discussion of these follows.

## Resource Constraints

Certain resources might be available in a fixed or limited quantity and the firm has to take the most appropriate decision. For example, for a small entrepreneur, capital would be a constraint. Similarly, a raw-material might have to be imported and there may be an import restriction or it might be locally available, but only in a limited quantity, in which case, the firm has to decide within this constraint. Also, sometime skilled labour, plant and machinery, or electricity, or even the factory space could be a binding constraint for the expansion of a firm. Such constraints are of particular significance in a resource-scarce country like India.

### Output Quantity or/and Quality Constraints

Production of some goods require a license from a competent authority. Since the licenses are issued for instilling a certain unit of maximum capacity, they become a constraint to the firm and, such constraints are sometimes dictated by the availability of the market. Similarly, the licensing authority might prescribe certain quality norms, which the organization must adhere to. Some common examples of output quality constraints are nutritional requirements for feed mixtures, audience exposure requirements for media promotions, reliability requirements for measurement devices, and requirements for minimum customer service levels.

### Legal Constraints

The legal constraints on firms' behaviour take the form of laws that define minimum wages and bonus, health and safety standards, pollution emission standards, fuel efficiency requirements, price controls including taxes and subsidies, import-export quotas and tariffs, fair marketing practices, priority lending and differential interest rates on loans, backward area subsidies, differential tax rates, and so on. All these constraints impose restrictions on managerial flexibility. Taking a decision require thorough understanding of all these restrictions and facilities.

### Environmental Constraints

No firm can afford to ignore the economic, social and political environment within which it has to function. It needs to understand these spheres not only within the economy but also in the world outside since most economies are open economies in the present day context. The manager must have a good knowledge about all the aspects of public policy, such as those dealing with plan objectives, taxation, subsidies, industrial policy, import-export policies, technology, foreign exchange regulations, etc. Besides, he must understand the trends in all pertinent aggregate variables, such as national income, price level, unemployment, poverty level (both absolute and relative), export-import, money supply, interest rate and foreign exchange balances. In addition, he should keep track of the changing political situation and social environmental in the country. It is almost impossible for the top management to be well informed about all these matters. They therefore employ the service of economists and other social scientists to constantly monitor the economy vis-a-vis their particular firm and appraise the management about forthcoming changes.

Due to the co-existence of objectives and constraints, decisions have to be arrived at by solving, what is called, the constrained optimization problems. In the real world, such problems are resolved both through non-scientific methods as well as scientific methods. The non-scientific techniques can be understood through case studies and their solutions, the scientific techniques would require deeper understanding of important methods such as the Lagrangian method and linear programming, both of which have been discussed in the Appendix on economic optimization.

## 1.4 DECISION PROCESS

Managerial economics is concerned with decision making at the level of the firm. Decisions have far reaching effects on the firm and are often not easy to make. In the circumstance, it is recommended that

the systematic efforts are made to arrive at the right decisions. The following steps should be followed for the purpose:

(a) establish objective (s)
(b) specify the decision problem
(c) identify alternatives
(d) evaluate alternatives
(e) select the best alternative
(f) implement the decision
(g) monitor the performance

Since decisions depend on the objectives of a firm, it is important to be clear about them from the outset. If a doctor is unable to diagnose the disease, his prescriptions may not cure the patient. If there are no alternatives, there is no decision problem. However, in today's complex world, there are many alternatives available to a firm and as many constraints facing them. A clear understanding of these necessitates a thorough scanning of the environment-opportunities and constraints. Evaluation of alternatives require more effort. It would involve collection of all relevant data and their analysis through appropriate techniques. Once these four steps are completed, choosing an alternative on the basis of objectives is simple. Again, implementation of a decision might require resources, an explanation to those who are affected by it, and the courage and ability to face the consequences of that decision. Even after implementing the project, its performance must be monitored vis-a-vis expectations, so that projection errors are minimized in future.

The time period under consideration will often be an important factor in our decision analysis. Ours is a dynamic economy and decisions would have to be made within a time constraint. Things do change over time and if undue delay is made in decision making, opportunities might turn into threats, and so on. There are a number of examples where firms have suffered significant losses due to delay in decision-making or/and implementation of decisions.

The text would pose many problems of decision making, and discuss the tools and techniques available for arriving at the right decisions.

## 1.5 BASIC PRINCIPLES

Before embarking on individual problems of decision-making and their analysis, it would be useful to explain the basic principles of managerial economics, such as the principles of opportunity cost, discounting and compounding, marginal or incremental, equi-marginal and time perspectives.

### Opportunity Cost Principle

The opportunity cost principle argues that a decision to accept an employment for any factor of production is good (profitable) if the total reward for the factor in that occupation is greater or at least no less than the factor's opportunity cost; the opportunity cost being the loss of the reward in the next best use of that resource. It should be noted here that the 'reward' includes both monetary as well as non-monetary and the true opportunity cost of a factor may not be exactly known but can be imputed as accurately as possible. Thus, for example, the opportunity cost of a professor's time when he launches

a full-fledged consulting firm would be the loss of his salary, perquisites in the form of residential accommodation, if any, provident fund, etc. and the academic environment for carrying on research projects and publications. According to the opportunity cost principle then, his profit, gross of the compensation for his time alone, should not be less than the sum of all the benefits he was deriving when he was a professor. In other words, his economic profit should not be negative.

This principle is emphasized by economists because most economic resources have more than one use and therefore have opportunity costs. Traditional business and accounting executives ignore the opportunity cost of a resource while computing their business costs or profits. In the above example of a professor turned into a consultant, it was not difficult to compute his opportunity cost. However, in most cases it does pose serious problems and one has to be content with some sort of imputation. For example, an entrepreneur who has never had a job or who has not had a paid job for several years would rarely know his opportunity cost. Similarly, an entrepreneur may not know the precise opportunity cost of his equity capital, for he does not know where he would have invested that money had he not done so in his own business. The cases of opportunity cost of family labour, own land, own buildings are similar. In all these situations, one must estimate the opportunity cost as accurately as possible and use it in decision making. No decision would be right if the opportunity costs were ignored.

## Discounting and Compounding Principle

The discounting and compounding principle (DCP) states that when a decision involves money receipts or payments over a period of time, all the money transactions must be valued at a common period to be meaningful for decision analysis. This is because money has time value for three reasons: earning power, changing prices, and uncertainty. Money has earning power, for it can earn at least an interest rate even if it is deposited in a bank. On this count, a rupee to day is worth more than a rupee at a future date. Money has a derived demand, in the sense that it is wanted not for its own sake but for its purchasing power, which depends inversely on the price level. Thus, during inflation, a rupee today is worth more than a rupee at a future date. Similarly, today's money is certain but a promise to give it tomorrow is uncertain, for the promise may not be honoured either because the payee has no money or because he is dishonest. This point could be driven home more forcefully through the proverb, "a bird in hand is worth two in the bush".

Only simple arithmetic is needed to apply this principle. Suppose an investment costs Rs l00 lakhs this year and is expected to yield net returns of Rs 30 lakhs, Rs 40 lakhs and Rs 60 lakhs, in the next three years, respectively. Assume further that the interest cost of the money is 10 per cent, there is no inflation/deflation and no uncertainty about these cash flows. Then, whether the investment should be made or not depends upon whether the following equation yields a positive or negative value:

$$\frac{30}{(1+.1)} + \frac{40}{(1+.1)^2} + \frac{60}{(1+.1)^3} - 100$$

The solution to this yields an amount of Rs 5.41 lakhs, and so the investment is desirable. Incidentally, it should be noted here that interest is compounded once a year but any frequency of compounding can easily be handled through this technique. Also, the method can easily be extended to take account of inflation or deflation and uncertainty. For example, if inflation is expected to be four per cent and uncertainty is considered worth one per cent, then in the formula above, the discount rate will be 15 per cent (0.15) instead of 10 per cent (0.1). The method is discussed in detail in the chapter on investment analysis.

## Marginal or Incremental Principle

The marginal or incremental principle (MIP) states that given the objective of profit maximization, a decision is sound if and only if it leads to increase in profit, which would arise in either of the following cases

(i) If it causes total revenue to increase more than the total cost.

(ii) If it causes total revenue to decline less than the total cost.

The MIP is significant, for some businessmen take an erroneous view that to make maximum profit they must make a profit on every job. The result is that they refuse orders that do no cover full cost plus some profit. This could be better explained through a numerical example. Consider a firm whose output-cost relationship is as follows.

(Rupees)

| *Output* | *Total cost* | *Marginal cost* | *Average cost* |
|---|---|---|---|
| 0 | 20 | | |
| 1 | 28 | 8 | 28 |
| 2 | 37 | 9 | 18.50 |
| 3 | 47 | 10 | 15.70 |
| 4 | 58 | 11 | 14.50 |
| 5 | 68 | 10 | 13.60 |

Suppose Firm A is producing three units and selling them at a price of Rs. 25 per unit, making a total profit of Rs. 28. If the customer for its fourth unit of output, is offering Rs. 14 only, should the firm accept this offer? According to the full cost principle, the offer must be rejected since the average cost of four units equals Rs. 14.50, which exceeds the offered price. However, the marginal or incremental principle would argue that the cost of the fourth unit (MC) equals Rs. 11, which is less than the price offered, thereby his profit would increase, and so the order must be accepted; profit would increase from Rs. 28 to Rs. 31. True, but there is a catch in this argument.

In the above example, marginal cost is less than the average cost i.e., there is excess capacity in production. In other words, the fixed resources are not optimally exploited. Thus, it is profitable to sell below the full cost (average cost) because of the existence of excess capacity in production. In the absence of excess capacity, such would not be the case even under the incremental principle.

The decision to accept the offer at a price below the average cost assumes that it does not lead to any long term repercussions. For example, if by selling the fourth unit at the price of Rs. 14, the customers of the first three units, who paid Rs. 25 per unit, get disturbed and decide to boycott the firm in the future, the offer of Rs. 14 for the fourth unit should not be accepted. Thus, while applying the principle of incremental analysis, one should bear in mind the long-run repercussion of the decision.

The foregoing analysis would have made it clear that the MIP is useful particularly in situations where there exists excess capacity in production and the long-term impact is insignificant. For example, one can conceive of a situation where airlines declare two fares, one regular, say, at Rs. 7,000 between Delhi and Mumbai, and the other standby at Rs. 3,000 for the same trip. On the regular fare, anyone can

reserve a seat while on the standby, accommodation would be made available if there is an empty seat at the last minute. Thus, on one hand, passengers whose trip were not urgent and who have less or no value of time (like senior citizens and students desiring to visit family during holidays) could save on air fare. Such passengers may, of course, spend some money on extra trips between home and airport should they not get a seat on the first attempt. On the other hand, the airlines could also benefit from such a pricing system as the standby rate would be considerably higher than the marginal cost even though this rate may be below its full cost. Similarly, a country might find it profitable to export a product at a price below its average cost, whereas its domestic price could be considerably higher than the average cost.

Incidentally, it is pertinent to note the difference between marginal and incremental concepts. The former has two salient characteristics. One, the marginal concept is applicable to change in revenue, cost or profit, etc. with respect to change in output only. Two, the concept requires that the change in output is infinitesimally small, to be approximated by one in case of discrete data. In contrast, the incremental concept is applicable with respect to any variable and for any extent of change. To illustrate the difference, let us examine the case of the carpenter again. The net profit was Rs 7425 when the carpenter sold 100 chairs, and, suppose, it would have been Rs 7625 if he had sold 101 chairs. Then the marginal profit = 200 [(7625 – 7425)/(101 – 100)], which is also the average incremental profit. Instead, if the profit was Rs 7425 when the carpenter charged a price of Rs 450 per chair and, suppose, Rs 8000 if he had charged a price of Rs 475, the incremental profits would equal to Rs 575 [8000 – 7425] and the average incremental profit = Rs 23 [(8000 – 7425)/(475 – 450). In the latter case, the concept of marginal profit is not applicable at all. Thus, while the incremental principle is versatile, the marginal concept is specific to changes in a particular variable brought about by small changes in output alone.

## Equi-Marginal (Rationality) Principle

The equi-marginal principle states that a rational decision-maker would allocate or hire his/her resources in such a way that the ratio of marginal returns and marginal costs of various uses of a given resource or of various resources in a given use is the same. For example, a consumer seeking maximum utility (satisfaction) from his consumption basket will allocate his consumption budget on various goods and services such that

$$\frac{MU_1}{MC_1} = \frac{MU_2}{MC_2} = \ldots = \frac{MU_n}{MC_n} \tag{1.3}$$

where, $MU_1$ = marginal utility from good 1, $MU_2$ = marginal utility from good 2, $MC_1$ = marginal cost of good 1, and so on. Similarly, a producer seeking maximum profit would use that technique of production (input-mix) which would ensure.

$$\frac{MRP_1}{MC_1} = \frac{MRP_2}{MC_2} = \ldots = \frac{MRP_n}{MC_n} \tag{1.4}$$

where $MRP_1$ = marginal revenue product of input 1 (e.g, labour), $MRP_2$ marginal revenue product of input 2 (e.g., capital), $MC_1$ = marginal cost of input 1, and so on. Incidentally note that if the output and input prices are fixed (given to the decision-makers), as is the case under perfect condition, $MC_1 = P_1$, $MC_2 = P_2$, and so on. This would be explained later in the relevant chapters.

The principle involves new concepts and we are not able to explain them in detail at this stage. Suffice it to say that it is easy to see that if the said equality were not true, the decision-maker could add to his utility/profit by reshuffling his resources/inputs. For example, if $\frac{MU_1}{MC_1} > \frac{MU_2}{MC_2}$ the consumer would add to his utility by buying more of good 1 and less of good 2.

### Time Perspective Principle

The time perspective principle argues that the decision-maker must give due consideration both to the short-and-long-run consequences of, his/her decision, giving appropriate weights to the various time periods before arriving at a decision. The principle can be well explained through recalling the example cited under the marginal or incremental principle. The order for the fourth unit at Rs. 14 in spite of an average cost of Rs. 14.50 was worth accepting by the producer on the short-run consideration for sure. But if that were to disturb the customers (market) in the long-run, it may have to be rejected. Similarly, we do come across many new products which are sold below cost or on relatively small margins in the beginning with the hope of commanding a good market and thereby making profits in the long-run. Nirma soap powder and Rin soap cakes perhaps fall in this category. If the managers did not have time perspective in their mind, they would never have resorted to such practices (i.e., of selling them at below their corresponding costs in the beginning), called 'price penetration'—the concept to be explained later under pricing strategies. The distinction between short and long-run would also be clarified in the chapters on production and cost analysis.

Some of the applications of these principles have already been indicated and many more will come through in the rest of the book.

## 1.6 CHAPTER PLAN

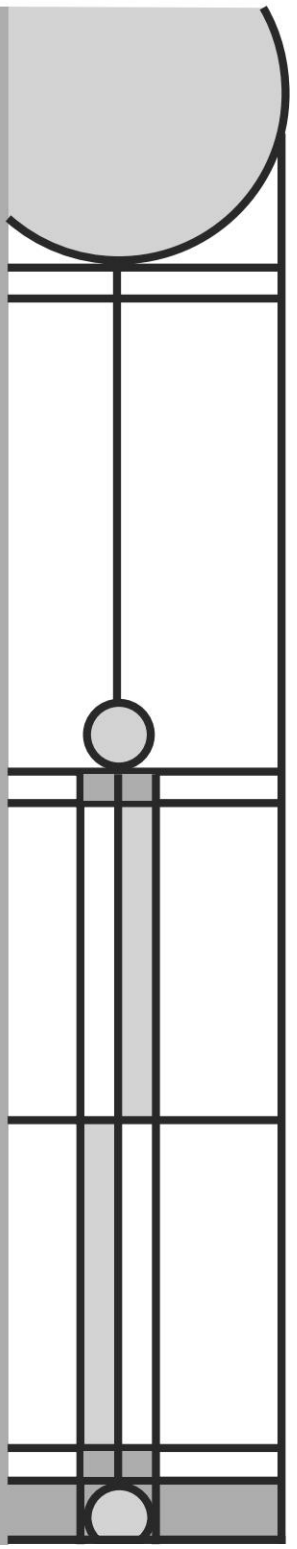
The text is divided into eleven chapters and four appendices. The first chapter is introductory; the second, third and fourth chapters deal with various aspects of demand analysis, theory of consumers' behavior, estimation and forecasting, respectively. Chapters 5 and 6 present production and cost analysis, and thus handle the supply side of the market. Chapters 7–10 integrate the demand and supply sides of the market, and discuss in detail the determination of product prices under various objectives and market conditions. A thorough blending of theory, policy and practices in product pricing is provided in these chapters. The last chapter focuses on investment analysis, providing due consideration to inflation, risk and uncertainty, and social benefit cost analysis. Since the principles and techniques of economic optimization are needed for almost all the managerial economics decisions, and since they involve a fair amount of mathematics, they are included in Appendix A. Appendix B presents some useful tables for carrying out investment analysis. For practice, chapter-wise multiple choice questions and fill-in blanks are provided in Appendices C and D, respectively. References for further readings and some caselets are provided at the end of each chapter.

## REFERENCES

1. Baumol, W.J. (1982); *Economic Theory* and *Operations Analysis*, 4th edition, New Delhi, Prentice-Hall.

2. Dean, J. (1976); *Managerial Economics*, New Delhi, Prentice-Hall.
3. Friedman, Milton (1953): The Methodology of Positive Economics, *Essays in Positive Economics*, Chicago, Univ. of Chicago Press.
4. Henderson, J.M. and R.E. Quant (1980); Microeconomic theory; *A Mathematical Approach,* 3rd edition, New York, McGraw-Hill.
5. Penrose, E. (1959); *The Theory of the Growth of the Firm,* Oxford, Basil Blackwell and Mott.
6. Simon, H.A. (1959); "Theory of Decision Making in Economics," *American Economic Review*, XLIX (June).
7. Williamson, O.E. (1963); The *Economics of Discretionary Behaviour; Managerial objectives in a Theory of the Firm*, Englewood Cliffs, N.J., Prentice-Hall.

The McGraw-Hill Companies



CHAPTER

2

# DEMAND AND REVENUE ANALYSIS

Demand for goods and services constitutes one side of the product market; supply of goods and services forms the other. It is needless to say that if there is no demand for a good, there is no need to produce that good. Also, if the demand for a good exceeds its supply, there may be a need to expand its production. Further, production generally takes time, and so one may like to know the likely demand for a relevant product at a future date to plan its production properly. Thus, a clear understanding of the relevant demand is imperative for any producer worth his name.

Economists prefer to speak in terms of **demands** rather than **needs**. This is because there are more than one products to satisfy a particular need. For example, need for food could be met by chapatti, rice, chicken, fish, etc.; while that for thirst by water, Pepsi, milk, orange juice, wine, etc. Thus, consumers face choices and accordingly describing demands by needs is misleading. Demand analysis seeks to identify and analyse the factors that influence the demand for various goods. As we shall see later in this chapter, a firm is not a passive taker of the demand for its product. It has the capacity to create demand as well. The study of demand is necessary for a decision-maker, for it has bearings on its production schedule, it is subject to manipulation by the producers, and it exerts influence on profit, among other critical variables. Before we pursue this further, it is imperative to explain certain concepts and jargons.

## 2.1 MEANING OF DEMAND

Demand in economics means effective demand, that is one which meets with all its three crucial characteristics, viz. **desire** to have a good, **willingness to pay** for that good, and **ability to pay** for that good. In the absence of any of these three characteristics, there is no demand. For example, a teetotaler professor may possess both the willingness to pay as well as the ability to pay for a bottle of liquor, yet he does not have a demand for it. This is because he does not desire to have an alcoholic drink. Similarly, a businessman might have the desire to have a television, he might be rich enough to be able to pay for it, but if he is not willing to pay for the television, he does not have a demand for this product. Also, a blue collar worker might possess both the desire for a scooter as well as the willingness to pay for it, but if

he does not possess enough money to pay for it, he does not have a demand for the scooter. In contrast, to these three situations, a doctor, who has the desire for a car, as well as both the will and ability to pay for it, has the demand for a car. Incidentally, sometimes there is a shortage of a commodity, meaning that there is no one from whom the product could be purchased at any/market price. For example, a doctor's demand might be for a new BMW car but there is no seller in India nor the government of India permits him to import it, then what? Well, there is a demand for a BMW car but it can not be met. Similarly, there are some items whose prices are fixed either due to government regulations (e.g. petrol/LPG cylinder) or some other factors. At such a fixed price, some prospective buyers (who do have effective demands) may not be able to buy the product, and then it is considered as a shortage of the product in the market.

Demand for a good depends on several factors and thus it varies as any one or more of these factors change. A detailed discussion of this will be taken up later. However, it is pertinent to recall here the two important determinants of demand, viz. own price and time. Demand in economics is defined as a schedule which shows various quantities of a product which one or more consumers are willing and able to purchase at various specific prices during a specific period of time. For example, demand for milk by a household per week may be described as follows:

| *Milk Price (Rs per litre)* | *Milk Demand (litres)* |
|---|---|
| 32 | 5 |
| 28 | 8 |
| 24 | 12 |
| 20 | 20 |
| 16 | 30 |
| 12 | 45 |

Evidently, the annual demand for milk would be more than (perhaps 53 times) the weekly demand. Daily demand would be less. The graphical version of this demand is shown in Fig. 2.1.

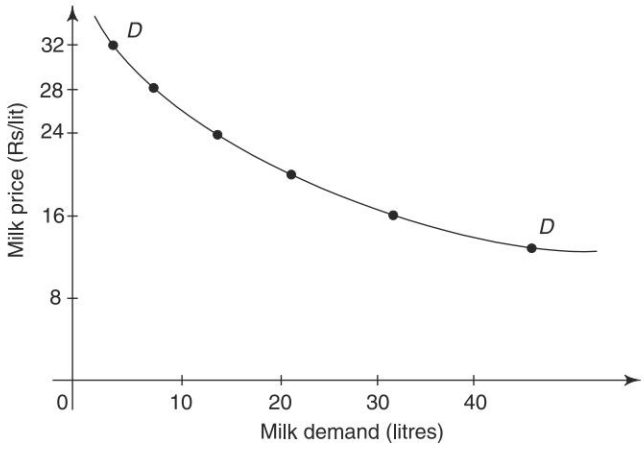



**Fig. 2.1** *Demand curve*

In Fig. 2.1, *DD* curve represents the demand for milk curve. It must be emphasized that the demand curve/schedule gives the **maximum price** that the consumer will pay for various quantities of the product, for he would, of course, be always happy to buy at lower prices. For example, in the above example, the consumer would buy 12 litres of milk if milk price were at most Rs. 24, and would be happy to buy the same quantity (and even more) even if the said price were less than Rs. 24.

## 2.2 TYPES OF DEMAND

There are a large number of goods and services available in every economy. Their classification is important in order to carry out a meaningful demand analysis for managerial decisions. Also, an understanding of demand at various levels of aggregation is inevitable for policy decisions. The significant classifications, in these two respects, are the following:

(a) Consumers' goods and producers' goods
(b) Perishable and durable goods
(c) Normal/superior and inferior goods
(d) Necessary, comforts and luxury goods
(e) Related goods: substitutes and complementary goods
(f) Autonomous (direct) and derived (indirect) demand
(d) Individual buyer's demand and all buyers' (aggregate/market) demand
(e) Firm and industry demand
(f) Demand by market segments and by total market

An explanation of each of these follows.

### Consumers' Goods and Producers' Goods

Goods and services used for final consumption are called consumers' goods. These include those consumed by human-beings (e.g., food items, drinks, clothes, kitchen utensils, residential houses, medicines, and services of teachers, doctors, lawyers, washermen, shoe-makers, bankers and transporters), animals (e.g., dog food and fish food), birds (e.g., grains), etc. In contrast, producers' goods refer to the ones used for production of the same or other goods. Thus, producers' goods consist of plant and machines, factory buildings, services of business employees, raw-materials, intermediate goods, etc.

The distinction is somewhat arbitrary. This is because, whether a good is consumers' or producers' depends on its use. For example, if a sofa set is used in the drawing room of a house, it is a consumer' good; while if it is used in the reception room of a business house, it is a producers' good. Similarly, sugar, computer and many items are both the consumers as well as the producer's goods. Nevertheless, the distinction is useful for a proper demand analysis for, as explained later, while the demand for consumers' goods depends on households' income that for producers' goods varies with the production level, among other things.

It is customary to deal with consumers' goods only under the chapter on demand and revenue analysis; the demand for producers' goods is handled under the chapter on production and cost analysis. We shall follow this tradition.

## Perishable and Durable Goods

Both consumers' and producers' goods are further classified into perishable (non-durable) and durable goods. In layman's language, perishable goods are those which perish or become unusable after sometime, the rest are durable goods. Thus, the former would include items like milk, fish, eggs, and paper cups and plates; and the latter would include furniture, cars, refrigerators, clothes and shoes. In economics, the meaning of these terms is different. Here perishable goods refer to those goods which can be consumed only once. In other words, these goods are themselves consumed while in the case of durable goods, their services alone are consumed. Thus, perishable goods include all services (e.g., services of teachers and doctors), food items, raw-materials, coal, and electricity, while durable goods include plant and machinery, buildings, furniture, automobiles, refrigerators, and fans.

The distinction is significant, for durable products pose more complicated problems for demand analysis than do non-durables. Sales of non-durables are made largely to meet current demands which depend on current conditions. In contrast, sales of durable goods go partly to satisfy new demand and partly to replace old items. Further, the latter set of goods are generally more expensive than the former set, and their demand alone is subject to preponement, and postponement, depending on current market conditions vis-à-vis expected market conditions in future.

## Normal/Superior and Inferior Goods

Normal goods, also called as superior goods, and inferior goods are economics jargons. The former are those whose demand increases as income increases, and the latter are those whose demand falls as income goes up, and vice versa. For example, milk, refrigerator, television, education, and the good quality of food grains and clothes are superior goods while the poor quality of food grains and clothes are inferior goods. In other words, the superior goods are the ones which the rich people consume while the inferior goods are for the poor people's consumption. Further, these are relative concepts. Thus, for example, scooter/motor bike is a superior good in relation to a pedal bike, while it is an inferior good relative to a car.

## Necessary, Comforts and Luxury Goods

In common sense, the necessary goods are essential for existence, comforts goods make the life comfortable and luxury goods are luxuries of life. However, in economics they have special meanings. These all are considered as superior goods but of different degrees. Thus, as the consumers income rise, more of each of these three kinds of goods is consumed but the proportion of the consumption budgets (treated same as income here) differ. In case of necessary goods, as income increases, while the consumption expenditure on them increases, the percentage of total expenditure/income spent on each of them goes down. In case of comforts, the said percentage remains the same, while in case of luxuries, it goes up. To illustrate this important distinction, consider the following example:

| *Example* | *Income (Y)* | *Demand (Q)* | *Price (P)* | *PQ/Y in % (% of income spent)* |
|---|---|---|---|---|
| (a) | 100 | 5 | 10 | 0.50 |
| (b) | 120 | 5.5 | 10 | 0.46 |
| (c) | 120 | 6 | 10 | 0.50 |
| (d) | 120 | 7 | 10 | 0.58 |

If the demand changes as in example (a) to (b), the good in question is a necessary item. If it changes as in example (a) to (c), the good is comfort and if it changes as in example (a) to (d), the good is a luxury. In general, ordinary foods, drinks, clothing, some education and medical aids are considered as necessary. Some means of transport, good quality of food, drinks and clothing, tourism, etc. are taken as comforts. Luxuries include foods in high end hotels, designers clothing, spacious residences, foreign tourism, and so on. It must be noted here that this classification of goods is quite subjective. This is because a particular good could be a necessary for some, comfort for some others and luxury for still some others. For example, a car in India could be a necessary good for a senior consultant/executives/ doctor/advocate; the same may be a comfort item for a junior executive and a luxury for a clerk. Also, note that while food cooked at home may be called as necessary, the food at a moderate restaurant a comfort and that at a five star hotel, the luxury.

## Substitute and Complementary Goods

As the term indicates, two goods are substitutes in a consumption basket if either of them could meet a given consumption need. For example, for many people, tea and coffee are substitute goods, and so are rice and chapatti, scooter and car, coke and Pepsi, orange and apple, and so on. Further, different brands of cars, scooters, cigarettes, soaps, toothpastes, computers, televisions, mobile phones, etc. are substitutes to each other. Also, note that the degree of substitution might vary from product to product. For example, a low end car may relatively be a good substitute for a two wheeler but a poor substitute for a pedal bike, and rice may be a good substitute to chapatti but a poor substitute to fish, and so on. Different brands of a good are generally (strong) substitute to each other, though the degree of substitution may vary. For example, Maruti Zen car may be a close substitute to the Tata's Indica car but a poor substitute to a Mercedes car. The distinction into substitutes and complementary goods is also somewhat subjective. For example, to many consumers, tea and coffee are substitutes, as are rice and chapatti, and yet to some they may not be.

The two goods *X* and *Y* are said to be complementary goods if consumer needs good *X* when he has good *Y, and vice versa.* For example, tea and sugar, car and petrol, cigarettes and match boxes are pairs of complementary items. Once again, the degree of complementaries might vary from one pair of goods to another. For instance, car and petrol are perfect complements while tea and sugar do not have such a strong relationship.

## Autonomous and Derived Demand

The goods whose demand is not tied with the demand for some other goods are said to have autonomous demand, while the rest have derived demand. Thus, the demands for all consumers goods are autonomous demands, for they are needed to satisfy consumers demands. In contrast, demands for all producers' goods are derived demands because they are needed in order to produce consumers or producers' goods. Likewise, the demand for money, which is needed not for its own sake but for its purchasing power (i.e., which can buy goods and services. Similarly, demand for car's battery or petrol is a derived demand, for it is linked to the demand for a car. There is hardly anything whose demand is totally independent of any other demand. But the degree of this dependence varies widely from product to product. For example, demand for petrol is totally linked to the demand for petrol driven vehicles, while the demand for sugar is only loosely linked with the demand for milk. Thus, the distinction between autonomous and derived demand is more of a degree than of a kind. Sometimes a distinction is also drawn between direct

and indirect demand, and that distinction is close to the difference between autonomous and derived demand, respectively. Goods that are demanded for their own sake have direct demand while goods that are needed in order to obtain some other goods possess indirect demand. In this sense, all consumers goods have direct demands while all producers' goods, including money, have indirect demand.

## Individual's Demand and Market Demand

The Demand for a good by an individual buyer is called individual's demand while the demand for a good by all buyers in a market is called market demand. For example, if the milk market consisted of, say, only three buyers, then individuals and market demand (weekly) could be as follows:

| *Milk Price* | | | | *Milk demand by (litres)* |
|---|---|---|---|---|
| *(Rs/litre)* | *Buyer 1* | *Buyer 2* | *Buyer 3* | *All buyers (Market demand)* |
| *(1)* | *(2)* | *(3)* | *(4)* | *(5)* |
| 32 | 5 | 10 | 0 | 15 |
| 28 | 8 | 12 | 4 | 24 |
| 24 | 12 | 15 | 7 | 34 |
| 20 | 20 | 19 | 12 | 51 |
| 16 | 30 | 25 | 20 | 75 |
| 12 | 45 | 30 | 30 | 105 |

In this table, columns (1) and (2) represent buyer 1's demand, columns (1) and (3) buyer 2's demand, columns (1) and (4) buyer 3's demand, and columns (1) and (5) the market demand for milk. The market demand is obtained as the sum of all buyers' demand at respective prices. The graphical derivation of market demand is illustrated in Fig. 2.2.

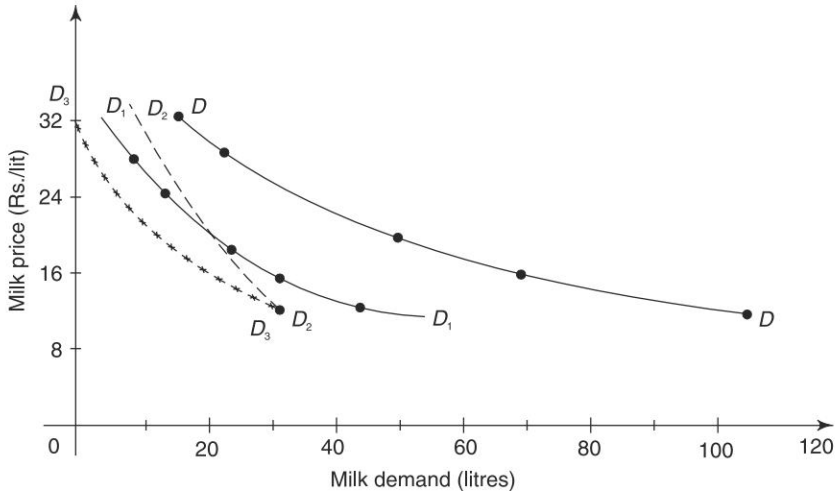



**Fig. 2.2** *Individuals and market demand curves*

Incidentally, it should be noted here that since tables are discontinuous and graphs are continuous, we have drawn graphs only to approximate table instead of exacting the two.

In Fig. 2.2, $D_1$ $D_1$ represents demand curve of buyer 1, $D_2$ $D_2$ of buyer 2, $D_3$ $D_3$ of buyer 3, and *DD* that of all three of them, called the market demand curve. The market demand curve can be obtained graphically through a horizontal summation of individual demand curves.

It needs to be mentioned that the above procedure of deriving the market demand assumes that the demands for a good by various consumers are mutually independent. This means demand of an individual consumer is not influenced by the demands of other consumers of the good in question. In real life, this may not be true due to at least two factors, viz. **bandwagon effect** and **snob/ego effect**. The former effect suggests that if my neighbour has a car, I too wish to have a car. In contrast, the latter effect indicates that if my neighbour has a Maruti zen, I do not wish to have the same but may be a Tata's Indica car. In other words, under the bandwagon effect, a consumer demands what his/her neighbour/friend has, while under the snob effect, one demands an exclusive product instead of a common good. Thus, the two effects work in the opposite direction, though they need not cancel out to render these effects superficial. Nonetheless, their combined effect on most goods is likely to be insignificant. For this reason and for the sake of simplicity, this interdependence of demands is ignored while deriving the market demand.

It should now be clear to the readers that the firm would be interested in the market demand for its products while each consumer would be concerned basically with only his own individual's demand.

## Firm and Industry Demand

Most goods today are produced by more than one firm and so there is a difference between the demand facing an individual firm and that facing an industry (all firms producing a particular good constitute an industry engaged in the production of that good). For example, cars in India are manufactured by Maruti Udyog, Hindustan Motors, Premier Automobiles, Tata Motors, and several other companies. Demand for Maruti car alone is a firm's (company) demand whereas demand for all kinds of cars is industry's demand. Similarly, demand for Godrej refrigerators is a firm's demand while that for all brands of refrigerators is the industry's demand.

The said distinction is very important because while there are close substitutes for firms' products, no such close substitute exists for industry's product. Thus, while a Maruti car is a close substitute for a Tata car, it is only a poor or distant substitute for a Bajaj scooter. The readers would appreciate this classification more when they go through a later chapter, covering market structure.

## Demand by Market Segments and by Total Market

If the market is large in terms of geographical spread, product uses, distribution channels, customer sizes or product varieties, and if any one or more of these differences were significant in terms of product price, profit margins, competition, seasonal patterns or cyclical sensitivity, then it may be worthwhile to distinguish the market by specific segments for a meaningful analysis. In that case, the total market demand would mean the total demand for the product from all market segments while a particular market segment demand would refer to demand for the product in that specific market segment. For example, one can talk about the domestic demand for Maruti cars versus the export demand for that product, demand for Maruti cars in Western India vis-à-vis demand for Maruti cars in each of Southern, Eastern and Northern regions, demand for steel for household (kitchen) vis-a-vis its demand for industrial uses,

demand for fish by households vis-a-vis that bulk buyers (e.g., hostels, restaurants, hotels, guest houses), and consumption (human) demand for fish versus industrial demand for fish (cattle feed etc.), and so on.

The distinction is useful particularly in finding out the problem areas. For example, if a classification is made between domestic and foreign demand, one can analyse the causes of poor foreign market, if that is the case, and finds out the ways of augmenting that market. Similarly, if the market for, say, Maruti car is excellent in Southern India and poor in the Northern part of the country, the company could study the classified data and discover for itself the most approximate ways of dealing with such a problem.

## 2.3 DETERMINANTS OF DEMAND

Demand analysis is needed basically for three purposes:

(a) to provide the basis for analysing market influences on the demand

(b) to provide the guidance for manipulating the demand

(c) to guide in production planning through forecasting the demand

To achieve the above objectives, demand analysis must include the factors which have bearings on the demand. Before we proceed with this, it will be interesting to look at the consumption pattern of an average household in our country.

**Table 2.1** Consumption Pattern in India.

(% of total private consumption expenditure in domestic market)

| *Item* | *1950–51* | *1970–71* | *1984–85* | *2000–01* | *2007–08* |
|---|---|---|---|---|---|
| Food, beverages and tobacco | 73.6 | 68.4 | 60.2 | 47.0 | 41.2 |
| Clothing and footwear | 4.6 | 7.2 | 10.7 | 5.9 | 4.3 |
| Rent, fuel and power | 8.4 | 7.1 | 7.0 | 12.5 | 11.0 |
| Furniture, furnishings, appliances and services | 2.2 | 2.9 | 4.1 | 3.4 | 4.0 |
| Medical care and health services | @ | @ | @ | 4.8 | 5.8 |
| Transport and communications | 3.6 | 5.0 | 8.8 | 14.4 | 17.0 |
| Recreation, education and cultural services | @ | @ | @ | 3.7 | 4.3 |
| Miscellaneous goods and services | 7.7 @ | 9.5 @ | 9.2 @ | 8.3 | 12.4 |
| All goods and services | 100 | 100 | 100 | 100 | 100 |

@ Means not available / included in

*Source:* Central Statistical Organisation; Various Publications

In Table 2.1, note that (a) data for 1950–51, 1970–71 and 1984–85 are at constant (1970–71) prices while those for the other two years are at current prices, and (b) since the earlier series did not have the separate data on medical care and health services and recreation, education and cultural services, they were included in miscellaneous. However, in spite of such variations, the above data do provide a

good input to understand the changing consumption pattern in the country. They suggest that, over the time, the relative share of expenditure on food group has gone down, while that each of medical group, transport group and recreation group has gone up. Since food (including drinks) is more essential than other items, and the per capita income in the country has gone up over time, the said finding is consistent with the a priori expectation.

The above is the consumption pattern of an average household in India. It differs from that of an average household in USA, UK., Japan, China, etc. It also differs from one family to another. Thus, consumption pattern varies not only over time but also across space and households. Why? An understanding of this requires a thorough study of the determinants of demand.

No list of demand determinants could be exhaustive. The major ones would include as shown in Chart 2.1, arranged in a convenient format.

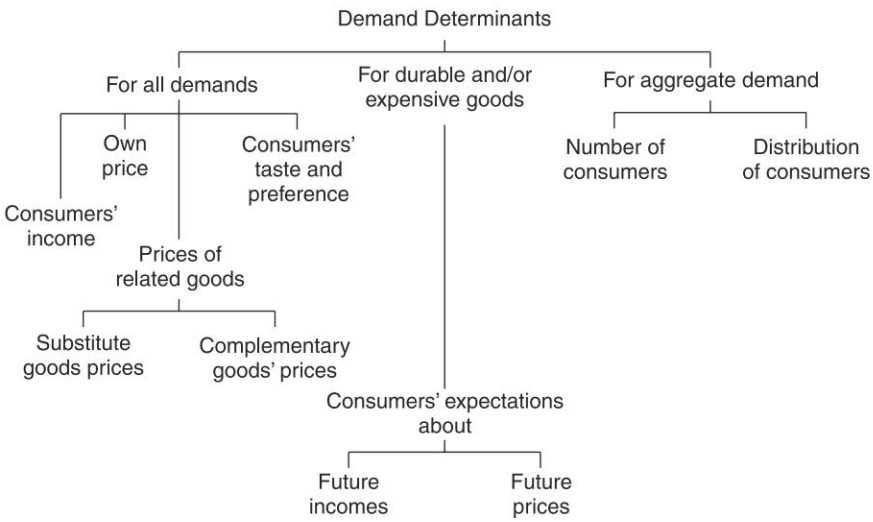



**Chart 2.1**

In what follows, the rationale for each of these factors would be offered and the influence analysed. Incidentally, it may be noted that the analysis will be partial in the sense that when the analysis is presented for demand with respect to any one particular determinant, the other determinants would be assumed to remain constant.

## Consumers' Income and Demand

Demand presupposes the existence of the ability to pay for the product. The ability to pay is in turn determined by the income, wealth or/and the credit worthiness of the consumer. Theoretically speaking, all these three factors should be included in the list of demand determinants. However, wealth is a stock variable, and the income would include that from work (labour) as well as property, and thus wealth is not taken as a separate argument in the demand function. Similarly, creditworthiness depends heavily on the income and wealth of a person, and therefore it is also excluded from the exclusive list of demand determinants. Thus, income acts as a surrogate, scale or constraints variable to take care of the ability to pay requirement in the demand function.

The relationship between demand and income is such that as the consumer's income increases, his/her purchasing power goes up and therefore he/she increases his consumption of superior goods, and if he was consuming some inferior items earlier, he would give them up either totally or partially in favour of superior items. Engel was the first person to study this relationship systematically and the curve reflecting the relationship between demand and income, *ceteris paribus,* is known as the ***Engel's curve***. Two hypothetical Engel schedules and their corresponding curves are given below in table.

The curvature of the Engel curve depends on the degree of the superiority or inferiority of the good in question. Thus, the curve could be linear (straight line), convex or even concave.

**Engel Schedules**

| *For Superior Goods* | | *For Inferior Goods* | |
|---|---|---|---|
| *Income* | *Demand* | *Income* | *Demand* |
| 100 | 5 | 100 | 10 |
| 150 | 8 | 150 | 9 |
| 200 | 10 | 200 | 7 |
| 250 | 11 | 250 | 4 |

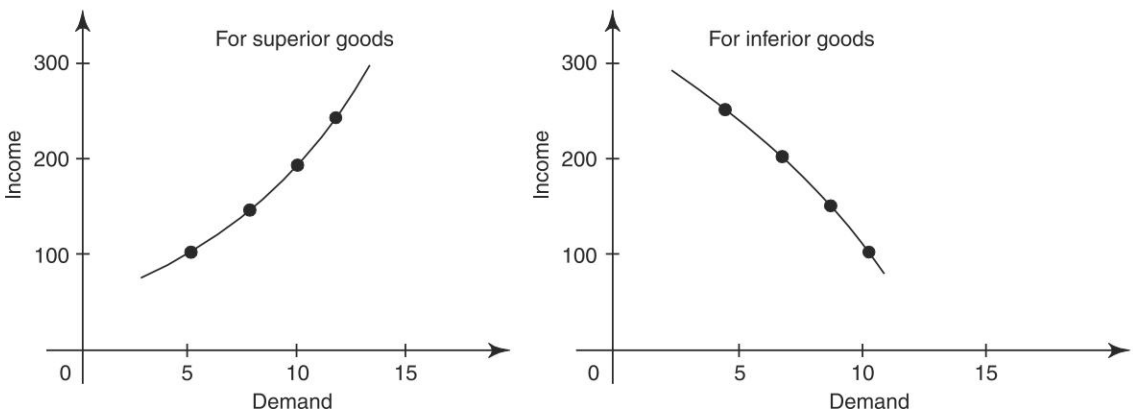



**Fig. 2.3** *Engel curves*

## Own Price and Demand (The Law of Demand)

Recall that demand is defined as a schedule indicating a relationship between demand for a good and price of that good. The said relationship is usually negative. Why? To provide a layman's answer, this is simply because price represents the cost of buying the product, and as the cost goes up, the consumer buys less of that product and vice versa. Further, the relationship between demand and price is partial because it assumes the "other things remaining the same" and these other things here refer to all the factors (like income as explained above) which affect the demand barring its own price. Thus, as the price of a Maruti car goes up, other things remaining same, the demand for Maruti car goes down, and vice versa. This is known as the law of demand. It must be emphasized here that the law, which describes the inverse relationship between quantity demanded and price, includes a rider, that is, other things remaining the same (or *ceteris paribus*). Some people make fun of economists by saying that while most

prices are going up, no demand is going down, meaning the law of demand does not hold good in real world. However, they forget that the 'other things' have not remained constant over time.

Why the law of demand? Several theories have been advanced to explain this law. The most convincing one goes through the concepts of price effect and its components, viz. income effect and substitution effect of a price change on its own demand. The ***price effect*** *(PE)* refers to the effect of a change in the price of a commodity, *ceteris paribus*, on the demand for that commodity. This effect is divided into ***income effect*** (*IE*) and ***substitution effect*** (*SE*):

$$PE = IE + SE$$

The income effect refers to the effect of a change in price of a product on the demand for that product, *ceteris paribus*, which arises due to the corresponding change in the (real) income of the consumer. For example, suppose the nominal income of the consumer is Rs. 40,000 per month, price of milk is Rs. 25 per litre and currently he buys 60 litres of milk per month. Now, if the price of milk falls to Rs. 20, *ceteris paribus*, he would save Rs. $60 \times 25 - 60 \times 20 = 300$ if he continues buying 60 litres of milk. Alternatively, with the earlier milk budget of Rs. $60 \times 25 = 1500$, he could now purchase 75 litres of milk instead of earlier 60 litres. Thus, a fall in the price of milk is tantamount to an increase in the consumers' income. Depending on whether milk is a *superior* or *inferior good*, its demand would increase or decrease. Thus, for a superior good, as its price falls, consumers' real income increases and so the demand for that product goes up. The resulting relationship between the price and demand is inverse and thus the income effect is negative for all superior goods. Quite the opposite would be true for the inferior goods and accordingly, the income effect would be positive for all inferior goods.

The substitution effect stands for the effect of a change in the price of a product on the demand *for that product, ceteris paribus*, which results due to a change in the relative prices of the goods in the consumption basket. For example, if the price of milk falls, other things remaining the same, milk would become relatively cheap (compared to other consumption goods), than it was before the price change occurred and therefore some people will substitute milk for, say, fruits, causing an increase in the demand for milk. Quite the opposite would happen when the price of milk goes up. Thus, a decrease in the price of milk, *ceteris paribu*s, would lead to an increase in the demand for milk, and vice versa, through the substitution effect. The substitution effect is thus always negative, in the sense that the price change and the change in demand through the substitution effect are in the opposite direction.

A price change, which causes both a change in consumers' real income as well as a change in the relative price of goods in the consumption basket, produces both the income and substitution effects, and thereby leads to a change in the demand for the good whose price has undergone a change. The total (sum of the two) effect is negative (i.e., the law of demand holds good) under two conditions:

(a) when the good in question is a superior good; and
(b) when the good in question is an inferior good but the negative substitution effect is stronger than the positive income effect.

When the product is a superior item, both the income and substitution effects are negative and hence this sum (total effect) is obviously negative. In a situation when a good in question is an inferior item, the two effects work in the opposite direction and their sum could either be negative or positive. But if the positive income effect is outweighed by the negative substitution effect, the price effect continues to be negative, substantiating the law of demand. The law of demand does not hold good for that inferior good (called the ***Giffen good****,* in the honour of Giffen, who first explained it) whose positive income effect

is stronger than the negative substitution effect. Thus, for a good to be Giffen, it must be an inferior good, it must have a strong income effect (which would be possible if and only if the consumer spends a significant part of his income on this good) and it must have a relatively weak substitution effect (which would happen if and only if there are no close substitutes for that product). The demand for a Giffen good varies directly with its price and thus it provides an exception to the law of demand. The Irish potato was considered one such good. Sir Robert Giffen noted that in 1845, 1846 and 1848, **Irish potatoes** were affected by fungus, destroying the crop significantly. Since potatoes happened to be the major crop and staple food, the event led to famine in Ireland. Finding themselves poorer and needing to maintain an adequate calories, Irish people gave up the more expensive food and turned increasingly to potatoes even in the face of rising price of potatoes. Under the situation, potatoes violated the law of demand and thus it was a Giffen good then. More recently (1987–89), some economists have opined that **Japanese SHOCHU** (an alcoholic beverages) may have also violated the law of demand. Soak and shochu are distilled from rice. While soak has three grades all of which are considered as high quality, shochu is of the lowest quality.

Exceptions to the law of demand could arise even for a non-Giffen good under the following situations:

(a) when the good in question is a luxury item, having some snob value.
(b) when the good in question goes out of fashion.

Rich people buy costly items such as diamonds in large quantities in spite of their high prices and sometimes they increase such purchases in the face of rising prices, for the acquisition of such expensive items distinguish them from common people who cannot afford them. Also, some consumers judge quality by high prices, and if so, they might purchase more of a product when its price is high than when its price was low. Similarly, with the emergence of television, the demand for radios has gone down even though their prices have declined. This is because radios are out of fashion these days.

The law of demand presented in the form of a table is called the ***demand schedule*** (vide section 2.1) and the one in the form of a graph is called the ***demand curve*** (vide Fig. 2.1). So long as the law of demand holds good, the demand curve is falling, and depending upon the degree of the relationship between price and demand, (vide Section 2.7), the curve could be linear, convex or concave. However, it should be noted that the *market* (aggregate) *demand curve does not touch either the price or quantity axis.* It does not touch the price axis because if it does, it would mean that there is some price at which demand is zero, which is never the case, for no matter what the price is, there would be at least some buyer in the market and no seller would set a price at which there is no demand. Similarly, the demand curve would not cut the quantity axis, because if it did, it would mean that there can be a zero price (free good), which is not possible for any economic good as there are no doubt some production costs, and that even at a zero price, the demand is limited. Incidentally, it may also be noted that since there is a unique quantity demanded at a unique price at a given point of time, all but one price-quantity combination on the demand schedule/curve are hypothetical ones.

## Prices of Related Goods and Demand

Recall that consumers' goods may have either of the two kinds of relationships: substitutes and complements. If the two goods *A* and *B* are substitutes, an increase in the price of good *B,* price of good *A* remaining constant, would induce consumers to substitute good *A* for good *B*, because good

*B* has become relatively more expensive now than before, and thereby increases the demand for good *A*. Thus, an increase in the price of a substitute good, say, Maruti Zen car, would lead to an increase in the demand for, say, Tata Indica Car. Quite the opposite would happen when the price of a substitute item falls. Thus, the relationship between the price of a good and the demand for its substitute items, is positive. In the case of complementary goods, the relationship is the other way round, that is negative. In particular, if the price of a complementary item (petrol) goes up, the demand for the parent good (car) goes down; for since the former item (petrol) has become relatively expensive now, the consumer would like to demand less of it, and since, he has less of it (petrol) now, he would need less of the latter good (car) as well. Similarly, when the price of a complementary good goes down, the demand for its parent good goes up. Thus, the relationship between the price of a good and the demands for its complementary items, is negative.

Since an item could have more than one substitute goods, and/or more than one complementary products, each of the substitute' prices and complementary' prices is a vector of prices rather than just one price.

## Consumers' Tastes and Preferences, and Demand

Consumers' tastes and preferences are an important determinant of the demands for all consumers' goods. If a person is a pure vegetarian, either because of religion, tradition or taste, his demand for meat is zero no matter what his income and the price of meat are. Similarly, if a product goes out of fashion, or taste, its demand goes down. On the other hand, if an item becomes popular (due to improved taste for it), its demand goes up. Television, car and most luxury items fall in the category of popular items. Also, while some people prefer to consume rice, some others prefer to eat chapattis, and so on.

Producers' spend a lot of money on advertising their products primarily because they can influence the tastes and preferences of the consumers in their favour, and thereby achieve an increase in the demand for their own products. It is essentially this activity which enables firms to manage the demand for their products.

## Consumers' Expectations and Demand

Consumers' expectations with regard to their future income, and future prices of the good in question in relation to its substitutes and complementary products exert influence on the demand for many goods. For example, graduate students are often observed to spend beyond their means, for their future incomes are high, while the service people in their late fifties try to cut on non-essentials, for their future incomes are low. Thus, an increase in expected future income leads to an increase in the demand for some consumers' goods and vice versa. Similarly, an increase in the expected future price of a product leads to an increase in the demand for that product in the current period. To cite an example, during the inflationary situation, there is generally an upward movement in the demands for durable goods (e.g., refrigerators, air conditioners, televisions, computers, housings), for there is always a fear that their prices might go up after wards when they would really like to have such goods. Also, when consumers fear shortage of a commodity, they are often found to buy and stock under panic. Quite the opposite holds good when there are rumours of bumper crops and deflation.

The expectations' variable plays a more significant role in the case of demand for durable and expensive items than it does in the case of demand for perishable and cheap products. This is because

the purchases of durables can be postponed and advanced more easily than those of non-durables, and the price changes matter more in the case of expensive goods than non-expensive ones. For this reason, the expectations' variables are often left out from the list of demand determinants for non-durable and cheap goods.

### Number of Consumers, their Distribution and Demand

The aggregate (firm, industry and market) demand for a good obviously depends also on the number of consumers. Other factors remaining the same, the larger the number of consumers, the greater is the demand, and vice versa. Thus, the demand for almost all the products in India as well as in the world as a whole is increasing over time, partly because the population is increasing over time. Incidentally note that the effect of the number of consumers on the aggregate demand could be taken care of in either of the two ways, viz. (a) through using the per capita income and the number of consumers (population size), (b) through using the aggregate (national) income instead of an individual consumer's income, in the demand function.

Since demand for most products vary from consumer to consumer, distribution of consumers among appropriate categories also exert an influence on the aggregate demand. For example, demand for a car depends on the distribution of households into rich and poor ones. If the proportion of rich households to the total number of households increases, demand for cars would increase and vice versa. Similarly, demand for cosmetics would increase if the proportion of women in the total population increases, and demand for baby food would fall if the proportion of babies in the total population falls, and so on. The other relevant distributions in this respect could be male and female, smokers and non-smokers, vegetarians and non-vegetarians, literates and illiterates, Hindus and non Hindus, and so on. Obviously, all these distributions are never simultaneously important in the case of a study of the demand for anyone product. Thus, the researcher has to try to choose only the appropriate ones for a particular demand.

Needless to say, neither the population nor its distribution is a relevant determinant of the demand for a good by an individual consumer.

This is not an exhaustive list of the determinants of demand for consumers' products. However, it does include all the significant ones. The remaining factors are basically psychological, climatical, traditional, cultural, habitual, demographic (like age, sex, education and caste), etc. and therefore difficult to quantify. To account for such factors, researchers often resort to the use of dummy variables in their empirical works. It must also be emphasized here that all the determinants are not equally important in the demand for various goods and services. Some variables are significant in the demand for some goods, while other variables are important in demands for other goods. We have discussed this briefly here because only experience can teach the rest.

## 2.4 DEMAND FUNCTION

In a nutshell, the relationship between a variable and its determinants is often described through a mathematical function. The demand function for a good relates the quantities of a good which consumers demand during some specific period to the factors which influence that demand. Thus, mathematically, the demand function for a good $X$ can be expressed as follows:

$$D_x = f(Y, P_x, P_s, P_c. T; E_p. E_y; N, D, u)$$

$$f_3,\ f_6,\ f_7,\ f_8 > 0 > f_2,\ f_4$$
$$f_1,\ f_5,\ f_9,\ f_{10} >,\ <\ \text{or} = 0 \qquad (2.1)$$

where,

$D_x$ = demand for good $x$

$Y$ = consumers' income

$P_x$ = price of good $x$

$P_s$ = prices of substitutes of good $x$

$P_c$ = prices of complements of good $x$

$T$ = measure of consumers' tastes and preferences for good $x$

$E_p$ = consumers' expectations about future price of good $x$

$E_y$ = consumers' expected future incomes

$N$ = number of consumers of good $x$

$D$ = distribution of consumers in some specific classification

$u$ = 'other' determinants of the demand for $x$

$f$ = unspecified function, to be read as "function of" or "depends on"

$f_i$ = partial derivative of $f$ with respect to the $i^{th}$ variable

The function (2.1) was rationalised above (and the full theory behind it, known as the Theory of Consumers Behaviour is presented in Chapter 3) and there is no point in repeating that here. It would suffice to point out here, $f_1$ is positive if $X$ happens to be a superior good and negative if it were an inferior good; $f_5$ is positive if consumers develop taste and preferences in favour of $X$ and negative if against it, the sign of $f_9$ depends on the way an appropriate distribution of consumers undergo a change, and the sign of $f_{10}$ would be subject to the specific determinant chosen/included in the function. Incidentally, it may also be recalled here that the first five determinants ($Y$, $P_x$, $P_s$. $P_c$, and $T$) affect the demand for all goods, the next two ($E_p$ and $E_y$) exert an influence mainly on the demand for durable and expensive goods, and the next two ($N$ and $D$) are arguments only in the demand functions for some group of consumers.

From the management point of view it would be of interest to know the *degree of control* a management exercises on the various determinants of the demand for its product. The management enjoys practically no control over the variables such as income (both current and expected), prices of related goods and the distribution of consumers. Depending upon the market structure in which it operates (vide chapters on pricing), it might have some role in determining the product price, and it does play a significant role in the determination of the number of consumers and their tastes and preferences through advertisement, goodwill, product quality, service, etc.

Recall that in Section 2.2, we made an important distinction between goods, viz. consumers and producers' goods. Function (2.1) above denotes the demand function for consumers goods. The one for producers goods [like labour, capital (structure, equipment and inventories), raw-materials and intermediate goods], would have a little different set of determinants. In particular, the differences in the determinants would be as follows:

(a) Instead of income, and expected income in consumers goods demand function, production level and expected future production, respectively would be the relevant arguments in the producers goods demand function.

(b) Prices of related goods (substitutes and complements) would remain even in producers goods demand function, but with a significant difference. Here the relationship is in production, not in consumption. For example, in production, capital is a substitute for labour and raw-material is a complement for labour and thus the demand for labour function would have the price of capital (capital rental) and price of raw-material as the determinants for the prices of related goods.

(c) Instead of consumers taste variable, production technology would be the relevant variable in the demand for producers goods function.

(d) Own price and expected own price would remain true even in the demand for producers goods demand function.

(e) Instead of number and distribution of consumers, the number and distribution of producers (firms) would be the relevant arguments in the producers goods demand function

In addition to the above classification, recall that in an earlier section to this chapter, we made a distinction between firm demand and industry demand. Here, it is pertinent to emphasize that though the two demand functions would have similar arguments, the direction and magnitudes of their effect would presumably be different. **The demand function facing a firm** (e.g. Maruti Udyog) would have rival firms' product prices as the prices of substitutes and thus 'other' firms' prices would be expected to exert a positive influence on the demand for a firm's product. In contrast, in the **demand function facing an industry** (car industry) all firms' prices are parts of its own price, thus each of these prices would exert a negative influence on the demand for the industry's product. Similarly, while the advertisement budget, which affects consumers' tastes and preferences, of 'other' firms would adversely affect the demand for a firm's product, the advertisement budget of all firms would be expected to favourably affect the demand for the industry's product. In terms of the magnitudes or strengths of the effect, consumers' income in the firm demand function would have a smaller coefficient than that in the industry demand function.

Another useful distinction is between **extension** and **contraction in demand**, and **increase** and ***decrease* (*change*) *in demand***. The former refers to the one caused by change in 'own' price while the latter to the one brought about by change in any argument other than 'own' price. For example, if demand for a Maruti car responds to a change in the price of the Maruti car, it is called extension or contraction in demand, while if it responds to changes in consumers' income, price of Tata's car, price of petrol or distribution of population into rich and poor, it is called increase or decrease in demand. To make it still clearer, movements along a demand curve are known as extension and contraction in demand, while movements from one demand curve to another are called increase or decrease in demand. It may be recalled that a demand curve describes the relationship between the quantity demanded of a good and its own price, other things remaining the same. Two hypothetical demand schedules for milk are given below and the corresponding demand curves are given in Fig. 2.4.

| *Consumer's income = 500* | | *Consumer's income = 1000* | |
|---|---|---|---|
| *Milk price* | *Milk demand* | *Milk price* | *Milk demand* |
| 32 | 40 | 32 | 50 |
| 24 | 50 | 24 | 70 |
| 20 | 60 | 20 | 90 |
| 16 | 80 | 16 | 120 |

$D_1D_1$ and $D_2$ $D_2$ are two different demand curves, corresponding to two levels of consumer's income. On $D_1D_1$, when one moves from demand of 50 at price 24 to demand 60 at price 20, it is an extension in demand, and when it moves backward from demand of 50 at price 24 to demand of 40 at price 32, it is contraction in demand. In contrast, when one moves from price of 24 and demand of 50 at $D_1D_1$ to price of 24 and demand of 70 to $D_2$ $D_2$, it is increase in demand; the reverse movement would be called decrease in demand. Thus, function 2.1 above indicates that the demand for good $X$ increases (or the demand curve shifts to the right) with an increase in consumers' income (assuming $X$ is a superior good); prices of substitutes of $X$, expected future prices and incomes and/or the number of consumers, while it decreases (or the demand curve shifts to the left) with increase in prices of complementary goods. Accordingly, all the 'non-own' price determinants of demand are together known as the "**demand shifters**".

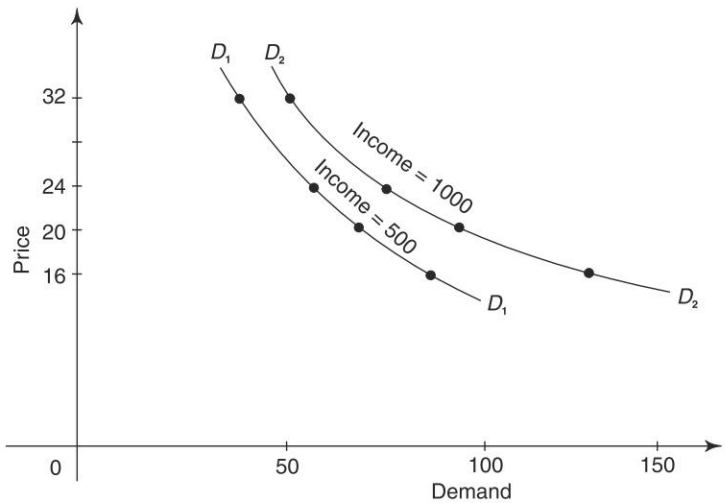



**Fig. 2.4** *Change in demand*

## 2.5 DEMAND ELASTICITIES

The foregoing section delineated the forces which impinge on demand as well as the directions of their effects. For managerial decisions, one also needs the extent or magnitude of these effects. One way to measure the magnitude is provided by the concept of elasticity. Elasticity is a general measurement concept and therefore, in what follows, we shall first discuss its general meaning and then go over to demand elasticities.

### Elasticity Concept

Elasticity is a measure of the sensitiveness of one variable to changes in some other variable. It is expressed in terms of a percentage and is devoid of any unit of measurement. For example, elasticity of a variable $X$ with respect to variable $Y$ ($e_{x'}$, $_y$) is expressed as

$$e_{x',y} = \frac{\text{Percentage change in } x}{\text{Percentage change in } y}$$

$$= \frac{\Delta x/x}{\Delta y/y} = \frac{\Delta x}{\Delta y}\left(\frac{y}{x}\right) \tag{2.2}$$

This is the formula if the data is discrete. For continuous functions, it is given as

$$e_{x',y} = \left(\frac{\partial x}{\partial y}\right)\left(\frac{y}{x}\right) \tag{2.3}$$

In this expression, please note that $X$ is a dependent (effect) variable and $Y$ is an independent (cause) variable. The term $\partial x/\partial y$ stands for partial derivative of $X$ with respect to $Y$, as $X$ could depend not only on $Y$ but also on some other variables, like $W$, $Z$ and so on, which are taken as given under partial derivatives. In other words, elasticity assumes *ceteris paribus*, as in the law of demand.

Elasticity could be measured at a point or on an arc (line). To explain this, consider a hypothetical schedule, describing a relationship between $X$ and $Y$:

| | $X$ | $Y$ |
|---|---|---|
| (a) | 5 | 50 |
| (b) | 3 | 100 |

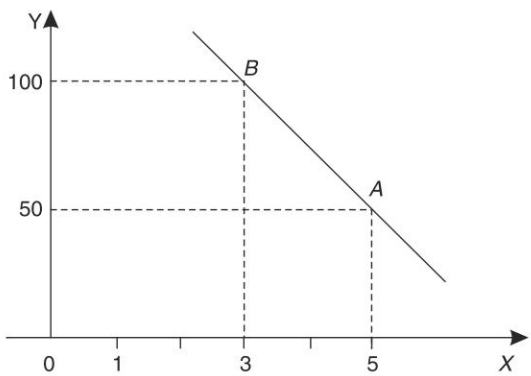



**Fig. 2.5**

The elasticity at point $A$ where $X = 5$, $Y = 50$ equals

$$e_{x,y}(A) = \left(\frac{-2}{50}\right)\left(\frac{50}{5}\right) = -0.4$$

The elasticity at point $B$, where $X = 3$, $Y = 100$ equals

$$e_{x,y}(B) = \left(\frac{2}{-50}\right)\left(\frac{100}{3}\right) = -1.33$$

Thus, if the set of values of the two variables at point $A$ are used, then it is the **point elasticity** at point $A$, and so on. In contrast, there is a concept which computes the **elasticity on the arc**. The formula for that is given by

$$e_{x,y}(\text{arc}) = \left(\frac{\Delta x}{\Delta y}\right)\left[\frac{\frac{1}{2}(y_1 + y_2)}{\frac{1}{2}(x_1 + x_2)}\right]$$

$$= \left(\frac{\Delta x}{\Delta y}\right)\left[\frac{y_1 + y_2}{x_1 + x_2}\right] \tag{2.4}$$

where subscripts 1 and 2 refer to the first and second values of the corresponding variables. Thus, the arc elasticity in the above example would be

$$e_{x,y}(\text{arc}) = \left(\frac{-2}{50}\right)\left(\frac{50+100}{5+3}\right) = -0.75$$

It would be seen that the arc elasticity lies between the corresponding two point elasticities and this is always true. However, the former is neither a simple average, nor any unique weighted average. The significant question is, which of the two elasticity concepts is more meaningful than the other for decision making? The answer depends upon the problem at hand. For small changes in the dependent variable, the point elasticity is relevant while for large changes, arc elasticity is the right concept to use.

## Demand Elasticities

Demand elasticities refer to elasticities of demand for a good with respect to each of the determinants of its demand. There is one demand elasticity with respect to each demand determinant and thus there are as many demand elasticities as the number of demand determinants. The important ones are the following:

(a) Price elasticity of demand
(b) Income elasticity of demand
(c) Cross elasticity of demand
(d) Promotional (advertisement) elasticity of demand

A discussion of each of these follows. At the outset, it must be emphasized that both the point and arc elasticity concepts are applicable to each of these elasticities. However, for the sake of simplicity, their formulae would be expressed in terms of point elasticity only. Also elasticities can be computed both for the discrete as well as continuous functions. However, again for the sake of simplicity, formula would be given for discrete functions only.

Incidentally, it is pertinent to note that since there is more than one determinant of demand, each of the *various demand elasticities are partial* in the sense that they assume all the 'other' determinants as held constant. For example, income elasticity of demand refers to the sensitivity of demand to changes in income, all the non-income determinants of demand remaining constant.

## Price Elasticity of Demand

Price elasticity of demand refers to elasticity of demand for a good ($D_x$) with respect to its own price ($P_x$). Symbolically, it can be expressed as

$$e_{D_x, P_x} = \left(\frac{\Delta D_x}{\Delta P_x}\right)\left(\frac{P_x}{D_x}\right) \tag{2.5}$$

If the law of demand holds good, the price elasticity of demand would be negative. Thus, if the two points on the demand curve were:

| $P_x$ | $D_x$ |
|---|---|
| 10 | 20 |
| 8 | 30 |

Then the price elasticity at the original point (first) would be $\left(\frac{10}{-2} \times \frac{10}{20}\right) = -2.5$ and at the new point would equal $\left(\frac{-10}{2} \times \frac{8}{30}\right) = -1.33$.

There are five critical values for this elasticity, viz. zero, less than one, one, more than one, and infinity. The price elasticity is zero (or demand is perfectly price inelastic) when demand is invariant to all changes in price. There is hardly any commodity in the world for which this is true. The closest example would be for salt. Salt is an inexpensive and yet an essential consumption item and its consumption can vary only within a small range. For this reason alone, its consumption hardly varies with variations in its price.

Price elasticity is less than unity (or demand is relatively price inelastic) when changes in price leads to a less than proportionate change in demand. A large number of goods and services, which include all the essential items, have inelastic demand. On the other hand, if percentage change in demand exceeds that in price, price elasticity is greater than unity (or demand is relatively price elastic). Most luxury items have elastic demands. Demand is unitary price elastic when percentage changes in both price and demand are the same.

In the situation where demand is unlimited or unbound at a given price and it vanishes altogether even if there is a small increase in its price, the demand is said to be perfectly or infinitely price elastic. Once again, there is perhaps no industry whose product possesses such a characteristic. Incidentally, as we shall see later in a chapter on pricing, demand for a firm's product in a purely competitive market is perfectly price elastic.

Figure 2.6 depicts some demand curves for constant price elasticities.

Elasticity is zero at every point on $D_1D_1$, is infinity on every point on $D_2D_2$ and it is unity at every point on a rectangular hyperbolic demand curve $D_3D_3$. On a straight line falling demand curve, elasticity varies from point to point (See Fig. 2.7).

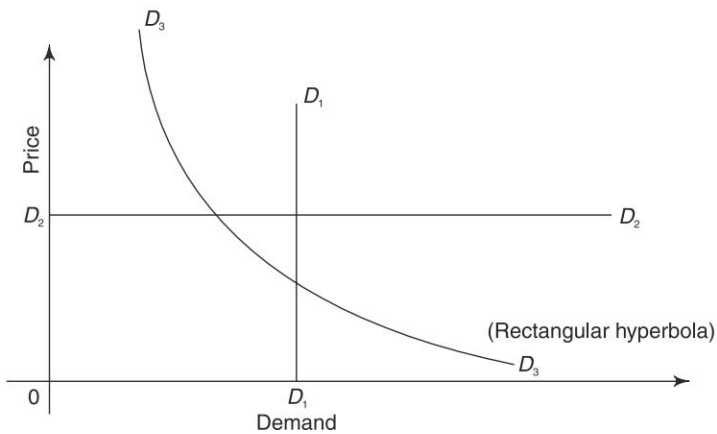



**Fig. 2.6** *Constant Elasticity Demand Curves*

It can be proved that on such a curve/line (See as in Fig. 2.7), elasticity (absolute value) on, say, point *P* equals *PB/PA*, point *R* equals *RB/RA*, and on point *S* equals *SB/SA*. By the same reasoning, elasticity at point *B* equals zero while that at point *A* equals infinity. Since point *P* is on the middle of the demand curve *AB*, the elasticity at this point is unity, point *R* lies between points *P* and *B*, elasticity at this point is less than unity (in absolute terms), and point *S* falls between points *A* and *P*, the elasticity at this point is greater than unity (in absolute terms). By recourse to the formula of the price elasticity of demand (Eq. 2.5), it would be easy to see that the said elasticity is high near point *A* because in this range while $P_x$ is high $D_x$ is low; $DD_x/DP_x$ is, of course, constant on a linear demand curve. In contrast, elasticity is low around point *B* because here while $P_x$ is low, $D_x$ is high.

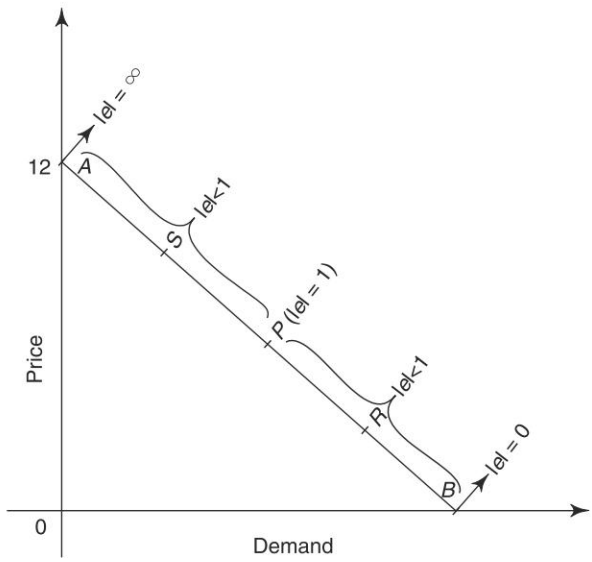



**Fig. 2.7** *Elasticity on a straight line falling demand curve*

The magnitude of the price elasticity of demand is governed by several factors, including the (a) nature of the good i.e., is it an essential item or a luxury one, (b) closeness of the substitute goods available, if any, (c) proportion of income spent on the good under question, and (d) time available for adjustment. The necessary items have low elasticity, for not only their consumption is essential but also it varies only within a small range. For example, a household requirement for wheat might vary with changes in wheat price but the response would be poor. In contrast, its purchase of petrol for its car for pleasure (luxury) would respond significantly to changes in petrol price.

The closer the substitutes available for a product, the greater would be the price elasticity, *ceteris paribus.* This is because the substitution effect would be strong in the case of close substitutes. For example, when price of good *x* goes up, *ceteris paribus,* people would switch from good *x* to its substitutes, if available. On this ground, price elasticity of demand for a brand of a product is higher than that for the product as a whole. In other words, the demand curve facing a firm is more elastic than the one faced by the corresponding industry. Thus, the Maruti car has high price elasticity of demand than cars in general.

The larger is the proportion of income spent on a good, the greater is the price elasticity of demand. This is because of the magnitude of the Income effect. A given change in price brings about a greater change in consumers' income if he were spending a larger percentage of his income on that good than if

he were spending a smaller share on it. Given the nature of the goods, a large change in income would produce a large change in the demand for a good and vice-versa. Thus, if a household spends 10 per cent of its income on rice and 2 per cent on wheat, *ceteris paribus,* the price elasticity of its demand for rice would be higher than that for wheat.

More the time a consumer gets for adjusting its consumption basket, the larger would be the price elasticity of demand for that product, and vice versa. This is because, consumers are reluctant to change from a good to its substitute but this reluctance gets weaker and weaker over time as the consumer realizes that he is irrational. For example, one may be using a Colgate tooth paste since birth and when its price goes up, other tooth pastes prices remaining unchanged, initially he may continue with the Colgate but eventually he would give it up in favour of some other brand.

Salt is a special commodity with regard to the price elasticity of demand. This is because the elasticity is perhaps the lowest for this product and this is because the first three determinants of the value of this elasticity (the fourth one, time, is irrelevant here) work uniformly towards this. It is a necessary item, it has perhaps no substitute and all households spend a rather small part of their income on this product.

The concept of price elasticity of demand is very useful for decision making. Its major significance lies in price determination. Since the topic of price determination is taken up later, a detail explanation of the managerial uses of price elasticity of demand would have to be deferred. However, this would be briefly discussed later in this chapter under the section on demand-revenue-relationships.

## Income Elasticity of Demand

Income elasticity of demand ($^{e}D_{x,y}$) measures the sensitiveness of demand ($D_x$) to change in consumer's income ($y$). Symbolically, it can be expressed as

$$e_{D_{x,y}} = \left(\frac{\Delta D_x}{\Delta y}\right)\left(\frac{y}{Dx}\right) \tag{2.6}$$

Recalling the meanings of various relevant concepts from an earlier section, it would be clear that the income elasticity of demand is (a) positive for superior goods and negative for inferior goods, (b) positive and less than one for all superior and necessary goods, (c) positive and around unity for all superior and semi-luxury (comforts) goods, and (d) positive and greater than one for all luxury (which are of course superior) goods.

An understanding of the income elasticity of demand is of great significance for both firms and the government. Firms whose products' demand functions have high income elasticities will have good growth opportunities in an expanding economy. Such firms therefore must take a close look at the aggregate economic activity and its likely growth rate in the future while planning their productions. In contrast, companies whose products have low income elasticities would neither gain much if the economy expands nor lose significantly if the economy retards. A corollary of this is that in a growing economy, while farmers suffer as their products have low income elasticities, industrialists gain as their products have relatively higher income elasticities. This will pose a significant problem to the policy makers and, in fact, most developing economies are plagued by this disease.

## Cross Elasticity of Demand

The cross (price) elasticity of demand ($e_{D_x}$, $p_y$) refers to the responsiveness of demand for product $x$ ($D_x$) to changes in the price of product $y$ ($P_y$). It can be computed as follows:

$$e_{D_x, P_y} = \left(\frac{\Delta D_x}{\Delta P_y}\right)\left(\frac{P_y}{D_x}\right) \tag{2.7}$$

It is positive if goods $x$ and $y$ are substitutes in the consumption basket, negative if they are complements, and zero if the two goods are unrelated. The greater the magnitude of this elasticity, the stronger is the relationship between two goods. Thus, the cross elasticity of demand between the demand for a Maruti car and the price of a Tata car would be positive, and that this elasticity would be greater than the cross elasticity of demand between the demand for a Maruti car and the price of a Bajaj scooter.

The cross-price elasticity concept is used for two main reasons. One, to assess the significance of changes in related products' prices on its own market, and thereby to formulate ones' own pricing strategy. For example, a knowledge of the cross elasticity between the Maruti car demand and Tata car price would enable Maruti Udyog to estimate the potential loss in its market in the face of a threat by Tata Motors to reduce its price by a given figure. Two, cross-elasticity is used in industrial organisations to measure the inter-relationships between industries and sectors. For instance, in an agriculture dominant economy like India, knowledge of the cross elasticity of demand for industrial products with respect to agricultural price would indicate the extent by which farmers' prosperity is the cause of industrial growth.

## Promotional Elasticity of Demand

Promotional elasticity ($e_{Dx,A}$) stands for the sensitiveness of demand for good $x$ to changes in advertisement budget of the producers ($A$) of good $x$. Its formula would be the following:

$$e_{D_x, A} = \left(\frac{\Delta D_x}{\Delta A}\right)\left(\frac{A}{D_x}\right) \tag{2.8}$$

This must obviously be positive, for advertisement expenditures are supposed to boost up the market. The higher this elasticity, the greater would be the incentive for the firm to go in for advertising its product.

Besides these four demand elasticities, as hitherto pointed out, there are other demand elasticities which are with respect to 'other' (other than the four corresponding demand determinants) arguments in the demand function. For example, the interest rate may be an important determinant in the demand for some expensive goods like housing, and if so, one might like to compute the interest elasticity of demand for housing. In the western world, where the climate is so cold, public utilities like electricity would like to know the weather elasticity of demand for their products, and so on.

Before we end this section, it is pertinent to explain the calculations of various demand elasticities through a hypothetical example. Let the demand function for coffee of a typical consumer be the following.

| *Year* | *Coffee Price (Rs/kg)* | *Quantity of Coffee Bought (kg)* | *Real Income (Rs)* | *Tea price (P.kg)* |
|---|---|---|---|---|
| 1 | 95 | 20 | 1000 | 35 |
| 2 | 98 | 18 | 1000 | 35 |
| 3 | 98 | 21 | 1050 | 35 |
| 4 | 95 | 21 | 1000 | 40 |

What are the values of various demand elasticities? There are in all three elasticities involved in this example, viz. price, income, and cross elasticities. Between years 1 and 2, while consumers' incomes and tea price have remained constant, coffee price and demand have undergone a change. Thus, price elasticity of demand ($e_{D,p}$) is computable here:

$$e_{D,P} = \left(\frac{\Delta D}{\Delta P}\right)\left(\frac{P}{D}\right)$$

At point year 1: $$e_{D,P} = \frac{-2}{3} \times \frac{95}{20} = -3.12$$

At arc (years 1 and 2): $$e_{D,P} = \frac{-2}{3} \times \frac{95+98}{20+18} = -3.4$$

Similarly, between years 2 and 3, both coffee and tea prices have remained the same, and income and coffee purchases have changed, thus income elasticity of demand for coffee ($e_{D,Y}$) is computable:

At point year 2: $$e_{D,Y} = \frac{3}{50} \times \frac{1000}{18} = 3.3$$

Between years 1 and 4, while coffee price and income have remained unaltered, there is a change in tea price and coffee purchases. This permits the calculation of cross elasticity of demand for coffee:

$$(e\ D_C, P_T) = \left(\frac{\Delta D_C}{\Delta P_T}\right)\left(\frac{P_T}{D_C}\right)$$

At point year 4: $(e\ D_C, P_T) = \frac{1}{5} \times \frac{40}{21} = 0.38.$

It should be noted that if one had the observations for years 1 and 3 alone, no such computation would have been possible. This is because between these two years, more than one variables (viz. coffee price and income) have undergone a change and hence the required *ceteris paribus* assumption for elasticity calculation is not valid.

## 2.6 DEMAND-REVENUE RELATIONSHIPS

There is a definite relationship between demand for a good and the revenue generated by that good. These two variables are connected by the price of the good in question. Before one dwells on this, it is pertinent to understand three revenue concepts, viz. total, average and marginal revenues. Total revenue (*TR*), by definition, equals quantity ($Q$) sold of a good times its price ($P$):

$$TR = Q \cdot P \tag{2.9}$$

Average revenue (*AR*) refers to revenue per unit of quantity sold. Thus,

$$AR = \frac{TR}{Q} = \frac{QP}{Q} = P \tag{2.10}$$

Marginal revenue (*MR*) stands for the change in total revenue brought about by an infinitesimal change in quantity sold:

$$MR = \frac{d(TR)}{dQ}$$

For a discrete function, no such calculation is possible. Instead, there is a concept called arc marginal revenue or average incremental revenue which is defined as the change in total revenue divided by change in quantity of product sold:

$$\text{Arc } MR = \frac{\Delta TR}{\Delta Q} \tag{2.11}$$

When a change in *Q* equals one, arc *MR* (henceforth called just *MR*) equals change in *TR*. Given a demand schedule/curve, one can easily obtain the *TR, AR* and *MR* schedules. For example, for the linear demand schedule in columns 2 and 3 below, the three revenue schedules would be as follows:

| **Point** | *P* | *Q* | *TR* | *AR* | *MR* |
|---|---|---|---|---|---|
| *a* | 10 | 3 | 30 | 10 | |
| *b* | 9 | 4.5 | 40.5 | 9 | 7 |
| *c* | 8 | 6 | 48 | 8 | 5 |
| *d* | 7 | 7.5 | 52.5 | 7 | 3 |
| *e* | 6 | 9 | 54 | 6 | 1 |
| *f* | 5 | 10.5 | 52.5 | 5 | –1 |
| *g* | 4 | 12 | 48 | 4 | –3 |

The *TR* column is the product of *P* and *Q* columns' corresponding rows (e.g. 10 × 3 = 30, and 9 × 4.5 = 40.5), *AR* column (which is identical to *p* column) is *TR* column divided by the corresponding number in *Q* column (e.g. 30/3 = 10 and 40.5/4.5 = 9), and *MR* column as the change in *TR* divided by the corresponding change in *Q* [e.g. (40.5 – 30) | (4.5 – 3) = 7 and (48 – 40.5) | ( 6 – 4.5) = 5].

The corresponding graphs of demand (= *AR*) curve, *MR* and *TR* curves are provided in Fig. 2.8.

In this example, the *TR* curve is inverted U-shaped. In particular, *TR* is zero when either *Q* = 0 (at point *A* on the demand curve) or *P*= 0 (at point *B* on *D* curve). *TR* is maximum when *MR* = 0 (at point *R* on *D* curve). The shapes of *MR* and *TR* curves are derived from that of the *AR* or demand curve. For instance, if the demand curve was parallel to the quantity axis (i.e., horizontal); as is the case for a perfectly competitive firm (vide chapter 8), the various schedules/curves would be as given below in table and Fig. 2.9.

| *P* | *Q* | *TR* | *AR* | *MR* |
|---|---|---|---|---|
| 10 | 0 | 0 | 10 | |
| 10 | 5 | 50 | 10 | 10 |
| 10 | 10 | 100 | 10 | 10 |
| 10 | 20 | 200 | 10 | 10 |
| 10 | 40 | 400 | 10 | 10 |

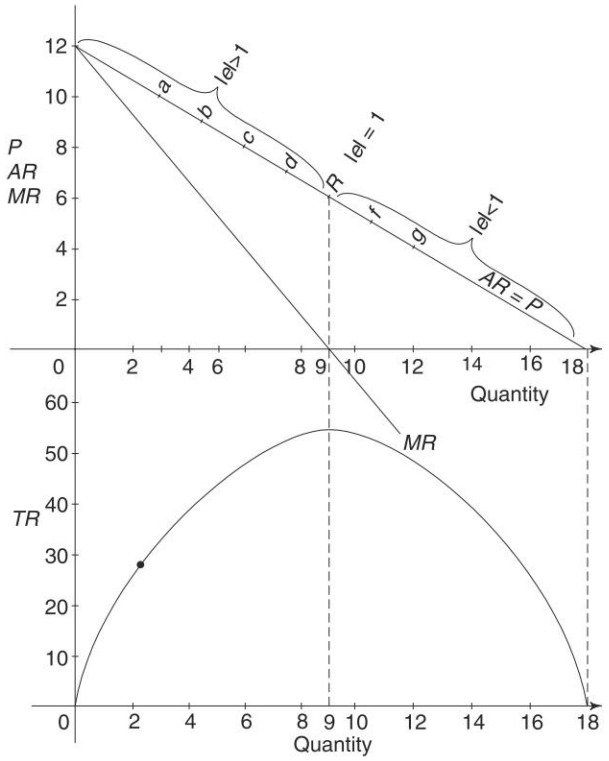



**Fig. 2.8** *Linear demand curve and corresponding AR, MR and TR curves*

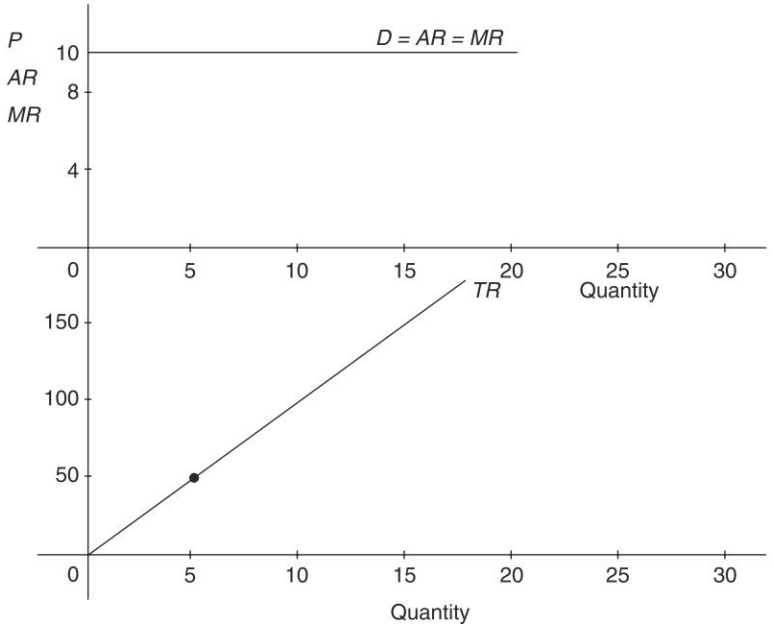



**Fig. 2.9** *Horizontal demand curve and corresponding AR, MR, and TR curves*

A careful examination of the linear and falling demand schedule and the corresponding *TR* schedule (See Fig. 2.8 and the corresponding Table above) would reveal that in the beginning (when *P* is high) as price falls, *TR* increases but after a while *TR* starts declining. The reason of this can be found in the magnitudes of the price elasticity of demand. The said elasticity at

(a) point *a* on arc *ab* = $\left(\frac{1.5}{-1}\times\frac{10}{3}\right) = -5$

(b) point *b* on arc *bc* = $\left(\frac{1.5}{-1}\times\frac{9}{4.5}\right) = -3$

(c) point *c* on arc *cd* = $\left(\frac{1.5}{-1}\times\frac{8}{6}\right), = -2$

(d) point *d* on arc *dR* = $\left(\frac{1.5}{-1}\times\frac{7}{7.5}\right), = -1.4$

(e) point *R* on arc *Rf* = $\left(\frac{1.5}{-1}\times\frac{6}{9}\right), = -1$

(f) point *f* on arc *fg* = $\left(\frac{1.5}{-1}\times\frac{5}{10.5}\right) = -0.71$, and at

(g) point *g* on arc *fg* = $\left(\frac{1.5}{-1}\times\frac{4}{12}\right) = -0.50$

It would be observed that so long as the absolute value of the price elasticity of demand is greater than unity, a decrease in price leads to an increase in total revenue and vice versa. Quite the opposite holds good when the said elasticity takes an absolute value of less than unity. In other words, if the absolute value of the price elasticity is greater than one (e.g. –1.5), *P* and *TR* more in the opposite direction, and if the said elasticity is less than one (e.g. –0.6), *P* and *TR* move in the same direction, A corollary of this, called the *total revenue test of the price elasticity*, is that if *P* and *TR* move in the opposite direction, the said elasticity (in absolute sense) is greater than unity and vice versa. Looking at the same thing through marginal revenue, one could say that if *MR* is positive, the price elasticity of demand is greater than unity, if *MR* is negative, the said elasticity is less than unity, and if *MR* equals zero, the elasticity takes an exact value of minus one.

For the readers with a mathematical background, the above relationship could be derived as follows:

$TR = PQ$ (by definition)

Differentiation of this with respect to quantity ($Q$) gives

$$MR = P + Q\,\frac{dP}{dQ}$$

$$= P\left(1+\frac{Q}{P}\frac{dP}{dQ}\right)$$

or, $$MR = P\left(1+\frac{1}{e}\right) \qquad (2.12)$$

From Eq. (2.12) it is obvious that

(a) if $|e| = -1$, $MR = 0$ and $TR$ = constant as $P$ and $Q$ change

(b) if $|e| > 1$, $MR > 0$, and as $P$ falls, both $TR$ and $Q$ increase, and vice-versa

(c) if $|e| < 1$, $MR < 0$, and as $P$ falls $TR$ decreases and $Q$ increases, and vice-versa

The above fact has an important implication for decision making: no profit-maximising (or even simply revenue maximising) *firm should operate* on the *inelastic* portion $|e| < 1$ of its demand curve. To explain this, consider the above example of falling linear demand curve (See Figure 2.8). Suppose, a firm were operating at point *f*, where it sold 10.5 units at a price of 5, and the price elasticity of demand was – 0.71. This is easy to see that it cannot be a profit-maximising or even sales maximising point. For, by raising the price from 5 to 6, it would ensure an increase in total revenue (as $|e| < 1$) and at the same time a decrease in total cost, for its sales (production) would come down from 10.5 units to 9 units, and thus an increase in profit. Thus, one could conclude that the firm must increase its price if $|e| < 1$. However, no such conclusion is possible if $|e| > 1$. For example, let the firm were at point *b* where $P = 9$, $Q = 4.5$ and $e = -3$. Now, if the firm raises its price, both $TR$ and $TC$ fall ($Q$ decreases) and if it lowers the price, both $TR$ and $TC$ increase ($Q$ increases); and thus unless the relative change in both is known, no conclusion is possible with regard to the direction of change in profit. This is thus considered as a significant use of the price elasticity of demand. The other uses would come by as we proceed with details on product pricing. Incidentally, note that formula (2.12) above is very useful in decision making, which would become obvious as we proceed further in the study of managerial economics.

## REFERENCES

1. Baumol, W.J. (1982): *Economic Theory and Operations Analysis*, 4th edition, New Delhi, Prentice-Hall.
2. Gupta, G.S. (1975): "Demand For Cement in India," *Indian Economic Journal*, XXII (Jan–March).
3. Gupta, G.S. and D. Chawla (1977): "Demand For Tea in India," *Dynamic Management*, II (Sept.)
4. Henderson, J.M. and R.E. Quant (1980): *Microeconomic Theory : A Mathematical Approach*, 3rd edition, New York, McGraw Hill.
5. Truett, L.J. and D.B. Truett (1984): *Managerial Economics*, 2nd edition, Cincinnati, South-western Publication.
6. Rangarajan, C. and G.S. Gupta (1973): "The Demand For Financial Assets—A study in Relation to Bank Deposits, *Prajnan*, II (Jan.–March).
7. Pappas, J.L., E.F. Brigham and M. Hirschey (1982) : *Managerial Economics*, 4th edition, Chicago, Dryden Pr.

## CASELETS

1. The data on tea demand and tea price in India for two years were as follows:

| *Year* | *Demand ('000 tonnes)* | *Price (1993–94 = 100)* |
|---|---|---|
| 2000–01 | 669 | 128.1 |
| 2007–08 | 871 | 130.9 |

The comment of a prominent politician on this data is as follows:

"This clearly shows that the law of demand is not operating in the Indian tea market. The price went up yet consumers bought more. We cannot rely on outdated economic concepts from the previous century for an analysis of current problems".

Do you agree with this observation? How would you interpret the above given data?

2. Ashok and Company (AAC) is currently selling 1,000 suits at a price of Rs. 2,500 per suit. Total cost is Rs. 22, 50,000 of which Rs. 7, 50,000 represents fixed costs. Average variable cost is constant. The Company aims at achieving maximum profit.

The marketing manager of the company believes that if the company could reduce the price to Rs. 2000 per suit its sales would increase by 20%. In contrast, the consultant of the company argues that a 20% reduction in the current price would bring 60% increase in sales.

(a) If you were the proprietor of AAC and were totally convinced by the demand estimates of your marketing manager, would you reduce the price of your product to Rs. 2000? If yes, why?

(b) If, instead of agreeing with the demand estimate of the marketing manager, you agree with that of the consultant, other things remaining the same as in (a) above, would you entertain the price reduction? If yes, why?

(c) Compute the price elasticity of demand and the average incremental revenue under each of the two likely changes in demand situations. Compare your findings of this part with those of parts (a) and (b) above and comment.

(d) Assume that the demand curve is linear. Determine its equation under each of the two likely changes in demand. Given these equations, would you recommend a price change (increase or decrease)? Why or why not? If yes, to what level and why?

3. Sharma Brothers is a multi-product firm engaged in the production of consumption goods. Two of its products, *X* and *Y,* are related in consumption. The past experience of the firm indicates the following relationships:

| *Price of **X** (Rs.)* | *Price of **Y** (Rs.)* | *Quantity of **X** Demanded* | *Consumers' Income (per capita) (Rs)* |
|---|---|---|---|
| 150 | 200 | 5,000 | 10,000 |
| 175 | 225 | 5,500 | 10,500 |
| 175 | 250 | 6,000 | 10,500 |
| 200 | 250 | 5,500 | 10,500 |
| 200 | 250 | 6,500 | 12,500 |

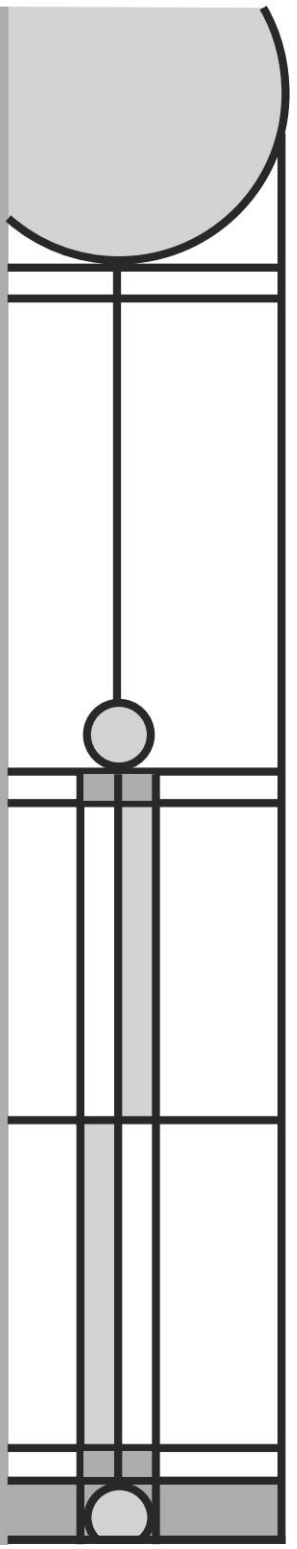
(a) Could the (own) price elasticity of demand for *X* between prices Rs. 150 and Rs. 175 be computed? Why or why not?

(b) Compute all the relevant elasticities of demand from the above data.

(c) What is the relationship between products *X* and *Y*?

(d) If Sharma Brothers' aims at maximum sales revenue from product *X*, would you recommend to raise the price of *X* from Rs. 175 to Rs. 200? Why or why not?

The McGraw-Hill Companies



CHAPTER

3

# THEORY OF CONSUMER BEHAVIOUR

The demand function as explained in the previous chapter is derived through an analysis of the theory of consumer behaviour. The said theory is best presented through three steps:

(a) Consumer preferences
(b) Budget constraints
(c) Consumer choices

## 3.1 CONSUMERS' PREFERENCES

Under consumer preferences, one looks at the way a consumer values various goods and services that are available to him/her for consumption according to the utility (which is synonymous to satisfaction or welfare) he/she derives from them. Generally, **a consumer prefers more to less of any good** until he/she reaches the point of satiation (the point beyond which more would be harmful) from that good. Assuming there are $n$ number of goods and services that are available in the market, the preference function may be described as follows:

$$U = f(X_1, X_2, X_3, \ldots\ldots, X_n)$$
$$f_1, f_2, f_3, f_4, \ldots, f_n > 0 \qquad (3.1)$$

where $U$ = consumers' utility from the consumption basket

$X_i$ = quantities of good $i$ consumed

$f$ = to read as "function of" (it denotes the consumers' taste)

$f_i$ = partial derivative of function $f$ with respect to variable $X_i$

The utility function varies from consumer to consumer and thus function (3.1) is for a specific consumer. Conceptually speaking, the function indicates that the consumers utility increases as he/she consumes more of any of the $n$ goods that are available in the market. However, since the functional form (linear, log-linear, quadratic, etc.) is not specified, it is silent with regard to the rate at which the utility increases with additional consumption of the same good or with consumption of other goods. In this respect,

there are laws in economics which through some light on such questions. One such law is known as the **law of diminishing marginal utility**. According to this, the addition to the total utility (called the marginal utility) goes on declining as one goes on consuming more and more units of the same good. In other words, as more units of a good are consumed, the total utility from that good increases but at a diminishing rate. This law is considered as the foundation of the law of demand. The other relevant law/theory here is one particular relationship between/among goods, i.e., two or more goods may be complimentary, as discussed above. By this theory, utility which a consumer gets from a good (say, milk) depends positively not only on the units of that good consumed but also on the consumption of the units of the other goods (like sugar and/or bread) which are complimentary in consumption to that good. To account for these two features, the utility function would have to be either a multiplicative (like Cobb–Douglas or double log form) or a quadratic or of higher degree polynomial.

The market offers a large number of goods to consumers. However, no individual consumer demands all those goods. Thus, many of the goods would have zero demands for any individual consumer. Still, every consumer would have multiple goods in his/her consumption basket. To ease the explanation, instead of considering goods separately, we may treat them as alternative bundles of potential consumption baskets and then look for consumers choice among them. Each such basket would have several over lapping goods, with different quantities. Assuming that **consumers' preferences are complete**, he/she should be able to rank all such baskets in the order of his/her preferences. If so, against a given basket, the consumer would distinguish all the other baskets in three categories, viz. better (preferred) ones, inferior ones and indifferent ones. The curve joining the indifferent baskets is called as the Indifferent Curve. The derivation of an indifference curve is explained in Fig. 3.1.

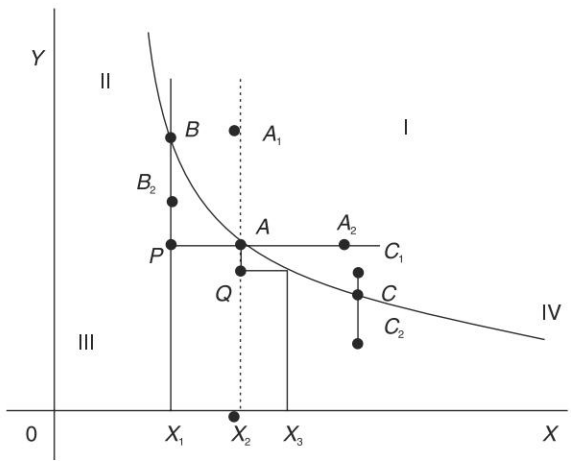



**Fig. 3.1** *Derivation of indifference curve*

For simplicity, let us assume that there are just two goods, call them $X$ (food) and $Y$ (clothing). The various baskets of these two goods are indicated by points $A$, $A_1$, $A_2$, $B$, $B_2$, $P$, $Q$, $C$, $C_1$, $C_2$, and so on in Fig. 3.1. Against the basket $A$, the particular consumer ranks baskets $B$ and $C$ as indifferent, and the rest as the better (e.g. baskets $A_1$ and $A_2$) and inferior (e.g. baskets $P$ and $Q$). Since consumer always prefer more to less of any good, the just indicated preferences are obvious. In particular, all points in the graph which fall in Quadrant I (North–East of point $A$) are preferred over point $A$, for they have more of either food or clothing or of both. Similarly, all points falling in the South–West region of

point *A* (Quadrant III) are inferior to the bundle at point *A*. In Quadrants II and IV, while some bundles are better than, some others are inferior than, and remaining are same as, the bundle at point *A*, By a repeated interview of the concerned consumer, one can identify the various bundles among which the consumer is indifferent in relation to the bundle at point *A*. Joining of those points of indifference yield an indifference curve like the one passing through the points *B*, *A* and *C*. Thus, an **indifference curve (also called as the iso-utility curve) is defined as the locus of all such combinations of two goods (or two bundles of goods) which yield the same level of utility (satisfaction) to the consumer concerned**. Accordingly, the consumer is indifferent among various points along an indifference curve. Since the utility that a consumer derives from consuming a good vary from consumer to consumer, indifference curves are subjective. Through iterating the above procedure, the consumer would be able to generate what is called as a map of indifference curves for him/her. The said map would delineate the consumer preferences fully. One such map is depicted in Fig. 3.2.

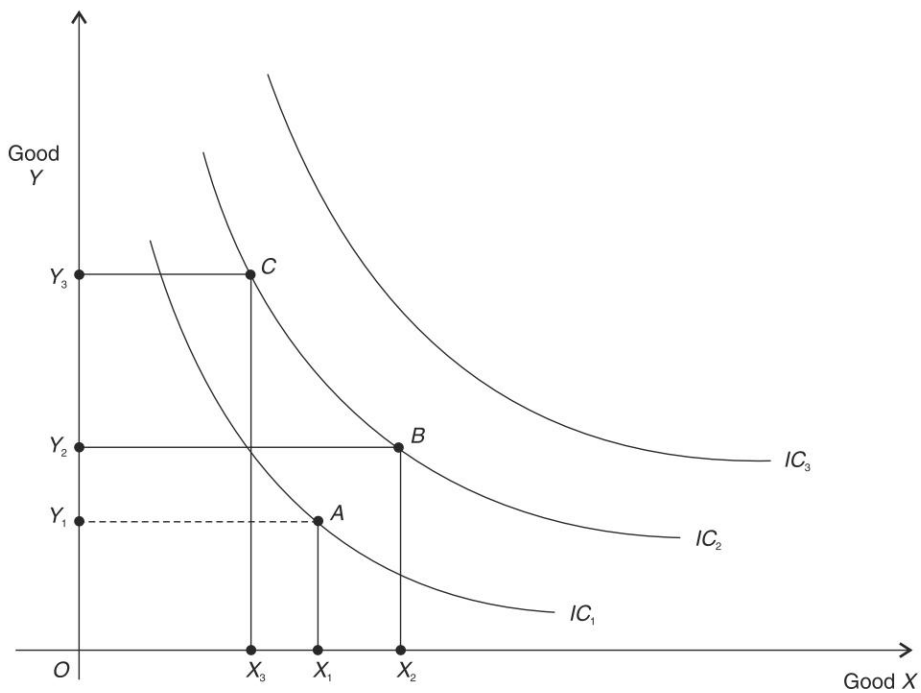



**Fig. 3.2** *Indifference curves map*

The indifference curves have four unique features, viz.

(a) they slope downward from left to right
(b) the higher the indifference curve, the more is the satisfaction
(c) no two indifference curves ever intersect
(d) they are convex to (bow towards) the origin

The first two features follow from the assumption that a consumer prefers more to less of any good. Thus, if the curve were not sloping downward, it would be vertical, horizontal or sloping upward from left to right. Thus, if it were vertical, it would be like the one passing through points *A* and $A_1$ in Fig. 3.1.

However, the latter could not be an indifference curve, for at $A_1$, the consumer has more of Clothing and same units of food as at point $A$, which violates the assumption that the consumer prefers more to the less of all goods. Alternatively, if indifference curves were horizontal, one such would be like the one passing through points $A$ and $A_2$ in Fig. 3.1, which would also violate the assumption just mentioned; for at point $A_2$, the consumer has more of food and same units of clothing as at point $A$. Further, an upward sloping curve would have two points, at one of which he would have more of both the goods than at the other, which is subject to a similar violation. Thus, indifference curves always slope downward from left to right. To understand the property that the higher is the indifference curve, the more is the satisfaction; consider the consumers' satisfactions at points $A$, $B$ and $C$ in Fig. 3.2. Point $B$ gives him/her more utility than point $A$, for at point $B$ he has more of both the goods than at point $A$. The Consumer is indifferent between points $B$ and $C$ as they fall on the same indifference curve. Thus, the consumer is better off at the higher indifference curve $IC_2$ than at the lower indifference curve $IC_1$.

Consider the third property. If indifference curves were to intersect, the situation would be like the one depicted below in Fig. 3.3.

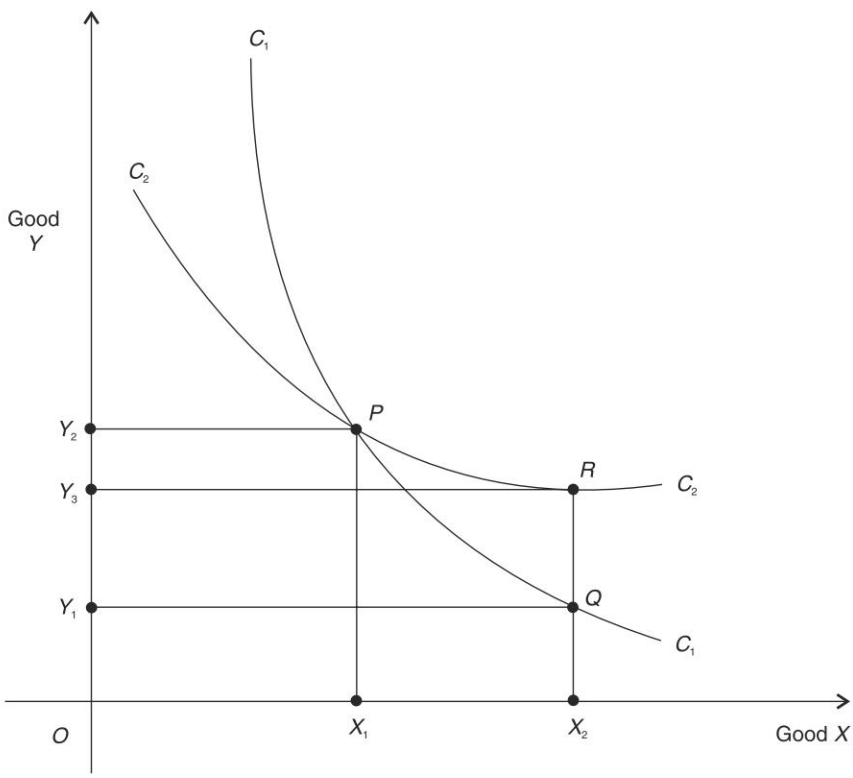



**Fig. 3.3** *Two intersecting curves*

The two curves in Fig. 3.3 cannot be indifference curves. This is because, by the definition of indifference curve, the consumer should be indifferent between points $P$ and $Q$ (as both lies on the same indifference curve), and $P$ and $R$ (as both lies on the same indifference curve), and if so, he/she ought to be indifferent between points $Q$ and $R$ by the **principle of transitivity**. However, at point $R$, has more

of both the goods than point *Q*, thus, inconsistent with the assumption of "more is preferred to less". Accordingly, no two indifference curves would ever intersect.

The last property of indifference curves (convexity to origin) follows from the assumption/belief in the **law of diminishing marginal rate of substitution (MRS)**. The MRS is defined as the maximum amount of a good (clothing) that a consumer would give up to get an additional unit of some other good (food). Accordingly, the law of diminishing MRS suggests that as the consumer substitutes one good (say food) for the other good (clothing), the sacrifice of clothing goes down with every unit increase in food. This is very reasonable because in the said substitution process, as one keeps going further and further, the consumer is left with the less and less of clothing, and more and more of food; and the less is left, the more valuable the marginal unit gets, and more one has the less useful the marginal units remain. Another way of looking at this is that consumers usually prefer a balanced diet (basket) over a single product diet (so called a corner solution), and even a well mixed diet over a concentrated diet (tilted to one good). It is easy to see that if the indifference curves were straight lines (i.e. linear), the MRS would be a constant; if they were concave to origin, MRS would be increasing; and if they were convex to origin, MRS would be declining. For the convex curves case, the MRS principle is illustrated in Fig. 3.4.

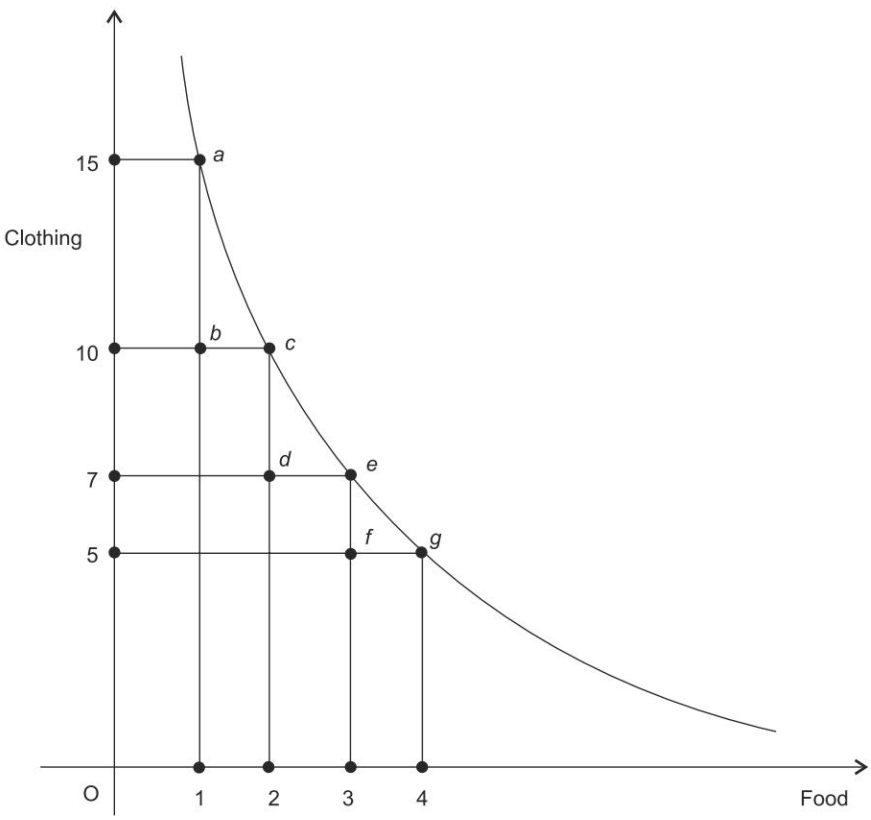



**Fig. 3.4** *Convexity and the diminishing marginal rate of substitution*

Suppose, initially the consumer is at point *a* on the indifference curve with 15 units of clothing and 1 unit of food (See Fig. 3.4). When he substitutes food for clothing, he moves to point *c*, where he has

10 units of clothing and 2 units of food. Thus, he gives up 5 units of clothing for 1 additional unit of food. When he takes one extra unit of food, he moves to point *e*, having 7 units of clothing and 3 units of food; thus sacrificing 3 units of clothing for one additional unit of food. As he moves further down along the indifference curve, the sacrifice of clothing for each extra unit of clothing keeps going down. To summarize, the various points along the indifference curve have the following consumption baskets:

**Table 3.1** *Consumption Baskets and Indifference Curve*

| *Point on I.C.* | *Units of Clothing* | *Units of Food* | *MRS of Food for Clothing* |
|---|---|---|---|
| *a* | 15 | 1 | – |
| *c* | 10 | 2 | 5 |
| *e* | 7 | 3 | 3 |
| *g* | 5 | 4 | 2 |

If indifference curves were concave to the origin, the MRS would be increasing and it would be constant if they were a straight line. Thus, due to the law of diminishing marginal rate of substitution, the indifference curve is convex from below.

This is usual shape of indifference curves, which is valid under most situations. However, there are two extreme (rare) cases where their shapes would be different. One, if the **two goods were perfect substitutes in consumption, the indifference curves would be straight lines** (linear) instead of convex to the origin. Two, if the **two goods were perfect complements in consumption, indifference curves would be shaped like the right angles**. As is obvious, two goods would be perfect substitutes only when the consumer is absolutely indifferent between those goods, and some other two goods would be perfect complements when any one of them is of no use, unless the other is there (like left shoe and right shoe). The shapes of indifference curves under each of these two extreme cases are illustrated in Fig. 3.5 below.

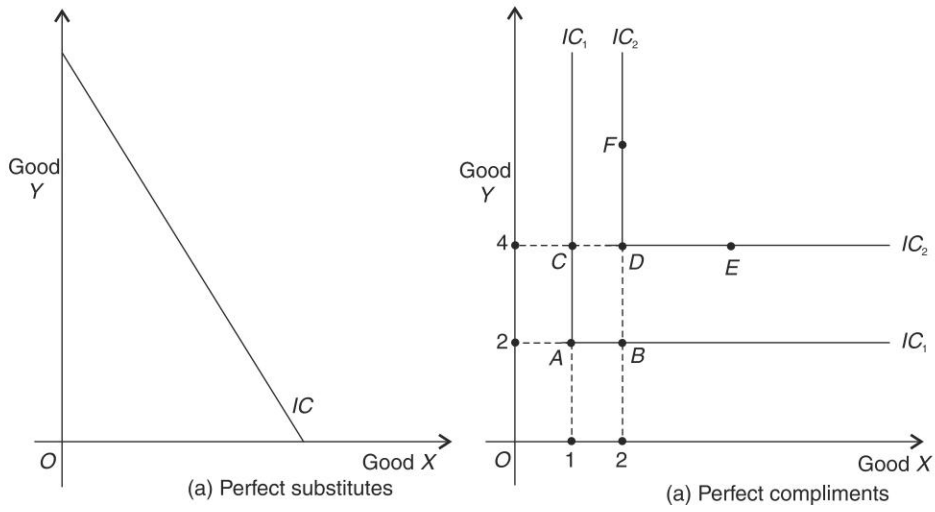



**Fig. 3.5** *Indifference curves under perfect substitution and perfect complements of goods*

Under perfect substitutes case, the MRS is a constant (need not equal one), while that under the perfect complements case is zero. This is obvious from the meanings of these two terms. Both these cases are rare, for such relationships are hard to find.

At this stage, one must understand the distinction between the Ordinal and Cardinal utility functions. Under the former, a consumer merely orders (or ranks) various consumption baskets while under the latter approach, he/she is able to exactly quantify the utility levels (i.e., gives numbers to each consumption basket) that he/she gets from the different baskets. As seen above, the indifference curve approach falls under the ordinal utility theory. The two approaches differ in some other respects as well. Suffice here to mention that since the utility from various baskets is hard to quantify while it is easy to rank the two or more baskets, the cardinal approach has basically been discarded in favour of the ordinal approach. There is yet one more approach for consumers' preferences, viz. Revealed Preference Theory. Under this, consumers' preferences are inferred from his/her choices, given his/her budget constraint, and thus it follows a reverse process. This theory, while useful in several areas, provides a relatively lesser convincing rationale than the theory based on indifference curves. Accordingly, the same is not covered in the present volume. This completes the discussion on consumers' preferences.

## 3.2 CONSUMERS' BUDGET CONSTRAINTS

No one (even if, one among the world's wealthiest persons, William Gates) has unlimited budget for his (family) consumption expenditure; the said budget is treated equal to the consumers income here for simplicity. Accordingly, a consumer faces a fixed income constraint, which he allocates on his consumption basket. Besides, assuming a large number of consumers for any good (which is generally true), a particular consumer faces given prices in the market for all goods and services that he buys.. Combining these two constraints, the budget constraint could be expressed as follows:

$$I = P_1 X_1 + P_2 X_2 + P_3 X_3 + \ldots\ldots\ldots\ldots + P_n X_n \qquad (3.2)$$

Where $I$ = Consumers income

$P_i$ = Price of good $X_i$

$X_i$ = quantity of good $i$ that consumer buys

$n$ = Number of goods

Equation (3.2) indicates that the consumers income equals his/her expenditure on all goods and services. This, in turn, suggests that the consumer faces a trade-off, meaning if he spends more on one good he is left to spend less on all other goods. In order to explain the consumers' choice making through geometry, let us consider, as above, that there are just two goods and let them be, food ($X$) and clothing ($Y$). Under this simplification, the budget constraint reduces to Eq. (3.3):

$$I = P_x X + P_y Y \qquad (3.3)$$

Suppose the consumers' weekly income equals Rs. 10,000 and the market prices of food ($P_x$) and Clothing ($P_y$) equal Rs. 100 and Rs. 200, respectively. Under these constraints, the consumer could choose any combination of Food and Clothing. The alternatives facing consumers would be infinite and some of those would be the following:

**Table 3.2** *Market Basket and Budget Constraint*

| *Market Basket* | *Units of Food (X)* | *Units of Clothing (Y)* |
|---|---|---|
| *A* | 100 | 0 |
| *B* | 80 | 10 |
| *C* | 60 | 20 |
| *D* | 40 | 30 |
| *E* | 20 | 40 |
| *F* | 0 | 50 |

Each of the consumption baskets in Table 3.2 is associated with the fixed income of Rs. 10,000 and fixed prices of food and clothing of Rs. 100 and Rs. 200, respectively. The same could alternatively be depicted in a graph as in Fig. 3.6:

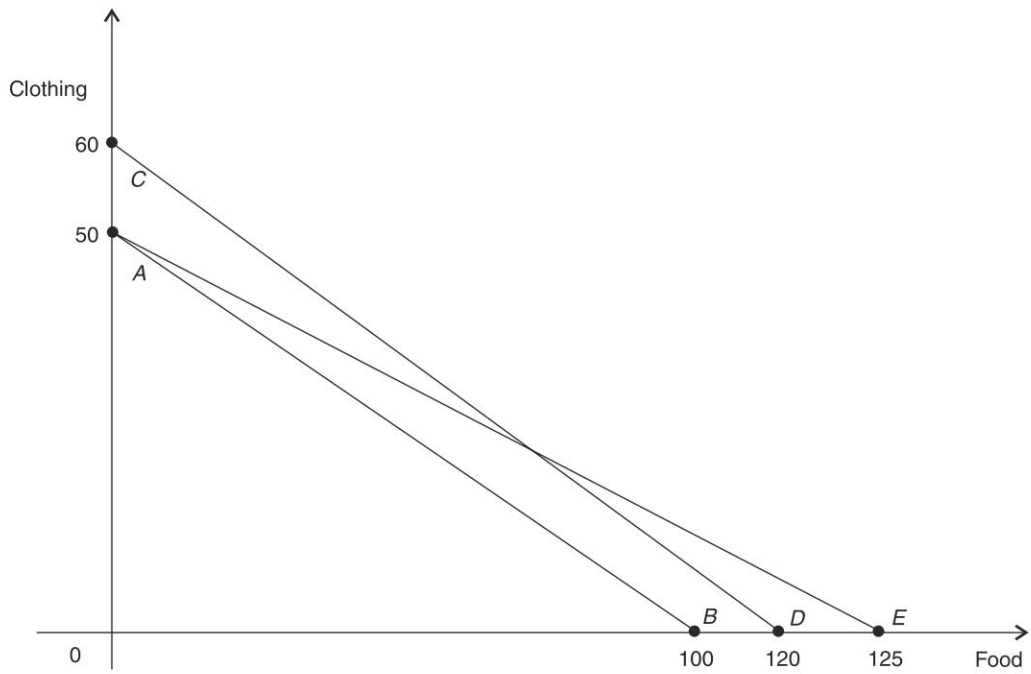



**Fig. 3.6** *Consumers' budget constraints*

The line marked *AB* represents the consumers' budget constraint (or the **purchasing power**) with income equals Rs. 10,000, and food and clothing prices equal Rs. 100 and Rs 200, respectively. Incidentally note that the consumers' purchasing depends not only on the income but also on the prices of goods and services that he/she buys. Since the expenditure budget (income) is fixed, the budget line slopes downward; and because goods prices are fixed, the said line is a straight line. The slope of this line is negative and equals the ratio of two prices:

$$\frac{dY}{dX} = -\frac{Px}{Py} \tag{3.4}$$

The budget line shifts to right parallelly as the consumers' income goes up, other things (both prices) remaining the same, and vice versa. This so happens because with the increased income, prices of food and clothing remaining unchanged, the consumer is able to buy more of both the goods or more of one good and the same of the other good, and so on. For example, if his income goes up to Rs. 12,000, he could buy 120 units of food with nothing of clothing, or nothing of food and 60 units of clothing, and so on. The new budget line would be *CD*, which is to the right of the initial line *AB*.

The slope of the budget line depends on the relative price and thus it undergoes a change as any one price change, the other remaining the same or both change but in different proportions. For example, if food price falls from Rs. 100 to Rs. 80 per unit, income and clothing price remaining the same as before, then the budget line rotates anti clock wise from *AB* to *AE*. Quite the opposite would happen if food price goes up.

What would happen if both the consumers' income as well as prices changes? It will depend on the rates of change in each of the three variables, viz. income and two prices. For example, if all the three variables move up (or down) proportionately, the budget line would stay the same. Thus, if the consumers' income as well as both food and clothing prices go up by 50%, from their levels at Rs. 10,000, Rs. 100 and Rs. 200; to Rs. 15,000, Rs. 150 and Rs. 300, respectively, the budget line would remain invariant at AB. But if they move up and down either in different directions or at different rates, the budget line would change.

## 3.3 CONSUMERS' CHOICE

Given the consumers objective of utility maximization, his/her utility (called preference) function (viz. Eq. 3.1), and the budget constraint (vide Eq. 3.2), one merely needs the optimization technique to determine the consumers' choice. Here the choice is in terms of choosing the best consumption basket from the unlimited possible baskets as represented by the budget constraint (equation 3.2). Algebraically, the problem is solved through the **Lagrangian multiplier technique**. Under this method, the consumer must maximize the following Lagrangian (L) expression with respect to $X_1, X_2, X_3, \ldots X_n$:

$$L = f(X_1, X_2, X_3, \ldots\ldots, X_n) + \alpha\,(I - P_1X_1 - P_2X_2 - P_3X_3 - \ldots - P_nX_n) \tag{3.5}$$

(where $\alpha$ = Lagrangian parameter)

The necessary conditions (partial derivatives with respect to each of $X_1, X_2, X_3 \ldots$, and $X_n$ to be zero) for optimization would number $n$ and these would be the following:

$$f_1 = \alpha P_1, f_2 = \alpha P_2, f_3 = \alpha P_3, \ldots\ldots\ldots\ldots, \text{and } f_n = \alpha P_n$$

Besides, there are sufficient conditions for optimization which will be met if indifference curves are convex to the origin. The above expression contains n number of equations and in these $f_1$ stands for the marginal utility to the consumer from good 1, $f_2$ that from good 2, and so on. The solution of the above would yield the following n number of equations:

$$\frac{MU_1}{P_1} = \frac{MU_2}{P_2} = \frac{MU_3}{P_3} = \cdots\ldots\ldots = \frac{MU_n}{n} = \alpha \tag{3.6}$$

System of Eq. (3.6) is known as the **Rationality Rule** and it states that for consumers' optimization, the consumer must equate the marginal utility per rupee from various goods and thus it is known as the **Law of Equi-Marginal Utility (per rupee)**, as briefly discussed in Chapter 1. These $n$ number of equations and the (one) budget constraint (viz. Eq. 3.2) could be solved for $n + 1$ number of variables,

viz. $X_1, X_2, X_3, \ldots\ldots, X_n$, and $\alpha$, which are nothing but the quantities of *n* goods that the consumer must choose to maximize his/her utility, and the value of the Lagrangian parameter, $\alpha$. Here, it may be noted that $\alpha$ stands for the **marginal utility of income**, as this equals the partial derivative of L (vide Lagrangian expression (3.5)) with respect to income (*I*). Obviously, the solutions would suggest that the demand for any good would depend on the consumers' income and all prices. Further note that it is the real income and relative prices (price of one good in relation to prices of other goods) that matter rather than the nominal income and absolute prices, respectively. The utility function would vary depending on as to whether the particular good is a substitute or complement for the chosen good, and whether the chosen good is a superior or inferior good, and so on. Accordingly, the signs and magnitudes of income and related prices would be determined. The taste and preferences variable in the demand functions would come again from the specification of the utility function. The above discussion is for a static model and thus it does not demonstrate the relevance of the remaining two relevant variables, viz. expected price and expected income, in individual consumers' demand functions. The expectation variables become relevant when we hypothesize dynamic (time dependents) utility and budget constraints, which are beyond the scope of this text.

In order to explain the consumers' choice through geometry (indifference curve approach), we revert to the assumption of just two goods, calling them as food (*X*) and clothing (*Y*). Under this assumption, the utility function is reduced to a function like the one in Eq. (3.7).

$$U = f(X, Y) \tag{3.7}$$

And the budget constraint to a function like Eq. (3.3).

For a given level of utility, Eq. (3.7) is the equation of an indifference curve. As explained earlier in this Chapter, a family of indifference curves could be drawn for various levels of utility for a particular consumer. Further, given the consumers' income and market prices of food and clothing, the budget line could also be drawn and superimposed on the same graph, having food and clothing on the *X*-axis and *Y*-axis, respectively. Such a graph is given in Fig. 3. 7.

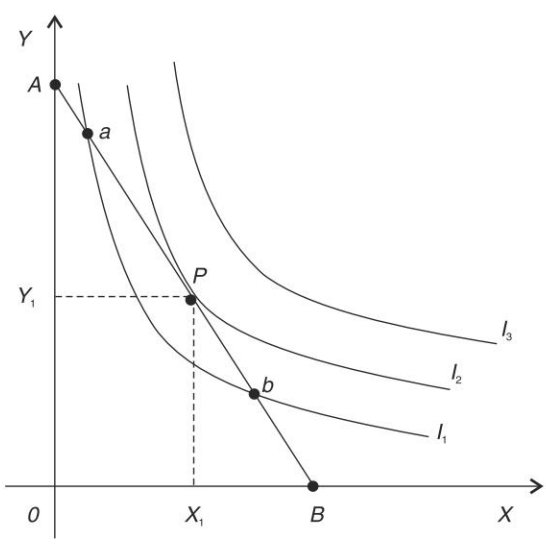



**Fig. 3.7** *Consumers' equilibrium*

In Fig. (3.7) $I_1$, $I_2$ and $I_3$ represent the consumers' preferences and the line $AB$ the consumers' budget constraint. As a utility maximizing consumer, he would like to attain as high an indifference curve as possible but his budget constraint forces him to stay on the line $AB$ only. Accordingly, his best choice would be at point $P$ where he would choose $OX_1$ units of good $X$ (food) and $OY_1$ units of good $Y$ (clothing). His budget constraint would permit him to choose a point like $a$ or $b$ but each of these points fall on indifference curve $I_1$, where the consumer has lower utility level than at point $P$ which lies on indifference curve $I_2$.. The consumer could even opt for any of the two corner points (viz. A and B), but as would be obvious, each of those would fall on a still lower indifference curve than $I_1$. Incidentally note that, the corner points are known as the corner solutions to the optimization problem, where a corner means choosing a single product (either food or clothing) instead of a mix of both the products. Recall that a balanced diet is usually preferred over a lop-sided (single product) diet!

At the consumers' optimum point $P$, note that the budget constraint is tangent to an indifference curve. This is an essential feature of the consumers' equilibrium position. For, if this were not so, the consumer could increase his/her satisfaction by moving to the tangency point. At the point of tangency, by definition, the slope of the budget line equals that of the indifference curve. The former, as explained above, is given by eq. (3.4). The latter can be derived using the definition of an indifference curve. At every point on an indifference curve, the utility that the consumer gets is a constant. Thus, along an indifference curve, the following holds good:

$$dY\,(MU_y) + dX\,(MU_x) = 0$$

where $dY$ = change in units of good $Y$, $dX$ = change in units of good $X$, $MU_x$ = marginal utility from good $X$ and $MU_y$ = marginal utility from good $Y$.

The solution of the above equation gives

$$\frac{dY}{dX} = -\frac{MUx}{MUy} \tag{3.8}$$

Equation (3.8) gives the slope of an indifference curve. Equating the slopes of the indifference curve with that of the budget line yields

$$\frac{MUx}{MUy} = \frac{Px}{Py} \tag{3.9}$$

Equation (3.9) denotes the equilibrium condition for the consumer in case of two goods economy. The left hand side term of equation (3.9) is known as the **marginal rate of substitution (MRS)** of good $Y$ for good $X$. The above equation could also be transferred into the following

$$\frac{MUx}{Px} = \frac{MUy}{Py}$$

which is a counterpart of the law of Equi-Marginal Utility (per unit of rupee) as expressed above in Eq. (3.6) for the $n$ goods economy. What would happen to the consumers' choice if either his/her income, or market price of one or both the goods, or all the three variables undergo a change? This is examined in what follows.

## Change in Consumers' Income

Recall that if the consumers income increases, ceteris paribus, the budget line shifts parallel to the right and if the income goes down the said line shifts parallel to the left. Since the utility function is assumed

to remain the same, there is no change in indifference curves when income changes. In consequence, as income increases, ceteris paribus, the consumers' equilibrium moves up to an higher indifference curve as shown below in Fig. 3.8.

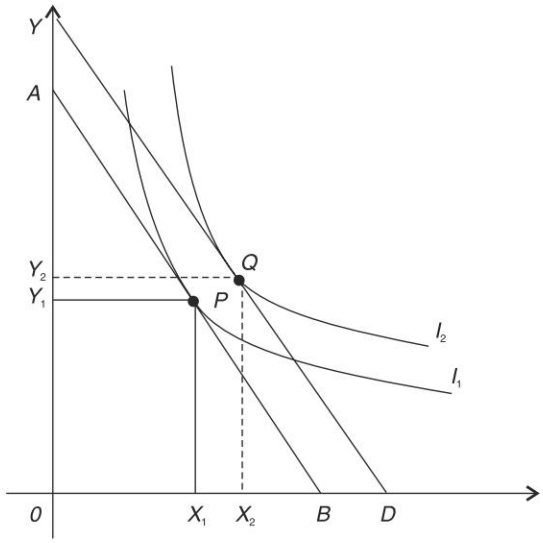



**Fig. 3.8** *Consumers' choice and increase in income*

With an increase in income, the consumer moves from point *P* to *Q* in Fig. 3.8. At point *Q*, the consumer has more of both the goods than at the old (less) income at point *P*. Thus, an increase in income leads to an increase in the demand for both goods. In this graph, the indifference curves assume that both the goods are superior. If one of the goods were inferior, the demand would have increased for the superior good and decreased for the inferior good. It must be noted that all the goods in the consumption basket can be superior but they all can not be inferior goods. To prove this assertion, let us continue with the two goods case. If both the goods were superior, the consumption of both would increase and that can come from splitting the increased income on both the goods. However, both the goods can not be inferior, for if the consumption of both the goods were to decline in the face of an increase in income, the expenditure on the two goods will fall rather than increase, thus leaving a part of the income unspent and there by violating the budget constraint.

The line joining points 0 (origin), *P* and *Q*, and so on (the equilibrium points with regard to different levels of income) is known as the **income consumption curve (ICC)**. The said line gives the relationship between the consumers' income and the optimum quantities of each of the two goods that the consumer buys. Mapping of this relationship between income and consumption of any good gives the so called, **Engel curve**. The curve would be upward sloping for superior goods and downward sloping for inferior goods. It is possible that a good could be superior up to a certain income level and inferior after that income level. If so, the ICC would slope upward up to that income level and downward there after.

## Change in Price

Recall that if one of the two prices change, ceteris paribus (i.e., no change in the other price, income or taste), the slope of the budget line changes. Thus, if the price of, say, Food (good *X*), falls, ceteris

paribus, the budget line rotates outward (anti-clockwise) from the point on the *X*-axis. This is illustrated below in Fig. 3.9.

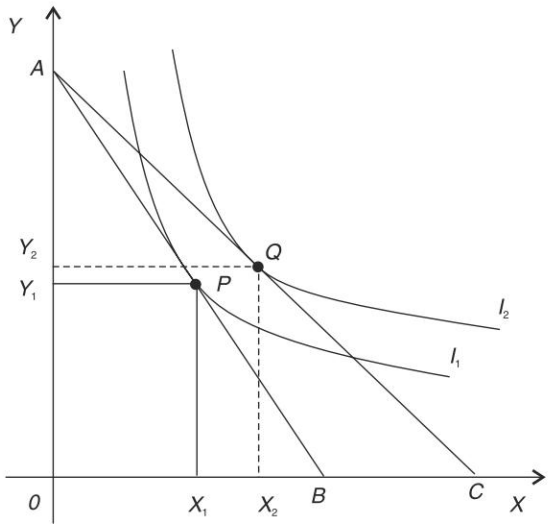



**Fig. 3.9** *Consumers' choice and decrease in price of good X*

Figure 3.9 illustrates that when the price of good *X* falls, price of good *Y* and the consumers' income remaining the same, the budget line changes from *AB* and *AC*. In consequence, the consumers' equilibrium moves from point *P* to point *Q*. At point *Q*, the consumer buys more of both the goods than at his initial point *P*. This is made possible through an increase in real (not nominal) income (due to fall in one price, ceteris paribus), and there by in his purchasing power, where in the consumer splits the increased purchasing power on each of both the goods. In Fig. 3.9, both the goods are assumed to be superior. If good *Y* were inferior and good *X* superior, demand for *X* would have increased and that of *Y* decreased. If good *X* were inferior (not Giffen) and *Y* superior, demand for both would have increased; for the income effect in case of the non-Giffen inferior good is weaker than the substitution effect (vide Chapter 2). However, if good *X* were Giffen good, its demand would have declined as its price fell.

The line joining the points 0 of origin, and equilibrium points *P* and *Q* in Fig. 3.9 is known as the **Price Consumption Curve (PCC)**. This curve slopes upward for all non-Giffen goods. The PCC line shows the demand for good *X* at its various prices, and thus it reflects the demand curve for good *X* as well. The line suggests that as the price of good *X* falls, its demand increases (quite the opposite is seen when one moves from point *Q* to point *P*), and vice versa, and thus it demonstrates the derivation of the demand curve.

The above analysis indicates that the demand for a good by a consumer varies with the consumers' income and the prices of all goods and services in his/her consumption basket. The effect of consumers taste and preferences is contained in the utility function (indifference curves). As hitherto mentioned, the relevance of expectations (expected income and prices) could not be demonstrated here due to the static nature of our analysis; the discussion of dynamic model is beyond the scope of the present volume. This rationalizes the demand function for an individual consumer. The market demand function contains two additional determinants, viz. number of consumers/size of population and the distribution of consumers in one or more specific groups. The relevance of these two variables is not a part of the theory of consumer behaviour, and, in any case, the same has been explained already in Chapter 2.

## REFERENCES

1. Baumol, W.J. (1982): *Economic Theory and Operations Analysis*, 4th edition, New Delhi, Prentice-Hall.
2. Henderson, J.M. and R.E. Quant (1980): *Microeconomic Theory: A Mathematical Approach*, 3rd edition, New York, McGraw-Hill.
3. Pindyck, Robert S., Daniel L. Rubinfeld (2009): *Microeconomics*, 7th edition, Pearson.
4. Samuelson, P.A. (1948): *Foundations of Economic Analysis*, Cambridge, Harvard University Press.
5. Stigler, G.J. (1966): *The Theory of Price*, 3rd edition, New York, Macmillan.

## CASELETS

1. Mohan has a budget of Rs. 5,000/week to spend on food ($F$) and clothing ($C$). The market prices of food and clothing (per unit) are Rs.125 and Rs 250, respectively. His utility function is $U = 25\ F\ C$. Based on this information, advise Mohan on his following questions:
   (a) What demand functions for each of the two goods, Mohan faces in the market?
   (b) Are food and clothing related goods for him? How?
   (c) Mohan is rational in choosing his consumption basket. If so, what consumption basket he must choose? Determine his marginal rate of substitution between the two goods at his equilibrium point.
   (d) Suppose due to good monsoon, food price falls to Rs. 100, ceteris paribus, how his optimum consumption basket would change?
2. Three consumers for ice cream ($I$) and apple pie ($A$) have the following utility functions between these two goods:

   (i) $U_1 = 10\ I\ A$ (ii) $U_2 = 2\ I + 5\ A$ (iii) $U_3 = 20\ I + 70\ A\ - I^2 - A^2$

   The market prices of ice cream and apple pie are Rs. 15 and Rs. 30, respectively. Advise the consumers on the following matters:
   (a) How different they are with regard to the relationship between the two goods?
   (b) Do they have different kinds of demand functions for the two goods?
   (c) Do their indifference curves differ? Why?
   (d) For utility level of 100 ($U = 100$), draw the indifference curve for each consumer.
   (e) Examine the validity of the law of diminishing marginal rate of substitution for each consumer.
3. Rama is pursuing her MBA education at a business school. She receives utility from the grades (marks) she receives in her exams and the leisure (entertainment) she gets. Her utility function from the two goods is described by the following function:

   $$U = 25\ G\ L$$

   where, $U$ = utility, $G$ = grade (GPA) she earns and $L$ = leisure she enjoys. At her business

school, grade is assigned in a letter grade which is convertible in to a numerical value ranging from zero to 10. As one would expect, the grade depends on time spent on studies ($T_S$) and the relationship between the two is described by the following:

$$G = 0.8\ T_S$$

Her constraint is the time at her disposal, which is 24 hours in a day, thus

$$24 = T_S + L$$

or, $$24 = 1.25\ G + L$$

Accordingly, Rama faces a trade-off between grade and leisure. Advice her on the following decision issues:

(a) What is her indifference curve if she wants to earn a utility level of 1000? Also, for utility levels of 1500 and 2000? Draw her indifference curves map.

(b) Determine her budget line and super impose the same on the graph of part (a)

(c) Suggest the optimum allocation of time between studies and leisure for Rama.

(d) Determine the maximum attainable utility level for Rama.

4. Radha has to decide on the allocation of her life-time income between consumption and saving, call them as present consumption (C) and future consumption (S). She stands good credit worthiness and thus is able to borrow or lend at an interest rate of 5 per cent/annum. Her present income is Rs. 100 thousands and she knows for sure that her future income would be Rs. 125 thousands. Radha's utility function from present consumption and saving is expressed by the following function:

$$U = 10\ C^{0.5}\ S^{0.5}$$

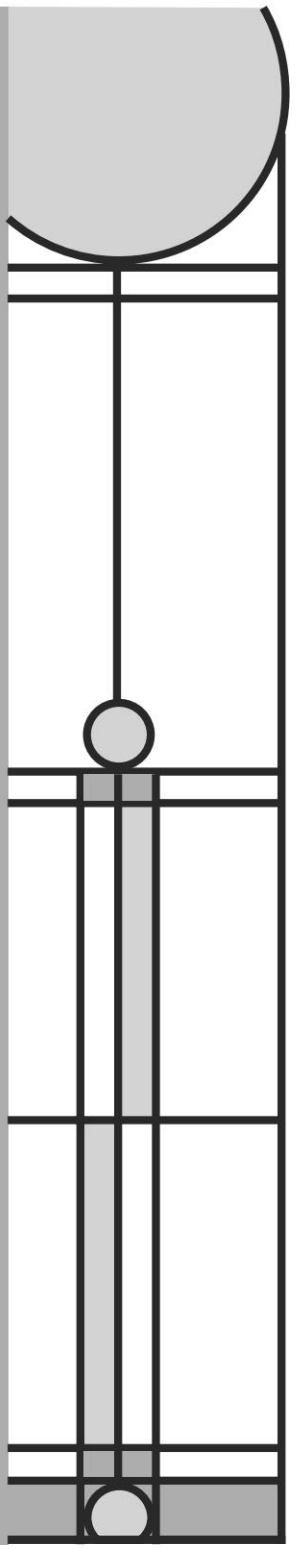
Given the above information, advice Radha on the following matters:

(a) Is her utility function consistent with economists' famous law of diminishing marginal rate of substitution? (*Hint*: Draw several indifference curves for various utility levels and analyse the marginal rates of substitution at different combinations of *C* and *S*).

(b) What is her budget constraint in the form of an equation?

(c) What combination of *C* and *S* Radha must choose if her objective is to maximize her satisfaction from the two goods, *C* and *S*?

The McGraw-Hill Companies



CHAPTER

4

# DEMAND ESTIMATION, ANALYSIS AND FORECASTING

Forecasting of demand for its products is an essential activity for all businesses, and particularly for those engaged in the production of goods and services that have some gestation period. The viability of a project involved in production of some goods or services hinges on the quantum of its expected sales, among other factors, and thus demand forecasting exercise even precedes the acceptance stage of any productive activity. While there are several approaches to forecasting, there is hardly any one that does not require estimation of the relevant demand function. Further, all estimations proceed through equations. Thus, an understanding of an equation, its estimation procedures, and the use of the estimated equation for deriving forecasts in imperative for all managers.

A function can be expressed in the form of a table (schedule), or graph or equation. Chapter 2 provided examples of demand function in terms of tables and graphs. The present chapter will deal with the demand function in terms of equations, which provides the much needed tool in the hands of decision makers.

To demonstrate the superiority of equations over tables and graphs, consider the example of a falling linear demand curve (Chapter 2, Fig. 2.7 and the Table in section 2.6). The table provides various quantities the consumer demands at various specific prices: 10, 9, …,4. If one were to determine demand at, say, $P = 5.5$, how would he get from the table? By interpolation! Similarly, he would have difficulties to quantify demand at prices higher than 10 and below 4. These problems arise because the table represents a discrete function and in the limited (finite) ranges of the values of the variables. Coming to the graph next, which, of course, is a continuous function, but it is so only within the ranges of the variables plotted on the graph. Fig. 2.7 covers all possible values of $P$ and $Q$, and so there is no problem in this case. But usually no graph is so exhaustive. If so, one would have problems of finding the values of demands at prices beyond the ones covered in the graph. For the sake of reasoning alone (ignoring economics), the graph is for $P$ between 0 and 12. What about demand if $P$ could take a negative value or a value greater than 12? Thus, a graph has its limitations. Equations are devoid of such difficulties. For example, the relationship of the table and graph just discussed could be written in the form of an equation as follows:

or,
$$\left.\begin{aligned} Q &= 18 - 1.5P \\ P &= 12 - \frac{2}{3}Q \end{aligned}\right\} \quad (4.1)$$

Function (4.1) is continuous for all values of $P$ and $Q$ and thus it would give unique values for $Q$ ($P$) for every specific value of $P$ ($Q$). Besides this factor, equations are better than tables and graphs, for they are easy to handle and analyse, particularly when there are several variables in the function. To elaborate this point, consider a demand function with two arguments ($P$ = Price, $Y$ = Income).

$$\left.\begin{aligned} Q &= f(Y, P) \\ f_1 &> 0 > f_2 \end{aligned}\right\} \quad (4.2)$$

The above is an unspecified function, in the sense its functional form is not indicated in it. Its three linear versions are shown below.

**Equation Version**

$$Q = 3 + 0.3Y - 1.5P \quad (4.3)$$

**Table Version**

| $P$ | $Q$ ($Y$ = **50**) | $Q$ ($Y$ = **60**) |
|---|---|---|
| 10 | 3 | 6 |
| 9 | 4.5 | 7.5 |
| 8 | 6 | 9 |
| 7 | 7.5 | 10.5 |
| 6 | 9 | 12 |

**Graph Version**

The equation is relatively neat. Plug in any values for $Y$ and $P$, you have the corresponding value for $Q$. The superiority of equation over table and graph increases as the list of explanatory variables increases.

The question now is, how to get the equation form of a function? This is the subject matter of the following section.

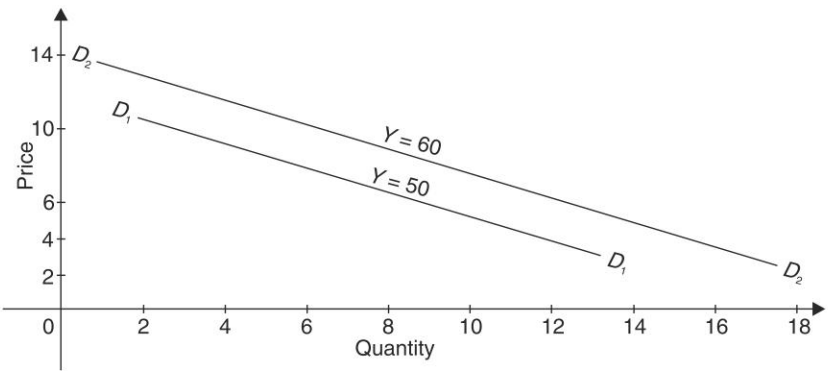



**Fig. 4.1** *Linear demand function in two causal variables*

## 4.1 DEMAND ESTIMATION

There are three methods for estimating a demand function

(a) Consumer interviews

(b) Market experiments

(c) Regression analysis

A discussion of each of them follows.

## Consumer Interview Method

Under this method, consumers are interviewed with regard to their consumption habits. Interviews could be conducted on the census basis or sample basis. Under the former, all past and prospective consumers are interviewed, while under the latter, only a subset of them, called the sample, are interviewed. The interviews could be planned orally or through pre-designed questionnaire, depending upon the complexities of the problem. These interviews, called surveys, aim at obtaining the relevant information on a variety of variables, useful for estimating the demand function for the product under study. For example, the survey could include information on the quantities of the concerned good bought at different periods at various prices of the product and its related goods, income of the consumer, his expectations, his socio-economic profile, and probe into likely changes, if any, in the taste and preferences of the consumer in the future vis-a-vis present and past. Given this information on a systematic basis from a sufficiently large sample of consumers, the researcher would be able to quantify the demand function for the good in question. This method is dealt more in marketing texts than in economics ones, and so is the tradition followed here. Suffice it to say here that interested readers might look up this part in some good marketing text or/and in a study co-authored by the present author (Gupta, 1984).

## Market Experiments Method

Market experiments provide an alternative method to estimate the demand function. It has two versions: actual and simulated. Under the actual experiment, shops are opened in different localities (places) and then consumers' reactions are observed and recorded. Different localities would include consumers with varying levels of income, caste and religion, sex, age group, tastes and preferences, etc. Further, during the experiments, various prices could be tried to elicit consumers' reactions to price changes. If such an exercise is carried out with sufficient care with regard to the sample of locations and probable prices, the researcher should have no difficulty in coming out with a demand function, indicating quantities that consumers would demand at various levels of incomes, prices, and other relevant variables in the function. Educational institutions organize marketing fairs to elicit such data from the visitors.

The market simulation method, also called consumer clinic or laboratory experiment technique, involves providing token money to a set of consumers and asking them to shop around in a simulated market. The prices of various goods, their quality, packaging, etc. vary during the experiments to observe consumers' reactions to such changes. This generates information which could be sufficient to estimate the demand function. Finance students conduct such simulations to practice stock market transactions and results.

Of these two versions, the actual experiment method is more reliable than the simulated experiment one. This is because the consumers have no stake in the latter and may not take the experiment seriously. As a result the data generated will not be reliable, which would render the whole exercise futile. However, actual experiments, though desirable might be too costly. Thus, the firm which is interested in following either version would have to debate the pros and cons before deciding.

## Regression Method

The most commonly used method for demand estimation followed by economists is the regression method. The method involves four steps:

(a) Identification of variables which influence the demand for the good whose function is under estimation

(b) Collection of historical data (time series and/or cross section) on all the relevant variables

(c) Choosing alternative functional forms (linear, reciprocal, double log, quadratic, etc.) for the function

(d) Estimation of the function

Specification of causal variables comes from the underlying economic theory of demand, discussed in the previous two chapters. For example, if we were to estimate the demand function for groundnut oil in India, the relevant causal variables would be national income at constant prices, prices of groundnut oil, prices of vanaspati ghee and pure ghee (substitute items) and prices of eggs, fish, meat, gram flour and vegetables (complementary items). By a prior reasoning, demand for groundnut oil might be affected by several other variables also, such as consumers' tastes and preferences, distribution of population into rich and poor or between South Indian and North Indian and consumers' expectations about future price of groundnut oil. However, one should note that a model is a simplified version of a true structure, and the model builder faces a lot of constraints such as availability of data, costs of their collection and time constraint within which one would like to complete the work of demand estimation. For these reasons, a model builder might be content with the important causal variables only. Here a note of caution is in order. No important determinants must be ignored, for otherwise the model would be a wrong tool for decision making. Similarly, inclusion of irrelevant variables may also jeopardize the accuracy of estimates. Thus, the model builder has to take utmost care in choosing the causal variables for estimating the demand function. Needless to say, the units of measurements for various variables would have to be carefully specified.

The demand for groundnut oil function may thus be hypothesized as follows:

$$\left.\begin{aligned} D_g &= f(Y, P_o, P_v, P_g, P_e, u) \\ f_1, f_3\ f_4 &> 0 > f_2\ f_5 \end{aligned}\right\} \tag{4.4}$$

where $D_g$ = Demand for groundnut oil (tens of thousand tonnes)

$Y$ = national income (Rs. billions at 1970–71 prices)

$P_o$ = price of groundnut oil (1970–71 = 100)

$P_v$ = price of vanaspati (1970–71 = 100)

$P_g$ = price of pure ghee (1970–71 = 100)

$P_e$ = price (composite) of egg, fish and meat (1970–71 = 100)

$u$ = 'other' determinants of groundnut oil demand

Once the model has been hypothesised, efforts would have to be put in data collection. It must be emphasised here that data forms the raw-material for estimation, just as wood forms the raw material for wooden furniture. If wood is of poor quality, the furniture would be of poor quality, thus please note that if data are poor (inaccurate), estimated function would have poor reliability. Data, which are historical, are of two types: **time series and cross-section**. Time series data are consecutive observations on a variable of a given population over time. For example, annual national income data of India for the

period 1951 through 2009 are the time series data. Here it is important to note that the population (India), frequency of data (annual) and the order in which they are recorded (1951, 1952, 1953, …, 2009) are very important. You can have time series income data for any other country, individual or state as well. Similarly, one can have time series data for any frequeney, annual, quarterly, monthly, weekly, daily or hourly, but all of them must be of a uniform frequency. Lastly, they must be arranged in a particular order (1951, 1952, … ; January, February, …, Monday, Tuesday …, and so on) to be meaningful. In contrast, cross-section data refers to observations on a variable at a point (over a period) of time across different populations. For example, income data for 2009 by various states in India are the cross section data. Here the year is fixed and the states vary. Yet another example would be the data on the age of all students in a class as of today. In the latter example, today is fixed, and the student (population) is variable.

The estimation could be based on either time series or cross-section data or even on the pooled data of both. Depending upon the availability of data and the problem in hand, the researcher would choose the appropriate data series. Yet another important thing at this stage is the sample size for data. As in all other matters, the greater the sample size, the more reliable are the estimates. Nevertheless, the quality of data is very important, and so is the cost of data collection both in terms of time and money. This is yet another area where the researcher would have to break his head. Notwithstanding this, there is a minimum size for the sample, which has to be greater than the number of demand determinants. In the absence of this minimum, the regression method of estimation would not be available. If the data on an important variable are not available either at all or for an adequate sample size, the researcher may have to resort to *proxy* or *dummy variables* instead. For instance, no data on consumers' tastes and preferences are generally available and researchers usually use time variable as a proxy for it. Similarly, if the demand for groundnut oil in India were affected by the Government policy with regard to import of edible oil in the country and that this policy underwent a significant change once during the sample period, then the effect of this variable could be incorporated into the model through a dummy variable, which could take a value of one (unity) when the policy is of one kind and a value of zero when the policy happened to be of the other kind. Thus, there are ways by which a careful researcher could salvage the model even in the presence of some data difficulties.

The next stage pertains to the choice of a functional form for the function. The function could take any, one of the several forms: linear, quadratic, cubic, double-log, semi-log, reciprocal, etc. Economic theory, which might give a rough idea about some functional forms, might never be able to identify a unique form. Under such a situation, the researcher has to experiment with all theoretically plausible forms and then to choose the one which is the most ideal on the grounds of both theory and empirical tests. A detail discussion of this is beyond the purview of this book. Suffice it to say here that the researcher could estimate the function in a few alternative forms, and then with the aid of some statistical tests and a prior knowledge of the signs and magnitudes of the coefficients, choose the most appropriate function. The linear and double-log forms are the most popular form in the literature. If a linear formulation is assumed, the function (4.4) can be written as

$$D_g = a_0 + a_1 Y + a_2 P_0 + a_3 P_v + a_4 P_g + a_5 P_e + u$$
$$a_1, a_3, a_4 > 0 > a_2, a_5 \qquad (4.5)$$

Alternatively, if the relationship was assumed to be of double-log type, the function would be as

$$\log D_g = A_0 + A_1 \log Y + A_2 \log P_0 + A_3 \log P_v + A_4 \log P_g + A_5 \log P_e + v$$
$$A_1, A_3, A_4 > 0 > A_2, A_5 \qquad (4.6)$$

In these equations, $a_0$, $a_1$, …, $a_5$ and $A_0$, $A_1$ …, $A_5$ are parameters, and $u$ and $v$ error terms.

Equations 4.5 and 4.6 are inexact (stochastic) equations as they contain an error term ($u/v$). The error term here accounts for three things:

(a) Effects of 'omitted' determinants of groundnut oil demand
(b) Errors in the measurement of variables in the function
(c) Mis-specification of functional form error, if any, arising due to the assumption of linear (double-log) relationship

It should be emphasized here that such errors are common in any function and thus all estimated functions are, in fact, stochastic.

Given the theoretical model and the data, the next and the last step is economic estimation. The most popular method available for this purpose is *the least-squares method of estimation*. It is based on an unconstrained optimization technique. Under the method, estimates of parameters are obtained such that the sum of the squares of the errors between the actual values of the dependent variable and its estimated value is minimised with respect to each of the parameters under estimation. For details on this, one should look up some econometrics text (e.g. Johnston, 1986, Gujarati, 2007). We shall illustrate the method briefly for a one independent variable (simple regression) model.

Let the model for estimation be

$$Y_i = a + b\,X_i + u_i \qquad (4.7)$$
$$i = 1, \ldots, n$$

Its estimated version be

$$\hat{Y} = \hat{a} + \hat{b}X_i \qquad (4.8)$$
$$i = 1, 2, \ldots, n$$

where hat (^) denotes the estimated value of the corresponding variable (parameter), and $n$ the sample size.

The optimization problem then is.

Minimize

$$\sum_i (Y_i - \hat{Y}_i)^2 = \sum Y_i - \hat{a} - \hat{b}X_i)^2$$

with respect to $\hat{a}$ and $\hat{b}$.

Solution of this problem will yield the following two formulae for $a$ and $\hat{b}$

$$\hat{b} = \frac{\sum_i (Y_i - \bar{Y})(X_i - \bar{X})}{\sum_i (X_i - \bar{X})^2} \qquad (4.9)$$

$$\hat{a} = \bar{Y} - \hat{b}\bar{X} \qquad (4.10)$$

where $\overline{X}$ and $\overline{Y}$ are arithmetic means of $Y$ and $X$, respectively. There is an important statistic associated with estimated equations. This is called the coefficient of determination and is denoted by $R^2$. Its formula is the following:

$$R^2 = 1 - \frac{\sum_i \hat{u}_i^2}{\sum_i (Y_i - \bar{Y})^2} \qquad (4.11)$$

where $\hat{u}_i = Y_i - \hat{Y}_i$

$R^2$ gives the percentage of variation in the dependent variable ($Y$) which is explained by the estimated equation or the independent variables in the function. Other things remaining the same, the higher the value of $R^2$, the better is the estimated equation.

Given the data on $Y$ and $X$, $\overline{Y}$ and $\overline{X}$ could easily be found, then through the use of formula (4.9) and (4.10), the values of parameters $\hat{a}$ and $\hat{b}$ could be obtained, and finally $R^2$ value could be found using formula (4.11). These calculations are illustrated below for a simple model of groundunt oil demand:

$$D_{gi} = a + b\,Y_i + U_i \tag{4.12}$$

The data on all the variables in the groundnut oil demand in India are provided in Table 4.1. The calculations for model (4.12) are given in Table 4.2 below:

$$\bar{D}_g = \frac{\sum D_g}{n} = \frac{2058}{27} = 7$$

$$\bar{Y} = \frac{\sum Y_i}{n} = \frac{7565}{27} = 280$$

$$\hat{b} = \frac{\sum (Y_i - \bar{Y})(D_g - \bar{D}_g)}{\sum (Y_i - \bar{Y})^2} = \frac{53931}{148443} = 0.36$$

$$\hat{a} = \bar{D}_g - \hat{b}_{\bar{y}}$$

$$= 76 - (0.36)(280)$$

$$= -25$$

Thus $\hat{D}_{gi} = -25 + 0.36Y_i$ (4.13)

Equation (4.13) is the estimated version of the model (4.12). Equation (4.13) could now be used to derive the series on estimated values of $D_g$ (See Table 4.2, column 8). Given the series on $D_g$ and $\bar{D}_g$, $\bar{u}$ could be obtained as the difference of these two (vide Table 4.2, column 10). Using these calculations, $R^2$ value is obtained as

$$R^2 = 1 - \frac{\sum \hat{u}^2}{\sum (D_g - \bar{D}_g)^2}$$

$$= 1 - \frac{5174.4}{23802}$$

$$= 0.78$$

Thus, the estimated equation (4.13) explains 78 per cent of the variation in the demand for groundnut oil in India. It should be noted here that the value of $R^2$ is necessarily positive and it can never exceed unity:

$$0 \leq R^2 \leq 1$$

The least-squares method just described is available for estimating equations with more than one cause variables (called **multiple regression equation**) as well. The principle remains the same, though mathematics becomes a little complicated and perhaps boring. Fortunately, computers are easily accessible these days and most softwares (including excel) have regression packages. The results of a few multiple regression models for alternative formulations of the groundnut oil demand in India model (4.4) are provided in Table 4.3. A careful study of the results in this table would reveal that the coefficients of variables might change as more or less variables are included in the function. For example, the co-efficient of $Y$ in Eq. 4.3.1 equals 0.36 while of it in Eq. 4.3.2 equals 0.30 and in Eq. 4.3.4

equals 0.24. Then the question is, on which coefficient to rely on? The answer is simple; rely on the ones which are in the chosen form. Of the four linear equations in Tables 4.3, equation 4.3.4 is perhaps the best, for it contains all the relevant variables, all coefficients have a prior signs and $R^2$ takes the highest value. Thus, the coefficients of variables in this equation alone should be accepted. Please note that the coefficient of $P_0$ is positive in Eq. 4.3.2, which is incompatible with a prior reasoning.

**Table 4.1** *Data on Consumption of Groundnut oil and Its Determinants*

| *Year* | *Consumption of ground oil (0000' tonnes)* | *National Income at 1970-71 prices (Rs. billions)* | *Wholesale Price Index of* | | | |
|---|---|---|---|---|---|---|
| | | | *Ground nut oil* | *Vanaspati* | *Egg, fish & meat* | *Ghee* |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| 1951–52 | 44 | 171 | 37.1 | 42.6 | 34.4 | 41.9 |
| 1952–53 | 34 | 177 | 30.1 | 35.0 | 34.7 | 42.6 |
| 1953–54 | 48 | 188 | 36.1 | 39.3 | 33.9 | 40.8 |
| 1954–55 | 50 | 193 | 23.2 | 31.2 | 33.2 | 37.9 |
| 1955–56 | 42 | 200 | 22.2 | 29.1 | 33.3 | 35.8 |
| 1956–57 | 58 | 210 | 31.2 | 36.9 | 33.5 | 40.1 |
| 1957–58 | 62 | 206 | 31.6 | 39.8 | 34.0 | 42.3 |
| 1958–59 | 78 | 223 | 32.3 | 40.3 | 37.6 | 42.9 |
| 1959–60 | 65 | 227 | 35.5 | 42.3 | 38.9 | 46.9 |
| 1960–61 | 73 | 242 | 41.5 | 48.2 | 42.9 | 48.4 |
| 1961–62 | 75 | 250 | 43.3 | 51.9 | 47.0 | 49.4 |
| 1962–63 | 71 | 254 | 38.2 | 50.0 | 54.1 | 53.4 |
| 1963–64 | 59 | 267 | 39.1 | 51.1 | 58.4 | 52.3 |
| 1964–65 | 76 | 288 | 49.3 | 69.0 | 66.5 | 60.4 |
| 1965–66 | 55 | 271 | 62.9 | 68.2 | 82.1 | 69.4 |
| 1966–67 | 57 | 273 | 89.4 | 93.5 | 90.8 | 78.6 |
| 1967–68 | 74 | 297 | 74.5 | 81.5 | 98.6 | 87.4 |
| 1968–69 | 58 | 305 | 67.7 | 74.3 | 97.5 | 92.4 |
| 1969–70 | 66 | 324 | 93.4 | 90.2 | 93.3 | 94.5 |
| 1970–71 | 79 | 342 | 100.0 | 100.0 | 100.0 | 100.0 |
| 1971–72 | 110 | 347 | 85.6 | 92.0 | 104.6 | 100.1 |
| 1972–73 | 98 | 342 | 99.5 | 100.1 | 116.3 | 103.3 |
| 1973–74 | 107 | 360 | 154.4 | 131.8 | 142.1 | 140.8 |
| 1974–75 | 116 | 365 | 173.5 | 171.5 | 169.5 | 152.1 |
| 1975–76 | 154 | 401 | 133.7 | 160.0 | 173.2 | 148.6 |
| 1976–77 | 120 | 403 | 134.2 | 148.0 | 176.2 | 153.1 |
| 1977–78 | 139 | 439 | 170.8 | 167.5 | 196.6 | 160.0 |

The double-log or linear in log version is provided by Eq. 4.3.5 in Table 4.3. This is also an alternative to each of the above four equations. Here all the coefficients have correct signs as they do in Eq. 4.3.4 and it has a value for $R^2$ which is higher than that in any other equation. Thus, on these grounds alone, Eq. 4.3.5 is preferred even to Eq. 4.3.4. There are several other statistical tests for choosing among alternative estimates but for that one would need a course in Econometrics.

Thus, from our point of view, Eq. 4.3.1 to 4.3.5 provide alternative estimated demand functions and

**Table 4.2** *Estimation of a Simple Demand Equation*

| *Year* | *Cons. of groundnut oil ($D_g$) ('000 tonnes)* | *National income at 1970–71 price (Y) (Rs. in billions)* | $D_g - \bar{D}_g$ | $Y - \bar{Y}$ | (3) × (4) | $(D_g - \bar{D}_g)^2$ | $(Y - \bar{Y})^2$ | $\hat{D}_g = (-25 + 0.36Y)$ | $(D_g - \hat{D}_g)$ | $(\hat{M})^2 = (9)^2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| 1951–52 | 44 | 171 | –32 | –109 | 3488 | 1024 | 11881 | 37.8 | 6.2 | 38.4 |
| 52–53 | 34 | 177 | –42 | –103 | 4326 | 1764 | 10609 | 39.0 | –5.0 | 34.8 |
| 53–54 | 48 | 188 | –28 | –92 | 2576 | 784 | 8464 | 43.8 | 4.2 | 17.6 |
| 54–55 | 50 | 193 | –26 | –87 | 2262 | 676 | 7569 | 45.5 | 4.5 | 20.2 |
| 55–56 | 42 | 200 | –34 | –80 | 2720 | 1156 | 6400 | 48.0 | –6.0 | 36.0 |
| 56–57 | 58 | 210 | –18 | –70 | 1260 | 324 | 4900 | 51.5 | 6.5 | 42.2 |
| 57–58 | 62 | 206 | –14 | –74 | 1036 | 196 | 5476 | 50.1 | 11.9 | 141.6 |
| 58–59 | 78 | 223 | 2 | –57 | 114 | 4 | 3249 | 56.0 | 22.0 | 484.0 |
| 59–60 | 65 | 227 | –11 | –53 | 583 | 121 | 2809 | 57.4 | 7.6 | 57.8 |
| 60–61 | 73 | 242 | –3 | –38 | –114 | 9 | 1444 | 62.7 | 10.3 | 106.1 |
| 61–62 | 75 | 250 | –1 | –30 | 30 | 1 | 900 | 65.5 | 9.5 | 90.2 |
| 62–63 | 71 | 254 | –5 | –26 | 1456 | 25 | 676 | 66.9 | 4.1 | 16.8 |
| 63–64 | 59 | 267 | –17 | –13 | 221 | 289 | 169 | 71.4 | –12.4 | 163.8 |
| 64–65 | 76 | 288 | 0 | 8 | 0 | 0 | 64 | 78.8 | –2.8 | 7.8 |
| 65–66 | 55 | 271 | –21 | –9 | 189 | 441 | 81 | 72.8 | –17.8 | 316.8 |
| 66–67 | 57 | 273 | –19 | –7 | 133 | 361 | 49 | 73.5 | 16.5 | 272.2 |
| 67–68 | 74 | 297 | –2 | 17 | –34 | 4 | 289 | 81.9 | –7.9 | 62.4 |
| 68–69 | 58 | 305 | –18 | 25 | –450 | 324 | 625 | 84.7 | –26.7 | 712.9 |
| 69–70 | 66 | 324 | –10 | 44 | –440 | 100 | 1936 | 91.4 | –25.4 | 645.2 |
| 70–71 | 79 | 342 | 3 | 62 | 186 | 9 | 3844 | 97.7 | 18.7 | 349.7 |
| 71–72 | 110 | 347 | 34 | 67 | 2278 | 1156 | 4489 | 99.4 | 10.6 | 112.4 |
| 72–73 | 98 | 342 | 22 | 62 | 1364 | 484 | 3844 | 97.7 | 0.3 | 0.1 |
| 73–74 | 107 | 360 | 31 | 80 | 2480 | 961 | 6400 | 104.0 | 3.0 | 9.0 |
| 74–75 | 116 | 365 | 40 | 85 | 3400 | 1600 | 7225 | 105.7 | 10.3 | 106.1 |
| 75–76 | 154 | 401 | 78 | 121 | 9438 | 6084 | 14641 | 118.3 | 35.7 | 1274.5 |
| 76–77 | 120 | 403 | 44 | 123 | 5412 | 1936 | 15129 | 119.0 | 1.0 | 1.0 |
| 77–78 | 139 | 439 | 63 | 159 | 10017 | 3969 | 25281 | 131.6 | 7.4 | 54.8 |
| Sum | 2058 | 7565 | – | – | 53931 | 23802 | 148443 | – | – | 5174.4 |



The McGraw-Hill Companies

Eq. 4.3.5 the best form of the model. Given the theoretical model and data on all the relevant variables, estimates could thus be obtained through the least-squares (regression) method of estimation.

**Table 4.3** *Estimated Groundnut oil Demand Models for India*

| | | |
|---|---|---|
| **Linear** | | |
| | 4.3.1 | $D_g = -25 + 0.36Y$<br>$R^2 = 0.79$ |
| | 4.3.2 | $D_g = -13 + 0.30Y + 0.09\ P_o$<br>$R^2 = 0.80$ |
| | 4.3.3 | $D_g = -17 + 0.26Y - 0.51\ P_o + 1.13\ P_v - 0.39P_e$<br>$R^2 = 0.84$ |
| | 4.3.4 | $D_g = -20 + 0.24Y - 0.67\ P_o + 1.18\ P_v + 0.50\ P_g - 0.66\ P_e$<br>$R^2 = 0.85$ |
| **Double Log** | | |
| | 4.3.5 | $\log D_g = -12.4 + 1.78 \log Y - 1.22 \log P_o + 2.20 \log P_v$<br>$+ 0.80 \log P_g - 1.62 \log P_e$<br>$R^2 = 0.90$ |

## 4.2 ANALYSIS OF ESTIMATED DEMAND FUNCTION

The theoretical demand function (Eq. 4.4) is useful for decision makers, for it tells them the factors which influence their market as well as the direction of their effects. For example, it informs them that if national income grows, the demand for their product would increase; and that if they want a growing market, one way to ensure that is through an increase in national income. However, it does not tell them, by how much? *The estimated demand function is preferred by managers to its theoretical counterpart*, for it is this form alone which answers the question, by how much while retaining all the virtues of a theoretical model? To explain this, let us go on with the analysis of the estimated functions.

In the linear form, regression coefficients stand for, what is called, the **multipliers!** For example, in equation 4.3.4, the coefficient of $Y$ (+ 0.24) indicates that if $Y$ increases by one unit (i.e. rupees one billion) *ceteris paribus* demand for groundnut oil will increase by 0.24 units (i.e. 0.24 × 10000 tonnes = 2400 tonnes), and vice versa. Similarly, in the same equation, if $P_o$ index increases by one point *ceteris paribus* groundnut oil demand would decline by 0.67 units (= 6700 tonnes). Note that these multipliers are additive. That is, if $Y$ increases by one unit and $P_o$ index also increases by one unit *ceteris paribus*, demand for groundnut oil would decline by 4300 (= + 2400 – 6700) tonnes.

In the double-log form, regression coefficients represent **elasticities**. For example, in Eq. 4.3.5, the coefficient of $Y$ (1.78) indicates that if $Y$ increases by 1 per cent (= log $Y$ increases by one unit), other things remaining the same, groundnut oil demand would increase by 1.78 percent (or log $D_g$ would increase by 1.78 units). Similarly, if both $Y$ and $P_o$ increase by 1 per cent each, *ceteris paribus*, demand for groundnut oil would increase by 0.56 (= + 1.78 – 1.22) per cent.

Taking up the total analysis of the estimated demand function 4.3.4, one can proceed as follows. First, the equation indicates that national income and prices of groundnut oil, vanaspati, ghee, egg, fish and meat are the determinants of groundnut oil demand in India. Secondly, the sign of income

coefficient indicates that groundnut oil is a superior item in the consumption basket of Indians, sign of $P_o$ indicates that the law of demand holds good in the case of the commodity under question, signs of $P_v$ and $P_g$ indicate that vanaspati and ghee are substitutes to groundnut oil, and the sign of $P_e$ reveals that egg, fish and meat taken together are complements to groundnut oil. Thirdly, they indicate various specific values of the multipliers (elasticities) as explained in the last but one paragraph and they could be used to analyse the effect of market forces on demand. Thus, for example, the income elasticity of demand ($e_{D,\ y}$) at

$Y = 400$, given $P_o = 150$, $P_v = 130$, $P_g = 140$ and $P_e = 120$ would be

$$e_{D,Y} = \left(\frac{\partial D}{\partial Y}\right)\left(\frac{Y}{D_g}\right) = (0.24)\frac{400}{D_g}$$

where, $D_g = -20 + 0.24\ (400) - 0.67\ (150) + 1.18\ (130) + 0.50\ (140) - 0.66\ (120)$

$= -20 + 96 - 100.5 + 153.4 + 70 - 79.2$

$= 119.7$

Thus, $e_{D,} = (0.24)\left(\frac{400}{119.7}\right)$

$= 0.80$

The income elasticity (arc) between income 400 and 500, given the values of all other variables as above, would be

$$e_{D,Y} = (0.24)\left(\frac{400 + 500}{D_{g_1} + D_{g_2}}\right)$$

where $D_{g_1}$ is the demand corresponding to $Y = 400$ and $D_{g_2}$ is the demand corresponding to $Y = 500$. The calculation for the first one, as just obtained, stands at 119.7 and for the second one, as can be obtained by replacing 400 by 500 in the above calculation, stands at 143.7. Thus,

$$e_{D,Y} = (0.24)\left(\frac{900}{119.7 + 143.7}\right)$$

$$= (0.24)\frac{900}{263.4}$$

$= 0.82$

Similarly, the price elasticity of demand at point $P_o = 150$, all other variables taking the above values ($Y = 400$), equals

$$e_{D,P} = (-0.67)\frac{150}{119.7}$$

$= -0.84$

The cross elasticity with respect to ghee price at $P_g = 140$, other variables taking values same as above, would be

$$e_D, P_g = (0.50)\frac{140}{119.7}$$

$= 0.58$

and so on.

Fourthly, the estimated demand function could be used to derive the equations of the demand and Engel curves for given values of other variables, which could then be used to derive equations for *AR*,

*TR* and *MR* curves. This is done by substituting the values of all other variables but the own price or income, depending upon whether the demand curve or Engel curve is to be derived, in the estimated demand function. For example, for $Y = 400$, $P_v = 140$ and $P_e = 120$ equation of the demand curve corresponding to Eq. 4.3.4 (Table 4.3) would be the following:

$$D_g = -20 + 0.24\,(400) - 0.67\,P_o + 1.18\,(130) + 0.50\,(140) - 0.66\,(120)$$

or, $$D_g = 220.2 - 0.67\,P_o \qquad (4.14)$$

Similarly, the equation of the Engel curve for $P_o = 150$, $P_v = 130$, $P_g = 140$ and $P_e = 120$, corresponding to equation 4.4 would be

$$D_g = -20 + 0.24\,Y - 0.67\,(150) + 1.18\,(130) + 0.50\,(140) - 0.66\,(120)$$

or, $$D_g = 23.7 + 0.24Y \qquad (4.15)$$

Given the equation of the demand curve, equations for average revenue (*AR*), total revenue (*TR*), and marginal revenue (*MR*) curves could be derived. For example, the *AR* equation corresponding to function (4.14) is obtained by solving the said equation for price, which equals *AR*:

$$P_o = \frac{220.2}{0.67} - \frac{1}{0.67}\,D_g$$

Writing $P$ for $P_o$ and $Q$ for $D_g$ (to use the common notation) and solving, we get,

$$P = 329 - 1.5\,Q = \text{AR} \qquad (4.16)$$

The *TR* equation then is obtained by multiplying both the sides of equation (4.16) by $Q$:

$$TR = PQ = 329\,Q - 1.5\,Q^2$$

or, $$TR = 329\,Q - 1.5\,Q^2 \qquad (4.17)$$

The *MR* equation is obtained simply by the first differentiation of *TR* equation:

$$MR = \frac{d(TR)}{dQ} = 329 - 3.0\,Q$$

or, $$MR = 329 - 3Q \qquad (4.18)$$

From Eq. (4.16) and (4.18), it is clear that the slope of *MR* curve (= –3) is twice that of the AR (demand) curve (= –1.5), measuring the slope from the quantity axis*, and the intercepts of both the curves (at price axis) are identical. Incidentally note that this is true for a linear demand curve only.

Fifthly, the estimated demand function could be analysed to given policy guidelines. For this purpose, the causal variables must first be classified into variables that can be controlled by the firm (industry/economy) and those which are outside the purview of firm's control. Since this is the market demand function, the question of the firm's control does not even arise. The industry (i.e., all groundnut oil producers) does exercise control on its product (groundnut oil) price, and perhaps on no other variable. If the industry desires to sell, say, 14 lakh tonnes (i.e., target $D_g = 140$, since $D_g$ is in tens of thousand tonnes) then it must set its product price at index 120 given the values of other variables as above:

$$D_g = -20 + 0.24\,(400) - 0.67\,(P_o) + 1.18\,(130) + 0.50\,(140) - 0.66\,(120)$$

i.e. $$140 = -20 + 96 - 0.67\,P_o + 153.4 + 70 - 79.2 = 220.2 - 0.67\,P_o$$

or, $$P_o = \frac{80.2}{0.67} = 120$$

*The slope of the MR curve would be half of that of the AR curve, if the slope is measured from the price axis.

Thus, in order to achieve the target of 14 lakh tonnes of sales, the groundnut industry must bring down its product price from the index of 150 to the index of 120, other things remaining the same.

The economy has a greater leverage for achieving a given target than does the industry. For example, if the planning commission has the target to sell groundnut oil in the amount of 14 lakh tonnes, then it can achieve this through fostering changes in anyone or more of the causal variables under its control. Suppose, the planning commission decides not to interfere with the prices of groundnut oil related goods and yet want to achieve the target. Then the alternatives open to it are three:

(a) Variations in $P_o$ alone

(b) Variations in $Y$ alone

(c) Variations in both $P_o$ and $Y$

Hitherto, it has already been seen that if the target were to be achieved through variations in $P_o$ alone, then it must be set at index 120. On the other hand, if the target is to be achieved through $Y$ alone, then the income level must be raised from earlier Rs. 400 billion to Rs. 485 billion:

$$140 = -20 + 0.24\,(Y) - 0.67\,(150) + 1.18\,(130) + 0.50\,(140) - 0.66\,(120)$$

or,
$$Y = \frac{116.3}{0.24} = 485$$

On the other hand, if the planning commission desires to obtain the target by a combination of variations in $Y$ and $P_o$, then the alternative policies would be inferred as follows:

$$140 = -20 + 0.24\,(Y) - 0.67\,P_o + 1.18\,(130) + 0.50\,(140) - 0.66\,(120)$$

or,
$$0.24\,Y - 0.67\,P_o = 15.8 \qquad (4.19)$$

Any combination of $Y$ and $P_o$ which satisfies the above equation would be the appropriate policy. A few of such policies (including the above two) would thus be the following:

| $P_o$ | $Y$ |
|---|---|
| 120 | 400 |
| 150 | 485 |
| 138 | 450 |
| 145 | 471 |

Thus, a simultaneous increase in national income from Rs. 400 billion to Rs. 450 billion and a decrease in groundnut oil price from index 150 to index 138 would ensure that the target is achieved. Any number of alternative policies could be inferred from the policy Eq. (4.19). Similarly, alternative policy models (like Eq. 4.19) could be derived for alternative sets of instruments and for alternative targets. All that one needs to do for this purpose, is to substitute the numerical values of the target variable and the other non-target, non-instrument variables in the estimated function, and solve the model for instrument variables.

The sixth and the *last* use of the estimated demand function would be for forecasting demand. This is discussed in detail in the following section. Suffice it to point out here that given the estimated function, if one knows the values of causal (independent) variables in the forecast period, the forecast for the dependent variable could be easy to infer. For example, if Eq. 4.3.4 (vide Table 4.3) were the estimated demand function for groundnut oil demand in India, and if the values of independent variables

in 2011 were $Y = 500$, $P_o = 180$, $P_v = 160$, $P_g = 150$, $P_e = 150$, the forecast for groundnut oil demand in 2011 would be

$$D_g\,(2011) = -20 + 0.24\,(500) - 0.67\,(180) + 1.18\,(160) + 0.50\,(150) - 0.66\,(150) = 144.2$$

At this stage it is necessary to emphasize that all the calculations that have been carried out in this section are subject to the errors, in the estimation of the demand function. For the reasons spelled out in Section 4.1 (Regression method sub-section), all the estimated equations are stochastic and, therefore, inferences from them should be taken as suggestive rather than definite. To minimize the error, researchers always endeavour to obtain the best possible estimates.

## 4.3 DEMAND FORECASTING

Let us begin this section by posing a set of standard questions and then move on to answer each one of them. What is demand forecasting, what are the various kinds of demand forecasts, who forecasts demand, why demand forecasts are made, how to make demand forecasts, and are accurate forecast possible and, if not, how to measure forecast inaccuracy?

### Meaning and Kinds of Demand Forecasts

Demand forecast means estimation of the demand for the good in question in the forecast period. For example, if the good is Maruti Car and the forecast period is the year 2011, then the forecasting problem is to estimate the demand for Maruti Cars in 2011. For genuine forecasts, the forecast period is a future period and they are referred to *ex-ante forecasts*. The forecasts for past and present periods, which are carried out to test the credibility of the forecasting model, are called *ex-post forecasts.* The *ex-ante forecasts* are often made for a number of periods in future. Thus, Maruti Udyog could very well be trying to estimate its market not only in 2011 but also in 2012, 2013, 2014, and so on.

Demand forecasts may be attempted not only for the *total market* but also for *market segments,* like domestic demand and foreign (export) demand. Similarly, attempts are made to forecast both the *firm demand* and *industry demand*. For example, the Planning Commission or the Car manufacturing association might engage itself in forecasting separately the demand for Maruti Cars, Tata Cars, Mahindra Cars and Toyota Cars, as well as the total demand for all cars in the country as a whole. Yet another classification of demand forecasts is between *passive* and *active* forecasts. The former refers to the estimation of future demand if things continue the way they have been in the past (status quo). In contrast, the latter kind of forecast takes into account the likely changes in the relevant variables in future in estimating the future demand. For example, if Maruti Udyog is contemplating an improvement in the quality of its car and at the same time it is committed to go in for a vigorous advertising campaign for its product, then while passive forecasts would ignore the impact of these changes on demand, the active forecasts would pay due consideration to these factors while estimating the future demand for Maruti car. Thus, it is clear that it is the active forecasts which are meaningful, though passive ones may be obtained to assess the impact of new policies on the market.

### Significance of Demand Forecasts

Demand forecasts are attempted by several organizations and individuals. For instance, the planning Commission undertakes systematic forecasts for the demands for all major goods and services in the

economy for the subsequent next five years or so. The industry organizations are engaged in forecasting the demand for their corresponding industrial products and firms in their corresponding brands. Researchers undertake forecasts of all kinds, including the worldwide forecasts, which are also carried out by international organizations like the World Bank, International Monetary Fund, United Nations' Organizations and Asian Development Bank.

The significance of these forecasting studies can hardly be exaggerated. In particular, these are needed to plan future production and thereby future needs for various resources, including manpower, raw-materials and funds. Production of most goods and services takes time, and sometimes this period, called the gestation period, is spread over decades. For example, production of hydro-electric power from scratch might well take over a decade and one very well knows that Maruti Car came to the market only about ten years after the project was conceived. The case of production of services may be similar. For example, if more M.B.A. graduates are to be turned out, it might take several years to achieve that target. New institutions might have to be started or/and more facilities in terms of buildings, furniture, faculty, staff, library, computers, etc. may have to be provided to enlarge seats in the existing institutions. Unless the future demand is known well in advance there may not be enough time to plan and execute the production to meet that demand. And if demand is not met, firms may not be able to attain their objectives. For example, if production were short of demand, some prospective consumers would have gone with little or no consumption of the good in question. On the other hand, if production had exceeded the demand; leading to a glut, firms would not be able to sell all their produce at reasonable or any price. In either case, the firms would make less profit than was possible if better forecasts were available. Thus, both under and over production are undesirable. To avoid such a situation, as accurate as possible estimates of likely future demands are essential for the success of an organisation.

There is no choice between forecasting and not forecasting. Not to forecast is to assume an indefinite continuation of the status quo. To expect no change, in these days of fast changing consumption habits in the face of emerging new products, seems very short sighted and highly unrealistic. The area of choice only concerns the way the forecast is made, who does it and what resources are devoted to it.

## Forecasting Methods

Forecasts (genuine or ex-ante), by its very definition, involves the future, which is uncertain. Thus, no forecast can be expected to be 100 percent correct. This together with the essential feature of forecasts argues for a paradox: forecasts are essential but there is no way to generate absolutely accurate forecasts. This is true but efforts have to be mounted to obtain as accurate a forecast as possible. The basic question then is; how to get good forecasts?

There is perhaps no unique method for anything, and there is not one for forecasting either. This is because methodology is always plural. Forecasting methods are no exception to this rule. As is generally the case, there are several methods of demand forecasting basically for three reasons:

(a) No method is perfect and no method (worth its name) is useless

(b) No method is the best under all circumstances

(c) The best method may not be available in a particular situation due to constraints from data and/or resources (time and money)

In what follows, we shall discuss briefly all the methods, bringing out clearly their pros and cons, as well as the situations in which a particular method is more appropriate than the others.

The various demand forecasting methods, which are quite versatile in the sense that they could be used to forecast most other variables as well, could, at a little risk of overlapping, be classified as in Chart 4.1.

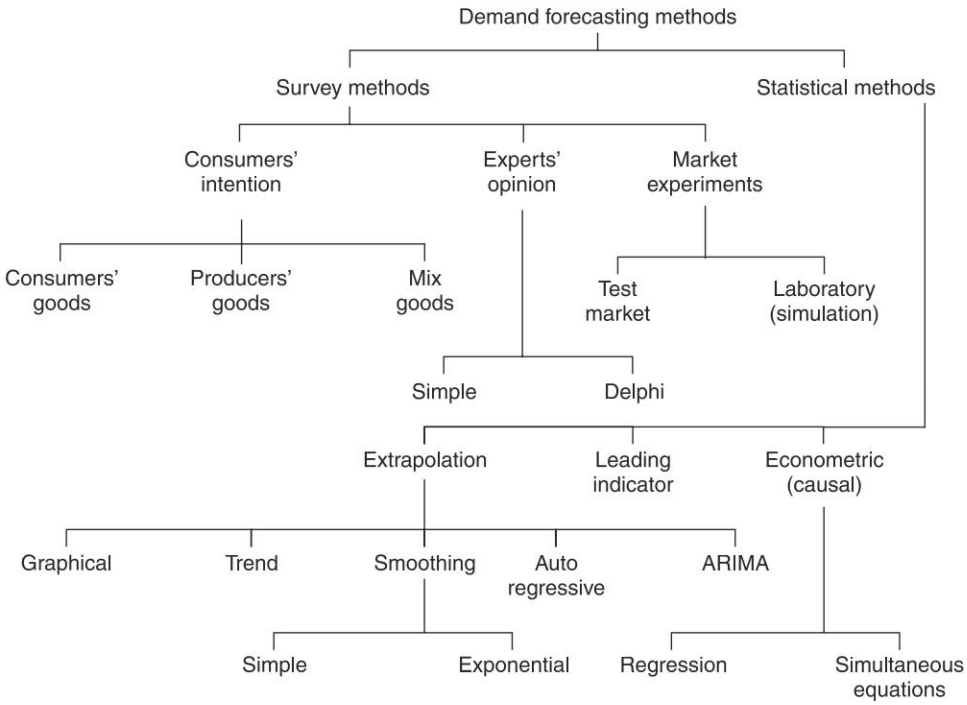



**Chart 4.1**

Broadly speaking, there are two approaches to demand forecasting: *Survey method* and *Statistical method.* Under the first, surveys are conducted about the intentions of consumers (individuals or/and industries), opinions of experts, or of markets, and through their analysis, forecasts on demand are made. Such surveys are of two kinds: census and sample. Under the former, all consumers/experts/markets are surveyed while in the latter only a selected subset of them are surveyed and through their study inferences about the whole population are drawn and forecasts made. Under statistical methods, historical data are extrapolated or analysed through econometric models and through them forecasts are squeezed out.

Before discussing each method in detail, we will discuss their special features. The survey methods are usually suitable for short-term forecasts and new products' demand forecasting. For short-term forecasts, because consumers' intentions are volatile over time, particularly in the present day world where new products get available year after year. The survey methods are suited to new products' demand forecasting because there are no historical data on their consumptions. Also, because consumers do not keep systematic data on their past consumption and it is difficult to recall their purchases after a lapse of time; researchers often survey the same group of consumers more than once in a year, particularly if the product under demand forecasting happens to be a seasonal one. In contrast, statistical methods are common for long-term forecasts and for forecasting demand for well established (old) products. This is because survey methods are unsuited for these purposes as just seen, and once the data exists, statistical

methods are capable of generating long-term forecasts. Theoretically speaking, both the groups of methods are at least available for forecasting demands for old products. For new products, only survey methods will do the job. This is because the data constitutes the raw-material for statistical methods, and if there is no raw-material for statistical methods there can be no finished product. We will now discuss the various methods in detail.

## Consumers' Intentions Survey Method

Under this method, demand forecasts are attempted through a survey of consumers' intentions. Who are the consumers' of a good? If the good in question happens to be purely a consumer product, then consumers are households. On the other hand, if the good is purely a producers' good, then consumers are industries using that product. However, if the product is used both for final consumption as well as a raw-material for some other goods' production, then its consumers come both from households as well as industries. Most goods fall under the last category and thus their demand forecasts calls for a survey of both the kinds of consumers.

For estimating the final consumption demand for a good in future, a survey of households is conducted. In the survey each potential household is asked its intentions about purchases of the product in question in the forecast period. If the survey happens to be a census, then the demand forecast for total household consumption is obtained simply by adding the intended demands of all households:

$$DF = ID_1 + ID_2 + \ldots + ID_n \tag{4.20}$$

where, $DF$ = demand forecast for all households. $ID_1$ = intended demand of household 1 in forecast period, $ID_2$ = intended demand of household 2 in forecast period, and so on, and $n$ = number of consuming households.

In case the survey was of a sample of households instead of all, then demand forecasts would have to be obtained differently, depending upon how the sample was drawn. For example, if the simple random sampling method was adopted for the purpose, the population consisted of $N$ households and sample stood at $n$, then demand forecast would be given by

$$DF = (ID_1 + ID_2 + \ldots ID_n)\left(\frac{N}{n}\right) \tag{4.21}$$

Alternatively, if stratified random sampling was resorted to and the population was divided into $K$ groups, then the forecast for aggregate demand by the whole household sector would be given by

$$DF = N_1\,(AID_1) + N_2\,(AID_2) + \ldots + N_k\,(AID_k) \tag{4.22}$$

where, $N_i$ = population of households in group $i$ ($i$ = 1, 2, …, $k$)

$AID_i$ = average intended demand of the surveyed households in group $i$ ($i$ = 1, 2, …, $k$).

The *end-use method* would have to be used if the product whose demand forecasts are being made is wholly or partly used in the process of production of some goods and services. Under this method, the industry demand for the product is obtained through the use of equations such as follows:

$$(DF)_I = a_1 X_1 + a_2 X_2 + \ldots + a_n X_n \tag{4.23}$$

where $DF_I$ = demand forecast for the product in question in the forecast period for industrial uses.

$a_i$ = input requirement of the product per unit of output of industry $i$ (input-output coefficient) ($i$ = 1, 2, …, $n$)

$X_i$ = production forecast of industry $i$ in the forecast period ($i = 1, 2, \ldots, n$)

The data for applying the end-use method would come through a survey of firms in all industries using the product. Once again the survey could be of all or selected firms. However, the choice in this respect would be restricted to firms and not to industries. Since different industries have different input-output coefficients, none could be left out. In this respect, documents of the Planning Commission would be very useful. These documents, in fact, could form an alternative to the survey of firms in various user industries, for they might provide reliable information both on the input-output coefficients as well as on the likely production of various industries in the forecast period.

When the product whose demand is under forecast happens to be both a consumers' good as well as a producers' good, then the two components have to be forecast separately through the methods just outlined, and the total demand is then obtained as the sum of the two components. Further, if the good in question has export demand and/or import supply, allowance will have to be provided for them as well. The export-import forecasts would come from a proper assessment of national policies with regard to them and the available foreign markets.

When consumers are not able to formulate their intentions about likely purchases in the forecast period, a careful surveyor could still arrive at them by collecting information on their past purchases if the product was an old one or by probing their minds and resources if the product happens to be a new one. Similarly, an able surveyor should be able to generate the relevant information when producers are not able to quantify their likely future purchases of the product whose demand is being forecast.

Needless to say, judgements do play a role even in such scientific methods. For instance, in this method, the researcher might modify the forecasts obtained by his evaluations of the errors, if any, in consumers' expressed intentions and the likely purchases as assessed by him in view of the emerging situation in the forecast period.

It is obvious that consumers' intentions' survey method of demand forecasting is tedious, time consuming and costly, particular for goods of common consumption. Yet, if data don't exist and forecasts are needed, there is no alternative to it. For example, the present author (with others) under the sponsorship of the Government of India, has used this method in forecasting the demand for fish in India during the next five years. For producers' goods or for goods which are essentially producers' goods, this method of demand forecasting is quite ideal. The National Council of Applied Economic Research (NCAER) has applied this method in forecasting demand for steel in India.

## Experts' Opinion Survey Method

Under this method, the researcher identifies the 'experts' on the commodity whose demand forecast is being attempted, and probes with them on the likely demand for the product in the forecast period. The word 'expert' is a high-powered term but it should be taken to stand for those who possess the requisite expertise on the subject. Thus, if one were to forecast the demand for say, car, the list of experts would include the Chairman and Managing Directors of various car manufacturing and two-wheeler manufacturing firms, the important research organisations and individuals engaged in such research, and the relevant Government officials concerned with the automobiles industry and import and export of cars. Similarly, if the demand for MBA's was under forecasting, opinions of major employers of MBA's and the heads of institutions producing them could be sought.

How to use experts' opinions in forecasting demand? If the number of experts is just one, then

whatever he/she tells, could be your forecast, subject to, of course, your own well found modifications. When the number of experts is more than one, there are two approaches:

(a) Take a simple or weighted average of the numbers given by various experts, temper it with your own well conceived judgement and arrive at the needed forecasts. Under this situation, the more weight would obviously be attached to the persons having greater expertise than those having limited expertise. Needless to say, the average of the individual guesses/estimates is typically better than all but one or two of the individual guesses; and in the latter case one does not know in advance who these one or two individuals were and, of course, an expert once may not prove to be an experts again and again.

(b) Follow the ***Delphi method:*** Under Delphi method, opinions are first collected from experts and then instead of mathematical averaging, efforts are made to remove or narrow the differences. It is true that no two experts agree, but something could be done to narrow the gap. This is done by the researcher (coordinator) entrusted with forecasts through collection of opinions, reviews of the various estimates so obtained, and feeding the same to the experts for their possible modifications of their earlier forecasts. This procedure could be iterated a few times until there was some scope for arriving at some consensus, for a narrow range, if not for a unique estimate. The narrow range could, of course, be used to give the *interval forecast* directly (i.e., the lower and upper limits within which the demand is forecast to be) and for arriving at a *point* (specific number) *forecast* by tampering it with the overall assessment of the researcher. For example, if the demand forecast for a product in India is being attempted for the year 2011 through this method, one could proceed through the following three steps: One, request all the experts of product *X* to give their individual estimates for the likely demand for *X* in 2011with reasons for the same. Suppose there are 15 experts and their requisite forecasts were the following:

100, 110, 190, 200, 210, 225, 240, 240, 250, 250, 260, 275, 300, 310, and 400

Two, if the differences in forecasts were significant; feedback the results and rationale for the differences to the experts and ask them if they would like to revise their earlier forecasts. In the second round, the experts who gave either too small numbers (e.g. 100 and 110) or too large numbers (e.g. 400), called "outliers", whose identity should of course, not be disclosed, would generally find themselves out of place and they, including the others, would perhaps see some new reasons to modify their hitherto given forecasts.

If the researcher feels that the range of variation is still large and there was some scope for further narrowing, he/she could iterate the procedure. Iteration of this would continue until the coordinator is able to arrive at an acceptable range or beyond which he does not see any scope for reduction. Suppose the exercise leads to the interval of 240 to 280. Three, declare the so arrived range (240 to 280) as the interval forecast for the demand for product *X* in the year 2011 and the simple average of the lower (240) and upper (280) values of the forecasts (i.e. 260) as the point forecast for the variable under forecasting. It should be emphasized that the coordinator would have, of course, also participated in the whole procedure with his/her own inputs at every stage of the exercise and accordingly his/her judgement would have already been reflected in the consensus arrived at stage two above. Needless to say, the interval forecast is more likely to be true than the point forecast.

It is obvious from the above discussion that the Delphi method is quite sound but it could be tedious and costly. Thus in the situations where the number of experts is not too large and they are co-operative,

and the researcher has the necessary fund and the expertise/authority to perform the task, the Delphi method could be appropriate for demand forecasting. Incidentally note that its name, viz. "Delphi", comes from the place where it was first practiced.

A slight variant of the Delphi method is the **Brain Storming method**. Under this, instead of dealing through mail or phone (of Delphi method), the researcher convenes one or more meetings with the experts and allow the experts to freely discuss and debate the subject. Whatever they arrive through such debates constitutes the forecasts.

### Market Experiments' Survey Method

Under this method, market experiments (actual or simulated) are performed to generate demand forecasts. The method of market experiments for demand estimation has been discussed earlier in this chapter (Section 4.1) and it is essentially the same for demand forecasting. The readers are thus advised to go through the relevant pages above for the purpose.

### Graphical Method

Under this method, a graph of historical data on the variable under forecasting is drawn, it is then extrapolated visually up to the forecast period, and finally the value of the variable in the forecast period is read out from the graph to yield the requisite forecasts. To illustrate the method, consider forecasting the groundnut oil demand in India. The historical data (annual for the period 1951–52 to 1977–78) for the variable under forecasting are available in Table 4.1, column 1. Its graph is drawn in Fig. 4.2. If the forecast period is 1987–88, then the graph is extrapolated up to that period as $AB_1$, $AB_2$ or $AB_3$. If the extrapolated graph were as $AB_1$, then the forecast for the groundnut oil demand in India in 1987–88 would be 1680 thousand tonnes. Needless to say, the longer the period beyond the last sample data, the larger is the extrapolation needed and the less reliable the forecast are likely to be.

It is obvious that the forecasts obtained through the graphical method suffer from an element of subjectivity in the extrapolation of the curve. For example, if the above graph were extrapolated to $AB_2$ or $AB_3$, the forecasts would have been different. To minimize this error, efforts would have to be made to collect data as up to date as possible so as to minimize the extrapolation period. However, since historical data of no variable when plotted, usually lie on any smooth curve, extrapolation would never be unique and thus the method would always suffer from subjectivity.

### Trend Method

Under this method, extrapolation of historical data is attempted through estimation of alternative trend equations. A trend equation is one in which the variable under forecast is made simply as a function of time:

$$D_g = f(T) \tag{4.24}$$

where $D_g$ = demand for groundnut oil (forecast variable),

$T$ = time.

Time variable is measured in any one of the three ways. For example, if the historical data were annual for the period 1951–52 through 1977–78 ($n$ =27), then $T$ could be quantified as 0, 1, 2, … 26; 1,

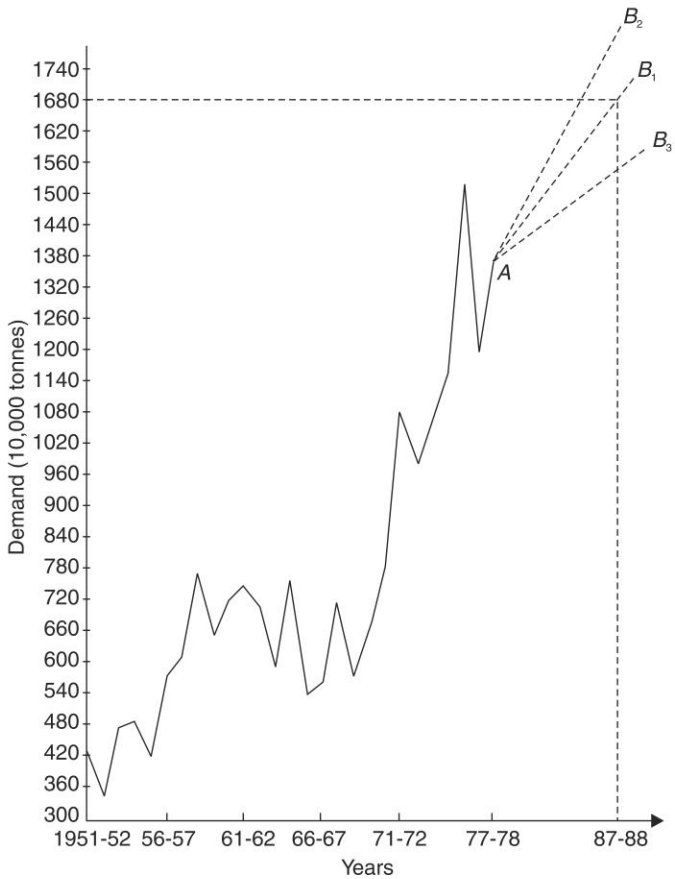



**Fig. 4.2** *Demand for groundnut oil over time*

2, 3, … 27; or –13, –12, …, –1, 0, 1, 2, … 13, where the year 1951–52 is normalized to take a value of 0, 1 or –13, respectively. The middle approach is the most popular and that is the one followed in the text.

As seen above, the functional form of Eq. 4.24 has to be specified before it can be estimated. There are several alternative specifications for the trend equation, but the most popular ones are: linear and semi-log:

$$D_g = a + bT \tag{4.25}$$

$$\log_e D_g = A + BT \tag{4.26}$$

Readers with some mathematical background would easily understand that while in Eq. 4.25, the slope (= $b$) indicates the absolute change in the dependent variable over time, that in Eq. 4.25 (= $B$) indicates the rate of growth in the said variable over time. A careful researcher would estimate both the forms (in fact, even other functional forms) of the function through the least-squares method of estimation (See section 4.2: Regression method) and choose the one which provides the best fit as judged by the value of the co-efficient of determination ($R^2$). The higher the value of $R^2$, the better is the trend equation. A low value of $R^2$ would mean poor or no specific trend. The best fit thus obtained for the groundnut oil demand was the following:

$$\log_e D_g = 6.0 + 0.0406\ T \qquad (4.27)$$
$$R^2 = 0.74$$

Equation (4.27) reveals that 74 percent of the variation in groundnut oil demand is explained by the trend variable alone. In this function, time variable is serving as the "proxy" for all the determinants of tea demand in India.

Once the appropriate trend equation has been estimated, derivation of forecasts involves simple arithmetic. The appropriate value for time variable is first found and then fed in the estimated trend equation, to get the forecast. For example, demand forecast for groundnut oil in 1978-79 would be obtained as ($T = 28$).

$$\log_e D_g = 6.0 + 0.0406\ T \qquad (4.27)$$
$$\log_e D_g = 6.0 + 0.0406\ (28)$$
$$= 7.1368$$

or, $$D_g = \text{anti} \log_e 7.1368$$
$$= 1257$$

Similarly for the year 1987–88 the requisite forecast would be given by ($T = 37$).

$$\log_e D_g = 6.0 + 0.0406\ (37)$$
$$= 7.5022$$

or, $$D_g = \text{anti} \log_e 7.5022$$
$$= 1812$$

For the year 2011–12, the trend forecast for groundnut oil demand in India would be

$$\log D_g = 6.0 + 0.406\ (61)$$
$$= 30.766$$

or, $$D_g = \text{anti} \log_e 30.766$$
$$= \text{Very large number}$$

It must be emphasized here that all post sample forecasts are "**genuine**" or "ex anti" forecasts. Thus, in the above example, the sample period is 1951–52 through 1977–78, and accordingly the forecasts for all periods beyond 1977–78 are the genuine ones. As would be clear by now, the trend method is easy to apply. In fact, to its credit, it is said "wonder goes where the knowledge fails". This is because in some cases the method yields forecasts which turn out to be better than the ones obtained through methods which are highly sophisticated. It is clear that the trend method is based on the premise that "history repeats itself" and if that is true, it does wonders. Thus, in cases where the future is not going to be significantly different from the average of the past, that is, like the forecasts of most aggregative variables and demand for stabilized goods and services, the trend method is quite useful. The trend method is excellent in forecasting India's population, demand for bank deposits, textiles, cement, paper, steel, etc.

The trend method is basically an objective method. However, if experiments with alternative formulations of the trend equation are not made, two researchers could come out with different forecasts through the use of the same method.

Some variables are seasonal and thus their magnitudes tend to fluctuate significantly over seasons. For such variables, the trend method would be highly inappropriate. In such cases, the alternative procedure is the **Decomposition of Time Series Method**. Under this, the time series is classified into four components, viz.

- Trend
- Seasonal
- Cyclical
- Random

The trend in a variable is caused by steady (long-term) changes over time through changes in population, prices, productivity, growth, etc. Seasonal variations are fluctuations over time caused by the differences in the length of the season (e.g. January has 31 days while February has 28 or 29 days), weather (monsoon, winter and summer), festivals (busy and slack seasons), etc. Cyclical variations over time arise due to business cycles (prosperity, peak, recession and trough), and the random changes are caused by random events, like revolutionary changes in political regime/system and/or economic policies, wars, epidemics, earthquakes, accidents, fires, and so on. Statisticians have designed methods to separate these four components, and there by contributed this technique of forecasting. Obviously, annual series are devoid of seasonal fluctuations. Busy-slack season data have two seasons, quarterly series have four seasons, monthly have twelve seasons, and so on. This is not a place to go into a detailed discussion of this method of forecasting. However, a brief discussion is in order.

The trend component is identified through fitting the trend equation as outlined above under the trend method. The seasonal component is obtained through four steps. One, take a simple moving average (as explained just after this) of the order of the season, i.e., of order four for a quarterly series, and so on. If the order of the moving average is even, then take the centered moving average and place the data against the middle of the season. Two, divide the original series by the centered moving average series and obtain the seasonal series. Three, arrange the series by each season (quarter) and compute the average values for each season. Find the overall average value per period (quarter) in the season series. Divide each of the season average by the overall average value and obtain the seasonal indices. Four, multiplication of corresponding seasonal indices and the actual (original) data would give the seasonal series. Once the trend and seasonal series are obtained, the residual procedure is used to de rive the combined cyclical and random series. The later series is further bifurcated using the simple smoothing whose order would be hard to decide but could be estimated through leading indicators.

Under the decomposition method, each of the four components are separately forecasted and then added to generate the forecasts for the combined series. Since separation of cyclical and random series is somewhat arbitrary, the method is useful basically for variables which have significant seasonal fluctuations. For others, the trend method will suffice. For details on the method, readers may refer to Gupta (2008), and Hanke and Wichen (2009).

## Smoothing Methods

If the variable under forecast does not follow any significant trend (upward or downward), the trend method is inappropriate. In such cases, the smoothing method could be more useful. The smoothing methods are useful even if the data contain a trend and/or seasonal fluctuations. There are two versions

of the smoothing method: *simple smoothing* (averaging) and *weighted smoothing*. In the former, a simple average of the specific number of observations (called the 'order') is taken, while in the latter, a weighted average is taken out. Since the recent observations are more relevant than the older ones for estimating the future, the weighted smoothing is preferred to the simple smoothing and the weights are assigned in a descending order as one goes from the current observations to the past ones. Further, since it is hard to decide a priori weights, they are often assumed to follow geometrical progression, i.e. like $a$, $a(1-a)$, $a(1-a)^2$, … where $a$ is the weight attached to the most current observation. $a(1-a)$ to the one period back observation, $a(1-a)^2$ to the two period back observation, and so on. Incidentally, note that the sum of all these weights equals unity and the value of $a$ lies between zero and one. When the weights are assigned this way, the weighted smoothing method becomes the geometrical or ***exponential smoothing method***.

Smoothing of a time series could be performed just once or it could be repeated. Thus there are single simple, double simple, single exponential, double exponential and triple exponential smoothing methods in literature. Under all these smoothing methods, forecasts are obtained through two steps:

(a) obtain the specific smoothen series from the observed time series; and
(b) obtain the desired forecasts from the specific smoothen series.

The specific formula for deriving various smoothen series are the following:

$$SS_t = \frac{1}{k}(Y_t + Y_{t-1} + \ldots + Y_{t-k+1}) \tag{4.28}$$

$$DS_t = \frac{1}{k}(SS_t + SS_{t-1} + \ldots + SS_{t-k+1}) \tag{4.29}$$

$$SE_t = a\,Y_t + a(1-a)\,Y_{t-1} + a(1-a)^2\,Y_{t-2} + \ldots \text{ (upto } \infty)$$

$$= a\,Y_t + (1-a)\,SE_{t-1} \tag{4.30}$$

$$DE_t = a\,SE_t + a(1-a)\,SE_{t-1} + a(1-a)^2\,SE_{t-2} + \ldots \text{ (upto } \infty)$$

$$= a\,SE_t + (1-a)\,De_{t-1} \tag{4.31}$$

$$TE_t = a\,DE_t + a\,(1 - - - a)\,DE_{t-1} + a\,(1-a)^2\,Se_{t-2} + \ldots \text{(upto } \infty)$$

$$= a\,DE_t + (1-a)\,TE_{t-1} \tag{4.32}$$

where, $SS_t$ = single simple smoothing value in period $t$
$DS_t$ = double simple smoothing value in period $t$
$SE_t$ = single exponential smoothing value in period $t$
$DE_t$ = double exponential smoothing value in period $t$
$TE_t$ = triple exponential smoothing value in period $t$
$k$ = simple smoothing (moving average) order
$a$ = exponential smoothing parameter
$Y_t$ = observed value of the variable under smoothing in period $t$

The second parts of the last three equations are obtained through algebraical manipulation.

The application of these formulae requires the values for the order of the moving averages ($k$) and the exponential smoothing parameter ($a$). The choice of each of $k$ and $a$ depends on the time path of the time series. A large $k$ and a small $a$ should be used when there is a lot of randomness in the data, i.e., when time series are relatively stable. Conversely, if the series changes rapidly, a small $k$ and a large $a$ are desirable. In practice, values of $k$ from 3 to 6 and of $a$ from 0.1 to 0.4 usually work best. The procedure to choose their values is to simulate the historical data set using alternative values for $k$ and $a$. The value which yields the best forecasts becomes the most appropriate value to be employed in the forecasting exercise. The relationship between $k$ and $a$ is approximately described by the following expression:

$$a = \frac{2}{k+1}$$

The calculations of the various smoothing series for the chosen values of $k = 3$ and $a = 0.4$ are illustrated in Table 4.4. To elaborate these calculations details on a few are provided below.

$$SS_{53-54} = \frac{1}{3}(438 + 345 + 481) = 421$$

$$SS_{77-78} = \frac{1}{3}(1538 + 1199 + 1386) = 1374$$

$$DS_{55-56} = \frac{1}{3}(421 + 441 + 466) = 443$$

$$DS_{77-78} = \frac{1}{3}(1257 + 1300 + 1374) = 1310$$

$$SE_{52-53} = 0.4\,(345) + (1 - 0.4)\,(392) = 373$$

$$SE_{77-78} = 0.4\,(1386) + (1 - 0.4)\,(1227) = 1291$$

$$DE_{52-53} = 0.4\,(373) + (1 - 0.4)\,(392) = 384$$

$$DE_{77-78} = 0.4\,(1291) + (1 - 0.4)\,(1130) = 1194$$

$$TE_{52-53} = 0.4\,(384) + (1 - 0.4)\,(392) = 389$$

$$E_{77-78} = 0.4\,(1194) + (1 - 0.4)\,(1015) = 1087$$

Incidentally, note that one loses certain observations under the simple smoothing methods. In particular, under the single smoothing the loss of observations equals $k - 1$ and that under double simple smoothing equals $2\,(k - 1)$.

To generate demand forecasts from various smoothing methods, the following formulas would have to be applied.

**Table 4.4** *Actual and Smoothen Time Series on Demand for Groundnut Oil ('000 tonnes)*

| Year | Actual Values | Simple: Single | Simple: Double | Exponential: Single | Exponential: Double | Exponential: Triple |
|---|---|---|---|---|---|---|
| | | Smoothen Values | | | | |
| (l) | (2) | (3) | (4) | (5) | (6) | (7) |
| 1951–52 | 438 | — | — | 392* | 392* | 392* |
| 1952–53 | 345 | — | — | 373 | 384 | 389 |
| 1953–54 | 481 | 421 | — | 416 | 397 | 392 |
| 1954–55 | 498 | 441 | — | 449 | 418 | 402 |
| 1955–56 | 419 | 466 | 443 | 437 | 496 | 440 |
| 1956–57 | 578 | 498 | 468 | 493 | 495 | 462 |
| 1957–58 | 627 | 541 | 502 | 547 | 516 | 484 |
| 1958–59 | 782 | 662 | 567 | 641 | 566 | 517 |
| 1959–60 | 649 | 686 | 630 | 644 | 597 | 549 |
| 1960–61 | 726 | 719 | 689 | 677 | 629 | 581 |
| 1961–62 | 753 | 709 | 705 | 707 | 660 | 613 |
| 1962–63 | 711 | 730 | 719 | 709 | 680 | 640 |
| 1963–64 | 586 | 683 | 707 | 660 | 672 | 653 |
| 1964–65 | 764 | 687 | 700 | 702 | 684 | 665 |
| 1965–66 | 546 | 632 | 667 | 640 | 666 | 665 |
| 1966–67 | 568 | 626 | 648 | 611 | 644 | 657 |
| 1967–68 | 739 | 618 | 625 | 662 | 651 | 655 |
| 1968–69 | 577 | 628 | 624 | 628 | 642 | 650 |
| 1969–70 | 662 | 659 | 635 | 642 | 642 | 647 |
| 1970–71 | 788 | 676 | 654 | 700 | 665 | 654 |
| 1971–72 | 1097 | 849 | 728 | 859 | 743 | 690 |
| 1972–73 | 984 | 956 | 827 | 909 | 809 | 738 |
| 1973–74 | 1068 | 1050 | 952 | 973 | 875 | 793 |
| 1974–75 | 1164 | 1072 | 1026 | 1049 | 945 | 854 |
| 1975–76 | 1538 | 1257 | 1126 | 1245 | 1065 | 938 |
| 1976–77 | 1199 | 1300 | 1210 | 1227 | 1130 | 1015 |
| 1977–78 | 1386 | 1374 | 1310 | 1291 | 1194 | 1087 |

*The first number in exponential smoothing has to be computed through some approximate method. In this case these were computed simply as the simple average of the first two figures of actual demands i.e. $\frac{1}{2}$ (438 + 345) = 392.

$$F_{t+n}(SS) = \frac{1}{k}\left(Y_{t+n-1} + Y_{t+n-2} + \ldots + Y_{t+n-k}\right) \tag{4.33}$$

$$F_{t+n}(DS) = (2\,SS_t - DS_t) + \frac{2n}{k-1}\left(SS_t - DS_t\right) \tag{4.34}$$

$$F_{t+n}(SE) = a\,Y_{t+n-1} + (1-a)\,SE_{t+n-2} \tag{4.35}$$

$$F_{t+n}(DE) = (2\,SE_t - DE_t) + \frac{a.n}{1-a}(SE_t - DE_t) \tag{4.36}$$

$$F_{t+n}(TE) = (3\,SE_t - 3DE_t + TE_t) + \frac{a.n}{2(1-a)^2}[(6-5a)\,SE_t$$

$$-(10-8a)\,DE_t + (4-3a)\,TE_t] + \frac{a^2n^2}{2(1-a)^2}(SE_t - 2DE_t + TE_t) \tag{4.37}$$

where $F_{t+n}(SS)$ stands for the forecast values of the variable $Y$ in period $t+n$ under the single simple smoothing method, $F_{t+n}(DS)$ that under the double simple smoothing method, and so on, $t$ for the present period and $n$ for the number of time periods from the present period.

The single simple and single exponential smoothing methods are appropriate for long-term ($n > 1$) forecasting only if the historical data are stationary, i.e., the time series contains no trend or that it oscillates around a constant mean. Consequently, if any of these two methods is employed to a non-stationary series, the single smoothing values will lag behind the actual data, if the trend is positive and ahead of the actual data if the trend is negative. Since the groundnut demand in India contains the trend, these two methods are not appropriate for the forecasts beyond a year. The other three methods do not suffer from this limitation.

Using the above formulae, and the data and results of Table 4.4, one period forecasts from the single smoothing methods and multi-period forecasts from the double and triple smoothing methods were obtained. The result are provided in Table 4.5. It may be noted that under the single smoothing methods, the smoothen value for period itself is the forecast for the period $t+1$. To further explain calculations of forecasts under the other three methods, details on a few computations are provided below.

$$F_{56\text{-}57}(DS) = [(2)(466) - (443)] + \frac{(2)(1)}{3-1}[466 - 443] = 512$$

$$F_{80\text{-}81}(DS) = [(2)(1374) - (1310)] + \frac{(2)(3)}{3-1}[1374 - 1310] = 1630$$

$$\mathrm{F}_{53\text{-}54}(DE) = [(2)(373) - (384)] + \frac{(0.4)(1)}{1-0.4}[373 - 384] = 355$$

$$F_{1980\text{-}81}(DE) = [(2)(1291) - (1194)] + \frac{(0.4)(1)}{(1-0.4)}[1291 - 1194] = 1583$$

$$\mathrm{F}_{53\text{-}54}(TE) = [(3)(373) - 3(384) + (389)] + \frac{(0.4)(1)}{2(1-0.4)^2}[(6 - 5(0.4)(373)$$

$$-(10 - 8(0.4))\,384 + (4 - 3(0.4))\,389)] + \frac{(0.4)^2(1)^2}{2(1-0.4)^2}[373 - 2(384) + 389] = 338$$

Thus, the forecasts for groundnut oil demand for the year 1987–88 stands at 2078 thousand tonnes by the double simple smoothing method, and at 2038 and 1668 thousand tonnes by the double and triple exponential smoothing methods, respectively. Which of these three forecasts is reliable? Obviously none in this case, for the gap between last year where actual data have been used (1977–78) and the year for which forecast is made (1987–88) is ten years, which is too long. These long-term forecasts have been

provided in the table basically to explain the methodology. A true researcher would use up to date data for generating forecasts. Of the three methods' forecasts, to decide which one is better depends on the kind of trend present in the actual series. If the series contains the linear trend ($Y = a + b\,T$) either of the double smoothing methods will be good. The triple exponential smoothing method is appropriate when the series is subject to the quadratic trend ($Y = a + b\,T + c\,T^2$).

**Table 4.5** *Forecasts for Groundnut Oil Demand by Various Smoothing Methods*

| Year | Actual Values | Forecasts Through | | | | |
|---|---|---|---|---|---|---|
| | | Single | Double | Single | Double | Triple |
| | | Simple Smoothing Method | | Exponential Smoothing Method | | |
| (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| *(i) Sample Period* | | | | | | |
| 1951–52 | 438 | — | — | — | — | — |
| 1952–53 | 345 | — | — | 392 | 392 | 392 |
| 1953–54 | 480 | — | — | 373 | 355 | 338 |
| 1954–55 | 498 | 421 | — | 416 | 448 | 487 |
| 1955–56 | 419 | 441 | — | 449 | 501 | 543 |
| 1956–57 | 578 | 466 | 512 | 437 | 339 | 17 |
| 1957–58 | 627 | 498 | 557 | 493 | 490 | 400 |
| 1958–59 | 782 | 541 | 620 | 547 | 599 | 596 |
| 1959–60 | 649 | 662 | 852 | 641 | 766 | 827 |
| 1960–61 | 726 | 686 | 799 | 644 | 722 | 720 |
| 1961–62 | 753 | 719 | 779 | 677 | 757 | 757 |
| 1962–63 | 711 | 709 | 718 | 707 | 785 | 786 |
| 1963–64 | 586 | 730 | 751 | 709 | 757 | 727 |
| 1964–65 | 764 | 683 | 634 | 660 | 640 | 537 |
| 1965–66 | 546 | 687 | 661 | 702 | 732 | 720 |
| 1966–67 | 568 | 632 | 561 | 640 | 597 | 521 |
| 1967–68 | 739 | 626 | 581 | 611 | 557 | 500 |
| 1968–69 | 577 | 618 | 603 | 662 | 680 | 723 |
| 1969–70 | 662 | 628 | 636 | 628 | 605 | 586 |
| 1970–71 | 788 | 659 | 707 | 642 | 642 | 656 |
| 1971–72 | 1097 | 676 | 719 | 700 | 758 | 826 |
| 1972–73 | 984 | 849 | 1091 | 859 | 1052 | 1229 |
| 1973–74 | 1068 | 956 | 1214 | 909 | 1016 | 1158 |
| 1974–75 | 1164 | 1050 | 1247 | 973 | 1136 | 1181 |
| 1975–76 | 1538 | 1072 | 1164 | 1049 | 1222 | 1259 |
| 1976–77 | 1199 | 1257 | 1519 | 1245 | 1545 | 1694 |
| 1977–78 | 1386 | 1300 | 1481 | 1227 | 1389 | 1335 |

**Table 4.5** *Continued*

| (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|
| *(ii) Post Sample Period* | | | | | | |
| 1978–79 | | 1374 | 1502 | 1291 | 1453 | 1425 |
| 1979–80 | | | 1566 | | 1518 | 1468 |
| 1980–81 | | | 1630 | | 1583 | 1507 |
| 1981–82 | | | 1694 | | 1648 | 1542 |
| 1982–83 | | | 1758 | | 1713 | 1573 |
| 1983–84 | | | 1822 | | 1778 | 1600 |
| 1984–85 | | | 1886 | | 1843 | 1623 |
| 1985–86 | | | 1950 | | 1908 | 1642 |
| 1986–87 | | | 2014 | | 1973 | 1657 |
| 1987–88 | | | 2078 | | 2038 | 1668 |

### Auto-Regressive Method

Under this method forecasts are obtained through estimation of an equation which expresses the forecast variable as a function of its own lagged value (s):

$$Y_t = f(Y_{t-1}, Y_{t-2}, \ldots\ldots Y_{t-s}) \tag{4.38}$$

Alternative values for $s$ and alternative functional forms of this function have been tried to obtain the best possible fit (as judged by value of $R^2$) and reasonable degrees of freedom, which equals the number of observations (= sample size) less the number of parameters to be estimated. Forecasts are then obtained through simply feeding the past values of the variable under forecast in the estimated equation. The method is thus purely a statistical one.

### ARIMA Method

The auto-regressive integrated moving average (ARIMA) method has been given by Box and Jenkin and thus, in their honour, it is also called the **Box-Jenkin method**. The method combines the moving average (smoothing) and auto-regressive techniques, and thus is considered as the most sophisticated statistical method of forecasting. Since it involves complicated mathematics, a discussion on this is left out of this text. Computer centres of relevant institutions do maintain ready packages for using this technique and interested readers could use them. To give a brief idea, the forecasts under this method are obtained under three steps, viz.

(a) Original time series (variable $Y$) is first transformed into a stationary series (the one which fluctuates around its mean value, with a constant mean and variance) through one or more first differences or logarithms, depending on the data pattern over time (call it variable $Z$).

(b) The so obtained series ($Z$) is used to forecast its own values, within and outside the sample period. This is done through an appropriate integration of auto regressive and moving average (ARIMA) method. Since the method is non-linear, a specially designed technique is applied to estimate the parameters of the ARIMA model. The estimated model is then applied to obtain the forecasts for the transformed variable $Z$.

(c) The transformation of step (a) is subsequently reversed to derive the forecasts for the original variable ($Y$).

## Leading Indicator Method

The method involves three steps

(a) Identification of the leading indicator for the variable under forecasting.

(b) Estimation of the relationship between the variable under forecasting and its leading indicator.

(c) Derivation of forecasts.

In order to identify the leading indicator, it is imperative to understand the three kinds of time series: leading, coincident and lagging. The leading series is data on the variable which move up or down ahead of some other series, coincident series moves along with some other series, and the lagging series moves up or down behind some other series. For example, the Bank rate (the rate at which the Reserve Bank of India lends to its member banks) is the leading interest rate; the rates at which commercial banks accept deposits from and lend to the private sector are more or less the coincident series with regard to the Bank rate; and the rate at which private money lenders accept deposits and lend to individuals is lagging series with reference to both the Bank rate and commercial banks' deposit and lending rates. Similarly, the construction plans of buildings leads the demand for fans, requests for loans from financial institutions lead the actual capital expenditures, children's birth lead the admission in kindergarten, number of students in schools leads admissions in colleges, and so on. It is not that for every variable, there is a leading variable but for some they do exist. Thus, through this kind of search, one may be able to find an appropriate leading variable for the variable under forecast. If no such variable is available, this method of forecasting is also not available. For groundnut oil demand, perhaps no such lead variable is available and so it is not possible to explain this method through this case. Illustrating through a general example, let $X$ be the leading indicator for $Y$, whose values are to be forecast.

The second step involves the estimation of the relationship between $Y$ and $X$:

$$Y_t = f(X_{t-s}) \tag{4.39}$$

where, $s$ = length of the lag.

The Least-squares method of estimation could be used to estimate Eq. (4.39) for its various functional forms (linear, double-log, semi-log, quadratic, etc.) and for various values of $s$. The one which yields the highest value for $R^2$ could then be selected to represent the relationship between the two variables. Thus, suppose the best form were the following.

$$y_t = 150 + 2.5\, x_{t-3}$$

The forecast then would be easy to derive:

$$y_{t+1} = 150 + 2.5\, x_{t-2}$$
$$y_{t+2} = 150 + 2.5\, x_{t-1}$$
$$y_{t+3} = 150 + 2.5\, x_t$$

If the data is annual, the lag is of three years, the forecasts can be obtained for the next three years. In period $t$, data on $x_{t-2}$, $x_{t-1}$ and $x_t$ are available, which could be plugged in the above equations, and thus derive the desired forecasts. Forecasts for periods which are ahead by more than the lag period can just not be obtained through this method.

The leading indicator method of forecasting is also called the *Barometric method*, where the leading indicator is taken as the barometer. A slight variant of this method is the forecasting through anticipatory data. Rangarajan (1981) has used the latter method in forecasting investment in the corporate sector

through informations on capital consents, loan requests pending with the financial institutions, and the past record on loan requests and loan disbursements.

As hitherto seen, the method is available for short-term forecasts only. Further, there is a dearth of the lead-lag relationships. For these reasons, its use has been quite limited.

## Regression Method

The most popular method of forecasting among economists is perhaps the regression method. The method employs both the principles of economic theory and appropriate statistical methods of estimation in forecasting demand or any other variable. It requires historical data (time series and/or cross section) on the variable under forecasting and its determinants.

Under this method, forecasts are obtained through six steps. The first four are the same as discussed above under the regression method of demand estimation (See section 4.1) viz., identification of causal variables, collection of historical data on the variable under forecasting and on its determinants, selection of appropriate functional form for the demand function and estimation of the function; the details on how to carry on these steps could be read in the section just referred to. The fifth step consists of derivation of forecasts for the independent (causal) variables in the function. This is a difficult step, for it means one needs forecasts on some variables in order to be able to obtain forecasts on one (dependent or effect) variable. Hopefully, the former should be an easier task than the latter lest it be tantamount to aggravating the problem rather than solving it. Generally speaking, the cause variables are either aggregative in nature such as national income and general prices, on which forecasts may be available from organizations like the Planning Commission and/or the Central bank of the country (RBI), or the variable which are relatively easy to forecast through simple methods like the trend method. The last step in the regression method comprises of the derivation of forecasts for the variable under forecasting. This is accomplished through simple arithmetic. To illustrate the application of the regression method, demand forecasts for groundnut oil are derived in what follows.

Recall that Eq. (4.3.5) (See Table 4.3) was considered as the best estimated equation to represent the estimated demand function for groundnut oil in India. For convenience, the equation is recalled here:

$$\log D_g = -12.4 + 1.78 \log Y - 1.22 \log P_o + 2.20 P_v + 0.80 \log P_g, - 1.62 \log P_e$$

$$R^2 = 0.90 \qquad (4.3.5)$$

The equation has a high value of $R^2$ (0.90) and thus it is quite suitable for forecasting. It contains five independent variables, whose forecasts have to be first obtained. This is done through fitting alternative trend equations for each of these variables. The best for each, decided on the basis of the highest value of $R^2$, are given below in Table 4.6.

**Table 4.6** *Selected Trend Equations for Independent Variables*

| | *Independent Variable* | *Equation* | $R^2$ |
|---|---|---|---|
| (*i*.) | National Income ($Y$) | $\log_e Y = 9.731 + 0.0339T$ | 0.99 |
| (*ii*.) | Groundnut oil price ($P_o$) | $\log_e P_o = 3.018 + 0.0753T$ | 0.88 |
| (*iii*.) | Vanaspati Price ($P_v$) | $\log_e P_v = 3.262 + 0.0660T$ | 0.92 |
| (*iv*.) | Ghee Price ($P_g$) | $\log_e P_g, = 3.374 + 0.0615T$ | 0.93 |
| (*v*.) | Egg, fish & meat price ($P_e$) | $\log_e P_e = 3.163 + 0.0760T$ | 0.95 |

Sample period : 1951–52 to 1977–78 ($n = 27$)

$$T = 1, 2, \ldots, 27$$

By feeding the appropriate values for $T$ in the trend equations of Table 4.6 ($T = 28, 29, \ldots, 37$) forecasts for the five independent variables were derived for the period 1978–79 through 1987–88, and the same are reported in Table 4.7.

**Table 4.7** *Forecasts for the Independent Variables*

| *Prediction Period* | *National income (Y) (Rs. crores)* | *Forecasts of Price of* | | | |
|---|---|---|---|---|---|
| | | *Groundnut Oil ($P_o$)* | *Vanaspati ($P_v$)* | *Ghee ($P_g$)* | *egg, fish and meat ($P_e$)* |
| | | *(Wholesale Price Index)* | | | |
| 1978–79 | 43486 | 168.4 | 165.7 | 163.4 | 198.5 |
| 1979–80 | 44986 | 181.6 | 171.0 | 173.7 | 214.2 |
| 1980–81 | 46537 | 195.8 | 189.0 | 184.7 | 231.1 |
| 1981–82 | 48142 | 211.1 | 201.9 | 196.5 | 249.4 |
| 1982–83 | 49801 | 227.6 | 215.7 | 208.9 | 269.1 |
| 1983–84 | 51519 | 245.4 | 230.4 | 222.2 | 290.3 |
| 1984–85 | 53295 | 264.6 | 246.2 | 236.3 | 313.2 |
| 1985–86 | 55133 | 285.3 | 263.0 | 251.3 | 338.0 |
| 1986–87 | 57034 | 307.6 | 280.9 | 267.2 | 364.7 |
| 1987–88 | 59000 | 331.7 | 300.1 | 284.1 | 393.5 |

The forecasts for the groundnut oil demand in India for a particular year are then obtained simply by feeding the forecasts of the independent variables for that particular year in the selected estimated equation for the function (4.3.5). To illustrate the calculations, forecast for 1978–79 were thus obtained as

$$(\log_e D_g)\ 78\text{–}79 = -12.4 + 1.78 \log_e (43486) - 1.22 \log_e (168.4) + 2.20 \log_e (165.7) + 0.80 \log_e (163.4) - 1.62 \log_e (198.5)$$

$$= -12.4 + 1.78\ (10.68) - 1.22\ (5.13) + 2.20\ (5.11) + 0.80\ (5.10) - 1.62\ (5.29)$$

$$= 7.116$$

or, $$(D_g)\ 78 - 79 = \text{antilog } 7.116$$

$$= 1232$$

The so obtained forecasts for various years are given in Table 4.8.

**Table 4.8** *Demand Forecasts for Groundnut Oil by the Regressing Method (000' tonnes)*

| *Prediction period* | *Demand forecasts* |
|---|---|
| 1978–79 | 1232 |
| 1979–80 | 1281 |
| 1980–81 | 1331 |
| 1981–82 | 1385 |
| 1982–83 | 1442 |
| 1983–84 | 1500 |
| 1984–85 | 1562 |
| 1985–86 | 1625 |
| 1986–87 | 1690 |
| 1987–88 | 1757 |

Thus, the forecast for groundnut oil demand in India for the year 1987–88 by the regression method stands at 1757 thousand tonnes. Since the above forecasts are for the post sample period, these are genuine forecasts.

The principal factor behind the regression methods popularity is that it is not only prescriptive as the other methods discussed above are, but it is also descriptive/analytical. In other words, the method besides giving forecasts also explains as to why the forecast is what it is. The explanation is provided through the regression equation, which as seen in section 4.1 (regression method) above, gives the list of causal variables and various demand elasticities/multipliers. For example, the demand forecast for 1987–88 stands at 1757 vis-a-vis the actual demand of 1386 in 1977–78, because of the following two forces:

(a) Groundnut Oil demand varies with variation in five variables in a specific way (vide equation 4.3.5).

(b) Each of the five variables have undergone a certain change (See Table 4.6).

The other points in favour of the regression method are that it is neither a very mechanical method as the extrapolation methods are, nor it is basically a subjective method as the survey methods are. Though there is a possibility of two researchers choosing two different equations for forecasting, the difference will not be significant, assuming both the researchers have good mastery over the subject.

The major limitation of this method is that, as hitherto mentioned, it requires the use of some other forecasting method to estimate the values of the explanatory variables in the forecast period. To the extent forecasts of the values of explanatory variables are wrong, the forecasts based on this method would be wrong. To minimize the dangers of such errors, researchers often consider all possible scenarios with regard to the determinants' values in the forecast periods, and provide alternative forecasts depending on all likely scenarios expected in future. Thus, if income is expected to increase either at 7 per cent or at 8 per cent, ceteris paribus, the researcher uses each of these values of future income levels and generate two sets of forecasts, one for each value of income, *ceteris paribus*. In addition to the above limitations, as is true in cases of all statistical methods, the regression method forecasts on the basis of the past average relationships. Thus, to the extent the future relationship deviates from the past average, in terms of the parameters' estimates (regression coefficients) and the error term (which is zero on average), the

forecasts would be wrong. The latter errors could be minimized by a careful researcher by using sound economic theory, and accurate and up to date historical data.

## Simultaneous Equations Method

The simultaneous equations method, also called the complete system approach to forecasting, is the most sophisticated econometric method of forecasting. Since it involves complicated mathematical and statistical tools, its detail discussion is beyond the scope of this text. Suffice it to say here, it involves specification of a number of economic relationships, one for each behavioural variable, estimation and solution of which yield the forecasting equations similar to the estimated regression equation presented above. The equation here would differ from the regression equation talked above in one special way, i.e. this one would contain the variable under forecasting and a few other variables, all of which would be truly exogenous (policy or non-policy) variables and thus they would pose little problem in forecasting their own values. Thus the simultaneous equations method reduces the major problem of the regression method, viz., forecasts for the independent variables.

## Evaluation of Forecasts

Since no forecast is expected to prove absolutely correct, researchers would be interested in evaluating the accuracy of their forecasts. The ex-post (sample period) forecasts can be evaluated right when forecasts are made, while evaluation of ex-ante (genuine) forecasts has to wait for the forecast period to come.

There are two approaches to such an evaluation. Under the one, forecasting error for each period is carefully examined, while under the other, various errors for the entire period are summarized in one surrogative measure. In order to explain these approaches for the regression method, the true values of groundnut oil demand during the sample period and the forecast (ex-post) values of this variable through the regression method are provided in Table 4.9. The last column of this table contains the difference between the two, called the (forecasting) error, expressed as a percentage of the corresponding true demand. Thus, the magnitude of error varies from +18 per cent to –40 per cent. Alternatively, these errors could be expressed in terms of geometry. One could draw a four quadrant graph, *X*-axis of which would measure the true change in the variable under forecasting and *Y*-axis the forecast change in the said variable. If the various points of the graph lie on a 45° line, then there is no error in forecasting. The farther away the points are from this line, the greater is the error. Further, if the points lie either in quadrant I or III, there is no error in forecasting the direction of change. However, if they fall either in quadrant II or IV, there will be an error even in forecasting the direction of change (See Fig. 4.3).

The most popular surrogative measure, which combines all errors into one measure, is called the ***mean absolute percentage error*** **(MAPE)**. It is defined as

$$\text{MAPE} = \frac{1}{n}\left[\sum \frac{\left|Y_t - \hat{Y}_t\right|}{Y_t}\right]100 \tag{4.40}$$

where, $n$ = sample size

$Y_t$ = true value of $Y$ in period $t$

$\hat{Y}_t$ = forecast value of $Y$ in period $t$

**Table 4.9** *Forecasts of Groundnut Oil Demand by Regression Method ('000 tornnes)*

| *Period* | *Actual demand* | *Estimated/forecasted demand* | *% Error* |
|---|---|---|---|
| 1951–52 | 438 | 435 | 1 |
| 1952–53 | 345 | 382 | –11 |
| 1953–54 | 481 | 444 | 8 |
| 1954–55 | 498 | 466 | 6 |
| 1955–56 | 419 | 425 | –1 |
| 1956–57 | 578 | 564 | 2 |
| 1957–58 | 627 | 642 | –2 |
| 1958–59 | 782 | 638 | 18 |
| 1959–60 | 649 | 561 | –2 |
| 1960–61 | 726 | 718 | 1 |
| 1961–62 | 753 | 744 | 1 |
| 1962–63 | 711 | 695 | 2 |
| 1963–64 | 586 | 674 | –15 |
| 1964–65 | 714 | 750 | 2 |
| 1965–66 | 546 | 526 | 4 |
| 1966–67 | 568 | 652 | –15 |
| 1967–68 | 739 | 667 | 10 |
| 1968–69 | 577 | 682 | –18 |
| 1969–70 | 662 | 859 | –30 |
| 1970–71 | 788 | 1022 | –40 |
| 1971–72 | 1097 | 981 | 11 |
| 1972–73 | 1097984 | 827 | 16 |
| 1973–74 | 1068 | 897 | 16 |
| 1974–75 | 1164 | 1139 | 2 |
| 1975–76 | 1538 | 1502 | 2 |
| 1976–77 | 1199 | 1266 | –6 |
| 1977–78 | 1386 | 1256 | 9 |
| Mean absolute % error | | | 9.3 |

This measure has been computed for all the statistical methods for the case of groundnut oil demand in India. The computed values are provided in Table 3.10.

**Table 4.10** *MAPE of the Groundnut Oil Demand Forecasts by Different Methods*

| | *Method* | *MAPE* |
|---|---|---|
| (*a.*) | Trend (Mathematical extrapolation) | 15.9 |
| (*b.*) | Smoothing | |
| | (i) Double simple | 14.1 |
| | (ii) Double exponential | 16.5 |
| | (iii) Triple exponential | 22.2 |
| (*c.*) | Regression | 9.3 |

Thus, the regression method performs the best on this criterion.

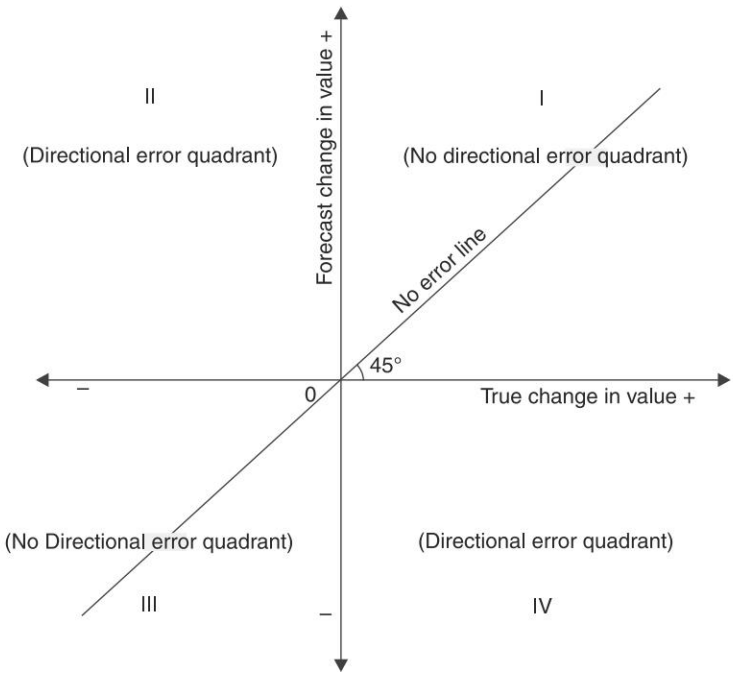



**Fig. 4.3** *Evaluation of forecasts*

The foregoing paragraphs provided the test procedure for ex-post or within the sample period forecasts. In a way, it is a testing procedure for ex-ante forecasts as well. This is because, if the model is not able to reproduce history, it can rarely be useful to predict the future. However, the reverse is not true, i.e. if the model has an excellent power to reproduce the history, it does not mean that it would give equally excellent predictions for the future. To cite an analogy, a palmist may well tell you all or 90 per cent correct about the past events in your life but his/her predictions about your future may be totally out of line. Thus, a test for ex-ante forecasts is also needed. This, however, is simple. Just examine the deviation from the true value, both the direction as well as the magnitude. For a forecast to be good, the direction must be predicted right and the magnitude of error must not be too high. As mentioned above, such a test would, of course, be possible when the true figures become available. It is needed not for correcting or really evaluating the forecasts but for avoiding such errors in future, i.e., for post-mortem.

## Conclusions

Demand forecasting is a necessary exercise for all decision makers. There are a number of alternative methods to perform this task. Since no method is perfect, it is recommended that more than one method is used in practice. This would at least serve the purpose of cross-checking and thereby improve the credence of forecasts. To this extent, the various methods are not merely substitutes but they are complementary as well.

In general, the survey methods are appropriate for short-term forecasts and for forecasting the demands for new products. In contrast, the statistical methods are suitable for long-term forecasts and for products having a long history.

The various methods explained above could be used to derive both macro as well as micro forecasts. In case the requisite data for micro forecasts are not available, the forecasts for demand of a firm's product could be derived through three steps. One, derive the forecasts for industry product's demand. Two, derive the forecasts for the firm's share in industry demand through applying some statistical methods (e.g. trend method) to the past data on this variable. Three, multiply the results of steps one and two to obtain the needed forecast. Similarly, the likely regional demands could be derived from the forecasts for aggregate demand in the country as a whole.

The last point to be emphasized here concerns the use of judgement. The judgement of the researcher is important not only in the choice of forecasting technique, but also in tampering with the so obtained forecasts by his own subjective evaluation of the forecasts period vis-a-vis history with reference to the commodity whose demand is being forecast. When he/she is not sure about some factors' behaviour, he might like to generate alternative forecasts, one for each possible outcome.

## REFERENCES

1. Gujarati, D. (2007): *Econometrics*, McGraw-Hill Publishers.
2. Gupta, G.S. (1974): "Forecasting Techniques", *Management Annual,* IV (Nov).
3. Gupta, G.S. (1975): "Demand For Cement in India," *Indian Economic Journal*, XXII (Jan–March).
4. Gupta, G.S. and D. Chawla (1977): "Demand For Tea in India," *Dynamic Management,* II (Sept.)
5. Gupta, G.S. (1993): "ARIMA model and Forecasts on Tea Production in Inida", *Indian Economic Journal*, 41, 2 (October), 88-110.
6. Gupta, G.S. (1994): "Survey Research: Data Collection, Analysis and Presentation" in K. Puttaswamaiah, edited, *Econometric Models, Techniques and Applications,* Indus Publishing Company, New Delhi, 26-43.
7. Gupta, G.S. (2008): *Forecasting through the Decomposition Method: Demand for Tea Production in India*, Technical Note # 326, Indian Institute of Management, Ahmedabad.
8. Gupta, V.K., et al (1984): *Marine Fish Marketing in India,* Vol. III (*Consumer Behaviour and Demand Forecasts*), Ahmedabad, Indian Institute of Management.
9. Hanke and Wichern (2009): *Business Forecasting*, 9th edition, Prentice Hall.
10. Markridakis, S. and C. Steren (1978): *Forecasting Methods and Applications*, New York, John Wiley and Sons.
11. Rangarajan, C. (1981): "Corporate Investment in 1981: A Forecast", *Economic and Political Weekly*, XVI, 9 (Feb.).
12. Rangarajan, C. and G.S. Gupta (1973): "The Demand For Financial Assets–A study in Relation to Bank Deposits, *Prajnan*, II (Jan.–March).
13. National Council of Applied Economic Research (1971): *Demand for Steel, 1975 and 1980*, Delhi, NCAER.
14. Pal, S.P. (1985): *Long-Term Demand for Iron and Steel*, Delhi, NCAER.

## CASELETS

1. Ashok Sharma, a market analyst at a multinational, was asked to conduct a market analysis for tea demand in India. For the purpose he collected time series data on the relevant variables and estimated the demand function through the least-squares technique of multiple regression analysis). The data and results are provided in Table and equation.

| *Year* | *Tea demand (thousand tonnes)* | *Real GDP at factor cost (Rs. hundreds of crores)* | *WPI for Tea (1993-94 =100)* | *WPI for Coffee (1993-94 =100)* | *WPI for Sugar (1993-94 =100)* | *Population (Crores)* | *Estimated demand (thousand tonnes)* | *Estimated error (actual–estimated demand) (thousand tonnes)* |
|---|---|---|---|---|---|---|---|---|
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| 1981–82 | 337 | 6780 | 29.6 | 28.4 | 48.3 | 69 | 389 | –53 |
| 1982–83 | 387 | 6979 | 35.5 | 29.2 | 45 | 71 | 401 | –14 |
| 1983–84 | 438 | 7527 | 52.5 | 35.4 | 45.5 | 72 | 415 | 22 |
| 1984–85 | 439 | 7825 | 58.8 | 44.5 | 47.9 | 74 | 429 | 10 |
| 1985–86 | 407 | 8150 | 45.9 | 47.3 | 55.5 | 76 | 442 | –35 |
| 1986–87 | 469 | 8502 | 50 | 47.7 | 56.7 | 77 | 455 | 14 |
| 1987–88 | 498 | 8803 | 51.5 | 50 | 57,4 | 79 | 468 | 30 |
| 1988–89 | 484 | 9697 | 53.9 | 51.7 | 61.4 | 80 | 486 | –2 |
| 1989–90 | 510 | 10292 | 83.8 | 54.5 | 68.5 | 82 | 501 | 9 |
| 1990–91 | 556 | 10836 | 88.9 | 87.4 | 68.5 | 84 | 521 | 35 |
| 1991–92 | 516 | 10991 | 80.3 | 82.7 | 75.8 | 86 | 532 | –16 |
| 1992–93 | 582 | 11580 | 91 | 78.7 | 85.7 | 87 | 545 | 36 |
| 1993–94 | 599 | 12238 | 100 | 100 | 100 | 89 | 565 | 34 |
| 1994–95 | 604 | 13021 | 81.6 | 167.4 | 112.6 | 91 | 594 | 10 |
| 1995–96 | 613 | 13970 | 102.8 | 178.4 | 109.5 | 93 | 615 | –2 |
| 1996–97 | 643 | 15084 | 105.5 | 169.2 | 116.7 | 95 | 633 | 10 |
| 1997–98 | 662 | 15733 | 160.3 | 210.2 | 133.8 | 96 | 653 | 9 |
| 1998–99 | 627 | 16784 | 149.9 | 187.2 | 135.4 | 98 | 670 | –43 |
| 1999–2000 | 667 | 17865 | 152.8 | 146.5 | 137.4 | 100 | 684 | –17 |
| 2000–01 | 669 | 18643 | 128.1 | 126.7 | 142.6 | 102 | 699 | –30 |
| 2001–02 | 655 | 19726 | 117.4 | 101.2 | 135.4 | 104 | 718 | –63 |
| 2002–03 | 712 | 20483 | 118 | 91.6 | 123.5 | 106 | 734 | –21 |
| 2003–04 | 725 | 22228 | 108.6 | 96 | 124.4 | 107 | 760 | –34 |
| 2004–05 | 859 | 23888 | 131.4 | 110.3 | 148.1 | 109 | 784 | 75 |
| 2005–06 | 873 | 26161 | 117.1 | 153.8 | 163.8 | 111 | 821 | 52 |
| 2006–07 | 865 | 28711 | 139 | 183.7 | 164.9 | 112 | 857 | 7 |
| 2007–08 | 871 | 31297 | 130.9 | 202.3 | 142.7 | 114 | 894 | –24 |
| Mean | 602 | 14955 | 95 | 106 | 100 | 91 | 602 | 0 |
| Std Dev. | 151.2 | 6966 | 38.1 | 58.3 | 40.3 | 13.9 | 14.8 | 33 |

*Source:* Centre for Monitoring Indian Economy, Various publications.

**Note:** Tea demand data were obtained as production + import – export of tea = domestic demand

Selected equation: $D_T = -101.1 + 0.00845\ Y - 0.040\ P_T + 0.166\ P_C - 0.0834\ P_S + 6.26$ POP

(–0.50) (1.91) (–0.08) (0.66) (0.09) (1.81)

$R^2 = 0.953$

Where, $D_T$ = Tea demand, $Y$ = Real GDP, $P_T$ = Tea price, $P_C$ = Coffee price, $P_S$ = Sugar price and POP = Population. Numbers in parentheses are *t*-values of the corresponding coefficients. Columns 8 and 9 in Table above give the estimated values of tea demand and the estimated error as obtained using the above estimated equation. Based on these data and results, attempt the following exercises:

(a) Determine the equations for the demand curve and the corresponding total revenue, average revenue and marginal revenue curves for year 2007–08. At what level of output, would the total revenue be maximum?

(b) Assume that tea industry had the absolute power to set the price of its product and nothing else. If the industry desires to sell 900 thousand tonnes of tea, what price it must set to achieve its target?

(c) Assume that India's growth rate was 6.4 % in 2008–09 and 7.4 % in 2009–10 and it expects to achieve a growth rate of 8 % during 2010–11; that the average (annual) inflation rate in each of the three relevant prices during the three years of 2008–11 at 7 %; and population growth rate (annual) at 1.5%. Determine the forecasts for tea demand in India in 2010–11. Comment on the reliability of your forecasts.

(d) Forecast tea demand in 2010–11using the trend method. Compare these with the forecasts of regression method and comment.

(e) Forecast tea demand in 2010–11 using the double simple smoothing and exponential smoothing methods. Compare the two results and comment.

(f) Examine the suitability of the estimated regression equation for forecasting tea demand (Hint: Evaluation of ex post forecasts).

(g) Compute various mean demand elasticities (viz. income, price, cross and population) and analyse the inter-relationships between tea, coffee and sugar industries.

2. Rajan Raju and Group (RRG) conducted a project course on demand estimation and forecasting at a management school. As a part of this course, they collected the following time series data on the registration of non-taxi cars in India and likely determinants.

Suppose you were a member of the RRG, what uses would you make of this data to achieve your task? You must ensure, at least the following:

(a) Estimate the demand function for non-taxi cars in the country. If you have estimated more than one form, indicate your choice for one equation and attempt the next questions for the chosen one only.

(b) Compute the various demand elasticities at the last observed values of the variables, i.e. for 1980-81, and give an economic interpretation of the same.

| *Year* | *Non-taxi cars registration ('000 numbers)* | *Per capita income (Rs. at 1970–71 prices)* | *Car price index* | *Motor cycle price index* | *Scooter price index* | *Petrol price index* | *Urban population index* |
|---|---|---|---|---|---|---|---|
| | | | | *(1965–66 = 100)* | | | |
| 1965–66 | 359 | 559 | 100 | 100 | 100 | 100 | 100 |
| 1966–67 | 385 | 552 | 107 | 109 | 103 | 109 | 99 |
| 1967–68 | 417 | 587 | 115 | 115 | 111 | 111 | 105 |
| 1968–69 | 458 | 589 | 115 | 115 | 110 | 124 | 112 |
| 1969–70 | 490 | 613 | 118 | 126 | 110 | 136 | 118 |
| 1970–71 | 539 | 633 | 119 | 131 | 110 | 151 | 119 |
| 1971–72 | 585 | 627 | 130 | 132 | 110 | 152 | 127 |
| 1972–73 | 580 | 603 | 133 | 135 | 112 | 187 | 129 |
| 1973–74 | 622 | 621 | 140 | 143 | 114 | 200 | 133 |
| 1974–75 | 596 | 717 | 166 | 152 | 137 | 429 | 140 |
| 1975–76 | 605 | 664 | 210 | 160 | 151 | 452 | 158 |
| 1976–77 | 698 | 652 | 190 | 158 | 153 | 456 | 163 |
| 1977–78 | 729 | 695 | 201 | 168 | 158 | 454 | 175 |
| 1978–79 | 788 | 717 | 231 | 177 | 164 | 470 | 177 |
| 1979–80 | 836 | 665 | 297 | 212 | 202 | 574 | 178 |
| 1980–81 | 898 | 698 | 344 | 236 | 226 | 680 | 195 |

(c) Forecast the demand for non-taxi cars in India in 1990-91 using the trend method, an appropriate smoothing method and the regression method. Compare your three forecasts and comment on their relative reliability.

(d) Suppose the Car Manufactures' Association has a target of 15,000 thousand non-taxi cars in 1990–91, suggest at least two policy prescriptions to achieve this target.

3. ABC enterprises manufactures a desk designed as a micro computer work station. Anil Goyal, its marketing manager, has graduated from a prestigious school of management, majoring in marketing and economics. Anil was anxious to apply some of the tools he has learnt and so estimated the demand function for desks in India. His results were:

$$Q_d = -2.8 + 2.5\,Y - 8.5\,P_d + 3.5\,P_o + 0.19\,A$$

$$R^2 = 0.87$$

where, $Q_d$ = annual sales of desks (thousands of numbers)

$Y$ = average household annual income (thousands of rupees)

$P_d$ = desk price (thousands of rupees)

$P_o$ = price of other (related) goods (thousands of rupees)

$A$ = annual advertising budget (thousands of rupees)

The current values of the independent variables are $Y = 16.5$, $P_d = 4$, $P_o = 2$ and $A = 200$.

(a) Name at least two important causal variables that Anil seems to have ignored in his estimation.

(b) Is product desk a normal good or a Giffen good? Explain.

(c) What do the coefficients + 2.5 and –8.5 represent in the function? What does the value of $R^2 = 0.87$ mean?

(d) Find the point (own) price elasticity of demand at the current price. If the firm contemplates a price increase, would its total revenue rise or fall? Why?

(e) Find the equation of the firm's demand curve. If the company's goal is to maximize total revenue, what price should it charge?

(f) Find the equation of the Engel curve for desks. What is the income elasticity of demand between $Y = 16.5$ and $Y = 20.0$?

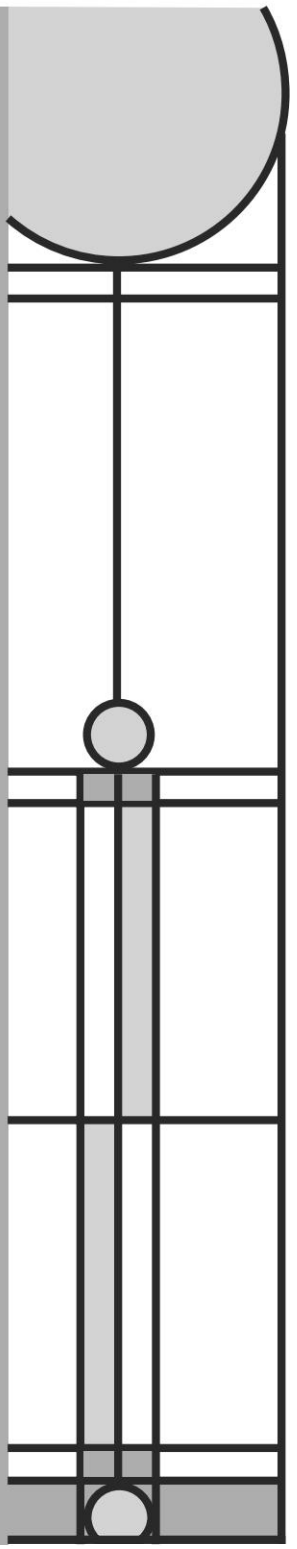
(g) What is the relationship between the desk and the other (related) good?

(h) If the firm has an annual sales target of 54,000 desks, indicate two policy options for attaining the target.

(i) Economic forecasters think that there is a possibility of a major recession next year, which will reduce the average household annual income to Rs. 15,000 without affecting any other relevant variable. Forecast the company's sales for the next year.

(j) Comment on the reliability of your answers for questions (h) and (i) above.

The McGraw-Hill Companies



CHAPTER 5

# PRODUCTION ANALYSIS AND INPUTS' DEMAND

The basic function of a firm is to produce one or more goods and/or services and sell them in the market. Production requires employment of various factors of production, which are substitutes among themselves to certain extent. Thus, every firm has to decide what combination of various factors of production, also called inputs, to choose to produce a certain fixed or variable quantities of a particular good. The problem is referred to as "how to produce?" The present chapter deals with this and the related problems of decision-making.

## 5.1 MEANING OF PRODUCTION

Production has a broad meaning in economics. It means presenting an item for sale, where the item could be a tangible good or an intangible good. Tangible goods could be presented for sale through either their manufacturing or through just trading in them. The latter would include activities such as transporting, storing, and packaging of goods. Thus, people who are engaged in transporting, say, wheat from Haryana to Kerala, are also considered as producers of wheat. So are the people who buy wheat at the time of harvest when its price is low, and sell it at a later date when its price is high. Firms which procure wheat from the market, convert it into wheat flour and then pack it in bags under their brand name are also producers, as they produce wheat flour. In the case of services, however, intermediaries do not exist and their production comes through manufacturing (rendering) only.

## 5.2 PRODUCTION FUNCTION

A production function expresses the technological or engineering relationship between output of a good and inputs used in the production, namely land, labour, capital and management (organization); besides raw-materials and intermediate goods. Both the quantities and qualities of these inputs have

bearing on the output. Traditionally, production functions are defined in terms of quantities of output and inputs. The quality of inputs is accounted for by introducing a variable called, technology. This is a separate input variable in the production function. The technology variable consists of all improvements in technology, including introduction of computer which permits a firm to produce a given output with fewer raw-materials, energy or/and labour, and training programmes which increase the productivity of labour. Thus, a production function could be written as

$$Q = f(Ld, L, K, M, T)$$
$$f_1, f_2, f_3, f_4, f_5 > 0 \qquad (5.1)$$

where

$Q$ = output in physical units of good $X$

$Ld$ = land units employed in the production of $Q$

$L$ = labour units employed in the production of $Q$

$K$ = capital units employed in the production of $Q$

$M$ = managerial units employed in the production of $Q$

$T$ = technology employed in the production of $Q$

$f$ = unspecified function

$f_i$ = partial derivative of $Q$ with respect to $i$th input

Output or production is measured in two ways, viz. gross output and net output. The difference between the two is given by the consumption of raw-materials and intermediate goods in production. Thus, sugar industry produces sugar, which is its gross output. This gross output is produced using certain quantities (and qualities) of primary inputs ($Ld$, $L$, $K$ and $M$) and secondary inputs (like sugarcane). If the contribution of these secondary inputs is netted out from the gross output, one gets the net output of the sugar industry. In function (5.1), $Q$ stands for the net output, which is also known as the value added, and accordingly it depends on primary inputs only. The said function assumes that output is an increasing function of all inputs. This is generally true. However, it is conceivable that if an input is excessively applied in relation to other inputs, an increase in it, other inputs held constant, might lead to a decrease in output. For example, consider a piece of land, with certain doses of fertilizers, water, ploughing, etc. and a certain quantity of labour used to cultivate it. If one goes on employing more and more units of labour (or any other input) on it without increasing the units of other inputs, including land, it is possible that after a certain point, the quantity of output produced would decline. This is because the labour input becomes relatively excessive, thereby the extra labour, instead of extending helping hands, prohibit the earlier labour to work. A provision for such a saturation of output would be included in the discussion later.

It should be noted that function (5.1) gives the maximum possible output that can be produced from a given amount of various inputs, or, alternatively, the minimum quantity of inputs necessary to produce a given level of output. Also, note that all the output and input variables are in their corresponding physical quantities and not in their value (rupee) terms. Thus, if it were the production function for MBA graduates, output consists of a number of MBA graduates turned out and inputs stand for the number of faculty engaged in MBA teachings (or faculty time), number of supportive staff, time spent by the management, the number of pages of teaching materials used, use of class rooms, dorms and offices expressed in physical units (e.g. hours), consumption of electricity in kilowatts, etc.

The McGraw-Hill Companies



Function (5.1) describes a general production function. A specific production function may not have all these inputs or may have some inputs in disaggregated terms. Besides, the relative importance of various inputs in production varies from product to product. For example, production functions for agricultural products generally have inputs as land, organic fertilizer, inorganic fertilizer, rainfall, irrigation, high-yield variety seed, etc. In contrast, the production function for industrial products, have inputs as labour, capital, management and technology. Thus, while land is an important factor of production in agriculture, due to its little significance, it is included in capital in industry. Quite the reverse is true with regard to labour and management inputs.

A production function, like any other function, can be expressed and analyzed through any one or more of the three tools: table, graph, and equation. As mentioned in Chapter 4, tables and graphs are hard to handle and equations get complicated when the number of variables in a function exceed three (one dependent variable and two independent variables). Besides, all the significant results in production theory can be explained through a two input variables (and one output variable) function and the same can easily be generalised for any number of inputs. For these reasons, a two input production function is taken up for further analysis. If the two inputs considered are labour and capital, the production function reduces to

$$Q = f(L, K)$$
$$f_1,\ f_2 > 0 \qquad (5.2)$$

It must be emphasised that in function (5.2), variables $Q$, $L$ and $K$ are measured in physical units. In particular, $Q$ stands for output in units like tones, metre, numbers, and so on; $L$ is measured in terms of number of workers, man/human days or man-hours; and capital in monetary (rupee) value (at constant prices if the data is time series) of all the structures, equipments and inventories used in production.

An important distinction for production analysis is between short-run and long-run. In microeconomics, short-run is defined as the period during which at least one of the factors of production is available in a fixed quantity, and at least one in a variable quantity. In contrast, long-run is the period during which all the factors of production are available in variable quantities. Thus, in the short-run, the production could be increased or decreased through changes in variable inputs only; while in the long-run, increased or decreased in production could be accomplished through changes in any one or more of the inputs. For example, in a given season, a farmer might have a fixed quantity of land to grow wheat, but he could still produce more or less wheat by using more or less fertilizer or irrigation. Similarly, Maruti Udyog has a fixed capacity to manufacture cars (i.e., fixed quantity of capital in the forms of structure, equipments and inventories) but it could still increase its production by employing more workers and scheduling work on two shifts instead of one shift or on all the three shifts instead of two shifts. How short is the short-run? It is not defined in terms of the time duration and depends on the nature of the product. In the case of a farmer, the short-run could be less than a year, while in the case of Maruti Udyog, it could well be around five years. This is because, land is not fixed for the cultivation of wheat beyond a particular season, and the next season would come after one year only. In contrast, Maruti Udyog would need around five years to add to its capacity, for this has to go through hurdles of licensing (if required), fund raising, assets purchasing and installation, testing, etc. The short-run version of production function (5.2), assuming capital as the fixed input ($K$) and labour as the variable input, could be written as

$$Q = f(L, \overline{K}) \qquad (5.3)$$

In the long-run, since all the inputs are variable; in just the two inputs case, function (5.2) represents the long-run production function. Functions 5.1 and 5.2 are examples of unspecified production functions. To be specified, they must be expressed in the form of a table, graph, or an equation. However, before one could do that, one has to remember certain special features of a production function. These are

(a) labour and capital are both inevitable inputs to produce any quantity of a good; and

(b) labour and capital are substitutes, though not perfect substitutes, to each other in production.

These features imply that one need some quantity of both the inputs to produce any quantity of a good but there are alternative combinations of these two inputs to produce a given quantity of output. Bearing these features in mind, a two input long-run production function for quantities of labour and capital up to 10 units each could be expressed like the one in Table 5.1.

**Table 5.1** *Long-Run Production Function*

| ***Labour*** | ***Capital (K)*** | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| (*L*) | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 5 | 15 | 35 | 47 | 55 | 62 | 61 | 59 | 56 | 52 |
| 2 | 0 | 12 | 31 | 49 | 58 | 66 | 72 | 77 | 75 | 74 | 71 |
| 3 | 0 | 35 | 48 | 59 | 68 | 75 | 82 | 87 | 91 | 89 | 87 |
| 4 | 0 | 48 | 59 | 68 | 72 | 84 | 91 | 96 | 99 | 102 | 101 |
| 5 | 0 | 56 | 68 | 76 | 85 | 92 | 99 | 104 | 108 | 111 | 113 |
| 6 | 0 | 55 | 72 | 83 | 91 | 99 | 107 | 112 | 117 | 120 | 122 |
| 7 | 0 | 53 | 73 | 89 | 97 | 104 | 111 | 117 | 122 | 125 | 127 |
| 8 | 0 | 50 | 72 | 91 | 100 | 107 | 114 | 120 | 124 | 127 | 129 |
| 9 | 0 | 46 | 70 | 90 | 102 | 109 | 116 | 121 | 125 | 128 | 130 |
| 10 | 0 | 40 | 67 | 89 | 103 | 110 | 117 | 122 | 126 | 129 | 131 |

Table 5.1 gives a hypothetical production function in the matrix form. The numbers inside the table indicates **maximum** quantities of output that a firm could produce with given quantities of labour and capital. For example, if the firm employs two units of labour and three units of capital, the maximum that it could produce equals 49 units of output. Since the function contains only two inputs and both are variable in the table, the production function represented by the table is a long-run one.

If capital was the fixed input in the short-run, then each column (together with the column of labour input) of the table represents a short-run production function with respect to a specific quantity of the fixed (capital) input. For example, for $K = 2$, the short-run production function would be the following:

**Table 5.2** *Short-run Production Function (K = 2)*

| Labour (*L*) | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Output (*Q*) | 0 | 15 | 31 | 48 | 59 | 68 | 72 | 73 | 72 | 70 | 67 |

Both the above functions could be expressed in terms of geometry. The long-run function has three variables, namely *L*, *K* and *Q*, and thus one needs a three dimensional diagram, which is complicated, to present it on a single figure (curve). Alternatively, it can be plotted on a two dimensional diagram,

but with a family of production curves, one for each production level. Figure 5.1 provides such a representation for two selected levels of production: $Q = 91$ and $Q = 122$.

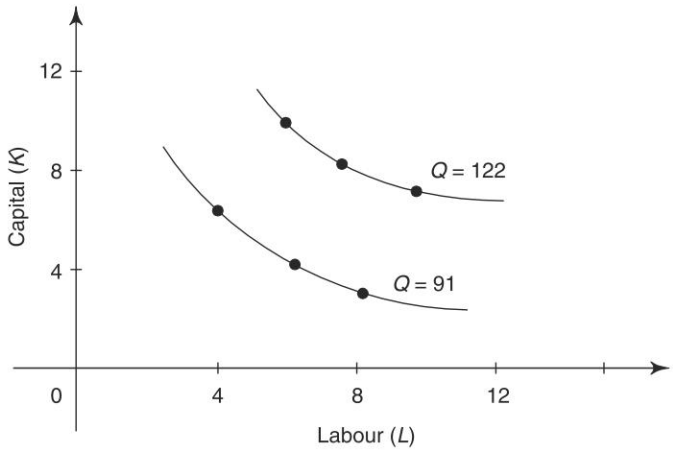



**Fig. 5.1** *Long-run production function (Isoquants)*

Table 5.1 indicates that there are four alternative ways of producing 91 units and three for producing 122 units of output:

| $Q = 91$ | | $Q = 122$ | |
|---|---|---|---|
| $L$ | $K$ | $L$ | $K$ |
| 3 | 8 | 6 | 10 |
| 4 | 6 | 7 | 8 |
| 6 | 4 | 10 | 7 |
| 8 | 3 | | |

In Fig. 5.1, various alternative input combinations for $Q = 91$, have been marked and joined to form a curve, identified as $Q = 91$. Similarly, $Q = 122$ graph was drawn by joining alternative input combinations for $Q = 122$. The graphs are continuous, and so every point on them represents a technology (input-mix) for a given level of output. Incidentally, note that one needs at least three points to draw a non-linear graph. Since this requirement was met for only two output levels by Table 5.1, graphs of $Q = 91$ and $Q = 122$ alone have been drawn in Fig. 5.1. Of course, if one goes for interpolation of data in Table 5.1, more such graphs could be drawn. Note that table represents a finite number of technologies (input mix) while a graph has infinite number of them. In deriving graphs from tables above, we have implicitly assumed patterns as given in the corresponding tables.

Each of the two curves in Fig. 5.1 is called an *isoquant* for a given level of output. Thus, an isoquant (equal output curve) is the locus of all those combinations of two inputs which yields a given level of output. These isoquants have the following properties:

(a) They are falling
(b) The higher the isoquant, the higher is the output
(c) They do not intersect each other.
(d) They are convex from below.

The above properties are the same as those of indifference curves as explained in Chapter 3 and thus their reasons are also similar. The isoquants are falling, for production is an increasing function of both the inputs. If they were rising, one would have a constant level of output for increasing levels of both the inputs, which is not true. Similarly, if they were a straight line (horizontal or vertical) increase in one input while the other is held constant, would leave the production level unaltered, which make no sense. For the same reasoning, a higher isoquant represents a higher level of output. An isoquant does not intersect another isoquant, for if it did, it would mean that with the same unit of labour and capital, two different levels of (maximum possible) output could be produced, which is absurd.

Isoquants are convex from below, for the substitution of labour for capital or vice versa become more and more difficult as more of one factor is substituted for another factor. If labour and capital were perfect substitutes, the isoquants would be straight lines as in Fig. 5.2(a) and if they had no substitutability at all, the isoquants would be rectangular as in Fig. 5.2 (b).

A short-run production function, like the one in Table 5.2, could also be represented through a graph as shown in Fig. 5.3.

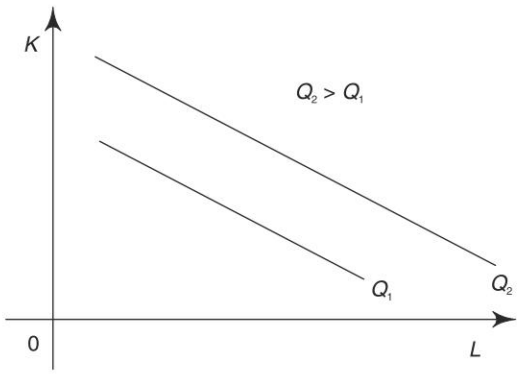



**Fig. 5.2** *(a) Isoquants under perfect substitutability*

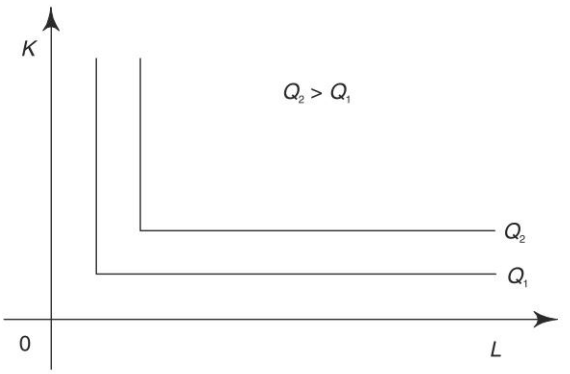



**Fig. 5.2** *(b) Isoquants under zero substitutability*

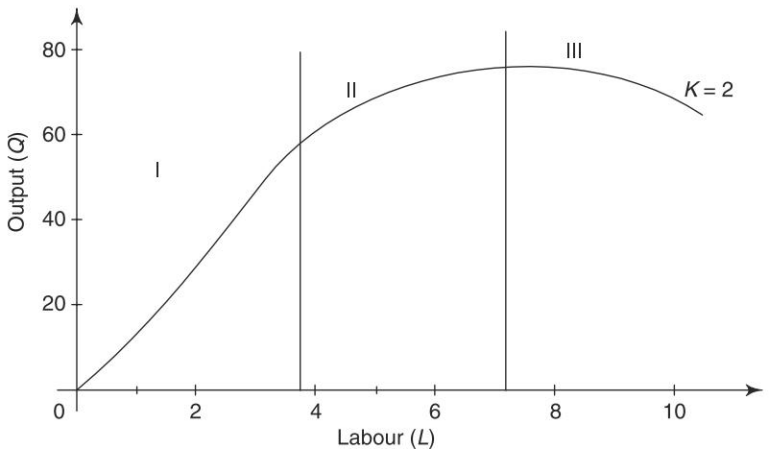



**Fig. 5.3** *Short-run production function*

The shape of a short-run production curve is such that it is first convex from below (bow down) and then concave from below bow up). This is because of the operation of the *Law of Variable Marginal* ***Returns***. Under this law, as more and more units of the variable input are employed in the production, fixed inputs remaining unaltered, production first increases at an increasing rate and then at a diminishing rate, leading to a decline in total production eventually. Since generally, the first stage is short, the second stage a long one and the third stage never entertained by a prudent production manager, the law is popularly known as the *Law of Diminishing Marginal Returns*. As this law happens to be an important law of economics, its formal definition is in order. The law of diminishing returns states that, as the more and more units of labour are employed in production, holding all other inputs constant, the marginal product of labour first increases, then remains constant, and eventually goes on falling indefinitely. The law generally holds good, for (a) in the beginning as more labour is used, fixed capital is utilized better and more efficiently than before, thereby output increases at all increasing rate (or marginal product increases), and this continues until optimal utilization of fixed capital is achieved; (b) after this point, new (additional) labour finds the fixed capital inadequate and hence increment in output is at a diminishing rate (i.e., marginal product declines); and (c) eventually labour input becomes so much that there is no work for new labour and so they disturb the earlier labour from carrying out their work, thereby leading to a decrease in total output (i.e., marginal product becomes negative). Economists talk of the three stages of production, viz. total production increasing such that average product is rising (I), total production increasing but average product is falling (II), and total product decreasing (III). These three stages are demarked by vertical lines in Fig. 5.3. The point at which stage I ends, APP of labour starts falling and the point at which stage II ends is known as the point of Saturation. Further details on this are presented in section 5.4.

Finally, a production function could be expressed in the form of an equation. It is pertinent to note here that a production function is never linear, for a linear function implies that (a) one can produce any quantity of a good with just one input, which is not usually possible, (b) various inputs are perfect substitutes in production, which is absurd, and that (c) it is inconsistent with the law of diminishing marginal returns. The usual functional forms specified for a production function are (i) double-log or linear in log (called the Cobb-Douglas form in honour of the persons who popularized this form) and (ii) quadratic:

**Equations of Long-Run Production Functions**

**Double-log**

General: $Q = A\, L^{\alpha}\, K^{\beta}$ (5.4)

Specific: $Q = 1.01\, L^{0.75}\, K^{0.25}$ (5.5)

**Quadratic**

General: $Q = a + b\, L + c\, K + d\, L\, K + e\, L^2 + f\, K^2$ (5.6)

Specific: $Q = 549.92 + 12.98\, L + 26.72\, K + 0.196\, L\, K - 0.104\, L^2 - 0.319\, K^2$ (5.7)

In short-run production functions, the fixed factor ($K$) takes a specific value. Thus, if $K = 16$, short-run counterparts of specific functions (5.5) and (5.7) would be functions (5.8) and (5.9), respectively:

$$Q = 2.02\, L^{0.75} \quad (5.8)$$

$$Q = 895.78 + 16.12\, L - 0.104\, L^2 \quad (5.9)$$

Having discussed various aspects of the production function, including its representation through a table, graph and equation, we proceed to production analysis for decision-making in long and short-runs, in that order.

## 5.3 PRODUCTION ANALYSIS: LONG-RUN

The decision problems which a production manager faces in the long-run could be summarized as follows:

(a) What are the optimum quantities of labour and capital that he should hire and employ? The answer to this question might depend on his specific problem. For example, the manager may have a fixed production target and wish to find out the least-cost input combination for that level of output. Alternatively, he could have a fixed rupee budget for production and wish to determine that input combination which maximizes his output for a given cost. Lastly, neither the production target nor the production budget may be fixed, and accordingly the manager might seek that input combination which maximizes the firm's profit.

(b) What is the expansion path? Are returns to scale increasing, constant or decreasing? Answers to these two related questions would guide the manager in his future expansion or contraction plan.

Answers to the above questions are the subject matter of firm behaviour. Recall that to understand consumer's equilibrium, we had to identify his/her objective function and the constraints. Thus, at the outset let us note that the firm's objective is assumed to be profit maximization and its constraints include the given technology (or production function), and given factor prices. In what follow, we shall present approaches to answer the above and related questions under these objectives and constraints.

### Least-Cost Input Combination

The least-cost input combination for a given output could be explained through any of the three forms of the production function: table, graph and equation. However, since the table provides only a limited number of alternative input combinations for any given level of output, this could easily be evaluated. The text attempts to examine the subject mainly through graphs and equations.

For determining the least-cost technology, one needs, besides the production function, the factor prices. If $P_L$ and $P_K$ were the prices of labour and capital, respectively, then the firm's total cost equation would be the following:

$$C = L\,P_L + K\,P_K \tag{5.10}$$

where, $C$ = total cost of production, and $L$ and $K$ are units of labour and capital, respectively. In function (5.10), while $P_L$ and $P_K$ are fixed factor prices (an individual producer is assumed to have no control over $P_L$ and $P_K$), $L$, $K$ and $C$ are variables. While $P_L$ stands for the **wage rate**, $P_K$ stands for **capital rental** (also known as the **user cost of capital**). In case the production firm hires all the capital, then the rent it pays for the use of the capital measures $P_K$. However, if the firm owns its capital, as is usually the case, capital rental is given by

$$P_K = i + d - \text{inflation rate}$$

where, $i$ = interest rate (weighted average rate of the one paid on the borrowed part of $K$ and the other which is forgone on the equity component of $K$), $d$ = the depreciation rate of $K$ (wear, tear and obsolescence of equipment, structures and inventories), and inflation rate = the rate at which the price of capital has increased during the period.

As isoquant map was drawn from the production function, a map of isocost lines could be drawn from Eq. 5.10, where a specific isocost line would be for a given value of $C$. Thus, if factor prices were $P_L = 3$ and $P_K = 3$, an isocost line for $C = 30$ would have input combinations such as the following:

| $L$ | $K$ |
|---|---|
| 10 | 0 |
| 8 | 2 |
| 5 | 5 |
| 0 | 10 |

Similarly, more isocost lines could be identified and drawn, one for each value of total cost, $C$. An isocost map for the factor prices corresponding to above prices is given below in Fig. 5.4.

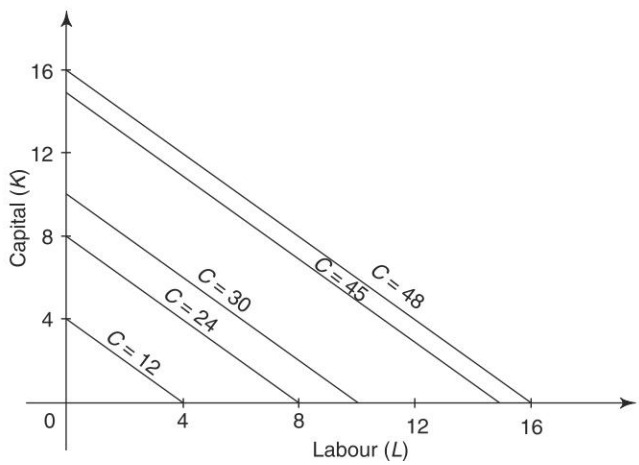



**Fig. 5.4** *Isocost lines*

Thus, an isocost line gives alternative combinations of labour and capital that a firm could hire with a given cost. Since factor prices are assumed constant, isocost lines are linear, and they are all parallel to each other. Further, the higher the cost line, the greater is the cost.

Given the isoquant map and isocost lines, the least cost input combination could easily be derived. Super impose the isoquant specific to the target production level (See Fig. 5.1, $Q = 122$) and the isocost lines map (See Fig. 5.4) on a graph, as given in Fig. 5.5. Fig. 5.5 indicates that $Q = 122$ could be produced by any one of the input combinations given by points $A$, $B$, and $E$. However, the production cost associated with points $A$ and $B$ is higher than the one at point $E$, for the former points lie on a higher cost line ($C = 48$) then does the latter point ($C = 45$). Thus, in this case, point $E$ is the least-cost point for $Q = 122$, and so it is called the equilibrium point. The input combination corresponding to this point ($L = 7$, $K = 8$) is then the least-cost input combination for $Q = 122$. The special feature of the equilibrium point is that it is the point where the specific isoquant is tangent to an isocost line. Tangency implies equal slope. The slope of the cost line is given by

$$dK / dL = - P_L / P_K$$

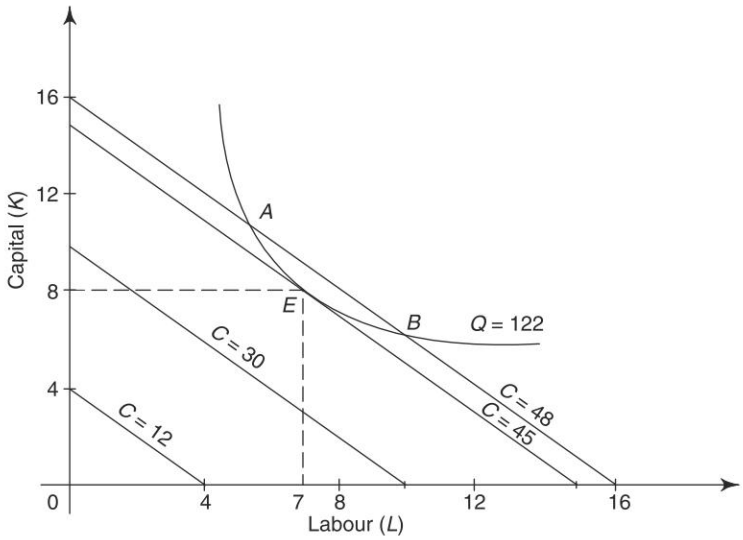



**Fig. 5.5** *Least cost input combination*

On an isoquant, output is a constant, and thus

$$dK\,(MPP_K) + dL\,(MPP_L) = 0$$

The above result holds, for the first expression gives the change in output due to a change in capital and the second term gives the change in output due to a change in labour, the sum of the two must equal zero for the two points to be on an isoquant. Solution of the above equation yields

$$dK / dL = - MPP_L / MPP_K$$

This measures the slope of the isoquant, and is called the **Marginal Rate of Technical Substitution (MRTS)** between labour and capital. Thus, the MRTS measures the rate at which the firm must reduce (add) units of capital to keep output unchanged when labour is increased (decreased) by a small amount. Equating the slopes of the cost line and isoquant, and rearranging the resultant expression gives

$$\frac{MPP_L}{P_L} = \frac{MPP_K}{P_K} \tag{5.11}$$

The above expression denotes the condition for optimization. Thus, it can be concluded that the least-cost input combination is that combination at which the slope of the isoquant equals the slope of isocost line or at which the *MPP* per unit of input price are equalised across inputs. The said equality condition is the necessary condition for the least cost input combination. There is also a sufficient condition for the least-cost combination, which is met if the isoquant is convex from below.

It should be clear that the least cost input mix depends both on the production function (isoquant) and factor prices (isocost lines). If any one of them undergoes a change, the equilibrium point would change. To illustrate this, let us analyse the effect of a fall in the price of labour ($P_L$), *ceteris paribus*. As labour price falls, the isocost lines change. Since $P_K$ has not changed, their coordinate at vertical axis will remain unaltered but since $P_L$ has fallen, their coordinate at horizontal axis would increase. An illustration of this is provided in Fig. 5.6.

The original isocost line is *AB* and original equilibrium point is *E*. As $P_L$ falls, *ceteris paribus*, the isocost line rotates anticlockwise to the one like line *AC*. The original output ($Q_0$) is now producible at original cost by choosing the input combination of point *F*. However, this is not the least-cost technology. The least-cost input-mix for original output level is given by tangency point *G*, where the cost is lower than at point *F* for the same quantity of output. Thus, due to a fall in $P_L$, equilibrium moves from point *E* to point *G*, showing a decrease in the employment of capital, whose price has remained constant, and an increase in the employment of labour, whose price has fallen. This results through the working of *output* and *substitution effects*, which are components of the factor price effect on input combination. As $P_L$ falls, *ceteris paribus*, cost of producing a given output declines, and this causes firms to expand its production, leading to an increase in the demand for both labour and capital. This is the output effect of a fall in $P_L$. Under the substitution effect, as $P_L$ falls, labour becomes relatively cheaper to capital, which induces firms to substitute labour for capital, thereby demand for labour expands while that for capital contracts. The total effect, which is the sum of the two, of a fall in $P_L$ is thus an increase in the employment of labour and uncertain change in the employment of capital. The movement from point *E*

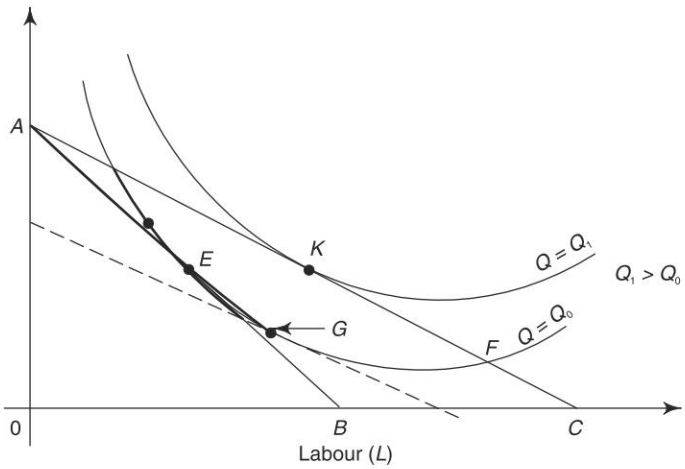



**Fig. 5.6** *Change in labour price and least cost input combination*

to point *K* indicates this total effect. But if output is held constant at the original level (by nullifying the output effect through an appropriate change in cost), a fall in $P_L$ would lead to an increase in labour input and a decrease in capital input, as indicated by a movement from point *E* to point *G*. The movement from point *E* to point *G* is the pure substitution effect.

The least-cost input combination may now be explained through equations. Under this, the firm's objective is to minimize the total production cost; subject to given output, given production function (or technology) and given inputs' prices. Suppose a firm faces a production function as in equation (5.7), factor prices as $P_L = 0.75$ and $P_K = 0.50$, and it is interested in producing 1400 units of output. What is its least-cost technology? The problem can be formulated as a constrained optimization problem:

Minimize $C = 0.75\ L + 0.50\ K$

Subject to $Q = 1400 = 549.92 + 12.98\ L + 26.72\ K + 0.196\ LK - 0.104\ L^2 - 0.319\ K^2$

$L, K \geq 0$

It can be solved through the Lagrangian multiplier technique. The Lagrangian expression for

minimising total production cost with respect to $L$, $K$ and $\lambda$ would be the following:

$$L^* = 0.75L + 0.50K + \lambda\,(1400 - 549.92 - 12.98L - 26.72K - 0.196\,LK + 0.104L^2 + 0.319\,K^2)$$

where, $L^*$ = Lagrangian expression

$\lambda$ = Lagrangian multiplier

The necessary condition for minimization requires that each of the three partial derivatives be zero:

$$\frac{\partial L^*}{\partial L} = 0.75 + \lambda\,(-12.98 - 0.196K + 0.208L) = 0$$

$$\frac{\partial L^*}{\partial K} = 0.50 + \lambda\,(-26.72 - 0.196L + 0.638K) = 0$$

$$\frac{\partial L^*}{\partial \lambda} = 1400 - 549.92 - 12.98L - 26.72K - 0.196LK + 0.104L^2 + 0.319K^2 = 0$$

Solving each of the first two equations for $\lambda$, we get

$$\lambda = \frac{-0.75}{-12.98 - 0.196K + 0.208L}$$

and

$$\lambda = \frac{0.50}{-26.72 - 0.196L + 0.638K}$$

Since the LHS of the last two equations are the same, their RHS must also be the same. Equating the RHS and re-arranging the term, we get

$$\frac{-12.98 - 0.196K + 0.208L}{-26.72 - 0.196L + 0.638} = \frac{0.75}{0.50} \tag{5.12}$$

Solution of Eq. (5.12) yields

$$0.251L - 0.576K = -13.55$$

or,

$$L = 2.29K - 53.96 \tag{5.13}$$

The third partial derivative gives the equation of the isoquant for $Q = 1400$:

$$549.92 + 12.98L + 26.72\,K + 0.196LK - 0.104\,L^2 - 0.319\,K^2 = 1400 \tag{5.14}$$

Solution of Eq. (5.13) and (5.14) yields

$$L = 18.6$$

$$K = 31.7$$

Which denotes the least-cost input combination for $Q = 1400$. There is a sufficient or second-order condition for the least-cost combination as well, verification of which, not attempted here, would show that it is also met by the above input-mix. The corresponding least-cost, obtained by substitution of the least-cost input magnitudes into the cost equation, would be

$$C = (18.6)\,(0.75) + (31.7)\,(0.50) = 29.80$$

Incidentally note that in Eq. (5.12), the LHS equals the slope of the isoquant, which is referred to as the Marginal Rate of Technical Substitution:

$$= \left[\frac{\text{marginal physical product of labour } (MPP_L)}{\text{marginal physical product of capital } (MPP_K)}\right] \text{and}$$

The RHS equals the slope of the iso-cost line ($= P_L/P_K$). Thus,

$$\frac{MPP_L}{MPP_K} = \frac{P_L}{P_K} \quad (5.11a)$$

Equation (5.11a) thus stands for the necessary condition for the least-cost input combination for a

given output, when the firm has no control over factor prices. This is identical to Eq. (5.11), the one we obtained above through the geometrical approach. If the firm were large enough (i.e. had some market power) to affect factor prices, the condition would change to

$$\frac{MPP_L}{MPP_K} = \frac{MC_L}{MC_K}$$

where, $MC_L$ = marginal cost of labour

$MC_K$ = marginal cost of capital

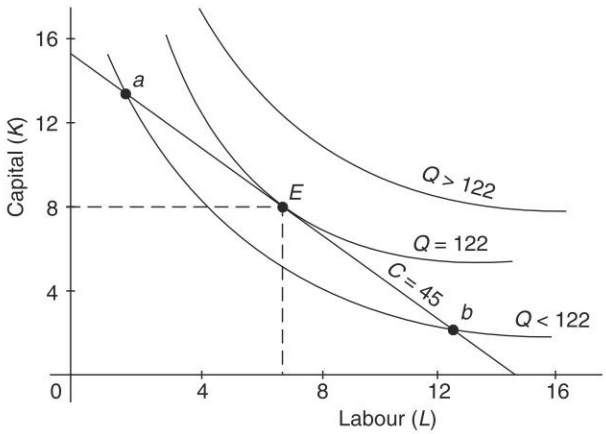



**Fig. 5.7** *Maximum-output input combination*

## Maximum-output Input Combination

When a firm has a fixed rupee budget for its production, its optimization problem is one of maximizing output for a given cost. This problem, like the least-cost problem, can be handled both through geometry and calculus. Since the cost is fixed, the isocost line will be unique and since output is a variable, there will be a family of isoquants. The equilibrium would be determined as in Fig. 5.7.

In Fig. 5.7, the cost line ($C = 45$) is fixed, and there are three isoquants, one, for $Q = 122$, second for $Q$ less than 122 and third for $Q$ greater than 122. With a budget of Rs. 45, less than 122 units of output is producible (e.g. at points $a$ and $b$) but obviously that is not the best the firm could do. Point $E$ would still mark the equilibrium point, where 122 units are produced by employing 7 units of labour and 8 units of capital.

The optimization problem in terms of calculus for the equation form could be formulated as follows:

Maximize $Q = 549.92 + 12.98L + 26.72K + 0.196LK - 0.104L^2 - 0.319\,K^2$

Subject to $C = 29.80 = 0.75L + 0.50K$

$L, K, \geq 0$

The Lagrangian expression would then be

$L^{**} = 549.92 + 12.98L + 26.72K + 0.196LK - 0.104L^2 - 0.319\,K^2 + \lambda_1 (29.80 - 0.75L - 0.50K)$

the solution of which would yield the same result as before, viz. $L = 18.6$, $K = 31.7$ and $Q = 1400$. Thus, given the factor prices, the least-cost input combination for a given output is the same as the maximum output-input combination for a given cost, provided the total cost or total output in the two problems is the same.

## Maximum-Profit Input Combination

Most firms today have neither a fixed output target nor a fixed rupee budget for their production. Instead, they seek that level of output and budget which serve their overall objective the best. For different objectives, there may be different input combinations even under such a situation. The determination of optimum technology is described in what follows for a profit-maximizing firm.

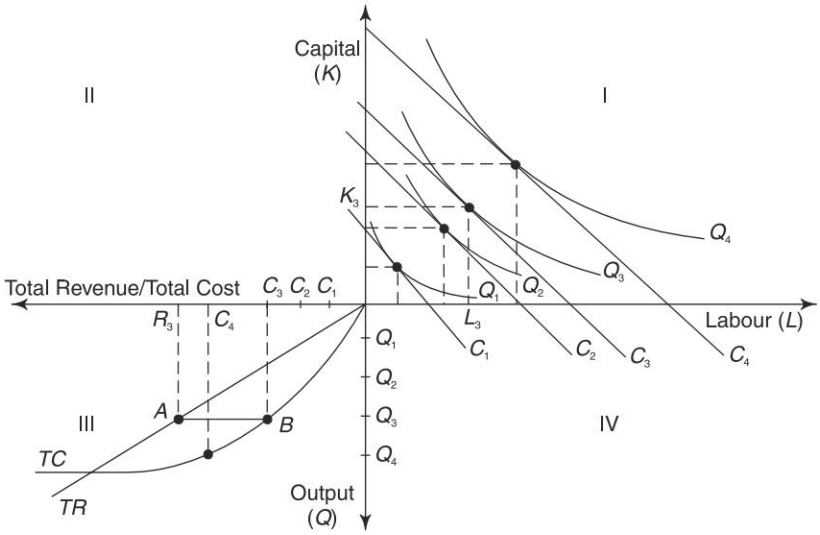



**Fig. 5.8** *Maximum profit input combination*

For the purpose, one needs, besides the production function and factor prices, the product price (if it is fixed) or the demand curve (if product price is a variable). For a constant product price case, the equilibrium is explained in Fig. 5.8.

Quadrant I gives the equilibrium input-combinations for various output levels. The results of this quadrant in terms of outputs and the corresponding least-costs are transferred in terms of the total cost curve in quadrant III. The total revenue (*TR*) curve is drawn on the basis of a given product price (= slope of the *TR* curve). The horizontal gap between *TR* and *TC* (total cost) curve represents profits for various output levels. The output at which profit is maximum denotes the profit-maximizing output and the input combination corresponding to this output is the profit-maximizing technology. Thus, $L_3$ units of labour and $K_3$ units of capital is the profit-maximizing input combination, and the line *AB* in quadrant III gives the measure of maximum profit.

In terms of calculus, the problem could be formulated as follows:

Maximize $\pi = TR - TC = PQ - LP_L - KP_K$

where, $\pi$ = profit

$P$ = product price

$TC$ = total cost

For the long-run production function of Eq. (5.7); $P_L = 0.75$,

$P_K = 0.50$; and $P = 0.1585$ (assume), the problem reduces to

Maximize $\pi = (0.1585)\ (549.92 + 12.98L + 26.72K + 0.196LK - 0.104L^2 - 0.319K^2) - 0.75L - 0.50K$

$$L, K \geq 0$$

This is simply an unconstrained maximization problem, and so it could be handled through simple calculus. The necessary conditions are that each of the two first-order partial derivatives be zero:

$$\frac{\partial \pi}{\partial L} = (0.1585)\,(12.98 + 0.196K - 0.208L) - 0.75 = 0$$

$$\frac{\partial \pi}{\partial K} = (0.1585)\,(26.72 + 0.196L - 9.638K) - 0.50 = 0$$

Simplifying these, we get

$$\frac{(0.1585)(12.98 + 0.196K - 0.208L)}{0.75} = 1 \tag{5.14}$$

and

$$\frac{(0.1585)(26.72 + 0.196L - 0.638K)}{0.50} = 1 \tag{5.15}$$

Solution of these two equations yield

$$L = 103$$

$$K = 68$$

The above technology denotes the profit-maximizing input combination. The corresponding profit-maximizing output is then obtained by substituting these input values into the production function:

$$Q = 549.92 + 12.98\,(103) + 26.72\,(68) + 0.196\,(103)\,(68) - 0.104\,(103)^2 - 0.319\,(68)^2$$
$$= 2498$$

The corresponding maximum profit is obtained by the substitution of the output and input values into the profit function:

$$\pi = (0.1585)\,(2498) - 0.75\,(103) - 0.50\,(68)$$
$$= 395.9 - 111.2$$
$$= 284.7$$

Once again the verification of the second order conditions is not attempted, but the same are met by the above solution. If one studies Eqs. (5.14) and (5.15) carefully one would realize that they stand for

$$\frac{P(MPP_L)}{P_L} = 1$$

and

$$\frac{P(MPP_K)}{P_K} = 1 \tag{5.16}$$

or,

$$\frac{MRP_L}{P_L} = \frac{MRP_K}{P_K} = 1$$

where, $MPP_L$ = marginal physical productivity of labour

$MPP_K$ = marginal physical productivity of capital

$MRP_L$ = marginal revenue productivity of labour

$MRP_K$ = marginal revenue productivity of capital

Equation (5.16) thus stands for the necessary conditions for profit maximization, given the factor prices. If factor prices could be influenced by the firm, the conditions would change to

$$\frac{MRP_L}{MC_L} = \frac{MRP_K}{MC_K} = 1$$

incidentally note that $MRP_L = P(MPP_L)$ and $MRP_K = P(MPP_K)$ only when the product price ($P$) is constant. However, the condition (5.16) holds good even when $P$ is a variable.

Before we conclude this sub-section, it must be emphasized that a firm maximises its profit under all the three situations above. This is because, when output is fixed (and product price is fixed), total revenue is fixed; thus, minimizing total cost for a given output is tantamount to maximizing profit, subject to an output constraint. Similarly, when cost budget is fixed (and input prices are given), total cost is fixed; thus maximising output (with given product price) is the same as maximizing total revenue or maximising profit with a cost constraint. Accordingly, the constrained least-cost and constrained maximum output cases vary with the corresponding constraint, while the unconstraint profit maximisation is a unique situation. In other words, while the solution of the unconstrained profit maximisation problem is also a solution for both a constrained least cost as well as a constrained output maximisation problem, only one of the many possible constrained least-cost or possible constrained output maximisation solutions will be a solution for the unconstrained profit maximisation problem.

## Expansion Path and Returns to Scale

The foregoing sections have examined the determination of the optimum input combination in a static sense, that is, what is the best input mix for today? However, no firm can ignore dynamic production decisions. These are concerns with the production levels that a firm must endeavour to attain over a long period of time. This requires an understanding of the relationship between output and production cost (in rupees) at their optimum levels, and between all inputs and corresponding output levels. The former is contained in the expansion path of the firm and the latter is described by returns to scale.

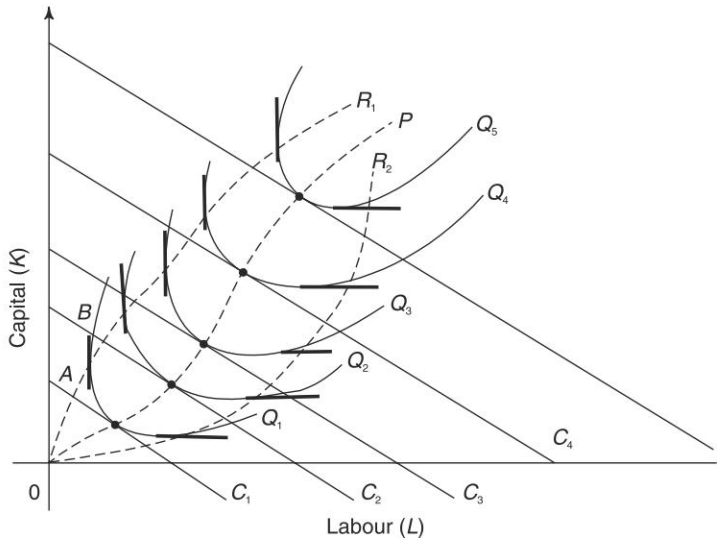



**Fig. 5.9** *Expansion path and ridge lines*

Given the production function (iso-quant map) and factor prices (iso-cost lines), a map showing the expansion path can easily be drawn as shown in Fig. 5.9.

The curve *OP* in Fig. 5.9 denotes the *expansion path*. It is the locus of the least-cost input combinations for various output levels, which is same as the locus of maximum-output input combinations for various cost constraints. In algebraic terms, it is given by the equilibrium condition (5.11). Thus, if a firm desires to expand its output from $Q_1$ to $Q_2$ its (minimum) total cost would increase from $C_1$ to $C_2$, and so on. This, together with the output-revenue relationship (demand curve) would help the firm to decide on its expansion strategy. Incidentally note that while all points along the expansion path are the least-cost for a given output as well as maximum output for a given cost input combinations, only one of those points denote the (unconstrained) profit maximisation point.

Figure 5.9 also includes the *ridge lines* $OR_1$ and $OR_2$. These lines separate the economic region from the non-economic region. All least cost input combinations are in the economic region and thus the expansion path necessarily falls between the two ridge lines. The ridge line $OR_1$ passes through those points on various isoquants, where the isoquants are either vertical or upward sloping. This is because the economic choice can not fall on the vertical or rising part of any isoquant, for in that region the firm would need to employ more of capital with a same amount of labour or more of both inputs than on a point immediately preceding such a point for producing the same level of output. For example, for producing $Q_1$ output, input combination of point *B* is inferior to that of point *A*, as the former requires more of both labour and capital than the latter for producing the same output. Similarly, line $OR_2$ passes through those points on various isoquants, where the isoquants are either horizontal or upward sloping. Remember, that the economic region must fall on the falling and convex (from below) parts of isoquant.

*Returns to scale* provide a measure of the direction of change in total factor productivity when all factors of production change in the same direction and same proportion. Thus, increasing returns to scale, which means, as all inputs increase in a given proportion (multiple), output increases by more than that proportion (multiple), implies increase in total factor productivity. If output increases by the same multiple as all inputs have increased, there are constant returns to scale and there is no change in total factor productivity. Finally, if output increases by a smaller multiple than have all inputs, there are decreasing returns to scale and a decrease in total factor productivity. For example, suppose a firm's production function is given by $Q = 10\ L^{0.8}\ K^{0.5}$. If the firm uses two units of *L* and two units of capital, it would produce 24.6 units of output. If labour and capital are doubled to four units each, $Q = 60.6$, which is more than double the original quantity. Thus, the firm has increasing returns to scale. On the other hand, if the production function is $Q = 10L^{0.5}\ K^{0.5}$, and $L = 2$, $K = 2$, then $Q = 20$. If labour and capital are doubled, output too doubles. In this case, the firm has constant returns to scale. Finally, if $Q = 10L^{0.5}\ K^{0.3}$, and $L = 2$ and $K = 2$, then $Q = 17.4$. If labour and capital were now doubled, *Q* would increase to 30.3, which is less than double the original output level. In this case the firm faces decreasing returns to scale. It must be emphasized that while examining the returns to scale, all inputs must change in the same direction and same proportion.

There is an elasticity concept which is related to the returns to the scale concept. This is called the all input elasticity of output and is defined as follows:

$$\begin{aligned} eQ, I &= \frac{\%\text{ change in output}}{\%\text{ change in all inputs}} \\ &= \left(\frac{\partial Q}{\partial I}\right)\left(\frac{I}{Q}\right), \text{ for continuous function} \\ &= \left(\frac{\Delta Q}{\Delta I}\right)\left(\frac{I}{Q}\right), \text{ for discrete function} \end{aligned} \tag{5.17}$$

where, $Q$ = output

$I$ = all inputs

$e_{Q,I}$ = all input elasticity of output.

If $e_{Q,I} > 1$, there is increasing returns to scale. Returns to scale are constant if $e_{Q,I} = 1$ and decreasing if $e_{Q,I} < 1$. In the Cobb-Douglas form of production function, this elasticity is given by the sum of the exponents in the function. The individual exponents in the function are measures of output elasticities with respect to corresponding input.

The returns to scale concept is of great significance in the theory of production. If an industry is characterised by increasing returns to scale, there will be a tendency for expanding the size of the firm, and thus the industry will be dominated by large firms. Quite the opposite would be true in industries where decreasing returns to scale prevail. Firms of all sizes would survive equally well in industries characterised by constant returns to scale. Incidentally, note that the type of returns to scale could change as one moves from one scale to another. For example, in the production function of Table 5.1, note the following inputs and output combinations:

| | $L$ | $K$ | $Q$ |
|---|---|---|---|
| (i) | 1 | 1 | 5 |
| (ii) | 2 | 2 | 31 |
| (iii) | 3 | 3 | 59 |
| (iv) | 4 | 4 | 72 |
| (v) | 6 | 6 | 107 |

A careful examination of these results would reveal that while there are increasing returns to scale between points (i) and (ii), (ii) and (iii), (i) and (iii), (i) and (iv), (ii) and (v); there are decreasing returns to scale between points (iii) and (v), and (iv) and (v). If we use the data in Table 5.1 and estimate the production function through the least-squares technique (Refer chapter 4), the following results would be obtained:

$$Q = 18.73\, K^{0.495} L^{0.449}$$

$$R^2 = 0.837$$

In this, the sum of the two exponents comes to 0.944, which is less than unity. Thus, over all input combinations, the Table 5.1 production function has decreasing returns to scale. Further, in the said function, elasticity of output with respect to capital = 0.495 and that with respect to labour = 0.449. The estimated production function could also be used for deriving the least-cost technology for a given output. Thus, for $Q$ =122, and $P_K$ = 50 and $P_L$ = 90, the least-cost input mix would be given by the solution of the following two equations:

(i) $122 = 18.73\, K^{0.495} L^{0.449}$

and, (ii) $\frac{MPP_L}{MPP_K} = \frac{P_L}{P_K}$, substitution of the values for MPPs as obtained through the above function and of the given factor prices, gives,

$(0.449)(18.73)(K^{0.495})(L^{-0.551})/(0.495)(18.73)(K^{-0.505})\,(L^{0.449})$

$= 90/50$

or, $0.907\,(K/L) = 90/50$ (iii)

The solution of equations (i) and (iii) will yield, $K = 9.34$ and $L = 4.72$, and total cost would equal Rs. 891.8. However, under the same input prices and the same production function, but in its discrete form (See Table 5.1), it could be easily verified that the least-cost technology for $Q = 122$ would be $K = 8$ and $L = 7$, and $TC$ = Rs. 1030. Obviously working with continuous function yields the accurate results, as under discrete data the equilibrium condition may not hold exactly.

## 5.4 PRODUCTION ANALYSIS: SHORT-RUN

The decision problem a firm faces in the short-run consists of deciding how much to produce given the capacity constraint or, to put it differently, how much of variable input(s) to employ, given the employment of fixed inputs. To answer this question, one needs a clear understanding of the relationships among the total, average and marginal factor productivities.

### Partial Factor Productivities

Factor productivities in the short-run consist of total, average and marginal physical products of each of the variable inputs. Since all inputs but one are taken as constants, these are referred to as partial factor productivities. These could be analysed for any production function in any of its functional form: table, graph and equation. Recall the short-run production function of Table 5.2. In the table, the outputs corresponding to various units of labour, holding $K = 2$, indicate the total physical product of labour ($TPP_L$) corresponding to various units of labour. The average physical product of labour ($APP_L$) and marginal physical product of labour ($MPP_L$) are defined as

$$APP_L = \frac{TPP_L}{L} \tag{5.18}$$

$$MPP_L = \frac{\partial(TPP_L)}{\partial L} \text{ (if function is continuous)} \tag{5.19}$$

$$= \frac{\Delta(TPP_L)}{\Delta L} \text{ (if function is discrete)}$$

For the production function of Table 5.2, the calculations of short-term factor productivities are illustrated in Table 5.3.

Incidentally note that once $TPP_L$ curve is known, the corresponding $APP_L$ and $MPP_L$ curves can be deduced geometrically. A point on $APP_L$ curve is given simply by the slope of the straight line passing through the origin and the corresponding point on the $TPP_L$ curve. For example, slope of line $OA = \left(\frac{L_2Q_2}{OL_2}\right)$.

$APP_L$ at $L = OL_2$. Similarly, a point on $MPP_L$ curve is given by the slope of the tangent at the corresponding point on $TPP_L$ curve. Thus, $MPP_L$ at $L = OL_1$ equals the slope of tangent at point $Q_1$ on $TPP_L$ curve, which equals $\frac{Q_1B}{BC}$. One additional point to note here is regarding the graphing of $MPP_L$ curve from discrete data on $MPP_L$ (arc) as in Table 5.3. Since $MPP_L$ in such a situation stands for arc $MPP_L$, the $MPP_L$ data must be marked on the midpoint of the corresponding arc. Thus, for example $MPP_L = 15$ of Table 5.3 must be marked at the midpoint of $L = 0$ and $L = 1$, $MPP_L = 16$ of Table 5.3 at the midpoint of $L = 1$ and $L = 2$, and so on. The $MPP_L$ graph in Fig. 5.10 is drawn on this basis.

**Table 5.3** *Short-run factor productivity (K = 2)*

| $L$ | $Q$ ($TPP_L$) | $APP_L$ | $MPP_L$ (arc) |
|---|---|---|---|
| 0 | 0 | — | — |
| 1 | 15 | 15.0 | 15 |
| 2 | 31 | 15.5 | 16 |
| 3 | 48 | 16.0 | 17 |
| 4 | 59 | 14.8 | 11 |
| 5 | 68 | 13.6 | 9 |
| 6 | 72 | 12.0 | 4 |
| 7 | 73 | 10.4 | 1 |
| 8 | 72 | 9.0 | –1 |
| 9 | 70 | 7.8 | –2 |
| 10 | 67 | 6.7 | –3 |

A graph of these various productivity is given in Fig. 5.10.

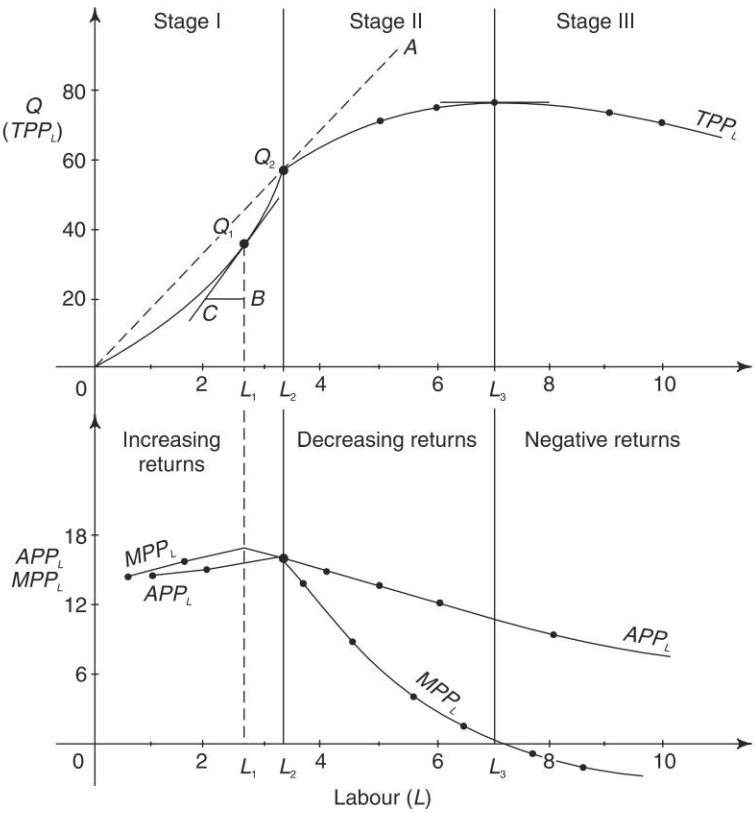



**Fig. 5.10** *Total, marginal and average physical product curves*

A careful examination of data in Table 5.3 and curves in Fig. 5.10 reveal the following interesting relationships

(a) As long as $TPP_L$ increases at an increasing rate, both $APP_L$ and $MPP_K$ increase. $MPP_L$ declines monotonically when $TPP_L$ increases at a decreasing rate.

(b) $MPP_L$ is maximum at the point of inflection (where the curvature of the $TPP_L$ curve changes from convex to concave or vice versa) on the $TPP_L$ curve ($L = OL_1$)

(c) $MPP_L = O$ when $TPP_L$ is maximum ($L = OL_3$)

(d) $APP_L$ is increasing when $MPP_L > APP_L$

(e) $APP_L$ is maximum when $MPP_L = APP_L$, or at the point where the slope of the straight line from origin to the $TPP_L$ curve is maximum ($L = OL_2$)

(f) $APP_L$ is decreasing when $MPP_L < APP_L$

(g) $MPP_L$ reaches its maximum value before $APP_L$ and $APP_L$ before $TPP_L$.

As seen in section 5.2 above, the $TPP_L$ curve is first convex from below and then concave from below because of the operation of the law of variable (diminishing) marginal returns. Since $APP_L$ and $MPP_L$ curves are simple mathematical derivations from $TPP_L$ curve, the shapes (close to inverted $U$) are also due to the working of the law of diminishing marginal returns.

Short-run factor productivities depend on the magnitudes of fixed inputs. If the quantity of any one or more of the fixed inputs employed by the firm increases, each of the three short-run productivities would increase. Thus, in the above example, if $K$ increases from $K = 2$ to $K = 3$, each of $TPP_L$, $APP_L$ and $MPP_L$ would increase, thereby there would be an upward shift in each of the three curves in Fig. 5.10. The same thing would result in the face of an improvement in the technology. Thus, the introduction of computer into business would result into an improvement in labour productivity. Note that the partial factor productivities of labour in advanced countries (like USA and Japan) are larger than those in developing countries (like India and China) due to the employment of more capital intensive techniques of production in the former than the latter. This explains why an Indian shifting to USA experiences a significant increase in his/her earnings.

## Optimum Output and Demand for Variable Inputs

In the short-run, employment of fixed resources is given, and thus there is a fixed production capacity. Within that capacity, the firm could decide to produce any output. Depending upon the amount a firm decides to produce, there will be some units of variable input(s) that the firm would hire. Thus, in the short-run production function of Table 5.3 and Fig. 5.10, the firm employs two units of capital, its capacity constraint in terms of output is $Q = 73$. The firm could produce any output in the range of 0 to 73, and depending upon its production level it would employ labour in the range of 0 to 7 units; employment of labour beyond 7 units leads to decrease in total production and that would clearly be non-economic. The question is, how much this firm should produce or how much labour should it employ? To answer this question, one must know the firm's objective, and if the objective is assumed to be that of profit-maximizing, then one needs the product's demand function and the labour supply function. In what follows, the optimization problem is first explained for the case where the product's demand function and labour's supply function are $P$ = constant and $P_L$ = constant, respectively, then for the case where product's demand varies inversely with price but labour price is still a constant, and finally for the case where both prices are variable.

Before individual cases are taken up, it will be useful to emphasize that a profit-maximizing firm

would go on employing more and more units of the variable input (labour) so long as the contribution of the additional unit exceeds its cost. The contribution of additional labour is denoted by incremental total revenue, called marginal revenue product of labour, and the cost of additional labour is measured by incremental total cost, called marginal cost of labour.

### Constant Product Price, Constant Input Price Case

This is the case where the firm could sell as much of its product as it wishes at a given price and could hire as many units of variable inputs as it wishes to at a given input price. In economics, such markets are called **purely competitive products and factor markets**, respectively. To explain the optimization under such a market, assume product price ($P$) equals 10, labour price ($P_L$) equals 90, and the short-run production function as in Table 5.3. The necessary calculations are presented in Table 5.4.

**Table 5.4** *Total, Marginal and Average Revenue Products of Labour (P and $P_L$ are Constants)*

| | $Q(TPP_L)$ | $P$ | $TRP_L$ | $MRP_L(arc)$ | $ARP_L$ |
|---|---|---|---|---|---|
| 1 | 15 | 10 | 150 | 150 | 150 |
| 2 | 31 | 10 | 310 | 160 | 155 |
| 3 | 48 | 10 | 480 | 170 | 160 |
| 4 | 59 | 10 | 590 | 110 | 147 |
| 5 | 68 | 10 | 680 | 90 | 136 |
| 6 | 72 | 10 | 720 | 40 | 120 |
| 7 | 73 | 10 | 730 | 10 | 104 |

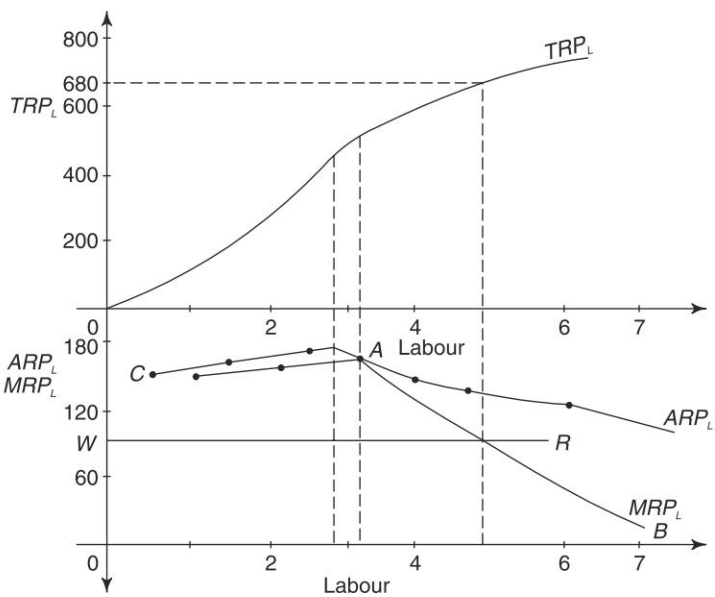



**Fig. 5.11** *Total, average and marginal products of labour curves*

Thus, if $P_L = 90$, the firm would maximize its profit at $L = 5$ and $Q = 68$. At this point, $TRP_L = 680$, total labour cost ($TC_L$) = 450 (90 × 5), and thus contribution = 230. From this contribution, the firm

has still to pay for the fixed factor (capital), the residual of which would denote the profit. Since the employment of fixed factor is constant, the point where the contribution is the most is also the point of maximum profit.

It is interesting to note that, the part of $MRP_L$ schedule where it is below the $ARP_L$ curve (i.e. where $MRP_L < ARP_L$) is also the labour demand schedule. This is because it will be easy to see that the profit maximizing firm would employ as many workers at the wage rate given by this schedule. Thus, if $P_L$ = 110, optimum $L = 4$; $P_L = 90$, optimum $L = 5$; $P_L = 40$, optimum $L = 6$; and if $P_L = 10$, optimum $L = 7$. The other part of $MRP_L$ schedule is not the labour demand schedule for if $P_L = 150$, optimum $L = 3$, and not one as given by the schedule. A graph of the various revenue products of labour is given in Fig. 5.11.

In Fig. 5.11, *AB* part of the $MRP_L$ curve denotes the **labour demand curve**. Thus, the condition for profit-maximization is $MRP_L = P_L$. There is also a sufficient condition for it but same is ignored here. Suffice to state here that the said condition would be met if the $MRP_L$ intersects the $ARP_L$ curve from above.

Thus, given the wage rate, the firm would be able to decide on its output and the corresponding labour demand. For instance, if the wage rate equals 90, *WR* would be the labour supply curve; and given a product price of 10, the firm would employ 5 units of labour, would have $TRP_L$ equal to 680, and would produce 68 units of output.

## Variable Product Price, Constant Input Price Case

This is the situation where the firm sells its products in a non perfectly competitive market (i.e., in monopoly, oligopoly or monopolistically competitive markets) but buys its labour input in a purely competitive market. For the short-run production function of Table 5.3, and the product demand function as assumed in Table 5.5, columns 2 [($TPP_L$ = Demand) and 3], the various useful calculations would be as shown in Table 5.5.

Thus, if $P_L = 150$, it would be optimum for the firm to employ $L = 1$; if $P_L = 129$, optimum $L = 2$, and so on. It should be noted that the $MRP_L$ here declines faster than the $MRP_L$ in the above case. This is because while in the earlier case, the $MRP_L$ declined just because of the law of diminishing marginal returns, now it is declining partly due to the law and partly due to the falling product price as more and more output is produced through employment of more and more labour. In terms of geometry, the $MRP_L$ of this case would be steeper than the $MRP_L$ of the case above. Also, note that while in the above case, $MRP_L = P\ (MPP_L)$, in this case the same is not true.

**Table 5.5** *Total, Average and Marginal Revenue Products of Labour (P Variable, $P_L$ Constant)*

| $L$ | $TPP_L$ | $P$ | $TRP_L$ | $MRP_L$ | $ARP_L$ |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | 15 | 10 | 150 | 150 | 150 |
| 2 | 31 | 9 | 279 | 129 | 139 |
| 3 | 48 | 8 | 384 | 105 | 128 |
| 4 | 59 | 7 | 413 | 29 | 103 |
| 5 | 68 | 6 | 408 | –5 | 82 |
| 6 | 72 | 5 | 360 | –48 | 60 |
| 7 | 73 | 4 | 292 | –68 | 42 |

## Variable Product Price, Variable Input Price Case

The large firms in an industry exercise influence not only on the product price but also on the wage rate. The equilibrium under such a situation (absence of perfect competition both from product as well as labour markets) for the short run production function of Table 5.3 is illustrated below with the help of a hypothetical product demand and labour supply schedule, as reflected in Table 5.6, columns 2 and 3, and columns 1 and 6, respectively.

**Table 5.6** *Total, Average and Marginal Revenue Products or Labour (P, $P_L$ are Variables)*

| $L$ | $TPP_L$ | $P$ | $TRP_L$ | $MRP_L$ | $P_L$ | $TC_L$ | $MC_L$ |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 1 | 15 | 10 | 150 | 150 | 90 | 90 | 90 |
| 2 | 31 | 9 | 279 | 129 | 95 | 190 | 100 |
| 3 | 48 | 8 | 384 | 105 | 102 | 306 | 116 |
| 4 | 59 | 7 | 413 | 29 | 110 | 440 | 134 |
| 5 | 68 | 6 | 408 | –5 | 120 | 600 | 160 |
| 6 | 72 | 5 | 360 | –48 | 135 | 810 | 210 |
| 7 | 73 | 4 | 292 | –68 | 160 | 1120 | 310 |

In this case the profit-maximizing firm would employ two units of labour and produce 31 units of output. The corresponding maximum contribution would equal 89 (279–190). The equilibrium condition is $MRP_L \geq MC_L$. In the case of a discrete function, the equality between $MRP_L$ and $MC_L$ may not be possible but in the case of a continuous function, the two would necessarily be equal at the optimum point. The equilibrium is illustrated graphically in Fig. 5.12.

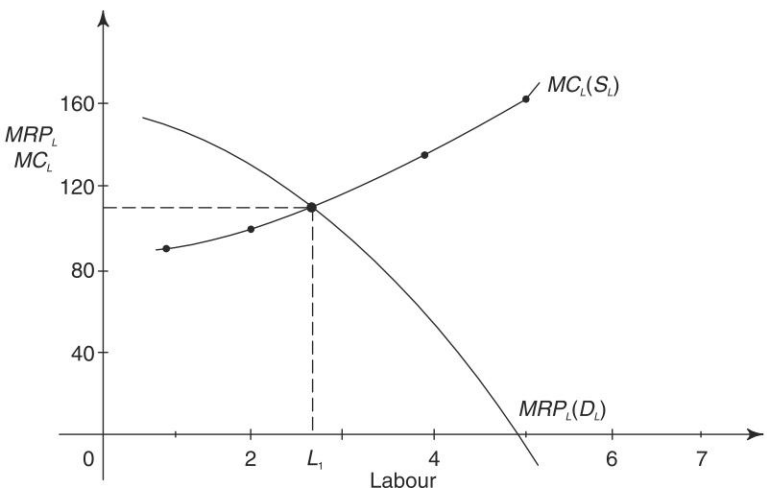



**Fig. 5.12** *Labour demand and supply curves under variable product and labour prices*

As mentioned above, the $MRP_L$ (labour demand: $D_L$) curve in Fig. 5.12 is steeper than that in Fig. 5.11. Unlike the horizontal labour supply ($S_L$) curve of Fig. 5.11, there is an upward sloping labour supply curve here, for the wage rate is assumed to increase as the firm hires more labour. The equilibrium

employment equals $OL_1$, and the equilibrium level of output would be given by the short-run production curve ($TPP_L$) corresponding to this level of output. Incidentally note that there is a slight discrepancy between the equilibrium employment level as found through a tabular approach ($L = 2$) and as found here through a graphical method ($L = OL_1 > 2$). This is only due to the discrete nature of the table in contrast to the continuous nature of the graph.

Although this is not a place to go into the detailed derivation of the **labour supply curve**, it may be mentioned that the same comes from the workers optimization procedure. Under this procedure, labour faces the following constrained optimization problem:

Maximize $U = f(Y, F)$ (5.20)

$U_1, U_2 > 0$

Subject to constraint $T = L + F$ (5.21)

And where $Y = LW$ (or $L = Y/W$) (5.22)

The notations are, $U$ = utility from income and free time (called leisure), $Y$ = income from labour (work), $F$ = free time (called leisure), $T$ = total time available for the allocation between work and leisure, $L$ = time worked, and $W$ = wage rate.

In the above optimization problem, $T$ is the constraint, $L$ and $F$ are variables to be determined; the values of $Y$ and $U$ would be given simply by Eqs. (5.22) and (5.20), respectively. Once the utility function is specified for the particular worker and his $T$ value known, the values of $L$ and $F$ (and subsequently of $Y$ and $U$) could be determined using the indifference curve approach or the constraint optimization (Lagrangian multiplier) technique, as discussed in Chapter 3. The so obtained solution would yield the labour supply curve as a function of the wage rate ($W$) and time ($T$). The labour supply curve would be an upward sloping curve up to a certain wage rate and backward bending beyond that rate, caused by the substitution and income effects of a change in the wage rate on labour supply.

**Factor prices** Given the labour demand curve and labour supply curve, the intersection of the two would determine the real wage rate. Through the similar procedure, the intersection of the demand and supply curves for loanable funds would determine the interest rate.

## 5.5 PRODUCTION ANALYSIS: LONG-RUN VERSUS SHORT-RUN

It will be interesting to compare the optimum input combinations of long-run with those of short-run. This is illustrated for the least-cost input combination in Fig. 5.13.

The long-run expansion path is given by the curve *OABCDE*. If capital input was fixed at $K = \overline{K}$, the short-run expansion path will be given by the curve *OXYCZ*. Thus, while point *A* denotes the least-cost input combination for output level of $Q_1$ in the long-run, point *X* denotes the same in the short-run. This is because if the firm operates at point *A* in the short-run it would use $OL_2$ units of labour and $OK_1$ units of capital, but since it has fixed units of capital, it would have to pay for $OL_2$ units of labour and $O\overline{K}$ units of capital, leaving $K_1\overline{K}$, units of capital unemployed. In contrast, if it operates at point *X*, it would pay for a fewer units of labour ($= OL_1 < OL_2$) and the same units of capital as at point *A*. Thus, assuming the firm has no other use of extra units of capital, the short-run least-cost input combination for $Q = Q_1$ is given by point *X* and not by point *A*. By the same reasoning, the short-run least-cost technologies

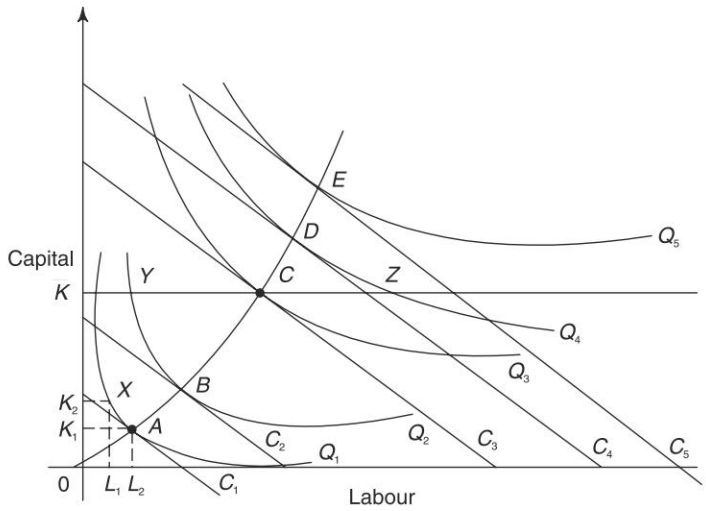



**Fig. 5.13** *Short-and long-run expansion path*

for output levels $Q_2$, $Q_3$ $Q_4$ are given by points *Y*, *C*, and *Z*, respectively. Note that there is no way by which the firm could produce $Q_5$ units of output in the short-run, for $Q_5$ isoquant lies above the capital constraint ($\overline{K}$). In other words, $Q_5$ is above the firm's capacity output in the short-run.

It is interesting to state that one and only one of the least-cost (or maximum output) input combinations of the long-run is also the least-cost input combinations of the short-run; here it is represented by point *C* and it is for $Q = Q_3$. At this point alone, long-run total cost equals short-run total costs. For all other output levels; short-run total cost exceeds long-run total cost. This point can be seen easily by observing that the cost line passing through point *X* would be above the one passing through point *A* (both cost lines would have the same slope, for factor prices are given), the one passing through point *Y* would be above the one passing through point *B*, and so on.

## 5.6 ELASTICITY OF FACTOR SUBSTITUTION

Various factors of production are substitutes to some extent and their relative use depends on their relative prices, among other things. Elasticity of factor substitution measures the degree of sensitivity of the relative factor demands to their relative prices. Thus, if the substitutability between labour and capital is examined, elasticity of factor substitution ($e_s$) would be given by

$$e_s = \frac{\%\text{ change in } K/L}{\%\text{ change in } P_L/P_K}$$

$$= \left[\frac{\Delta(K/L)}{\Delta(P_L/P_K)}\right]\left[\frac{P_L/P_K}{K/L}\right] \tag{5.23}$$

Since the elasticity is defined in positive terms, note that in equation (5.23), while the numerator has *K*/*L* the denominator has $P_L/P_K$. One could, of course, use the inverse of these ratios as well.

To illustrate its calculation, consider the following two input-factor price combinations

| | $K$ | $L$ | $P_L$ | $P_K$ | $K/L$ | $P_L/P_K$ |
|---|---|---|---|---|---|---|
| (i) | 6 | 4 | 5 | 15 | 3/2 | 1/3 |
| (ii) | 8 | 2 | 5 | 10 | 4 | 1/2 |

The elasticity of substitution for this example at point (i) would be

$$e_s = \left(\frac{4 - \frac{3}{2}}{\frac{1}{2} - \frac{1}{3}}\right)\left(\frac{\frac{1}{3}}{\frac{3}{2}}\right)$$

$$= \frac{5}{2} \times \frac{6}{1} \times \frac{1}{3} \times \frac{2}{3}$$

$$= \frac{10}{3}$$

Like demand elasticities, one can compute similarly the elasticity of factor substitution at point (ii) or on the corresponding arc. The concept is useful to measure the effect of a change in the relative price of inputs on their relative employment. Thus, the policy makers and firms managers could assess the likely impact of changes in the wage rate and/or the interest rate on employment (of labour).

## 5.7 PRODUCTION ANALYSIS AND INPUT DEMAND: A GENERALIZATION

The production analysis presented above assumed just two inputs, labour and capital. However, in real life, there are many inputs, partly because land and organization are certainly treated as two other factors of production and partly because some time one or more of these four factors of production is classified into two or more categories, such as skilled and unskilled labour, irrigated and unirrigated land. In what follows, the results of production analysis of two input case is generalized to $n$ input case.

The least-cost input combination could still be determined through a tabular analysis. To explain this, assume there are four factors of production and four alternative technologies to produce 20 units of a product:

| *Technology* | *Land* | *Units of Labour* | *Capital* | *Entrepreneur* |
|---|---|---|---|---|
| 1 | 2 | 1 | 5 | 3 |
| 2 | 4 | 2 | 2 | 1 |
| 3 | 4 | 1 | 1 | 4 |
| 4 | 4 | 3 | 2 | 1 |

Assume further that prices of land, labour, capital, and entrepreneur are fixed (the assumption of perfect competition in the factors markets) at 4, 5, 3 and 2, respectively. To determine the least-cost technology, one could simply compute total costs under each of the four input combinations and the one which is associated with the least-cost would be the desired input combination. Thus, the total cost under the four alternatives would be

| *Technology* | *Total Cost* |
|---|---|
| 1 | $(2 \times 4) + (1 \times 5) + (5 \times 3) + (3 \times 2) = 34$ |
| 2 | $(4 \times 4) + (2 \times 5) + (2 \times 3) + (1 \times 2) = 34$ |
| 3 | $(4 \times 4) + (1 \times 5) + (1 \times 3) + (4 \times 2) = 32$ |
| 4 | $(4 \times 4) + (3 \times 5) + (2 \times 3) + (1 \times 2) = 39$ |

Since technology 3 costs the least, it gives the least-cost input combination for the given output level ($Q$ = 20). The method is obviously applicable to any number of factors of production problem. In the same way, the maximum-profit input combination for a given production cost can be determined if the necessary information is available in a tabular form.

The graphical method of determining the optimum input combination is not available for many inputs cases. The equation method poses no problem. Under $n$ input case, the necessary condition for the least-cost input combination for a given output or the maximum-output input combination for a given cost, under the fixed product price (assumption of perfect competition in the product market), would be the following:

$$\frac{MPP_1}{P_1} = \frac{MPP_2}{P_2} = \ldots = \frac{MPP_n}{P_n} \tag{5.24}$$

where subscripts 1, 2, …, $n$ stands for various factors of production. Equation system (5.24) has ($n$ – 1) equations in $n$ input variables. Besides, there is an equation of the constraint output/or constraint cost. Thus, there are $n$ equations in $n$ variables, which can be solved to determine the desired optimum input-mix.

The necessary condition for profit-maximization under no constraint of output or cost and given factor prices would be the following:

$$\frac{MRP_1}{P_1} = \frac{MRP_2}{P_2} = \ldots = \frac{MRP_n}{P_n} = 1 \tag{5.25}$$

Equation system (5.25) has $n$ equations in as many variables (factors of production) and so unique solution for the profit-maximizing input combinations could easily be derived from them.

If product price as well as the factor prices are variables (assumption of the absence of perfect competition in both the product and factor markets), the condition for profit-maximizing input-mix would change to

$$\frac{MRP_1}{MC_1} = \frac{MRP_2}{MC_2} = \ldots = \frac{MRP_n}{MC_n} = 1 \tag{5.26}$$

Equation systems (5.25) and (5.26) are valid both for long-run and short-run profit-maximizing problems. In the case of the long-run, all factors of production are involved in these equations, while in the case of the short-run only the variable inputs are involved in them.

It is interesting to remind that while a profit-maximizing input combination is necessarily a least-cost (or maximum-output) input combination, a least-cost input combination may or may not be the profit-maximizing one. This is easy to see from Fig. 5.9, which shows several least-cost input combinations, one for each output level, and only one of them as a profit-maximizing input-mix. It can also be seen through equation systems (5.24) and (5.25). The two equation systems differ on two counts. One, while in the former the ratios of marginal products and prices of each factor must be in exact proportions only, in the latter they must be in the same proportion and must equal to unity. Second, in equations system (5.24), the marginal products are in physical terms while in equations system (5.25), they are in nominal (rupee) terms. If product price ($P$) is a constant; the two makes no material difference, for in that case ($MRP$) = ($P$) ($MPP$) and $P$ will cancel out from all equations in Eq. (5.25). However, if $P$ is a variable, $MRP \neq (P)\ (MPP)$ and this would mark another difference between Eqs. (5.24) and (5.25). Thus, Eq. (5.25) imply Eq. (5.24) but not necessarily vice versa.

## REFERENCES

1. Cobb, C.W. and P.H. Douglas (1928): "A Theory of Production," *American Economic Review* XVIII (Sept.).
2. Gupta, G.S. (1983–84): "Production Function and Optimum Input Mix in Fish Farming in India," *Vishleshan*, IX (Dec.–March).
3. Gupta, G.S. and K. Patel (1976): "Production Function in Indian Sugar Industry," *Indian Journal of Industrial Relations*, II (Jan.).
4. Hopper, W.D. (1962): "The Economics of Fertilizer Use—A case study in Production Economics," *Indian Journal of Agricultural Economics*, XVII (Oct.).
5. Kendrick, J.C. (1961): *Productivity Trends in the United States*, Princeton, Princeton University Press.
6. Pappas, J.L., E.F. Brigham and M. Hirschey (1982): *Managerial Economics,* 4th edition, Chicago, Dryden Press.
7. Walters, A.A. (1963): "Production and Cost Functions", *Econometrica*, XXXI.

## CASELETS

1. Patel Brothers own a large farm, on which they have been growing wheat. Their past experience suggests that production varies with the doses of fertilizers and irrigation that they use and with no other inputs. In particular, they have observed the following relationships among the two inputs and wheat production (Numbers in the box of the matrix denote output levels):

| *Units of fertilizer* | *Units of irrigation (I)* | | | | | |
|---|---|---|---|---|---|---|
| (F) | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | 10 | 20 | 29 | 37 | 44 | 50 |
| 2 | 20 | 38 | 46 | 53 | 60 | 62 |
| 3 | 29 | 55 | 60 | 68 | 75 | 77 |
| 4 | 37 | 61 | 75 | 82 | 89 | 88 |
| 5 | 44 | 75 | 89 | 95 | 102 | 98 |
| 6 | 50 | 89 | 102 | 107 | 114 | 107 |
| 7 | 55 | 102 | 114 | 118 | 120 | 115 |

The price of fertilizer is Rs. 2000 per unit and that of irrigation is Rs. 1500 per unit. Further, assume that the above input-output relationships are just illustrative points on a spectrum of continuous input combinations.

(a) Draw isoquants for at least two different output levels for Patel Brothers (smoothen the curves).

(b) Suppose the firm is currently producing 75 units of the output, employing $F = 4$ and $I = 3$. Is the firm using the least-cost input combination? Explain.

(c) If the firm intends to produce 102 units of its output, determine its least-cost technology.

(d) Choose some two sets of appropriate values for inputs and outputs from the table and examine the type of returns to scale within your selected scales.

(e) Is the production function described in the table a short-run or a long-run production function? Why?

(f) Suppose Patel Brothers decide to freeze irrigation at 6 units, answer the following questions:

(i) Where do diminishing returns to fertilizer set in?

(ii) If the price of wheat is Rs. 200 per unit and Patel Brothers aims at maximum profit, how many units of fertilizer must they employ in the cultivation of wheat?

(iii) Assume that there is imperfect competition in the wheat market, meaning that the more the firm wants to sell (by using more units of fertilizer), the less would be the price of its product. In particular, wheat price is Rs. 500 if it employs just one unit of fertilizer and it falls by Rs. 50 for every extra unit of fertilizer employed until the saturation point is reached, beyond which the price stays constant. Determine the units of fertilizer the profit-maximising firm must employ.

2. Prakash and Sons operates on a technology described by the following production function:

$$Q = L^2 + 5LK + 4K^2$$

where $Q$ = quantity of output

$L$ = quantity of labour

$K$ = quantity of capital

(a) Determine its isoquant for $Q = 100$.

(b) Derive the marginal physical product of labour function if $K = 10$.

(c) If $K = 10$, where do diminishing returns to labour set in?

(d) If the price of labour equals Rs. 5 and that of capital equals Rs. 10, and the company wishes to produce 45,000 units of its production, determine the least-cost input combination that the company must employ.

(e) If the input prices were the same as in part (d) above, and the company had a fixed production cost budget of Rs. 1,000, how many units of labour and capital must it employ, assuming that its objective were that of profit maximization and that the price of its product was fixed irrespective of the quantities it produces?

(f) Compare your results of parts (d) and (e) and comment.

3. Given below are inputs and output data for "All Industries" in India.

| *Year* | *Net Value Added (Rs. Billions at 1993–94 prices)* $Q$ | *Fixed Capital (Rs. Hundred Billions at 1993–94 prices)* $K$ | *Persons Employed (Numbers in lakhs)* $L$ |
|---|---|---|---|
| 1981–82 | 353 | 8.26 | 7.89 |
| 1982–83 | 391 | 9.49 | 8.17 |
| 1983–84 | 446 | 10.83 | 7.99 |
| 1984–85 | 432 | 11.62 | 7.98 |
| 1985–86 | 441 | 11.78 | 7.58 |
| 1986–87 | 481 | 12.57 | 7.55 |
| 1987–88 | 497 | 14.11 | 7.90 |
| 1988–89 | 556 | 14.08 | 7.86 |
| 1989–90 | 616 | 15.30 | 8.26 |
| 1990–91 | 685 | 17.65 | 8.28 |
| 1991–92 | 655 | 17.34 | 8.32 |
| 1992–93 | 768 | 19.90 | 8.84 |
| 1993–94 | 884 | 22.44 | 8.84 |
| 1994–95 | 967 | 26.22 | 9.23 |
| 1995–96 | 1144 | 31.71 | 10.22 |
| 1996–97 | 1265 | 32.85 | 9.54 |
| 1997–98 | 1300 | 36.69 | 10.07 |
| 1998–99 | 1089 | 33.72 | 8.66 |
| 1999–2000 | 1130 | 34.61 | 8.17 |
| 2000–01 | 1014 | 32.51 | 7.99 |
| 2001–02 | 1000 | 33.46 | 7.75 |
| 2002–03 | 1163 | 34.13 | 7.94 |
| 2003–04 | 1298 | 35,67 | 7.87 |
| 2004–05 | 1564 | 36.62 | 8.45 |
| 2005–06 | 1820 | 41.20 | 9.11 |

*Source:* Annual Survey of Industries, Various issues (www.mospi.nic.in)

Based on the above data, the following production function was estimated in the log (natural) linear (Cobb-Douglas) form:

$$Ln\ Q = 2.91 + 0.87\ Ln\ K + 0.52\ Ln\ L \qquad [\text{or, } Q = 18\ K^{0.87}\ L^{0.52}]$$

$$R^2 = 0.97$$

Based on the data and the said estimation results, attempt the following questions:

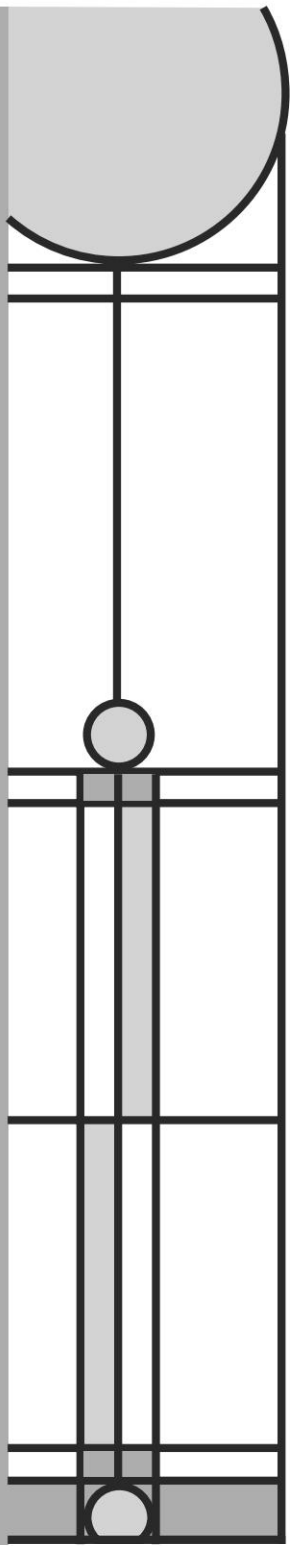
(a) What is the elasticity of output with respect to labour? What kinds of returns to scale your results suggest? Derive the estimates for a short-run production function with $K$ =10. Is the estimated function consistent with the law of diminishing marginal returns to labour? How?

(b) Estimate the function in its quadratic form, like $Q = a + bK + cL + dKL + eK^2 + fL^2$ (where small letters $a$, $b$, $c$, $d$, $e$, and $f$ are parameters). Examine the validity of the law of diminishing marginal returns to labour in this estimation results. Calculate the mean (i.e. at the mean value of $K$, $L$ and $Q$) elasticity of output with respect to capital. Compare the said elasticity result here with that obtained under the Cobb–Douglas form above and comment.

(c) Using the estimation results of Cobb–Douglas form, derive the production function in the form of a table like those in Table 5.1 in the text, for each integer value of $K$ and $L$ between 1 and 10.

(d) For the capital rental = Rs. 10 billion per unit (= 10 % of $K$) and wage rate = Rs 25 billion per year per one lakh of workers (= Rs 2.5 lakhs/person/year), and the estimated production function of Cobb–Douglas form as given above, find out the least-cost input combination for output = Rs.1820 billions (= actual output in 2005–06). How does your finding compare with the actual data? Comment.

The McGraw-Hill Companies



CHAPTER

6

# COST AND SUPPLY ANALYSIS

Production and cost analysis constitute the supply side of the market. While the production analysis, as presented in Chapter 5 deals with the supply side in terms of physical units of inputs and output, the cost analysis is concerned with the supply side in terms of physical units of output and the monetary (rupee) units of cost of production.

The significance of cost analysis in decision-making needs no exaggeration. The cost of production provides a floor to product pricing and no firm can afford to ignore its profits, which is the excess of total revenue over total cost. With the emergence of increasing competition over time, both from the domestic and foreign markets, there is a pressure on every firm to reduce its unit cost. A firm which is able to achieve some degree of success on this front is always able to shine in the market enjoying a large market share, high turnover and high profits. Further, production cost, looked from another side, stands for payments made to factors of production, which is of immense significance from the point of view of income distribution, national planning, etc.

The supply function is derived from cost and firms' objective functions. This function together with the demand function determines products price.

## 6.1 COST CONCEPTS

There are several useful cost concepts and a clear understanding of them must precede any discussion on cost analysis.

### Economic Cost

Cost of anything in economics means what one has to give up in getting that particular thing, which is same as the opportunity cost. Accordingly, economic cost that a firm incurs in the production of a good refers to the payments it must make to all the resources (factors of production) employed by it in the production of that good. The resources referred here include the ones owned by the firm itself as well as

those which it hires from outside for the purpose. To illustrate this, consider a firm which produces 10 tonnes of wheat by employing the following resources, costing the amounts as indicated against them:

| *Factors hired from outside* | *Cost (actual)* |
|---|---|
| | (Rs.) |
| Seed | 7,500 |
| Labour | 19,000 |
| Tractor for ploughing | 20,000 |
| Fertilizer | 11,000 |
| Tubewell for irrigation | 12,500 |
| Sub-total | 70,000 |
| *Self-owned Factors Employed* | *Cost (imputed)* |
| | (Rs.) |
| Family labour | 35,000 |
| Land | 50,000 |
| Sub-total | 85,000 |
| Total | 155,000 |

The true cost of hired resources is what one pays for them, and thus is known. The cost of self owned resources is not known exactly but would have to be imputed on the basis of their opportunity cost to the producer, as explained in Chapter 1, section 1.5. In the above example, data are derived accordingly. The economic cost of 10 tonnes of wheat thus stands at Rs. 155,000. Many small firms ignore the cost associated with self owned resources, and thus they underestimate their cost of production and thereby overestimate their economic profits. Even in the corporate world, cost of equity capital is ignored (due to the difficulty in estimating it) while arriving at net profit and thus the said profit is gross of the cost of equity.

Production cost does not include either the **research and development (R&D) cost** or the **selling cost**. The former is the cost incurred in developing the technology which enables the production of a particular good. This could be quite enormous and while some R&D leads to significant outcomes, lots of that brings no benefits. It is for this reason that companies who discover new products are granted **patent rights** so that they could recover the R&D cost and have incentives to carry out such activities. The selling cost is of two types, viz. cost incurred in getting orders and that incurred in executing orders. The former includes cost of advertisements and promotions, and the latter that of packing, transport, insurance, etc. In micro economics, we focus largely on the production cost.

## Explicit and Implicit Costs

Economic costs are classified into two types: explicit and implicit costs. The former stands for the payments that must be made to the factors hired from outside the control of the firm. In contrast, implicit cost refers to the payments due to (or opportunity costs of) the self-owned resources used in the production. Thus, in the above example, explicit cost of wheat production is Rs. 70,000 and its implicit cost is Rs. 85,000. The only difference between these two cost concepts is in terms of whether the amount spent is on hired factors or no self-owned ones.

A somewhat similar distinction is drawn between **cost and cash payments** (**out-of-pocket payments**). While some costs involve cash disbursements, others do not. For example, all implicit costs as well as depreciation cost are devoid of cash payments. Also, all cash payments are not costs. For example, dividend paid out in cash as well as cash advances (loans) to employees are not costs of production. Often, business people get confused in such differences and end up in wrong decisions.

### Opportunity Cost

The opportunity cost concept was discussed in detail in chapter 1, section 1.5, and the readers are advised to look for it there.

### Normal Profit

Normal profit is an economic Jargon. It is not profit but a component of economic cost. In particular, it stands for the opportunity cost of the entrepreneur's time. To extend this a little further, the cost of family labour is referred to as implicit wage, the cost of the use of self-owned premises in the business as implicit rent, the cost of self-owned money in the business as implicit interest, and the cost of owner's time as implicit cost of entrepreneurship or normal profit. Thus, the normal profit is a part of implicit cost. In the above example of wheat, the cost of family labour (implicit wage) is inclusive of normal profit, for family labour in that includes the time spent on entrepreneurial functions.

### Historical and Replacement Cost

The historical cost of an asset refers to the actual cost incurred at the time the asset was acquired. In contrast, the replacement cost stands for the cost which must be incurred if the asset is to be purchased today. The two concepts differ due to price variations over time. During inflationary conditions, as witnessed today, the replacement cost exceeds the corresponding historical costs, and quite the opposite holds good during deflationary situations.

Conventionally, balance sheets are cast on the basis of historical costs. Also, while calculating costs for use in completing a firm's income tax returns, accountants are required by law to list the actual rupee amounts spent to purchase the labour, raw-materials, capital equipments, etc., used in production. For managerial decisions, however, historical costs may not be appropriate. Instead, replacement costs are relevant for these purposes. For example, suppose an oil mill has an inventory of 10 tonnes of groundnut purchased at a price of Rs. 20,000 per tonne. Groundnut price now increases to Rs. 30,000 a tonne. If this firm is asked to bid a price for its groundnut oil, what cost should it assign to the groundnut to be used in the deal—the historical cost of Rs. 20,000 a tonne or replacement (current) cost of Rs. 30,000 a tonne? Obviously, the latter, for it must buy groundnut at the new price to replace the old stock; alternatively, it could sell the groundnut at the new price if it elects not to use it on the proposed project. Thus, Rs. 30,000 is the relevant cost of groundnut per tonne for purposes of bidding on the job. However, the cost of groundnut for tax purposes is still Rs. 20,000 per tonne.

Note that while historical cost of an asset is known as it is actually incurred, replacement cost may not be known unless proper inquiries are made in the market. Further, for fairly old plants/machines, exact replacement costs may not be known at all, for they may no longer be available in the market. In such circumstances, the management will have to be content with their approximate (imputed) values.

## Incremental and Sunk Costs

Incremental costs are costs incurred when output is increased while sunk costs are costs which do not depend on output. In particular, sunk cost is an expenditure that was made or committed to in the past and is now irrelevant. Thus, such costs are lost forever as soon as they have been incurred. Accordingly, sunk cost is non-avoidable and is irrelevant for decision-making. Suppose a firm hires a consultant at a fee of Rs. 100,000 to advise on the economics of fish farming on an erstwhile agricultural land. The consultant recommends that by switching over to fish farming, the firm would make a profit of Rs. 50,000 over and above what it could make in agriculture. For the sake of argument, assume that the extra profit of Rs. 50,000 is over the life time with due care of time value of money. Should the firm switch over to fish farming? The answer is yes, for the consultant fee of Rs. 100,000 is a sunk cost, which has been incurred irrespective of the decision. If sunk cost is added to the cost of fish farming, fish farming would turn out to be less profitable than agriculture. But this would be a wrong calculation. Consider another example. Suppose there is a university which runs all its activities during the day time. It is considering to launch an evening programme on commercial basis. The evening programme is intended to use faculty time, administrator's time and sometime of the clerk-cum-peon, for whom extra payments would have to be made for their services. In addition, there will be some cost on electricity, chalk, etc. Besides, the class room of the university would be utilized for the purpose. What are the incremental and sunk costs of the said programme? The answer is simple. The cost of the time of faculty, administrator, clerk-cum-peon, and the amount spent on electricity bill, chalk, etc. will be the incremental costs and the cost of the use of class room and blackboard would be the sunk cost. Similarly, if a movie flops, all costs of making it are sunk costs; and if someone writes a book and gets no market, all cost associated with it is sunk cost. However, factory building or a class room is not a sunk cost!

The costs relevant for decision-making are incremental costs only. Since sunk costs do not depend on the decision, they must be ignored in decision-making. Quite often, managements get mixed up between these two types of costs and, if so, they might err in decision making.

## Expenditure and Cost

At this point, it is important to understand the distinctions between expenditure and cost**.** While expenditures are incurred by firms, costs are associated with outputs. Thus, all the amounts that a firm spends on various items, including structures, equipments, inventories, raw-materials, wages, interest, taxes, etc. are expenditures of the firm but all such payments are not costs. To give a precise definition, only those expenditures which are consumed in the production process are costs. Since the fixed capital (structure, equipment and inventories) is only partly consumed in production, only the consumed part of it is cost. The consumed part is called the **capital rental**, which includes depreciation, interest paid or foregone on the amount of money invested in fixed capital minus the gain on the value of capital due to inflation or loss due to deflation in the price of capital. When fixed capital is on rental, all rent charges on those items are the cost of production. In contrast, since raw-materials, intermediate goods, etc. are fully consumed in the production, as also the services of labour; all expenditures on such items are costs of production. Even if some fixed capital item (like computer) is lost in theft or damaged in fire, expenditure on that is not the cost of production, for it has not been used in production. Accounts would classify the latter as "**loss**".

## Fixed and Variable Costs

The distinction between fixed and variable costs is peculiar in short-run. In the short-run, there are some costs which do not vary as output varies, while there are others which move up and down as output increases and decreases. The former are called the fixed costs and the latter the variable costs. Recall that under short-run production analysis (chapter 5), we made a distinction between fixed and variable factors of production. The costs associated with fixed factors are called the fixed costs and the ones associated with variable factors, the variable costs. Thus, if capital is the fixed factor, capital rental is the fixed cost and if labour is the variable factor, wage bill is the variable cost.

Coming to the usual cost items of a firm, one could say that fixed costs include interest on fixed capital employed (both borrowed and net worth), rental expense on leased buildings, plant and equipment, if any, depreciation charges, associated with technical obsolescence of assets, property taxes, if any, and wages and salaries of permanent employees, etc. In contrast, variable costs include cost of raw-materials and intermediate goods, interest on working capital, depreciation charges associated with wear and tear of assets, wages and salaries of temporary employees, variable portion of charges on electricity, water, etc. and sales commissions. Undoubtedly, there are difficulties in classifying all costs into fixed and variable. For example, depreciation cost is partly a fixed cost and partly a variable. However, it is not easy to see how much of it is due to obsolescence and hence fixed cost, and how much is due to the use of equipments and hence variable costs. Similar problems arise with regard to labour cost, for it is not easy to decide who is a permanent employee and who is a temporary employee, and with regard to sales commission. Nevertheless, as will be evident later through break-even analysis and product pricing, the distinction is very important and every firm observes it to the best of its ability.

The distinction between fixed and variable cost is similar to that between sunk and incremental costs. Like sunk costs are irrelevant for decision-making, fixed costs are irrelevant for decision-making but in the short-run only. In the long-run there are no fixed costs, and thus note that fixed costs are sunk costs in short-run only. The proportion of these costs in total production cost varies significantly across industries. For example, in software, salaries of IT (information technology) professionals is the major component and so the sunk cost is relatively high; in personal computers, cost of components (memory chips, micro processor, hard disk drivers, storage drivers, video and sound cards, etc.) is high and thus variable cost component is proportionately high; while in pizza, fixed cost component is relatively high due to the costs associated with owners' time, rent and utilities.

## Separable and Common Costs

Costs are also classified on the basis of their traceability. The costs which can be attributed to a product, a department, or a process are the separable costs, and the rest are non-separable or common costs. For example, in a multi-product firm, raw-materials cost is separable product-wise but the management cost is not separable that way. Similarly, in a university, while professors' costs are separable department-wise, vice-chancellor's cost is not separable that way. In an educational institution like Indian Institute of Management, even the Professors' costs are not separable by its academic activities for professors' work comprises of teachings in several degree and training programmes, research, consulting, and administration.

The distinction between separable and common costs (which is also referred to as **direct and indirect costs**, respectively) is of significance particularly in a multi-product firm. As we shall see later, costs play an important role in pricing and thus unless they are classifiable into this category, it will

not be possible to estimate production cost by product in a multi-product firm and consequently there would be difficulties in setting economic prices for different products. In actual practice, firms allocate common costs among various departments, processes or products through their approximate uses in various activities, which is often judged on the basis of their relative turnovers.

## Private and Social Costs

Private costs refer to the costs incurred by an individual firm while social costs stand for the costs incurred by the society as a whole. The former is the sum total of explicit and implicit costs that a firm incurs in the production of a good. These differ from the costs incurred by the society on two counts, viz. exclusion or inclusion and **externalities**. There are costs like taxes, which are costs to the firm but not to the society. Likewise, subsidies that a firm receives result in cost-saving to a firm but not to the society. Externalities refer to the side (also called the third party) effects which the working of a firm creates on the society. For example, it might lead to air, water or noise pollution, traffic congestion, and accidental hazards (Like that of Union Carbide in Bhopal). These are costs to the society, though not to the firm. A firm could also lead to some social benefits or negative social costs. For example, development of a swimming pool, dam or rose garden would result in an increase in the value of surrounding properties, and thereby create social benefits. Although, both exclusion/inclusion and externalities give rise to social costs and social benefits, they do not necessarily cancel out. Consequently, the distinction between private and social cost is significant. However, since managerial economics deals primarily with decision-making by firms, social costs are not quite relevant from our point of view.

## Total, Average and Marginal Costs

Total cost (*TC*) is the sum total of explicit and implicit costs. Thus, in the above wheat example, the total cost of 10 tonnes of wheat is Rs. 155,000. Just as there are average and marginal revenues, and average and marginal productivities, there are average and marginal costs. The former stands for per unit cost and can be computed simply by dividing total cost by the quantities of output produced. Thus, average cost (*AC*) for wheat is Rs. 15,500 per tonne. The marginal cost (*MC*) constitutes the change in total cost as output changes by an infinitesimally small unit. In case of discrete data, there is only arc marginal cost, which is defined as the change in total cost divided by change in output. For computing even arc marginal cost, one needs data on two different levels of output and corresponding total cost.

The total cost concept is useful for break-even and profit analysis. The average cost concept is of relevance for estimating profit margin per unit of sales. The marginal cost concept is of significance in deciding the optimum level of output. Chapters on pricing would provide ample illustrations on the significance of these three cost concepts.

## Long-Run and Short-Run Costs

Long and short-run costs are related to long and short-run production functions. The former refers to costs when all factors of production are subject to change, while the latter stands for costs when at least one of the factors of production is fixed. The fixed factor(s) is usually taken to limit the production capacity. Thus, long-run costs refer to costs across all possible production capacities while short-run costs stand for costs within a given production capacity.

In the long-run, all costs are variable costs while in short-run, some costs are fixed and some are variable. Combining this classification with the earlier one between total, average and marginal costs, there are three long-run cost concepts and seven short-run ones. The former consists of long-run total cost (*LTC*), long-run average cost (*LAC*) and long-run marginal cost (*LMC*); and the latter comprises of short-run total cost (*STC*), short-run average total cost (*SAC*), short-run marginal cost (*SMC*), short-run total fixed cost (*TFC*), short-run average fixed cost (*AFC*), short-run total variable cost (*TVC*), and short-run average variable cost (*AVC*). It is obvious that

$$STC = TFC + TVC$$
$$SAC = AFC + AVC$$

Both long-run and short-run costs are useful for decision-making. In the long-run, a firm is concerned with the optimum plant (firm) size and in the short-run, it is concerned with optimum output within a given plant size. Such decision issues will be examined in the forthcoming sections.

## 6.2 COST FUNCTION

Costs which a firm incurs in the production of a good or service depends basically on two functions:

(a) Firm's production function

(b) Market's inputs' supply functions

As seen in the previous chapter, production function specifies the technical relationship between combinations of inputs and the level of output. Given this relationship and input prices (if they are fixed for the firm), one can easily determine the costs associated with different levels of output. The costs would thus vary as output level varies, nature of production function varies, or factor prices change. The nature of a production function is tantamount to factor productivities (efficiencies). In addition to factors (a) and (b) above, there are some minor factors which exert influence on production cost. These include the extent of the firm's experience in the particular business (called **Learning Curve**) and the role of **scope economies**. Usually, one learns by experience and thus firms which have more experience, have relatively less cost compared to those having less experience. For example, cost of providing MBA education may be low in an older institute like Indian Institute of Management (IIM), Ahmedabad as compared to a relatively new IIM, Indore. Also, cost in a multi-product firm may be lower than a single product firm if the products of the former firm were complimentary in production (two goods are complimentary in production if their production requires similar technology and/or resources). Thus, cost of milk of AMUL at Anand, Gujarat, may be lower than that of Mehsana dairy, Gujarat because the former has many more complimentary products than the latter. Putting all this together, we have the following cost function:

$$\begin{aligned} C &= f(Q, E_I, P_I, L, S, Z) \\ & f_1, f_3 > 0 > f_2, f_4, f_5 \\ & f_6 = ? \end{aligned} \tag{6.1}$$

where

$C$ = Total (production) cost

$Q$ = Total output

$E_I$ = Efficiencies of inputs

$P_I$ = Prices of inputs

$L$ = Learning/experience curve effect

$S$ = Scope economies effect

$Z$ = Other (including government policy-taxes and subsidies) determinants

The total cost is obviously an increasing function of output, for "there is no free lunch". Increasing production, *ceteris paribus*, requires increasing units of inputs and all inputs carry price tags. However, as would be explained later, the relationship need not be linear. Improvements in factor productivities, other things remaining the same, have a depressing effect on input requirements per unit of output, and since inputs have price tags, it leads to decrease in total cost. It must be noted that factor productivities depends on the level of technology (use of computer, modern plants and equipments, etc.), the quality of the work force and management, which are influenced by education, training and health conditions, and sincerity and integrity of the labour and management, which are reflected in absentism, strikes, lock outs and fooling around during working hours. Thus, through factor efficiencies, many factors exercise influence on the cost of production.

Since no output is possible without an input (factors of production), an increase in input price, other things remaining the same, would lead to an increase in the cost of production. The price of an input, like any other price, depends on the demand and supply of that input, and on government regulations, if any. Generally, in the theory of firm behaviour, input prices are taken as parameters. This is because, a firm is usually an insignificant part of an economy, and its activities have no perceptible bearings on the total demand for and supply of inputs in the economy, and so, on factor prices. However, if the firm in question happens to be large in this respect, factor prices would be variables like output and factor productivities. As explained in the previous paragraph, each of learning and combination of complimentary products under one umbrella reduces the cost of production. Since no theory is perfect, no list of determinants is complete and accordingly the function (6.1) has a "catch all other determinants" in variable $Z$. In particular, all governments do regulate businesses to some extent and such regulations often come in the forms of (indirect) taxes (like excise duties, custom duties/tariffs) and subsidies. Taxes cause cost to go up while subsidies have dampening effects on costs. Thus, variable "$Z$" includes such regulations, among other left out determinants.

Before the section is closed, it must be emphasised that both factor productivities and factor prices are plural, though they have been argued in singular manner in the foregoing paragraphs. Thus, by an increase in factor productivities we mean, an increase in total factor productivity, or an increase in some productivity with other productivities held constant, or an increase in some productivities and decrease in some other but the effect of the former outweigh that of the latter on total cost. The arguments with regard to factor prices should be treated in a similar fashion.

Of the five sets of cost determinants, output assumes a special role. This is for two reasons. One, output is the only significant variable which is under the direct control of all the firms in an industry. For, factor productivity and input prices can hardly be influenced by an organization, experience is a past legacy, and decision pertaining to joint products is not a usual routine. Two, the relationship between total cost and output, though is unique in direction, is varying in terms of magnitude/proportion. That is, total cost increases as output expands, but the rate of increase varies from one set of output levels to the other for the same firm and for the same set of output level from one firm to another firm. For example, for firm $A$, between output levels 100 to 200, total cost might increase by just 100 percent but between output levels 200 and 400, the total cost might increase by 150 percent. Similarly, for firm $B$

even between output levels 100 to 200, the total cost might increase by 75 percent. Further, firms *A* and *B* may or may not belong to the same industry.

Due to high significance of the cost-output relationship, the same is discussed in detail separately for long-run and short-run.

## 6.3 COST-OUTPUT RELATIONSHIP: LONG-RUN

In the long-run, the firm has no fixed commitments and so all long-run costs are variable. The firm faces a long-run production function and a cost equation, which in the two-input cases of the previous chapter are of the following type:

$$Q = f(L, K)$$
$$TC = LP_L + KP_K$$

Given factor prices and a specific production function, one can draw an expansion path (Refer chapter 5, section 5.3), which gives the least-costs associated with various levels of output, which, in fact, yields the long-run total cost schedule/curve. Thus, Fig. 5.13 of chapter 5 gives the following *LTC* schedule.

| *Output* | *LTC* |
|---|---|
| $Q_1$ | $C_1$ |
| $Q_2$ | $C_2$ |
| $Q_3$ | $C_3$ |
| $Q_4$ | $C_4$ |

It is obvious from Fig. 5.13 that $Q_4 > Q_3 > Q_2 > Q_1$ and $C_4 > C_3 > C_2 > C_1$, which proves that *LTC* is an increasing function of output. The rates of change in these two variables are not known unless the qualitative relationship (directional) is quantified. If one recalls the concepts of returns to scale, and assumes fixed factor prices, one could see three kinds of relationship:

(a) When returns to scale are increasing, *LTC* increases as output increases but at a less than proportionate rate.

(b) When returns to scale are constant, *LTC* and output move in the same direction and same proportion.

(c) When returns to scale are decreasing, *LTC* increases at a faster rate than does output.

Thus, depending upon the nature of returns to scale, there will be a relationship between *LTC* and output, given factor prices. It is generally found that most industries and firms reap increasing returns to scale to start with, which are followed up by constant returns to scale, which give place to decreasing returns to scale eventually. This is primarily because of the indivisibility of the most efficient plants, equipments and personnel and the degree of specialization permitted by plant size. When the scale of operation is small, the firm is unable to take advantage of the most sophisticated technology, reflected in a large plant, highly competent management and a high degree of division of labour. For example, a farmer with one acre of land may not be able to take advantage of a tractor and a tube well, but as he grows bigger he will be able to introduce such efficient tools of farming, and thereby reap increasing returns to scale. However, once he reaches the scale of operation in the where about of 50 acres of land, no better technology remains to be introduced and management of large resources under one roof might pose

problems, thereby giving rise to decreasing returns to scale. Similarly, a small industrial enterprise may not find it worthwhile (or feasible) to go in for a sophisticated plant and management personnel but as it grows, it may go in for such things to get the advantage of increasing returns to scale, which, after a stage would take off to the region of decreasing returns to scale. In such a situation, the relationship between *LTC* and output will be of a changing character. As output expands, in the beginning *LTC* would increase but less than proportionately; after a while, *LTC* would increase and at a proportionate rate, and eventually *LTC* would increase more than proportionately. An hypothetical example could be the following.

**Table 6.1** *Cost-Output Relationship: Long-Run*

| *Q* | *LTC* | *LAC* | *LMC* (*arc*) |
|---|---|---|---|
| 0 | 0 | — | — |
| 5 | 25 | 5.00 | 5 |
| 10 | 45 | 4.50 | 4 |
| 15 | 60 | 4.00 | 3 |
| 20 | 85 | 4.25 | 5 |
| 25 | 120 | 4.80 | 7 |
| 30 | 180 | 6.00 | 12 |

*LMC* (arc) = $\Delta LTC/\Delta Q$

The graphs of the above relationship are provided in Fig. 6.1.

The *LTC* curve gives the least total cost for various levels of output when all the factors of production are variable. Its shape is such that the curve is first concave (bow up) and then convex as looked from the output axis. As seen above, its shape follows from the operations of the varying degrees of returns of scale, given the factor prices.

The relationship between total, average and marginal is mathematical in character and as with regard to $TPP_L$, $APP_L$ and $MPP_L$ curves (Refer chapter 5, section 5.4), the shapes of *LAC* and *LMC* follow from that of *LTC* curve. Both *LAC* and *LMC* are U-shaped. Further, the following relationships hold good.

(a) At the point of inflexion on *LTC* curve (*A*), *LMC* takes the minimum value.
(b) At the point of kink on *LTC* curve (*B*), where the slope of the straight line from origin to the *LTC* curve is the minimum—*LAC* assumes the minimum value.
(c) *LAC* is the least when *LMC* = *LAC*
(d) *LAC* curve is falling when *LMC* < *LAC*
(e) *LAC* curve is rising when *LMC* > *LAC*

The foregoing cost-output relationship assumes constant factor prices. However, one can conceive of a large firm, expansion of which could exercise some influence on factor prices, viz. wage rate and capital rental. If so, what would be the cost-output relationship? As the firm expands, it would hire more units of factors of production, which given the factor supply, would tend to increase factor prices*.

*It is possible that factor prices might in fact fall as the firm expands and hires more inputs. This is due to quantity discounts which bulk buyers are sometimes able to obtain. However, this possibility would only reinforce our argument in favour of the U-shaped long-run average cost curve.

Economists distinguish between internal and external economies/diseconomies of scale. The former results due to the firms' own expansion while the latter arise not due to its own expansion but due to expansion of some other firms/industry/economy. The text is dealing with internal economies/diseconomies only.

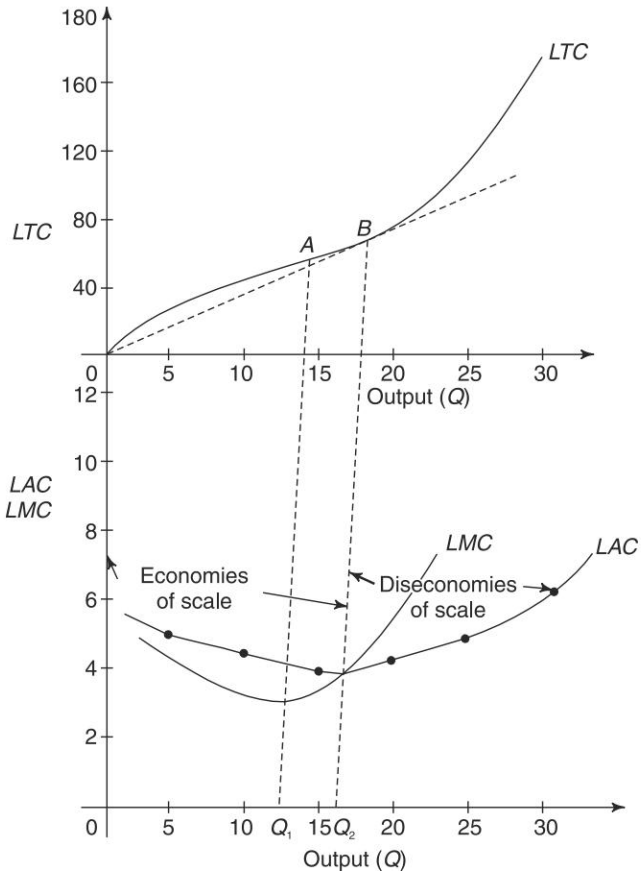



**Fig. 6.1** *Long-run cost curves*

Recall that in the beginning, there are increasing returns to scale, then constant and finally decreasing returns to scale. Super imposing both the changing factor prices and the nature of returns to scale together, the three stages appear as follows:

(a) In the beginning, because of increasing returns to scale, cost increases less than proportionately as output increases but because of rising factor prices, cost increases more than proportionately as output expands. Thus, the two forces work in opposite directions and so the net effect is ambiguous. However, there is an overwhelming evidence in favour of *LTC* increasing at a decreasing rate when the scale of operation is small.

(b) In the intermediate stage, where returns to scale are constant and factor prices are rising, *LTC* increases more than proportionately as output expands.

(c) In the third and final stage, where returns to scale are decreasing and factor prices are increasing, both the forces reinforce each other in favour of increasing *LTC* at an increasing rate.

Thus, the relationship between *LTC* and output under changing factor prices is similar to the one under constant factor prices. The difference between the two is in terms of extent (magnitude) and not in terms of kind. Under the changing factor prices, the decreasing rate of increase in *LTC* (or fall in *LAC*) is

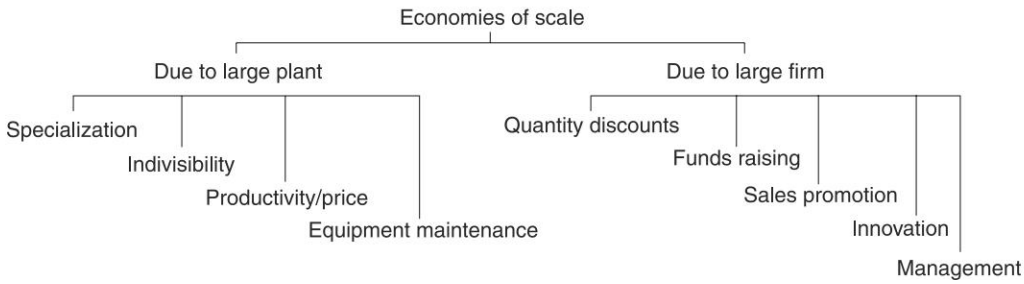
slower and increasing rate of increase in *LTC* (or rise in *LTC*) is faster than under constant factor prices. In consequence, *LAC* curve is U-shaped under both the situations.

The relationship between cost and output which one gets under general situation, that is, when all factors of production are variable (long-run) and so are factor prices, is often described under **economies and diseconomies of scale**. In other words, economies and diseconomies of scale combine the effects of returns to scale and factor prices on the relationship between cost and output. In what follows, we shall elaborate this rather important subject from the view point of decision-makers.

**Economies of Scale** A firm is said to be reaping scale economies when its *LTC* increases less than proportionately with increase in its scale of operation (output) or when its *LAC* falls as its output expands. Alternatively, it could be defined in terms of output elasticity of total cost ($e_{TC}$, $Q$):

$$e_{TC}, Q = \left[\frac{\partial(TC)}{\partial Q}\right]\left[\frac{Q}{TC}\right], \text{ if } TC \text{ is a continuous function}$$
$$= \left[\frac{\Delta(TC)}{\Delta Q}\right]\left[\frac{Q}{TC}\right], \text{ if } TC \text{ is a discrete function} \qquad (6.2)$$

If $e_{TC}, Q < 1$, there are economies of scale. The question is: what are the sources of such economies or why are some firms able to enjoy economies of scale? There are several reasons for this. The major ones are summarised in the chart below.




Since Adam Smith, it is well-known that specialization or what he called, division of labour, improves productivity as there is need for a limited training for a specialized job only, learning by doing, saving of time which otherwise gets lost in moving from one activity to the other, reduction in the period of idleness of plants and equipments, need for lesser equipments, etc. The bigger plants are able to introduce the desired level of specialization, which small plants can only ill afford. Thus, for example, big plants have a number of production managers, each looking after one particular process; small plants have just one or two production managers, looking after all the processes. Specialization in a specific process requires less training and by doing the same work repeatedly, the officer becomes more perfect than otherwise.

There are certain plants and equipments, and even personnel which are indivisible in small units/sizes: For example, tractor and tube-well for agriculture, large plants and machines for manufacturing, computers for computation work, and management personnel. Small plants can not employ such things and thus are unable to enjoy the fruits of up to date technology. In contrast, large firms are able to benefit from such things.

Another source of scale economics arises from the relative productivity and purchase prices of

different sizes of capital equipments. It is believed that the productive capacity of capital equipment rises much faster than its purchase price. In other words, a machine that costs twice as much than a smaller one will typically produce more than twice as much output. Thus, large plants which alone can use large machines get an edge over the small plants through this source as well.

It is well-known that large plants need to carry relatively less spare parts and personnel for attending to random breakdowns than do small plants. Thus, large plants spend relatively less on repairs and maintenance, and thereby reap the economies of mass production.

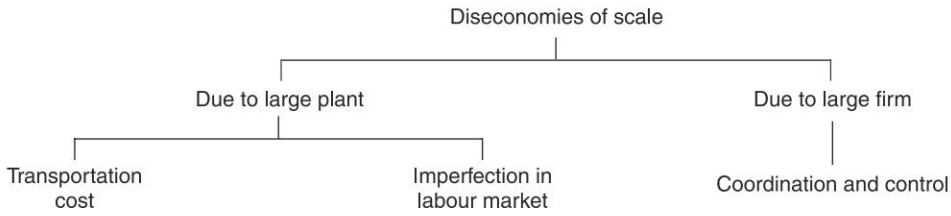
In addition to the above (plant) economies, there are some more which are associated with the overall size of the firm. It is believed that while manufacturing work is handled by individual plants, purchases of raw-materials, selling of products, advertisements, fund raising, and overall management activities are centralized at the firm level. If so, a large firm enjoys certain cost advantages over small firms in carrying out such activities. One such possible source is in materials procurement. Firms that purchase large quantities of goods and services from suppliers are often able to negotiate quantify discounts, which results in lower average costs in relation to small firms. Also, per unit voucher cost is less for large firms vis-a-vis small firms.

Although the interest cost on debt and dividend cost of equity may well be uniform for large and small firms, the administrative cost of negotiating debt and floating cost of equity capital and debentures per unit of funds raised vary inversely with the size of debt/capital issue. In addition, the securities of large firms are generally less risky than those of smaller firms, and thereby large firms get an edge over the small firms in fund raising. For these reasons, the cost of capital is often found to vary inversely with the size of the firm.

Large firms are able to secure quantity discounts even in securing space and time in various advertising media. Besides, advertisement outlay is like a fixed cost and so they create lesser burden per unit of output on larger firms than on smaller ones. More or less the same is true with regard to innovations. Large firms are in a better position to carry on research and development activities, and thereby to adopt the latest methods of material procurements, production and sales, than their small counterparts. Again with regard to overall management, large firms enjoy the benefit of top caliber management personnel, specialization, and so on, which are denied to small firms.

Due to all these reasons, there are economies of mass production. **Walmart** offers an interesting example of how a firm can benefit from such economies. Due to economies of scale, the long-run total cost curve is concave from the output axis, the output elasticity of total cost is less than unity, or the long-run average cost curve is downward sloping, until the output level where such economies dominate. It must be emphasized that the presence of economies of scale offers a forceful reason for firms to seek mergers with firms engaged in the production of identical goods at the same or different geographical areas (i.e., **horizontal mergers**). Banks and many manufacturing units are reaping enormous growth and earnings through this strategy all over the globe. Also, economies of scale available in an industry play a significant role in the size and number of firms in the industry. Industries where such economies are significant would have larger sized and smaller number of firms than the ones where such economies do not exist or are non-significant. Thus, most consumer durable goods' industries, including cars, refrigerates, air conditioners, etc. have large sized and small number of firms while most service industries, including educational institutions, law firms, medical practitioners, hair cutting, laundries, etc. have small sized and large number of firms.

**Diseconomies of Scale** There are certain cost disadvantages in mass production vis-a-vis small production. The reasons for this are summarised in the chart below.




A primary source of diseconomies of scale due to large plant size is transportation cost. If the raw-martial is spatially well spread, then the transportation cost of raw-material will be more in a large plant wherever located than in a number of small plants well dispersed geographically. This source is stronger in the case of distribution of output among consumers. Consumers are generally well distributed across the geographical region and thus a large plant would incur a lot more on transportation cost than would a number of small plants.

Labour markets are far from perfect, particularly in a developing economy like India. There is a high degree of immobility due to attachment to relatives and land, cost of mobility in terms of transport and residential accommodation, contendness, etc. Since small plants require fewer workers than large plants, the former are able to attract labour at a relatively lower wage rate than the large plants. Similar reasonings hold good for other factors of production.

The existence of diseconomies of scale for the firm results from problems of coordination and control experienced by the management. Today, we have labour unions and their force increases faster than the size of the firm. We have evidence of strikes, lock outs, and absentism. Besides, there is a dearth of capable and honest top management. Due to all these factors, small firms are better managed than the large firms, and accordingly on this count small firms have lower costs than do large firms.

Diseconomies of scale explain the upward sloping long-run average cost curve, convexity of long-run total cost curve, or the more than unity elasticity of total cost with respect to output. Both the economies as well as diseconomies of scale operate simultaneously at every stage of production. In the beginning, the former exceeds the latter and thus *LAC* curve slopes down. After a while, the two cancels each other and accordingly *LAC* curve stays flat (horizontal). Eventually, a stage is reached from where onwards diseconomies dominate over economies of scale and thus *LAC* curve slopes upward. Thus, the U-shape of the long-run average cost curve is fully rationalised.

Many economists have found empirically, and so have argued in favour of, the **L-shaped long-run average cost** curve. They content that the economies of scale outstrips diseconomies of scale as a firm expands from small size up to a certain size, beyond which the two just balance, leading to no further net economies or diseconomies of scale. If so, the *LAC* curve would be the L-shape as in Fig. 6.2 (a).

In Fig. 6.2 (a) *LAC* declines as output expands up to $Q_1$, beyond which it stays constant at 0*P*. The output level where the *LAC* is the least is called the ***optimum level of output from the supply side***. In case of Fig. 6.2 (a), the optimum level thus comprises of output level $Q_1$ or any level greater than that. To distinguish $Q_1$ output from other optimum output levels under such a situation, economists have coined yet another terminology, called the ***minimum efficient scale*** (MES). The *MES* is defined as the least volume of output at which *LAC* is the least. Thus, output level $Q_1$ marks the *MES* while output $\geq Q_1$ gives the optimum level of output.

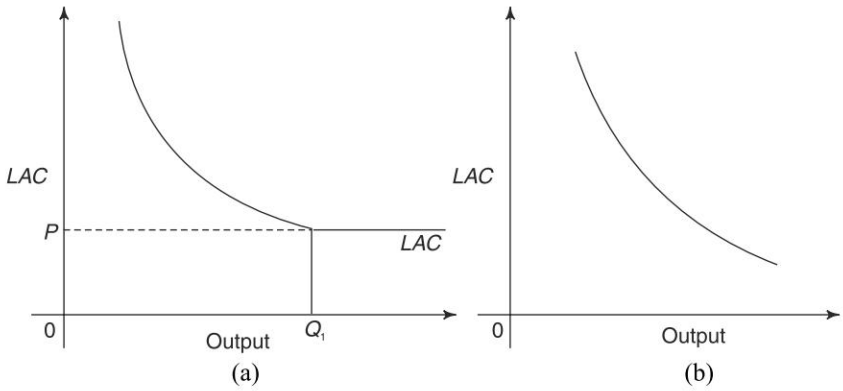



**Fig. 6.2** *L-shaped long-run average cost curve*

There is still one more hypothesis regarding the shape of the *LAC* curve, which is given in Fig. 6.2 (b). In this version, *LAC* falls monotonically as output expands, that is the economies of scale outweigh the diseconomies of scale at all levels of output. However, the hypothesis does recognize that the rate of fall in *LAC* declines as output expands, and thus the *LAC* curve is downward sloping but is convex from below. In such a situation, there is nothing like an exact optimum output level or *MES*, but one could approximate this by estimating the output level beyond which the fall in *LAC* is insignificant, like 1 or 2 percent.

What is the correct shape of the *LAC curve*, the one in Fig. 6.1, in Fig. 6.2 (a) or in Fig. 6.2 (b)? This is an empirical issue, which will be examined later in this chapter.

### Returns to Scale vis-a-vis Economies of Scale

There is often some confusion between returns to scale and economies of scale. Though both the concepts pertain to long run, there is no unique relationship between the two. Only, under the fixed factor prices, they have a definite relationship, viz. increasing returns to scale (*IRTS*) implies economies of scale (*EOS*), decreasing returns to scale (*DRTS*) indicates diseconomies of scale (*DOS*), and constant returns to scale (*CRTS*) suggests neither. However, when factor prices are variables, their relationship is uncertain. To understand this, consider the following hypothetical example:

| | *L* | *K* | *Q* | $P_L$ | $P_K$ | *LTC* | *LAC* |
|---|---|---|---|---|---|---|---|
| (i) | 2 | 2 | 31 | 30 | 50 | 160 | 5.16 |
| (ii) | 3 | 3 | 59 | 30 | 50 | 240 | 4.07 |
| (iii) | 3 | 3 | 59 | 50 | 60 | 330 | 5.59 |

In the above example, it can be seen that (a) there are always *IRTS*, but (b) from case (i) to case (ii) there are *EOS* while (c) between cases (i) and (iii), there are *DOS*. The reason is obvious to see.

## 6.4 COST-OUTPUT RELATIONSHIP: SHORT-RUN

In the short-run, at least one factor of production is fixed, and at least one is variable. Due to this constraint, the firm may not be able to achieve the best combination of inputs for its desired level of

output. Consequently, the short-run cost could exceed the long-run cost for a given output level. This was shown in the chapter 5, section 5.5.

The short-run cost curves are derived from the short-run production curves, given the factor prices. In fact, the one set is the inverse of the other. This can be explained as follows.

Recall that there are seven cost concepts in the short-run: total fixed cost (*TFC*), total variable cost (*TVC*), total cost (*TC*), average fixed cost (*AFC*), average variable cost (*AVC*), average total cost (*ATC*), and marginal cost (*MC*). Also, remember that given anyone of the three: total, average and marginal, the other two are simply arithmetic. We illustrate this through an example of two input case. In case of two inputs, we have

$$TC = L\,P_L + K\,P_K$$

If labour is the variable input and capital the fixed input, then

$$TVC = LP_L \text{ and } TFC = KP_K = C \text{ (constant)}$$

Then

$$AVC = \frac{TVC}{Q} = \frac{LP_L}{Q} = \left(P_L \frac{1}{Q/L}\right)$$

or,
$$AVC = P_L\left(\frac{1}{APP_L}\right) \tag{6.3}$$

and,
$$ATC = AVC + AFC$$

or,
$$ATC = P_L\left(\frac{1}{APP_L}\right) + AFC \tag{6.4}$$

Further,
$$MC = \frac{\partial TC}{\partial Q} = \frac{\partial (TVC + TFC)}{\partial Q} = \frac{\partial (TVC)}{\partial Q} + 0$$
$$= \frac{\partial (LP_L)}{\partial Q} = P_L \frac{\partial L}{\partial Q} = P_L\left(\frac{1}{\partial Q/\partial L}\right) \tag{6.5}$$

or,
$$MC = P_L\left(\frac{1}{MPP_L}\right)$$

Thus, if factor prices are constants, *AVC* is inversely proportional to $APP_L$, and *MC* to $MPP_L$, and *ATC* is inversely related to $APP_L$. These could be illustrated further through the duality principle.

**Duality Principle in Production and Cost** The duality principle states that given the production function and inputs' prices, the cost-output relations can be determined. To explain the principle, consider the short-run production function in table form of the previous chapter (Refer section 5.4, Table 5.3), where capital input is fixed at $K = 2$, and assume that factor prices stand at $P_K = 50$ and $P_L = 30$. Given these, cost calculations are easy to obtain using their definitions given above as well as in Table 6.2. Thus, the duality principle is established. The table includes the calculations for average and marginal factor productivities for the sake of understanding the inverse relationship between factor productivities and corresponding costs. A careful examination of Table 6.2 reveals the following:

(a) There is an inverse relationship between $APP_L$ and *AVC*, and $MPP_L$ and *MC*. Recall that the reasons for this are found in Eq. (6.3) and (6.5), under fixed inputs' prices.

(b) Average fixed cost falls monotonously as output expands. This is because total fixed cost is invariant with respect to output.

(c) Average variable cost first falls as output expands, but after a certain point the relationship is reversed. Since factor prices are held constant, this relationship is entirely due to the behaviour of $APP_L$, which as seen in the previous chapter, is drawn from the law of variable (diminishing) marginal returns. Thus, the shape of *AVC* schedule is due to this law and fixed inputs' prices.

(d) Average (total) cost first falls as output increases but again after a certain point the trend is reversed. Since $ATC = AFC + AVC$, the shape of *ATC* follows from those of *AFC* and *AVC*. In particular, *ATC* is U-shaped partly because of the law of variable (diminishing) marginal returns and partly due to the falling *AFC*. Note that *ATC* falls until a longer output than does *AVC* and this is because of the falling *AFC*.

(e) Marginal cost follows the same pattern as do *AVC* and *AC*. Its trend is derived from that of $MPP_L$ (and fixed inputs' prices) which behaves the way it does entirely due to the operation of the law of variable returns.

**Table 6.2** *Correspondence between Production Function and Cost Output Relationship (Duality Principle): Short-Run*

| *Labour Input* (*L*) | *Output* (*Q*) | $APP_L$ (*Q*/*L*) | *Arc* $MPP_L$ (*Q*/*L*) | *TFC* ($KP_K$) | *TVC* ($LP_L$) | *TC* | *AFC* (*TFC*) | *AVC* (*TVC*) | *AC* | *Arc MC* $\left(\frac{\Delta TC}{\Delta Q}\right)$ |
|---|---|---|---|---|---|---|---|---|---|---|
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| 0 | 0 | — | — | 100 | 0 | 100 | — | — | — | — |
| 1 | 15 | 15.0 | 15 | 100 | 30 | 130 | 6.7 | 2.0 | 8.7 | 2.0 |
| 2 | 31 | 15.5 | 16 | 100 | 60 | 160 | 3.2 | 1.9 | 5.1 | 1.9 |
| 3 | 48 | 16.0 | 17 | 100 | 90 | 190 | 2.1 | 1.9 | 4.0 | 1.8 |
| 4 | 59 | 14.8 | 11 | 100 | 120 | 220 | 1.7 | 2.0 | 3.7 | 2.7 |
| 5 | 68 | 13.6 | 9 | 100 | 150 | 250 | 1.5 | 2.2 | 3.7 | 3.3 |
| 6 | 72 | 12.0 | 4 | 100 | 180 | 280 | 1.4 | 2.5 | 3.9 | 7.5 |
| 7 | 73 | 10.4 | 1 | 100 | 210 | 310 | 1.3 | 2.9 | 4.2 | 30.0 |

(Table is based on the assumed values: $K = 2$, $P_L = 30$ and $P_K = 50$)

Graph of total output and labour productivities ($APP_L$, $MPP_L$) are not drawn here as they have been drawn before (chapter 5, Fig. 5.10). The graphs of all the seven short-run cost curves are presented in Fig. 6.3. Recall that while the table is discrete, the graph is continuous. Thus, while drawing the former from the latter, some continuity is introduced in the data.

As seen in Table 6.2, there is a definite correspondence between factor productivities curves and short-run total cost curves, given the factor prices. The *AVC* and *MC* curves are U-shaped just as $APP_L$ and $MPP_L$ are inverted U-shaped. Further, the output level at which $APP_L$ is maximum, is also the output level at which *AVC* is the minimum. The same thing holds with regard to $MPP_L$ and *MC*. The *AC* curve is U-shaped partly because the *AVC* curve is U-shaped and partly because the *AFC* is a monotonically falling curve.

There are a few important things with regard to the shapes of these cost curves, which must be emphasized.

(a) The *TFC* curve is horizontal.

(b) The *TVC* curve starts from origin, is concave from below in the beginning and is convex from below after a certain level of output.

(c) The *TC* curve starts from a point above the origin and then follows the shape of the *TVC* curve. The *TC* and *TVC* curves are parallel.

(d) The output at which *MC* is minimum ($Q_1$) is less than the output at which *AVC* is minimum ($Q_2$), which in turn, is less than the output at which *AC* is the least ($Q_3$). The first ($Q_1 < Q_2$) because of the mathematical property of the relationship between average and marginal (Refer chapter 5, section 5.4). The second ($Q_2 < Q_3$), for the *AFC* curve is falling monotonously. In other words, the output at which *AVC* is minimum precedes the one at which *ATC* is minimum, for $ATC = AVC + AFC$ and *AFC* is falling continuously, thereby when *AVC* falls, *AC* falls partly because *AVC* is falling and partly because *AFC* is falling; when *AVC* rises, *AC* could fall, stay constant or rise, depending upon whether the increase in *AVC* outweighs, just cancels or is out weighted by decrease in *AFC*.

(e) The *MC* curve passes through the minimum points of both the *AVC* and *ATC* curves. This is again due to the mathematical property of the relationship between average and marginal.

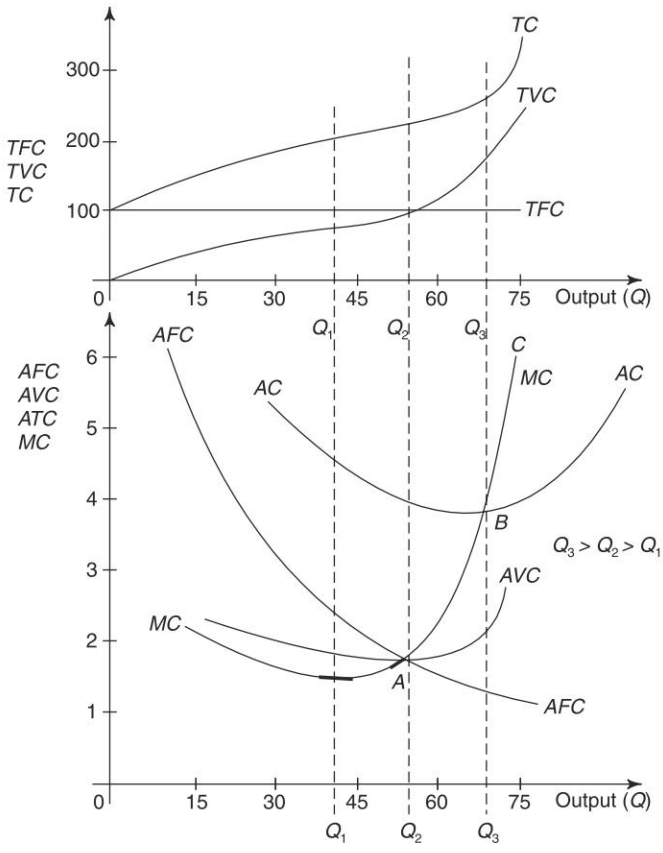



**Fig. 6.3** *Short-run cost curves*

## 6.5 COST OUTPUT RELATIONSHIP: LONG vs SHORT-RUN

Recall that for any given output, short-run total cost is greater than or equal to the long-run total cost. In fact, there is generally only one output level at which the costs in two periods are equal, and at all other output levels the short-run cost exceeds the long-run cost. Also, remember that in the short-run, the fixed cost depends on the quantities of fixed factors employed, which, in turn, determine the capacities of output which a firm could produce. The larger the fixed factors hired, the greater is the capacity output in the short-run. Keeping these factors in mind, one can draw short and long-run cost curves on the same graph, which, in turn, would reveal the relationship between the two. These are drawn in Fig. 6.4. Marginal cost curves too could have been drawn but they have been left out to avoid complicating the graphs.

In Fig. 6.4 there is one total/average cost curve for each magnitude of fixed resources, call it capacity. Thus, there is a family of short-run cost curves, one for each capacity which a firm might create in the short-run. In the long-run, capacity itself is a variable and the firm could choose its capacity, depending upon the output it wants to produce in the long-run. The long-run curve envelops on the short-run cost curves, for the long-run cost cannot exceed the short-run cost. The long-run cost curve would be

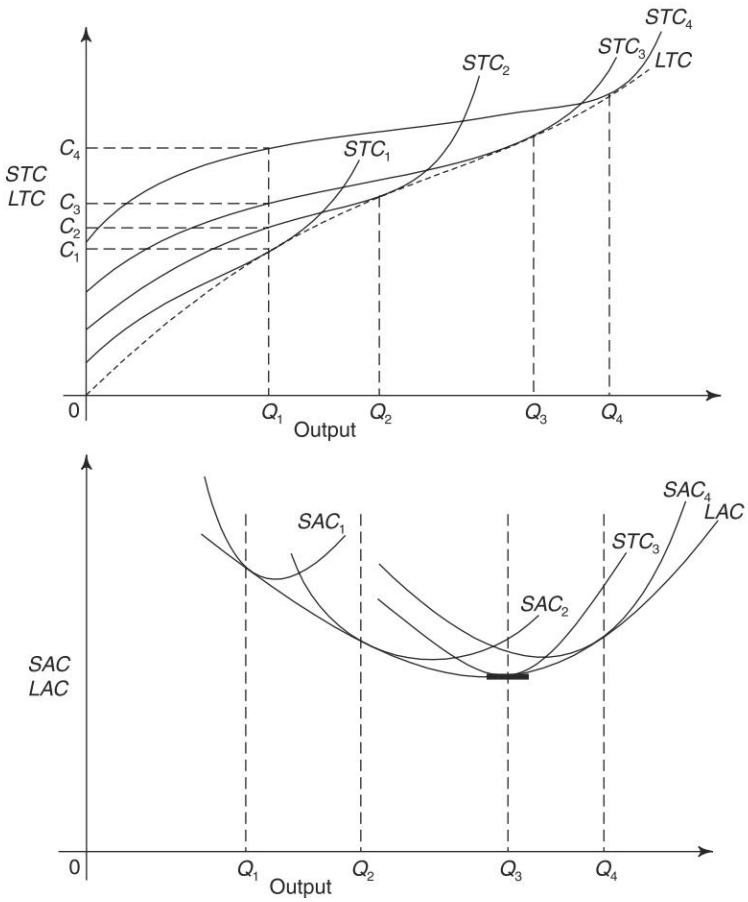



**Fig. 6.4** *Short and long-run cost curves*

smooth if there is a short-run cost curve for every size of the plant. In Fig. 6.4, $STC_1$, $STC_2$, $STC_3$ and $STC_4$ are four short-run total cost curves, where they assume fixed factors employed at $K_1$, $K_2$, $K_3$ and $K_4$, respectively ($K_4 > K_3 > K_2 > K_1$). It will be seen that the long-run and short-run costs are equal only at output levels $Q_1$, $Q_2$, $Q_3$ and $Q_4$, points of tangency between *LAC* and *SAC* curves. The *LAC* curve is U-shaped as *SAC* curves are, but the former is flatter than the latter.

An interesting point to note here is the total cost for a given output and the size of fixed resources (plant). The relationship is not uni-directional. If the output is small, the total cost is less for a small plant than for a large plant, and quite the reverse hold good for large outputs. This is because if a large plant is installed, it will remain under-utilized when output is small while a small plant will be inadequate for large outputs. It is for this reason that short-run total cost curves cut each other and they extend to the output axis up to different lengths. Thus, for example, for output $Q_1$, the smallest plant is the most appropriate, for the total cost under this plant size equals $C_1$, which is less than that under any other plant size ($C_2$, $C_3$, $C_4$). However, since $STC_1$ curve does not go even up to output $Q_2$, it is not even feasible to produce $Q_2$ or more output under such a small plant. The inference from this is that firms should create capacity on the basis of the likely demand for its product. If the capacity is larger than the output needed to meet the demand, some fixed resources would remain idle and the (total) cost would be more than the minimum possible for that output. It is because of this, fixed costs are also called the **overhead costs**. On the other hand, if the capacity is inadequate to produce output to meet the demand, either some sales would be lost or the (total) cost would be larger than it could have been if the plant were somewhat bigger.

The duality principle, as explained above for short run, holds, even for the long run. Thus, if long run production function and factor prices are known, the long run cost–output relations could be derived. To briefly explain the principle, let us calculate *LTC* for output equal, 72 assuming the production function of Table 5.1 (See chapter 5) and wage rate equals Rs. 30 and capital rental equals 50. In the table, there are three alternative technologies for $Q = 72$ and the same along with the cost associated with them are as follows:

| | *L* | *K* | *TC* |
|---|---|---|---|
| (i) | 6 | 2 | 280 |
| (ii) | 4 | 4 | 320 |
| (iii) | 2 | 6 | 360 |

In the long run, the firm is free to choose any values for *K* and *L*, and thus would opt for technology (i) where the *LTC* is minimum at *LTC*= 280. The same procedure could be repeated for any other output level and *LTC* calculated. This demonstrates the duality principle for a long run. Further, to show that *LTC* is less than (or at worst equal to) the *STC* for a given output, let *K* be fixed at $K = 2$ in the short run. Under short-run, the firm would have just technology (iii) where *STC* = 360 which exceeds *LTC* =280. However, if by chance the installed capital happened to be at $K = 6$, the *STC* would equal 280, same as *LTC*. Under no situation, the *STC* would be less than *LTC*.

## 6.6 ECONOMIES OF BIG BUSINESSES

Business is becoming increasingly competitive over time. During the early period, firms concentrated on increasing labour productivity through technological advancements, capital intensity and reorganization of workers. During the last century, the emphasis has shifted on to giant corporations, made feasible

through continuing progress in technology (particularly IT and transport network), availability of energy and materials, globalization, etc. These developments have enabled businesses to expand not only through multiplication of capacities in the existing production lines but also via integration of various kinds, viz. horizontal, vertical and conglomerates. Such integrations have enabled firms to cut their costs not only through economies of scale, but also through new cost saving avenues. The present section deals with such cost reduction strategies.

## Learning Curve

Like all else, firms learn through experience and thereby often reap the fruits in terms of some savings in their production costs. How this saving comes? Recall Adam Smith's theory of specialization under which a person who performs an activity repeatedly becomes an expert in that activity over time and thereby is able to save some time in subsequent attempts. Similarly a firm producing the same good year after year, develops expertise in the job and accordingly may be able to produce that product in a given quantity at a lesser cost in future than in the past. The saving may come through training/supervising the workers/managers, testing of product quality, defects reduction, wastage reduction, promotional cost, and negotiations with employees, customers, suppliers and competitors, etc. All such savings will obviously lead to reduction in production cost as the firm earns more and more experience. The question now is how this experience is quantified? There are two alternative measures for this, viz. years of experience (or age of the organisation) and the total quantity of output produced to date (i.e. cumulative output). Of the two, the cumulative output is considered as the better measure of experience. This is because the experience/learning depends not just on the period of work but on the quantum of the work handled as well. For example, a professor who has taught Managerial Economics for, say, 20 years but, on an average, one group of 20 students a year has less experience than a professor who has taught the course for 10 years but, on an average, 5 groups of students a year.

By this concept, the more is the experience, the more is the learning, and, in consequence, the less is the total cost for a given output. Accordingly, average cost falls as the experience (cumulative output) increases.

Further, the law of diminishing marginal returns would generally apply even to the learning process. Thus, the decrease in *AC* would go on falling as more and more learning takes place. Accordingly, the learning curve, which denotes the relationship between the *AC* and the experience (cumulative output = summation of all past outputs), would be falling and convex to the origin (cumulative output axis). Such a curve is given in Fig. 6.5.

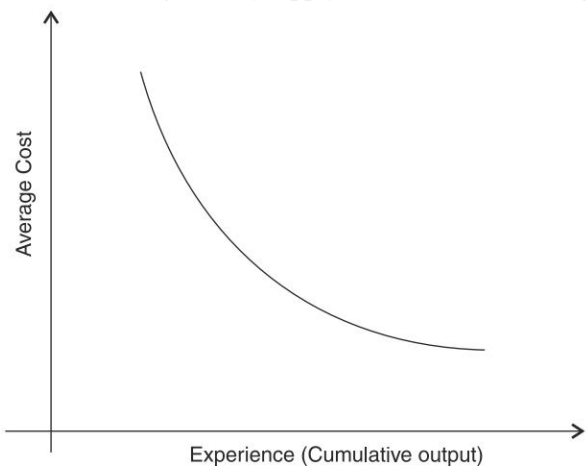



**Fig. 6.5** *Learning curve*

The learning curve effect varies across industries. It is more pronounced in knowledge-skill based industries than the rest. Thus, cost and consequently price of most knowledge based goods (like electronic goods) falls quite steeply as the good gets older while those of others keep rising over time.

## Economies of Scope

Economies of scope offers yet another avenue for a firm to lower its production cost. Economies of scope (*ESC*) means the total cost of producing two or more related goods in certain quantities is less to a multi-product firm than to the two or more firms each producing a single product. The concept is explained for a two goods case algebraically by the following inequality:

$$TC\,(Q_1 + Q_2) < TC\,(Q_1, 0) + TC\,(0, Q_2) \tag{6.6}$$

where, $TC\,(Q_1 + Q_2)$ = total cost of multi-product firm for outputs equal $Q_1$ and $Q_2$ of two related goods

$TC\,(Q_1, 0)$ = total cost of single product firm for output equals $Q_1$ of good 1 only

$TC\,(0, Q_2)$ = total cost of single product firm for output equals $Q_2$ of good 2 only

The scope economies arise due to saving in cost through using the same technology, same human resources, same inventories, and/or same suppliers, etc. in the production of two or more goods. For example, an automobile firm producing cars as well as two wheelers would have lower total cost for given units of two types of vehicles than the two firms producing those items independently. The multi-product firm saves in cost through saving in training cost, using the surplus staff/equipments/structure time of one product in other product where there is a crunch; saving in inventory carrying cost to the extent the two products have some common items, etc. By this reasoning, one can easily see the rationale for firms moving from single product to related multi-products and for them going into mergers in related product lines (horizontal mergers). IIMs have slowly moved from single product (PGDM=MBA course) to multi-products (PGDM, FPM, FDP, MDPs, Consulting, PGPX) firms for various reasons including the economies of scope. The various goods that IIMs produce use the same faculty and infrastructure, and require the similar teaching material, and thus save on cost. Federal Express combines ground delivery with air-based express and thus enjoys the economies of scope. Thus, a firm by **diversifying its product line in the related product** (**horizontal integration**) could reduce its production cost through economies of scope. Note that just as there are economies of scope, there could be **diseconomies of scope** as well. The latter would result when the production of one good conflicts with that of another.

The **degree of economies of scope** ($D_{ESC}$) may vary from firm to firm. To measure the said degree, the following formula may be used:

$$D_{ESC} = \frac{[TC(Q_1, 0) + TC(0, Q_2)] - TC(Q_1 + Q_2)}{TC(Q_1 + Q_2)} \tag{6.7}$$

A related concept here is the **cost complimentarity** between/among goods. Such a relationship exists between two goods when the marginal cost of one good falls as the output of the other goods increases. To illustrate this, assume that the cost function of the multi-product firm engaged in the production of goods 1 and 2 is the following:

$$TC = 50 - 0.15\,Q_1\,Q_2 + 0.10\,Q_1^2 + 0.25\,Q_2^2 \tag{6.8}$$

Under this function, *MC* of product 1 ($MC_1$) is given by

$$MC_1 = -0.15\,Q_2 + 0.20\,Q_1 \tag{6.9}$$

Equation (6.9) shows that the *MC* of good 1 falls as output of good 2 increases. Similar examination of $MC_2$ will reveal that *MC* of good 2 falls as output of good 1 increase. Thus, for the firm having the

cost function (6.8), the two goods have the advantage of cost complimentarity. In general, the two goods are complimentary in cost if the sign of the cross product term ($Q_1\ Q_2$) is negative.

Given the cost function of a multi-product firm and those of single product firms, it would be easy to see if the economies of scope exist. Thus, assuming that function (6.8) stands both for the multi product as well as the single product firms, the *ESC* exists if the following inequality holds good:

$$[50 + 0.10\ Q_1^2] + [50 + 0.25\ Q_2^2] - [50 - 0.15\ Q_1\ Q_2 + 0.10\ Q_1^2 + 0.25\ Q_2^2] > 0$$

or, $$[50 + 0.15\ Q_1\ Q_2] > 0 \qquad (6.10)$$

The inequality (6.10) holds good as both the terms are positive. However, if the coefficient of the cross product term (= – 0.15) were positive, the inequality may or may not hold. The above functions assume that *TFC* is the same both for a multi-product as well as each of the single product firm. However, if this were not true, the measure of *ESC* and *DSC* could still be computed. An inference from the above discussion follows: if cost complimentary exists, economies of scope exist as well, but not necessarily the vice versa.

As would be obvious from the above discussion, the economies of scope and cost complimentary concepts are peculiar to firms engaged in the production of more than one good. Further, to be able to reap benefits through these features, the two or more goods under production must be related in technology and/or resources' (inputs) requirements.

## Transaction Cost Economies

Saving in transaction cost offers yet another source of economies to firms. Transaction cost refers to the costs associated with acquiring an input that are in excess of the amount paid to the input supplier. Thus, it includes the search and information cost, bargaining and contracting cost, enforcement cost, and specialized investment cost (like those in establishing a division to handle this activity), if any. Firms spend enormous amounts in such costs. For example, an automobile firm needs a steady supply of engines, whose design and manufacturing will be tailored to the requirements of the finished product. To ensure this, the auto firm would have to spend money in identifying the best supplier and ensuring a steady supply. Such costs (per unit) could be reduced to some extent through scale (size) but more through **vertical integration**. Under vertical integration, a firm expands by adding new products which are related to their existing product (s) either as inputs or as output users (consumers). For example, if an erstwhile automobile firm starts producing engines or any other intermediary goods that it buys for its use in the production of cars, it is expanding through vertical integration of the type called **backward (vertical) integration**. In contrast, if the auto firm enters in transport (taxi) services, it goes into **forward** type of **vertical integration**. By these two kinds of expansion, the auto firm will have its own supply of engines and own buyer of cars, and thereby save on transaction costs. Incidentally note that under vertical integration of either type, the production of various goods appear like the different stages of production of a final good, particularly if the goods produced in pre-final stages are used entirely in the production of the final stage good. Thus, if engines and cars produced by the said firm are fully used in the taxi services of the company (i.e. not sold to outside market at all), then engines are produced in the first stage and cars in the second stage, and the company simply runs taxi services' business.

Transaction cost offers incentives for vertical integration. It is for this reason, among others, that there is a rush in this direction across all countries. ITC (Indian Tobacco Company) provides a good example in India. It started with cigarettes business, expanded in tobacco cultivation (backward integration),

in hotel business (forward integration), and so on. Many more such examples are available and surely readers can easily name a few. Before we conclude this section, it is pertinent to recall together that there are four kinds of economies which a firm could tap to lower its cost. These are (a) economies of scale, (b) learning curve gains, (c) economies of scope, and (d) economies of transaction cost. While the first three are reaped through horizontal (same or cost related products) expansion (diversification), the last one is received basically through vertical integration (expansion across products related through the input-output relationships).

## 6.7 ESTIMATION OF COST FUNCTION

Decision-making requires forward planning. Thus, a firm, be it a new one or an existing one, would like to know the cost function facing it. Of course, the function may not be available until the firm really goes for expansion of its output. However, there are methods through which the firm could get approximate information of its future cost-output relationship. As is usual with regard to methods, there are alternative methods available for this purpose. The three well-known are the following:

(a) Engineering method
(b) Survivorship method
(c) Statistical method

Discussion on each of these follows. A set of guidelines for the method to be followed in practice would be provided towards the end of this section.

### Engineering Method

The engineering method of cost estimation is based directly on the physical relationship expressed in the production function for a particular product. On the basis of the production function, and input prices, the optimum input combination for producing any given quantity of a given product is determined (Refer Chapter 5, section 5.3). The cost curve is then formulated by multiplying each input in the so obtained least-cost combinations by its price and summing, to develop the cost function. Since the estimates on the least-cost estimates are provided by engineers, it is called the engineering method.

The method is based on the currently available technology and the existing factor prices. The users must have a thorough knowledge of the production technology as well as of factor prices. Nevertheless, since technology as well as factor prices are highly volatile over time, the method may not yield accurate estimates.

### Survivorship Method

The survivorship technique was developed by George Stigler in 1958. Under this technique, the various firms of an industry are first classified into certain size groups, then the growth of firms over time in each size group is examined. The size group whose share in the industry grows the most is then considered as the most efficient (the least average cost) size group, and vice versa. To explain this approach, let us consider a hypothetical example. Let the firms be divided into three classes on the basis of their current outputs. Let the share of each groups in the industry output in the base year and the current year be as follows:

| *Size Group* | *Industry share in (%)* | |
|---|---|---|
| | *Base year* | *Current year* |
| Small | 10 | 12 |
| Medium | 30 | 50 |
| Large | 60 | 38 |

On the basis of the above data, the method would conclude that the medium size class is the most efficient one and the large size class the most inefficient. The rationale for this approach is that competition will tend to eliminate those firms whose size is relatively inefficient, leaving only those size firms with lower average cost to survive over time.

The survivor method is quite simple. However, it suffers from a major limitation. The method gives the optimum size of a firm only and that too in terms of output range or the size class. It does not yield the cost function. As we shall see later in this chapter, there are uses of cost function other than the determination of optimum size, to which the method is just silent.

## Statistical Method

Under this method, the cost function is estimated through the application of some statistical method (e.g. least-square method) to the historical data on the cost and its determinants. The data could be a time series data of a firm in the industry or of all firms in the industry or a cross-section data for a particular year from various firms in the industry. However, depending on the kind of data used, we would get a short-run or a long-run cost function. For example, if time-series data of a firm is used and the output capacity of the firm has not changed much during the sample period, the cost function would be the short-run one. In contrast, if cross-section data of many firms, whose size vary substantially, or the time series data of the industry as a whole, whose output has expanded enormously during the sample period, were used, the estimated cost function would be the long-run one.

Specification of the cost function must precede it estimation. It calls for two steps, viz. (a) identification of the determinants of cost and (b) specification of the functional form for the function. The determinants of cost, recall function (6.1) above, includes output, factor productivities (or the technology), factor prices, composition of output (if there are multiple products), a measure of experience, etc. Recall that factor productivities (which depend on the technology) variable is contained in the production function itself. In case of cross-section data, factor prices may not be relevant, for they might be uniform for all the firms at a particular point of time. In case of time-series data, factor prices must be included but there is another method to handle them. One could deflate the cost data by the inflation rate or the composite factor price index. Further, if one is using the historical (actual) data, the cost and output data include the effects of both factor productivities and factor prices on the cost-output relationship. Thus, the cost function for estimation through the statistical method could be hypothesized as follows:

$$C = F(Q, Z) \tag{6.11}$$

where, $C$ = Total cost

$Q$ = Output

$Z$ = 'other' determinants of cost

Regarding the functional form for equation (6.11), there are four common formulations:

Linear : $C = A_0 + a_1 Q + \sum_{i=2}^{k} a_i Z_i$

Quadratic : $C = A_0 + a_1 Q + A_2 Q^2 + .....$

Cubic : $C = b_0 + b_1 Q + b_2 Q^2 + b_3 Q^3 + .....$

Double Log : $\log C = B_0 + B_1 \log Q + \log Z_i$

By recourse to calculus and algebra, it can be proved that a linear total cost function would result into a constant marginal cost and monotonically falling average cost; a quadratic function to a U-shaped *AC* curve and a linear (straight line) *MC* curve; a cubic function to a U-shaped *AC* curve and a U-shaped *MC* curve; and double log function to a falling (or rising) *AC* and a falling (or rising) *MC* curve. Thus, if one were to test the hypothesis of U-shaped *AC* and *MC* curves, the cubic function alone should be postulated. Incidentally note that the cubic function incorporates both quadratic and linear forms under special cases, when $b_3 = 0$ and $b_2 = b_3 = 0$, respectively. Thus, it is recommended that the cubic form be hypothesized for the total cost function.

While collecting historical data on costs, care must be taken to ensure that all explicit as well as implicit costs have been properly taken into account and that all the costs are properly identified by time period in which they were incurred. This does create some problems, for implicit costs are often hard to quantify, and some costs like maintenance and repair expenses are not recorded in accordance to the time they were incurred.

Once the list of cost determinants is ready, the functional form is chosen and all the necessary data are collected. Estimation of cost function is a matter of running a regression programme on computer. Of course, a careful researcher would experiment with alternative functional forms and alternative sets of causal variables on computer, and then on the basis of the principles of economic theory and statistical inference choose the most appropriate function for his use.

Having described alternative methods of cost estimation, it would be useful to provide a couple of guidelines on their choice.

(a) There is no historical data on cost, output, etc. for new products. Thus, neither the survivor method nor the statistical method is available for estimating the cost function for such products. Further, if the product is old but has been in existence for a few years only, data may be inadequate for applying either of the two methods. Incidentally, if the product is new in the country, but not in the world, historical data from other countries might be available. If so, for such data to be useful, the country should use similar technology, have comparable input prices or appropriate adjustments made to take care of such differences. In the circumstances, one may have to resort to the engineering method only.

(b) Since no method is perfect, if permitted by resources, more than one method should be used. This would improve the credibility of estimates and thus their usefulness.

## 6.8 MANAGERIAL USES OF ESTIMATED COST FUNCTIONS

There are basically three uses of an estimated cost function:

(a) To determine the optimum scale or size of the firm (fixed plant and equipments)

(b) To determine the optimum output for a given plant size, and

(c) To determine the supply schedule/curve.

To illustrate these uses consider the following hypothetical estimated cost function:*

$$C = 0.04\,Q^3 - 0.9\,Q^2 + (11 - K)\,Q + 5K^2 \tag{6.12}$$

where, $C$ = Total cost (in rupees)

$Q$ = Output (in physical units)

$K$ = Plant size (in multiples of $Q$)

Function (6.12) represents a long-run cost function, for in it plant size is a variable.

**Optimum Scale** The optimum scale is given by that valve of $K$ (plant size) at which the total cost is the least. The necessary and sufficient conditions for that are the following:

Necessary condition : $\partial C/\partial K = 0$

Sufficient condition : $\partial^2 C/K^2 > 0$

Applying these conditions to function (6.12), we get

$$\frac{\partial C}{\partial K} = 0 - 0 + 0 - Q + 10K = 0$$

or, $K = 0.1Q$

and $\partial^2 C/\partial K^2 = 10 > 0$

Thus, at $K = 0.1Q$, total cost is the least. If the firm wished to produce 10 units of output, its optimum scale equals 1; if it wants to produce 50 units, the optimum plant size is 5; and so on.

**Optimum Output** The optimum output from cost angle is defined as the output level at which the average cost is the least. To determine this, let us first get the cost function entirely in terms of output. This can be done by substituting the optimum value of $K$ (= $0.1Q$) in function (6.12):

$$C = 0.04Q^3 - 0.9Q^2 + (11 - 0.1Q)\,Q + 5\,(0.1Q)^2$$

or,

$$C = 0.04\,Q^3 - 0.95\,Q^2 + 11\,Q \tag{6.13}$$

Dividing Eq. (6.13) by $Q$ on both the sides, we get the $AC$ function:

$$AC = 0.04\,Q^2 - 0.95Q + 11 \tag{6.14}$$

For minimum $AC$ with respect to $Q$ the first derivative must be zero:

$$\frac{d(AC)}{dQ} = 0.08\,Q - 0.95 = 0$$

or, $Q = 11.87 \cong 12$

The sufficient condition for minimization is met as the second derivative (=0.08) is positive. Thus, the output of 12 units denotes the optimum output from the cost angle.

*The cost function here incorporates the output and scale variables only, and thus it ignores the other determinants of total cost. The effect of other variables could be incorporated easily by introducing a constant term in the cost function. This is avoided here, for it would complicate mathematical derivations.

**Supply Schedule/Curve** The supply schedule gives the quantities of a good that a firm is willing and able to supply at various (minimum) prices, other things remaining the same. To get such a schedule, one needs firm's cost function and its objectives. If the objective is profit-maximization, the firm must equate its marginal cost to its marginal revenue; there is also a sufficient condition as well as an economic condition which are ignored here (Refer Chapter 7, section 7.3). Thus the supply schedule is given by

$$MC = MR$$

The *MC* schedule is obtained by differentiating the total cost equation, and *MR* by differentiating the total revenue function.

$$MC = \frac{d}{dQ}(0.04\,Q^3 - 0.95\,Q^2 + 11Q)$$

$$= 0.12\,Q^2 - 1.90\,Q + 11$$

and $$MR = \frac{d}{dQ}(TR) = \frac{d}{dQ}(PQ)$$

or, $$MR = P + Q\,\frac{dP}{dQ}$$

One needs the equation of the demand curve in order to express *MR* in terms of *P* or/and *Q*. Under conditions of perfect competition (Refer chapter 7), the price is independent of quantities a firm sells, i.e., $\frac{dP}{dQ} = 0$. Thus, if we assume perfect competition, $MR = P$. Recall that for profit maximization, $MR = MC$. Combining the two, note that under perfect competition, a profit maximizing firm must ensure $MC = P$. Equating *MC* to *P* (= *MR),* we get

$$0.12\,Q^2 - 1.90\,Q + 11 = P$$

or,

$$0.12\,Q^2 - 1.90\,Q + (11 - P) = 0$$

This is a quadratic equation in *Q,* a solution of which yields

$$Q = \frac{1.9 \pm \sqrt{3.61 - (0.48)(11 - P)}}{0.24}$$

or,

$$Q = \frac{1.9 \pm \sqrt{0.48P - 1.67}}{0.24} \qquad (6.15)$$

Equation (6.15) gives the equation of the supply schedule subject to two features of such a schedule. One, the supply is a positive function of own price. That is, the coefficient of *P* in the function must be positive. This rules out the negative sign preceding the second term of the numerator of function (6.15). Two, for all economic goods, there is a cost of production, and a firm has a choice of not supplying anything (i.e., shut down) if its price is below its cost. This cost for a profit-maximizing firm, happens to be the minimum *AC* in the long-run, and minimum *AVC* in the short-run. This is because if a firm closes down it would incur no cost and make no profit in the long-run, for all costs are variable. In contrast, if a firm shuts down in the short-run, it would still incur a cost equivalent to total fixed cost. Thus, its maximum loss in the short-run should not exceed total fixed cost. In other words, its shut down price equals the minimum *AVC*. Accordingly, a firm would supply a positive quantity in the long-run if and only if its *P* exceeds the minimum *ATC;* and in the short-run if and only if its *P* exceeds the minimum

*AVC*. To superimpose these conditions on function (6.15), we need to determine the minimum *AC* and minimum *AVC*.

The minimum *AC* and minimum *AVC* can be found out through the calculus of optimization. It was seen above that the *AC* is minimum when output equals 12 units. Setting $Q = 12$ in *AC* function (6.14), we get minimum *AC* as

$$\text{Min. } AC = 0.04\,(12)^2 - 0.95\,(12) + 11$$
$$= 5.36$$

The total cost function (6.13) has no intercept term, meaning there are no fixed costs. This is consistent with a long-run cost function alone. Consequently, we do not have a short-run cost function and hence there is no way to find the minimum *AVC* or a short-run supply schedule.

The long-run supply schedule is thus given by

$$Q = \frac{1.9 \pm \sqrt{0.48P - 1.67}}{0.24}, \text{ if } P \geq 5.36 \qquad (6.16)$$

$$= 0, \text{ if } P < 5.36$$

Function (6.16) contains quantities the firm would supply at various prices, *ceteris paribus*. One limited supply schedule and the corresponding supply curve would be the following:

**Table 6.3** *Supply Schedule*

| *P* | *Q* |
|---|---|
| 5 | 0 |
| 6 | 12.5 |
| 10 | 15.3 |
| 15 | 17.6 |
| 25 | 21.3 |
| 50 | 27.6 |

The above schedule was obtained by plugging in specific values of *P* in function (6.16).

In terms of geometry, the long-run supply curve for a purely competitive profit-maximizing firm is given by that part of its *MC* curve which lies above its *AC* curve. And, its short-run supply curve is given by that part of its *MC* curve which lies above its *AVC* curve. Thus, in Fig. 6.3 (vide section 6.4), the *BC* part of *MC* curve marks the long-run supply curve while the *AC* part of *MC* curve marks the short-run supply curve of the firm in question.

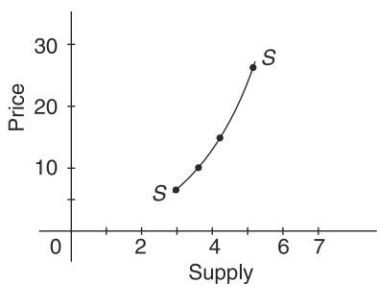



**Fig. 6.6** *Supply curve*

The cost function is thus a useful tool for determining the optimum scale, optimum output and the supply schedule. The knowledge of these fundamentals is crucial for the firm's plan for future expansion, today's operation and for setting the price for its products.

## 6.9 SUPPLY FUNCTION

Demand and supply are the two basic tools of economics. Demand for a good, as seen in chapters 2–4, draws from the theory of consumers' behaviour. Supply of a good is derived from the theory of firm behavior (and the behaviour of factors of production), which, as presented partly in chapter 1, and in chapters 5 and the present one, is based on the firms' objectives, their production functions (technologies), inputs' prices (or inputs' supplies), firms' organizations (single or multiple product firms), size and integration, government regulations, etc. In the demand theory, consumers were the only behavior unit and thus the demand function was relatively easy to derive systematically. However, in the supply theory, behavior of several units (such as firms, workers, savers, investors and governments) are involved and thus it is hard to derive the supply function in that unique fashion. Thus, at the cost of such a neat explanation, we present a theoretical link between the behaviour of all the relevant units to rationalize and present the supply function.

At the outset, note that, just as in demand theory, there is an individual firm's supply function as well as the industry' (i.e., all firms') supply function for each good and service. Further, the industry supply is merely a horizontal summation of supplies of all the individual firms (i.e., supplies by all firms at each of given prices) in the industry, assuming they are mutually independent in production. Also, supply means the quantities the firm/industry would (i.e., willing and able to) supply at various (**minimum**) prices. It must be emphasized that while in demand definition, the price at which the consumer demands was the maximum; here in supply definition, the price at which the firm supplies is the minimum. The rationale for this is simple, in that while the consumer pays the price, the firm/industry receives the price.

Recall that the production theory, and the duality principle of production and cost indicate that cost depends positively on the level of output and inputs' prices and negatively on the technology (factor productivities). Besides, recall that cost is affected by the learning curve, and economies of scope and transactions cost. Further, the profit theory argues that the necessary condition for profit maximization requires $MC = MR$ in general and $MC = P$ in perfect competition. These together suggest that supply of any product depends positively on output and factor productivities and negatively on inputs' prices and the extent of integration. Looked in other way, the higher the factor productivities, the lesser the input requirements and lesser is the total cost of production for a given level of output. Similarly, the lower is the input price, *ceteris paribus*, the lower is the cost of production. Also, the greater the integration a firm undertakes, *ceteris paribus*, the lower would be the cost for a given output. Further, a reduction in total cost of production lowers the cost curves, including the marginal cost curve, which, through the supply curve analysis of the previous section, leads to an increase in supply. Thus, increase in factor productivities, *ceteris paribus*, brings an increase in supply, and vice versa. Increase in factor prices, *ceteris paribus*, leads to increase in production cost, which, in turn, induces the firm to restrict their supply. Thus, factor prices exercise a negative impact on supply. Incidentally note that since there are more than one factors of production, each of factor productivities and factor prices is a set of variable (vector) rather than just one variable. Since integration brings cost saving, it tends to increase the supply.

Since price is the revenue per unit that a firm gets by selling its product, supply of a good varies directly with the price of the good, *ceteris paribus*. This is known as the **law of supply**. An alternative rationale for this important law could be provided through two sources, viz. (a) the law of diminishing marginal returns, and (b) entry/exist of new firms in the industry. Translated into cost terms, the law of diminishing marginal returns means that the addition to total cost for producing an additional unit of the good (i.e. marginal cost) rises as total output expands. Further, if marginal cost rises as output expand, a

firm would produce and supply more of a good only if it could command a higher price for its product. Thus, as the price goes up, *ceteris paribus*, supply goes up, and vice versa. The upward sloping market supply curve is further reinforced through the incentive/disincentive for new firms to enter/leave the industry in the face of increase/decrease in the price of their product. When the product price increases, *ceteris paribus*, firms' profitability increases; which, in turn, attracts the new firms to enter the industry. Since the number of firms influences the aggregate supply positively, the entry of new firms would tend to increase the market supply. Quite the opposite would happen when product price falls.

For simplicity, we detailed the production and cost theory, as well as the derivation of the supply equation [vide equation (6.16)] for a single product firm only. Accordingly, we ignored the effects of the prices of goods related in production on the supply of a good. Thus, for example, a multi product firm has the option of switching the production from one good to another if such an action improves its profitability. Just as prices of goods related in consumption exert influence on consumers' demand for goods and services, prices of goods related in production influence the firms' supply of goods and services. Thus, for example, an increase in the price of the colour television (CTV) *ceteris paribus*, would lead to a decrease in the supply of black and white television (BWTV) and vice versa. This is because, CTV and BWTV are substitutes in production—that is, the firm could choose the composition of the two types of televisions it wants to manufacture, given its plant. Thus, when CTV price goes up, the price of BWTV remaining the same, the firm would be induced to produce more of CTV and less of BWTV, which, in turn, would change the supply of the two types of televisions in the direction indicated above. To cite another example of two goods which are substitutes in production, one could talk about a dairy farm supplying both milk and ice cream. During summer season, the dairy may supply more of ice cream and less of milk, as the ice cream price goes up, *ceteris paribus* (i.e., price of ice cream relative to price of milk goes up); and may reverse the supplies during winter season when the opposite happens. In contrast, if the two goods were complimentary in production (like low-fat milk and butter for a dairy firm), quite the opposite would happen. For instance, if the price of low fat milk goes up, other things remaining the same, the supply of butter would increase, and vice versa. This happens because low fat milk price has increased, dairies are induced to go for expansion; and to increase low fat milk supply, since increased low fat milk supply is associated with the increased supply of butter, the supply of butter would increase as well. This assumes that the dairy has some excess capacity. Once again, since it is conceivable that a product has more than one related good in production, the price of related goods' variable is a vector of variables instead of a single one.

The objective(s) that a firm (or all firms in the industry) pursues has bearings on its supply. Generally speaking, firms seeking maximum profit supply less than those looking for service to the society with break-even or nominal profit, or the ones which look for maximum market share or those who look for maximum sales subject to a target profit, and so on. This would be pursued further in more detail in the next chapter. Since taxes increase and subsidies reduce cost, such government interventions affect supplies of concerned goods. The environment in which the firm/industry operates, like industrial peace versus strikes and lock-outs, political stability or otherwise, and the weather conditions obviously affect the cost as well as supply of goods and services. Suitable climate, stable governments and industrial harmony encourage producers to supply more and vice versa. For example, supply of woolen garment tends to increase in summer, as there are fewer buyers and there is a carrying cost of inventories. Unstable governments tend to increase uncertainty and thereby depress the supplies. Strikes and lockouts have depressing effects on supplies.

In the dynamic world, just as the consumers' expectations influence their demands, supplies of durable goods are affected by the suppliers' expectations about future prices of their goods. If firms expect that their product is going to be in short supply at a later date and that they could charge high price in future, they would sell less now and build up inventories for future sales to take advantage of price hike. Quite the opposite would happen if they fear a glut for their product and accordingly expect a fall in price in future. Similarly, sellers' expectations about future production costs vis-a-vis current production cost, and about the availability of raw-materials and intermediate goods would influence their current supplies. For instance, if the shortage of raw-material is feared, supplies would decline and vice versa.

In addition to the above factors, the industry supply would vary positively with the number of firms in the industry. The character of sellers, in terms of their strategies of competition in the market, also has a bearing on the industry's supply. If sellers act like rivals, they would enter into cut throat competition and supply more of their products than otherwise. In contrast, if they collude, formally (like the Organisation of Petroleum Exporting Countries : OPEC) or otherwise, they would tend to restrict the supply of their product.

Collecting all the above determinants of supply together would give us the industry supply function like the one in function (6.12).

$$Q_X = f(P_X, P_I, E_I, P_S, P_C, O, G, W, P_X^E, N, Z) \qquad (6.12)$$

$$f_1, f_3, f_5, f_9, f_{10} > 0 > f_2, f_4$$

$$f_6, f_7, f_8, f_{11} = ?$$

where,

$Q_X$ = Supply of good $X$
$P_X$ = Price of good $X$
$P_I$ = Inputs' prices
$E_I$ = Inputs' productivities (efficiencies) or the state of technology
$P_S$ = Prices of goods that are substitutes to good $X$ in production
$P_C$ = Prices of goods that are complements to good $X$ in production
$O$ = Objectives of firms in the industry
$G$ = Government policy towards good $X$
$W$ = Weather, industry environment etc.
$P^E$ = Industry's expectations about future price of good $X$
$N$ = Number of firms producing good $X$
$Z$ = "Other" determinants
$f_i$ = partial derivative of $f$ with respect to $i$th determinant in function (6.12)

Like demand analysis of Chapters 2–4, one could go for a detailed supply analysis here. This would involve discussion of changes (increase and decrease) in supply (as distinguished from extension and contraction in supply), supply elasticities, and estimation and analysis of supply function. The approach would be the same as adopted for demand analysis above, and the learning would be limited. For these reasons, no detail discussion on supply analysis is attempted here. Suffice it to point out, that supply of a good expands and contracts due to increase and decrease in that goods' price, supply increase and decrease due to changes in any one or more of the non-own price determinants of supply, and that there

are a number of supply elasticities, one with respect to each one of the supply determinants. Further, note that each of the non-own own price determinants of supply (viz. inputs' prices, inputs' productivities or technology, prices of related goods in production, etc.) is called as a "**supply shifter**". Accordingly, while the movements along the supply curve of, say, good *X*, are caused by a change in the price of good *X*, shifts in the supply curve of good *X* are caused by a change in any one or more of the shifters of the supply curve of good *X*. The said curve would shift to the right if, for example, the number of firms in the industry increase or the price of some input (factor of production) falls. The curve would shift to the left, if the opposite happens.

## REFERENCES

1. Dean, J. (1976): *Managerial Economics*, New Delhi, Prentice-Hall.
2. Gupta, G.S. (1975): "Economies of Scale in Indian Cement Industry", *Economic and Political Weekly*, X. 13 (March 29).
3. Gupta, V.K. (1968): "Cost Functions, Concentrations, and Barriers to Entry in Twenty-Nine Manufacturing Industries in India," *Journal of Industrial Economics*, XVII (Nov.).
4. Johnston, J. (1960): *Statistical Cost Analysis*, New York, McGraw-Hill.
5. McGuigan, J.R. and R.C. Moyer (1986): *Managerial Economics*, New York, West Publi. Co.
6. Tintner, G. (1965): *Econometrics*, New York, John Wiley and Sons.
7. Walters, A.A. (1963): "Production and Cost Function," *Econometric*, XXXI.

## CASELETS

1. Sharad Motors needs an economist who is conversant with the cost-output relationship. In order to test this knowledge, the company has designed a small test based on certain cost-output data from its own company. The same are given below:

| *Q* | *TC* | *Arc MC* | *TFC* | *TVC* | *AFC* | *AVC* | *ATC* |
|---|---|---|---|---|---|---|---|
| 0 | 250 | | | | | | |
| 10 | | 10 | | | | | |
| 20 | | | | | | | 21 |
| 30 | | | | 220 | | | |
| 40 | 510 | | | | | | |
| 50 | | | | | | 6 | |
| 60 | | | | | | | 10 |
| 70 | | 6 | | | | | |
| 80 | | | | 590 | | | |
| 90 | | | | | | 9 | |

where, *Q* = Output, *TC* = total cost, Arc *MC* = arc marginal cost, *TFC* =total fixed cost, *TVC* = total variable cost, *AFC* = average fixed cost, *AVC* = average variable cost and *ATC* = average total cost.

The company requires the candidates to complete the above table. Suppose you were a candidate for the post. Using your economics knowledge, complete the table.

2. The short-run total cost function of Jagdish Enterprises (JE) has been estimated as

$$TC = 100 + 6\,Q + 0.25Q^2$$

(where, $TC$ = total cost; $Q$ = output)

(a) Determine equations for JE's average fixed cost (*AFC*), average variable cost (*AVC*), average total cost (*ATC*), and marginal cost (*MC*).

(b) Choose a few values for $Q$ between $Q = 10$ and $Q = 25$ and draw the graphs of *AFC*, *AVC*, *ATC* and *MC* curves. Are these graphs consistent with the general properties of short-run cost curves as shown in Fig. 6.3 in the text? Explain.

(c) If you were the economist of JE, how much would you recommend it to produce, assuming the market (demand) is not a constraint?

(d) Determine the output level at which $ATC = MC$. Compare your answer to this part with that of part (c) above and comment.

(e) Determine the short-run and long-run supply functions of the firm, given that the firm sells in a purely competitive market and it aims at maximum profit.

3. Suppose *ABC* Ltd. faces the production function as given in Table 5.1 (vide chapter 5) and that it has installed a capacity at $K = 6$. The market prices of labour and capital are Rs. 90 and Rs 50, respectively. Based on these data, attempt the following questions:

(a) Using the duality principle of production and cost, determine the firm's short run cost schedules (all seven of them). Draw on a single graph the *AFC*, *AVC*, *ATC* and *MC* curves, and mark each of the short and long-run supply curves of the firm.

(b) Using the multiple regression software, estimate the total cost function in its quadratic form (like $TC = a + bQ + cQ^2$). Determine the firm's optimum output. Assuming that the firm was selling its product in the perfectly competitive market, derive its supply schedule/ curve.

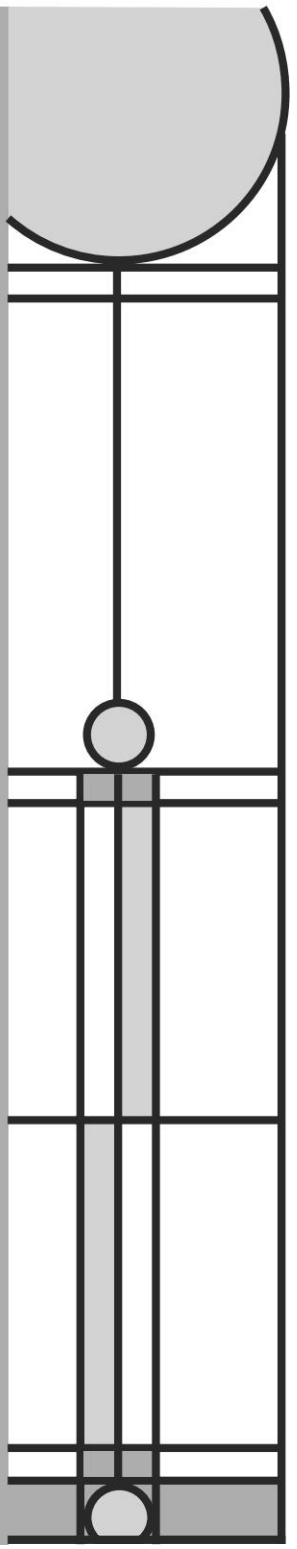
4. Consider the firm Radhe Enterprises (RE), which produces the two related goods in quantities $Q_1$ and $Q_2$, faces the following cost (*TC*) equation:

$$TC = 100 + 5\,Q_1 + 4\,Q_2 - 0.25\,Q_1\,Q_2 + 0.1\,Q_1^2 + 0.2\,Q_2^2$$

Verify if the two goods are complementary in cost.

RE's fixed cost remains at 100 irrespective of whether it produces just one or both the goods. Does the firm enjoy the benefits through economies of scope? How?

The McGraw-Hill Companies



CHAPTER 7

# PRELIMINARIES ON PRICE AND OUTPUT DETERMINATION

Pricing is an important, if not the most important, function of all enterprises. Since every enterprise is engaged in the production of some good(s) or/and service(s), incurring some costs to sell its produce in the market, it must set a price for its product. Here it must be noted that there are some extreme cases where the firm has no say in pricing its product. For example, as will be demonstrated later, under perfect competition and such situation in the market when the good happens to be of such public significance that its price is decided by the government, firm is merely a price taker. In an overwhelmingly large number of cases, the individual producer enjoys some market power and accordingly plays some role in pricing its product.

It is said that if a firm is good in setting its product price, it would certainly flourish in the market. This is because the price is such a parameter that it exerts a direct influence on the product's demand as well as on its supply, and through them on its turnover (sales) and profit, both of which are the important yardsticks for the success or otherwise of the firm (Refer chapter 1, section 1.3). Every manager endeavours to find the price which would best meet with its objective. On the one hand, if the price is set too high, the seller may not find enough customers to buy his product. On the other hand, if the price is set too low, the seller may not be able even to recover his costs. Thus, there is a need for the *right price*, and that is rather hard to know. However, we know that it must fall between its two extreme values viz. the product's value to the consumer (utility) and product's cost to the producer. In addition, it is understood that the right price must be high enough so that business makes some profit and yet low enough to keep customers coming through the door. Since utility (demand) and cost (supply) conditions are variable over time, what is the right price today may not be so tomorrow. Hence, pricing decisions must be reviewed and re-formulated from time to time.

## 7.1 PRICE CONCEPTS

Adam Smith suggested that a good has two kinds of values; viz. value in use and value in exchange. Value in use is the benefits that the buyer gets out of its use, and it could well vary from the buyer to

buyer. For example, the utility from a glass of milk to a thirsty (hungry) person will be much more than to another person who is not thirsty (hungry). In contrast, exchange value is the price at which the seller sells and buyer buys the good in the market. Accordingly, price denotes the exchange value of a unit of a good, expressed in terms of money. Thus, the current price of a Maruti Zen car is around Rs. 400,000, the price of a regular hair cut is Rs. 25, the price of a five minutes phone call from Ahmedabad to Delhi is Rs. 10, and so on. Nevertheless, if one gives a little thought to this subject, one would realize that there is nothing like a unique price for any good. Instead, there are multiple prices. The pertinent questions here are the following:

(a) Price of what?
(b) Price to whom?
(c) Price where?
(d) Price when?

Unless a good is properly defined, one can't talk of its price. This is particularly true with the case of durable goods. Consider the Maruti Zen car, for example. When one says, its price is Rs. 400,000, one is not quite precise. Is it the price of an ordinary (non-AC) or of a deluxe (AC) model? Does it include road (RTO) tax; if yes, for how long? Does it include, insurance; if yes, comprehensive or liability one? Does the price include free services; if yes, how many and when? Does it include a guarantee; if yes, for which parts and for how long? Depending on the answers to these and such questions, there are multiple prices of Maruti Zen car. However, if the product is unambiguously defined in all such respects, there could be a unique price, assuming the **law of one price** is practiced.

Price of a well defined product varies over the types of the buyer. Thus, there are prices to wholesalers/commission agents, to retailers and to consumers. Since wholesalers/commission agents buy in bulk, producers incur lower transaction costs (per unit) in such dealings than in dealings with retailers and consumers. The same is the case with regard to trading with retailers vis-a-vis consumers. Besides, the market intermediaries incur some costs (transport, storage and interest costs) and they too must make some profits. Thus, the three prices are different, for they include different things.

The price of a good or service also depends upon the place it is received, for there are transport and insurance costs. Thus, there is a price at the site of the factory/production centre, major rail or road head, the heart of the town, suburbs and at one's house (home delivery). Also, while some transactions take place on credit, others are on down payments. Consequently, for an item, there are two prices, viz. (a) price on buying on credit for a given period, and (b) buying against down payments. The former includes an element of interest cost, which the latter does not, and accordingly the former is higher than the latter.

It should be obvious to the readers, that the price differences on account of the above four factors are more significant, the more durable, expensive and heavy the good is, and vice versa. Thus, the case of multiple prices is more serious in the case of items like cars, refrigerators, coal, furniture and bricks, and is of little significance for items like shaving blades, soaps, tooth pastes, creams and stationeries.

Since differences in various prices of any good are due to differences in transport cost, storage cost, accessories, interest cost, intermediaries' profits, etc., they are explainable. One can still conceive of a basic price, which would be exclusive of all these items of cost, and then rationalise other prices by adding the cost (price) of special items attached to the particular transaction. In what follows, we shall explain the determination of this basic price alone and thus resolve the problem of multiple prices.

Before we move to the next section, it may be useful to understand that price (of goods and services) is known by different names; such as price, tariff, fare, toll and fee. Generally speaking, the term price

is used when the item under trade is a commodity, like car, computer, furniture, etc. It is called tariff in cases like use of telephone, electricity, etc. and also for import/export duties. The notion fare is applied in case of charges for transport services, like air fare, train fare, bus fare, taxi fare, and so on. The term toll refers to payment for the use of roads, bridges, etc. The last term, viz. fee is used for the payment for the use of services like education (school/college fee), consulting, insurance, etc. It must be emphasized that unlike the first four terms, fee has a special significance. It constitutes the payment for an item whose quantity is not exactly defined. Thus, when you pay your college tuition fee, you are free to attend all class sessions meant for you as well as the faculty time in their offices for your genuine queries. However, if you miss some class sessions or do not see your faculty for any doubts, you get no refund. Similarly, the fee you pay for the use of library, computers, park, club, etc.; and the ones you pay to an advocate, a doctor, an accountant, a management consultant, etc. may not define the product exactly.

## 7.2 PRICE DETERMINANTS

The list of price determinants is rather long and it is hard even to pin them down in some meaningful way. Accordingly, what we plan to do here is to highlight the important milestones in pricing a product. There are two kinds of rules in pricing which all firms and customers must obey. They are the (*a*) *market rules and* (*b*) *governments' rules*. The market rules are contained in the equilibrium between demand and supply. Demand and supply are influenced by factors such as, market structure, firms' objectives, practice of the law of one price vis-a-vis pricing strategies, presence of symmetric or asymmetric information, existence of externalities or otherwise, and the degree of risk and uncertainty in the product market. The government rules include the (indirect) taxes and subsidies, price ceilings and price floors, and pricing pegging. Even these rules operate on prices through their effects on the demand and/or supply of the relevant goods and services. However, to separate the two rules, government interventions are considered separately in the text. The details on all these factors are discussed in what follows.

### Market (or Demand-Supply) Rules

Since the publication of Alfred Marshall's "Principles of Economics" in 1890, it is well-known that the price of a product is determined by the *demand for and supply of* that product. According to Marshall, the role of these two determinants is like that of a pair of scissors in cutting cloth. It is possible that at times, while one pair is held fixed, the other is moving to cut the cloth. Similarly, it is conceivable that there could be situations under which either demand or supply is playing a passive role, and the other, which is active, alone appears to be determining the price. However, just as one pair of scissors alone can never cut a cloth, demand or supply alone is insufficient to determine the price. A simple illustration is provided in what follows.

Let the demand and supply schedules of a good be as shown in Table 7.1.

**Table 7.1** *Demand- Supply Schedule*

| *Price* | *Demand* | *Supply* |
|---|---|---|
| 5 | 100 | 200 |
| 4 | 120 | 180 |
| 3 | 150 | 150 |
| 2 | 200 | 110 |
| 1 | 300 | 50 |

Of the five possible prices in the above example; price ($P$) = 3 would be the market clearing price. No other price could prevail in the market. For, if $P = 5$, supply would exceed demand and consequently the producers of this good would not find enough customers for their product, thereby they would accumulate unwanted inventories of output, which, in turn, would lead to competition among the producers, forcing price to 3. Similarly, if $P = 1$, there would be excess demand, which would give rise to competition among the buyers of the product, forcing price to 3. At $P$ =3, demand equals supply and so both producers and consumers are satisfied. The economists call such a price the equilibrium price; equilibrium meaning, "in balance" or "at rest". The equilibrium price is thus determined by demand and supply.

It is explained in Chapters 2 and 4 that the demand for a good depends on a number of factors and so does, as discussed in chapter 6, the supply of a good. Thus, every factor which influences either demand and/or supply, is in fact, a determinant of price. Accordingly, a change in demand and/or supply causes price change.

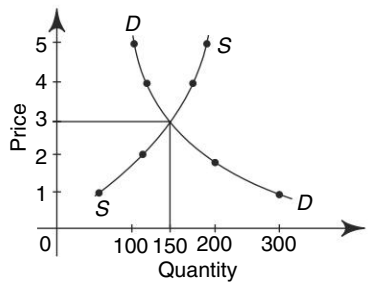



**Fig. 7.1** *Demand-supply curves*

The above explanation is so neat! Could it be practiced? Not quite! One reason is that it may not be consistent with the existing *government rules/interventions*. Other reasons are found in the explicit understanding of the demand and supply of the good in question. To take care of such rules and situations, demand and supply will have to undergo corresponding adjustments. In what follows, we discuss the various kinds of situations that exist in the products markets, as well as the varieties of government interventions that impinge on demand and supply, and thereby on the price-output decisions of the firms. The consequences of all such factors on demand and supply, and hence on product prices will be analysed in the next three chapters.

## Market Structure and Firms' Objectives

As will be clarified later, demand for a good facing the firm depends on the *market structure* in the industry (intra-industry competition), which varies across industries. Similarly, supply of a good from a firm/industry depends not only on the firms' *cost of production* and on whether there is *perfect competition in the market or not*, but also on the *objective(s) that a firm* is pursuing in the business. As studied in Chapter 6, only under perfect competition and profit maximization goal of the firm, a part of the marginal cost curve denotes its supply function. Further, the concept of supply curve of a firm, in fact, is valid merely in perfect competition. This is so because, recall that, a supply curve gives the quantities the firm would supply (produce and sell) at various (minimum) prices, *ceteris paribus*. Implicit in the previous sentence is that the *firm is a price-taker*, which is true only under the conditions of perfect competition in the industry the firm is operating. In all other kinds of market structure, *firm enjoys some market power* and hence in pricing its product. Again it would be demonstrated later in this chapter itself that the optimum price is influenced by the objectives that a firm pursues. Due to all such factors, in non-perfectly competitive markets, economists talk of *revenue and cost curves* (functions) instead of demand and supply curves, respectively while discussing firms' product's price-output decisions.

## Law of One Price

Traditional price theory is based on the assumption of uniform price of a good at any point of time. Under this theory, all buyers of a given good (with given quality) are charged the same price no matter

whether the customer is a large or small buyer, a regular or a first timer buyer, is rich or poor, crazy for the product or not, is your partner in the business or not (inter-divisional transactions in a vertically integrated firm), etc. Due to the increased competition and squeezing profits, and improved expertise and management tools, the law of one price in the present day world holds good in some industries only. Accordingly, the standard demand-supply model is not applicable in cases where the law does not hold.

### Strategic Pricing

With impending intensive competition and reduced profit margins; cost competitiveness, big businesses and pricing strategies have emerged. Firms are currently striving hard to cost cutting through downsizing, and all kinds of integration, such as horizontal, vertical and even conglomerates; and strategic alliances, franchises, and even joint ventures (with national as well as foreign partners). Such developments have induced firms to go for pricing strategies (known as *dynamic pricing*) under which they practice not only price discrimination but also strategies like, peak-load pricing, cross subsidization, transfer pricing, two-part pricing, block pricing, tying, commodity bundling, price matching, limit pricing, predatory pricing, penetrating and skimming pricing, etc. Such strategies affect demand and/or supply curves and thereby render the usual demand-supply model of little significance.

### Asymmetric Information

The standard demand-supply model assumes that the information set available to both the seller and buyer of a good is uniform and complete, both in terms of the content and time. This is not always true. In case of *new products* under "*commodities*", the assumption is approximately valid as the information technology (IT) is well developed and fairly accessed. However, in case of *old (prior owned or used) commodities* and *services*, the assumption of complete and uniform information may not hold water. For example, in case of, say, old car, the owner of the car knows more about the product than the prospective buyer. Similarly, for health insurance, the buyer of the said policy knows more about his/her health condition than the seller (insurance company) of the product. Also, for hiring an employee, the prospective employer knows less about the quality (expertise, sincerity, integrity, etc.) of the employee than the employee concerned. Accordingly, there is information asymmetry while trading in old goods and services, which violates a corresponding assumption of the demand-supply model. Due to such factors, product pricing method cannot be handled uniformly through the demand-supply model.

### Externalities

Recall from Chpater1, that while some goods are subject to externalities, others are not. For example, while dealing with products like car, house, food, drinks, health insurance, hair cut, etc., the costs and benefits fall squarely on the sellers and buyers of the car, and none others. In contrast, in goods like a pet dog in the house, a rose garden in the house, a television in the house, a chemical factory in the city area, basic research, etc., the benefits and costs are born, not just by the buyers and sellers of the product, but also to some extent by some "other people" (called "*third party*") as well. For example, if Mr. Ashok keeps a pet dog in the house, it could cause a nuisance (and hence cost) to the neighbours if it barks or stinks, and/or a benefit to the neighbour if it serves as a police/watchman to them in case of threats from thieves. In contrast, if a house has a rose garden, it beautifies the surroundings with fragrance, and thereby brings benefits to the neighbours. The example of Bhopal gas tragedy of 1984 is well

known to everyone in which the Union Carboid Company (UCC)'s chemical factory's gas leak rendered thousands of unconnected people (third party) dead and injured, causing serious costs to them. Thus, there are both positive (benefits) and negative (costs) externalities associated with some goods and, of course, the two may not just cancel. In such cases, the free market access to such products, while could bring optimal results for particular buyers and sellers (private), they could well lead to sub-optimal decisions for the society, besides creating the inter-personal problems in the society. Accordingly, externalities are considered as a source of market failure. Pricing of such products thus require some specific modifications to the standard demand-supply model. *Public goods* also cause such issues and, in fact, in a much larger way; for they have no price tags (non-excludable) are available for consumption to everyone (non-rival), and are supplied simply by the governments through tax-payers money.

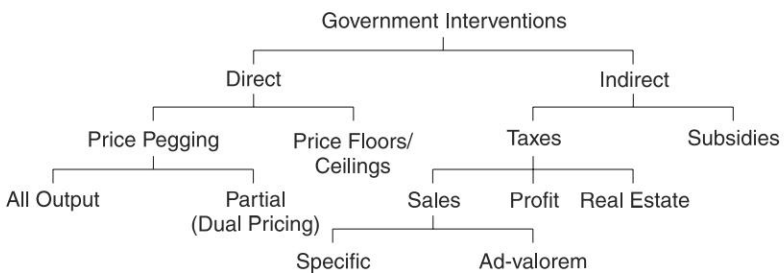
## Risk and Uncertainty

While the demand and supply theories are well developed and fairly sound, the knowledge about them for any product can never be treated as accurate. Thus, for a well established product, say, car, a competent researcher can very well estimate its demand and supply (cost function and firms' objective) function using the large data set and most sophisticated estimation technique. Nevertheless, neither the researcher him/herself will boast nor its users will ever take those estimates as absolutely accurate. As discussed in Chapter 4, all such estimates are subject to errors due to models' specification and measurement errors in data. Further, the said estimates may be valid for the past and not future, and while using them in pricing, one is dealing with future which is subject to risk and uncertainty. Thus, risk and uncertainty play role in pricing and accordingly require some adjustments in demand-supply model of pricing.

## Government Rules/Interventions

Governments in most countries today play an important role in product pricing. In this sub-section, government interventions in pricing are discussed with special reference to India. At the outset, we shall elaborate the types of price controls prevalent in India, and then go on their rationale and formulations.

**Types of Government Interventions** The several ways through which government influences product pricing are presented in the following chart under convenient heads.




In India as well as in most other countries, there are direct as well as indirect interventions by government in product pricing. The former includes price setting for the whole output (like petroleum

products, electricity, etc.) of a good or service or for a certain part of the product under the dual pricing system (like sugar, paper, etc.), and the price floors and ceilings found in case of a number of goods and services in the country. The Indian governments (central, states and locals), like most other governments, is itself an entrepreneur and thus is engaged in the production of many goods and services, whose prices for all outputs are obviously set by government agencies. These include services rendered by railways, post and telegraph networks, telephone company, electricity companies, road transport agencies, nationalized banks, etc., and commodities produced by nationalized textile mills, oil and natural gas networks, government owned iron and steel manufacturing units, cement firms, etc. Besides the government produced goods and services, government pegs the prices of certain other products which are produced under the private sector. These include major drugs, cement, paper, fertilizer, sugar, coal, school and college fees, and a few other essential goods and services. In case of some drugs and school fees in many institutions, government fixes the price for all their outputs, while in case of cement, paper and sugar, there is a dual pricing—a fixed part of the total output has to be sold at the government fixed price (called the *levy price*) while the remaining part is sold at the *free market price*. At one time, cement price in India was uniform throughout all rail heads and that price was decided on by the governments. Dual pricing has been practiced in cement, where a fixed part of it is to be sold at the levy price and the remaining at the free market price. The same is the case with regard to sugar, paper and a few other items of essential consumption. Under the liberation policy beginning 1991, this practiced has been restricted to certain extent but the basic principle still continues.

Price floors exist for many important agricultural goods in India and for the services of unskilled labour. Before the beginning of the sowing seasons each year, the Agricultural Prices Commission announces what they call the support, minimum or guaranteed prices for all important agricultural crops. Also, there is a minimum wage rate, below which no worker, however unqualified he or she may be, could be hired by any organization. Similarly, there are ceilings on a few prices. These include rent on residential or office accommodations, prices of life-saving and other basic drugs, etc. The indirect interventions are the means through which the government exerts influence on the product prices. These consist of various kinds of taxes excise duties, sales tax, value added tax (vat), custom duties, service tax, etc. and subsidies on goods and services. A large number of commodities and many services fall under one or more kinds of taxes, and there are subsidies available for the production of the selected goods across the country and for most goods if they are manufactured in the notified backward areas. The pertinent questions to ask are the following:

(a) Why government intervenes through various means in product pricing?

(b) How government determines the levy price and levy quantity, price floors and price ceilings, and tax and subsidy rates?

(c) What are the consequences of government interventions on price-output determination?

Consequences would be analysed later in Chapters 8–10, and discussion of each of the other two aspects follows:

**Reasons for Government Interventions** All government interventions are well founded, though may not be well implemented. Without going into the debate on the pros and cons of particular interventions, we shall discuss the theoretical rationale for them.

The government fixes/regulates the price for all output of a good when the good happens to be produced either exclusively or jointly in the public sector and/or it happens to be a basic good. These

include the prices of electricity, petroleum, cooking gas, services rendered by post and telegraph offices, telephones, railways, airlines, prices of life-saving and other basic drugs, cement, steel, fertilizer, pesticides, etc. The reasons for these measures are obvious. Some of these goods and services are manufactured by the government sector alone (e.g. post telegraph and railways services) and so it has to set their prices. The other such products are produced both by the public and private sectors, but as these happen to be essential goods either for human consumption itself and/or for the economic-social development of the country, their prices are managed by the governments. If governments did not fix their prices, they would either be not available for consumption to the vulnerable sections of the society due to high prices or would not be produced in desirable quantities due to low prices. Thus, if life-saving drugs were not governed by price controls, many patients will not be able to afford them. Similarly, if fertilizer prices were not fixed, many farmers would have gone without using it in at least optimum quantity, and consequently our agricultural production would have suffered a heavy jolt.

Levy prices are imposed on a part of the output of some goods under the dual pricing system in order to extend help both to the consumers and producers simultaneously. The dual pricing scheme exists for goods which have both essential as well as non-essential uses. Under the scheme, a part of the output is bought and sold at the levy price, which is low, and the other part at the free market price, which is high. The ration quota is so designed that the essential needs are met at the levy (low) price, and the residual demand is met through the free market price. Producers get the weighted average of the two prices, which leave adequate incentives for a proper flow of the good in question. By this system, thus, the interests of the vulnerable sections of the society and the economic development of the country, as well as the interests of the producers are well served. The levy price and quota vary from product to product as well as over time, depending on the need and sometime on political ground.

Price floors are imposed by government in order to safeguard the interests of producers in the event of a bumper crop. Thus, if there were no support or minimum prices of agricultural crops and if there was a favourable monsoon in a particular year, leading to a bumper harvest, prices of agricultural goods could be so low that farmers would have no profit, giving rise to what is called "*poverty in the midst of plenty*". In the circumstances, not only would the farmers' community be unhappy but the whole country might be taken to ransom as the farmers will retaliate by not cultivating such crops in the future, and thus creating a shortage. The commodities would then have to be imported at a grave risk to the country's goal of self-reliance, current account deficits, losses in foreign exchange reserves, political disturbances, etc. Thus, by declaring the minimum support prices in advance for all important agricultural goods, the Agricultural Prices Commission (APC) guarantees to purchase the respective produce at such prices should farmers not be able to sell their produce to others at prices equal to or above those minimum prices. By so doing, the APC ensures some incentives to farmers for cultivating such crops, and thereby help both the producers and the country in achieving its high goals. In advanced countries like USA, Canada, England and Germany, price floors are also imposed on the wage rate. These are meant to avoid exploitation of unskilled, child, female and other socially and economically backward classes of labour, and to ensure minimum remuneration to them, particularly at the time of high unemployment. In India, the minimum wage rules may exist but they are rarely honoured, particularly by the huge un-organised sector.

Price ceilings are found in the case of rent on residential and other accommodations, on the goods manufactured by monopolists and oligopolists, and on goods of essential consumption. Due to a high growth rate of population, there is a pressure on land and buildings. If their rents are not controlled, many people would not be able to afford to live in a house even with minimum facilities. Further, there

would be a class of landlords, who would make big profits with little efforts, enjoy numerous luxuries, and thereby giving rise to further inequalities. The prices of essential goods like basic and life-saving drugs, electricity, gas, diesel, petrol, kerosene, phones, etc. have to be contained in order to enable the vulnerable sections of the society to meet their requirements.

Government imposes taxes of various kinds on goods and services to (a) fetch revenue, (b) to regulate the consumption baskets of various segments of the population, and (c) to manage the production of various goods and services in the economy as per the requirements. Commodity (indirect) taxation is an important source of revenue to the government and no government can render its useful services to the society without it. By a differential rate system of commodity taxation, the governments encourage consumption of socially desirable goods through no or low tax rates on them, and discourage the consumption of luxury items by taxing them heavily. Subsidies are given in order to encourage production of essential goods in optimum quantities and at desirable locations, to encourage production of goods hitherto imported in undesirable quantities, and to encourage the use of essential inputs (like fertilizer, electricity) in the production of important goods and services.

**Formulations of Government Interventions** It is hard to rationalise government interventions in any sphere and product pricing is no exception. This is because they are usually integrated with the overall policy of the government, which is based on factors like economic, social, political, inter-country relations, etc. Nevertheless, some observations can still be made. As regards the price pegging for all outputs and parts of the output of the selected products are concerned, governments' assessment with regard to the goods and services that are essential either for the economic development of the economy and/or for human consumption is important. All the basic needs are intended to be met at reasonable (affordable) prices. The government either undertakes to produce such goods by itself and/or regulate the prices of such goods. When a product is seen to have both the essential as well as non-essential uses, governments resort to dual pricing. We have examples of the railways, post and telegram services, basic and life-saving drugs, electricity tariffs and petroleum products, whose prices are fixed for all outputs; and select food grains, kerosene, sugar, paper, cement, etc. where a part of the output is sold through ration shops at a levy price and the remaining at a free market price.

The levy prices and levy quota are governed by socio-economic-political factors. An important yardstick for the levy price is the cost of production. It is believed that the prices of goods produced in the public sector are primarily cost-based. Government prices its products and sets the levy price on other goods whose prices are directly regulated by it in such a way that all costs of production are recovered and some profit is made by the producers to leave incentives for them to continue in the business. However, this pricing method is not an easy one, as it must resolve the following issues:

(a) Which costs should be included in the cost of production for price-setting?
(b) Whose costs of production should form the basis for price-setting?
(c) What rate of profit should be used for fixing the cost-plus pricing?

As seen in Chapter 6, there are several kinds of costs, viz. variable and fixed costs, historical and replacement costs, cost of research and development (R&D), cost of bonus, private and social costs, etc. As regards the variable cost, there is no debate at all and all of it must be included in the cost. There is no such consensus with regard to most other costs. In case of a few basic goods and the ones that are produced entirely in the public sector, fixed costs may have to be ignored altogether or to a large extent

from the base cost for price-setting. Thus, if the government were to fix railway fares and freight charges on the basis of the cost inclusive of the overhead costs of putting up the railway tracks, railway stations, railway crossings, etc., one wonders how high these rates would have been and how many of us could afford to avail of railway services to the extent they are being used today. All forward looking businesses would like to use the replacement costs but the spoke persons of customers would argue for historical costs. Cost on research and development is also an important issue. While producers would argue for their inclusion, as these are really the costs incurred by them, consumers would oppose their inclusion, for these are costs associated with future production, which alone would benefit from R&D. Similar reasoning holds good with regard to the cost of bonus, particularly for that part of it which is voluntary. Consumers' spokesmen would vehemently argue against its consideration in pricing while producers' organizations would whole-heartedly ask for it.

In case of most goods and services, there are many firms. Since the factors of production are rarely homogeneous, cost of production of a good differs from firm to firm. Further, costs differ across firms in a given industry due to the age, size, location, technology, degree of integration, etc. of the firm. Due to these factors while some firms are more or less cost efficient, others are more or less cost inefficient. The problem then is, which firms' cost should form the basis for pricing? If the cost incurred by the most efficient firm is employed for the purpose, other firms would not be able to survive in the industry, which, in turn, would hamper the production of that good, which may hinder economic growth of the economy at large. On the other extreme, if the cost of the most inefficient firm was considered, the price would be too high and we would be encouraging inefficiency. For these reasons, the bulk line pricing is often recommended as the solution for this problem. By this method, the cost of the "bulk" of the firms is taken into account, but once again the term "bulk" is rarely quantified, and hence the trouble.

There is no unique profit rate. We have the rate of profit on capital employed, on net worth, on sales, etc. Once again there is no consensus as to which of these profit rates is the most appropriate from the point of view of setting the price. Similarly, no one knows the exact size of the "reasonable" rate of profit in a given industry; so as to leave enough incentives to the producers to continue in the business and at the same time ensures no exploitation of consumers through an unduly high price.

In view of these problems, the cost-based pricing is not an easy method of pricing. Consequently, there are no magic formulae for this method but rather we have price commissions, such as Agricultural Prices Commission, Bureau of Industrial Costs and Prices, etc. The indirect taxes and subsidies are formulated by the government in terms of the group of goods and services and are subjected to the above factors, and their rates of levy or grant again are determined on the basis of the socio-economic-political angles, and on the criteria of needs and availability of funds with the government.

To summarise the discussion of this long section, price of a good depends on the demand for and supply of that good, and on the government regulations applicable to the good in question. Thus,

$$P_x = f(D_x, S_x, G_x) \tag{7.1}$$

where, $P_x$ = price of good $x$

$D_x$ = demand for good $x$

$S_x$ = supply of good $x$

$G_x$ = government intervention in good $x$

For price function (7.1), it must be emphasized that behind demand lies, the customers' utility and constraints, market structure in the industry, applicability of the law of one price, pricing strategies, asymmetric information, externalities, and risk and uncertainty, etc. Behind supply, fall the production function, market structure in the factors market (input prices), extent of integration and corresponding cost of production, firms' objective, and applicability of strategic pricing, asymmetric information, externalities, and risk and uncertainty, etc. Government regulations pertaining the product pricing are in the forms of indirect taxes and subsidies, price ceilings and floors, and price pegging, if any. The previous chapters have presented a thorough analysis of demand and cost conditions. The remaining part of this chapter will analyse the impact of firms' objective on price and explain the market structure that prevails in various products in the country and globe. As a prelude to these, the conditions for profit maximization and the break-even analysis will be presented, for the former must hold when firms seeking maximum profit attempt pricing their products and the latter is relevant to the objectives that firms (particularly those firms which run on no-profit-no loss basis or the firms like non-government organizations—NGOs—who are motivated by social concerns) pursue as well as for benchmarking and some quick calculations.

## 7.3 CONDITIONS FOR PROFIT MAXIMISATION

Recall from Chapter 1 that, there are several theories of firm which postulate alternative objectives that a firm may pursue. However, in all those theories, profit enjoys a dominant place. After all, firms are made to make profits and their performance is often measured by the amounts of profit they make year after year during their existence. Further, the current profit (or the one a firm makes after a few years of the teething trouble) level serves as a major determinant even of the economic value of the firm (vide Eq. 1.2, Chapter 1). Accordingly, price theories have been developed on the assumption of profit maximization by firms. Thus, it is pertinent at this point to understand what this means in terms of the implied conditions. This could be pursued through table, graph or equation's approach. However, since equation method is the best as well as neat, and we have used that in previous chapters, we apply that in what follows.

Profit ($\pi$) is defined as the difference between total revenue ($TR$) and total cost ($TC$):

$$\pi = TR(Q) - TC(Q) \tag{7.2}$$

Each of $TR$ and $TC$ is treated as a function of output ($Q$) only. This is because the "other" determinants of each of demand (and $TR$) and cost are assumed to be given to the firm. For example, in case of $TR$ function, the firm is assumed to face given values of its customers' income, prices of related goods, consumers' taste (so no advertising on its part), number and distribution of customers, customers expected incomes and prices, etc. and thus the effects of all such variables is contained in the constant term, like $a$ if the demand function were

$$Q = a - bP$$

Further since $TR = PQ$ , the $TR$ could be expressed as a function of either $P$ or $Q$. Thus, $TR$ function in terms of $Q$ for the above demand equation could be obtained by substituting the value of $P$ in terms of $Q$ as derived from the demand equation in the $TR$ definition, as follows:

$$TR = Q\,[(1/b)\,(a - Q)]$$

$$= (a/b)\ Q - (1/b)\ Q^2 \tag{7.3}$$

In Eq. (7.3), *TR* is a function of *Q* only (*a* and *b* are parameters of the demand function)

One could simply apply the calculus of optimization technique to function (7.2) to determine the conditions for profit maximistion. If the firm decision variable were its output (since firm faces a given demand curve, it could decide either the price or output but not both), then maximizing profit with respect to output would tantamount to maximizing the value of function (7.2) with respect to *Q*. This has two conditions, viz. (a) the first derivative of the function with respect to *Q* must be zero and (b) the second derivative of the function with respect to *Q* must be negative. By their corresponding definitions, the said first derivative of *TR* equals marginal revenue (*MR*) and of *TC* equals marginal cost (*MC*). By calculus, the second derivatives of *TR* and *TC* would equal the first derivatives of *MR* and *MC*, respectively. Thus, the calculus conditions of profit maximization could be expressed as

$$MR = MC \tag{7.4}$$

$$d\,(MR) < d\,(MC) \tag{7.5}$$

(where, the operator *d* stands for the derivative with respect to output)

Calculus method of optimization ignores an important option which may be available to firms. This is the option of **shut down** in case the situation so demands. It is obvious that a profit maximizing firm would prefer to shut down if by operating it lends into a loss. Further, loss would occur if price is less than average cost. Also, recall that average cost varies with output in some way like the U-shaped or L-shaped curve. Accordingly, firms would shut down if price was below the minimum *AC*; the latter known as the **shut down price** in the long run. It is the long-run, because all costs (other than sunk cost, which is irrelevant once it has been committed) are avoidable in the long-run and so firm would take no loss. However, in the short run, fixed costs are unavoidable (fixed costs are fixed in the short run) and variable costs are avoidable. This means the firm would be willing to take a loss but no more than the total fixed cost in the short run. This argues that the shut down price in short run is given by the minimum average variable cost. Accordingly, we can write the third condition for profit maximization as follows:

$$P \geq P_s \tag{7.6}$$

(where $P_S$ = shut down price, which equals minimum *AC* in long run and minimum *AVC* in short run)

The above Eqs. (7.4), (7.5) and (7.6) represent the three conditions for profit maximization. The first condition simply ensures that profits are stationary, which could mean either maximum or minimum profit. Accordingly, it is called the necessary condition. The second condition distinguishes whether the stationary profit is maximum or minimum, and thus it is referred to as the sufficient condition for profit maximization. The last condition is relevant if the firm has the option to shut down.

## 7.4 PROFIT AND BREAK-EVEN ANALYSIS

Profit is the difference between total revenue and total cost, and the break-even point is defined as the one where profit = 0 or total revenue equals total cost. Given the demand and cost equations, the same could easily be derived.

The break-even analysis is often attempted on the basis of linear total revenue and total cost

functions. However, normally, as seen in earlier chapters, neither the total revenue nor the total cost function is linear for all possible levels of output, though for the relevant range of output, the two may be close to the linearity. Thus, in what follows, the profit and break-even analysis is first attempted on the linearity assumption and then for the non-linearity case.

## Profit and Break-Even Analysis—Linear Case

When total revenue ($TR$) and total cost ($TC$) functions are linear, their equations would be like the following:

$$TR = 10Q$$
$$TC = 12 + 7Q$$

where $Q$ = output. In this example, price ($P$) is assumed to equal 10, total fixed cost ($TFC$) to equal 12 and average variable cost ($AVC$) equals 7. The profit ($\pi$) function would then be given by

$$\pi = TR - TC = 10Q - 12 - 7Q$$

or
$$\pi = 3Q - 12 \tag{7.7}$$

Thus, given the linear total revenue and total cost functions, the profit function would also be linear. The break-even point is given by the point where $TR = TC$, or $\pi = 0$. Thus, applying this definition the profit equation gives

$$\pi = 3\,Q - 12 = 0$$

or
$$Q = 4$$

Geometrically, the concepts of profit and break-even are described in Fig. 7.2 below.

It is evident from Fig. 7.2 that in the linear analysis, profit is a monotonically increasing function of output. That is, the greater the output, the larger the profit, and *vice versa*. Since there is a fixed cost in the short-run (= 12), profit is negative at zero and small levels of output. The $TC$ curve is above the $TR$ curve for outputs up to $Q = 4$, indicating a loss, while quite the opposite is true beyond $Q = 4$, indicating a profit. The gap between $TR$ and $TVC$ is called the **total contribution**, which is always positive. The contribution per unit of output, called the **average contribution**, is defined as price minus average variable cost. In the above example, the values for all relevant concepts are as follows:

$$P = TR/Q = 10$$
$$TFC = 12$$
$$TVC = 7Q$$
$$AVC = TVC/Q = 7$$

Total contribution $= TR - TVC$
$= 10Q - 7Q$
$= 3Q$

Average contribution $= P - AVC$
$= 10 - 7$
$= 3$

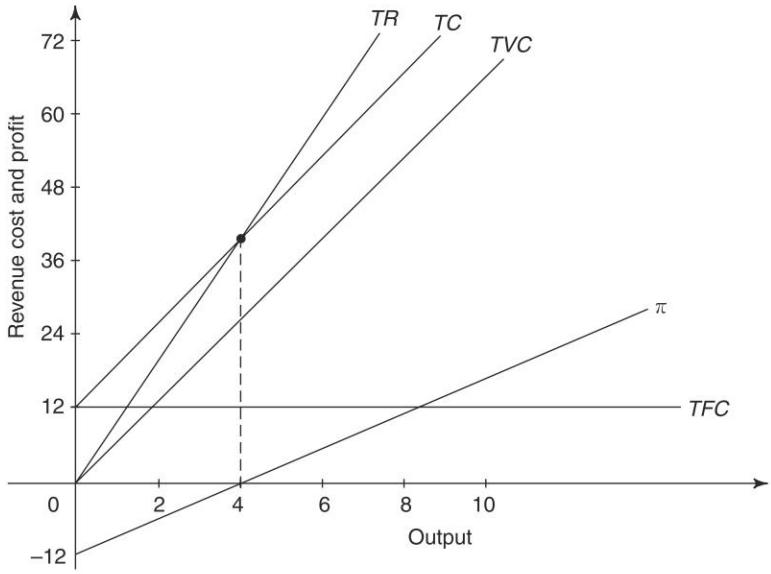



**Fig. 7.2** *Linear break-even graph*

## Profit and Break-Even Analysis—Non-Linear Case

The analysis will be non-linear, if either the total revenue function, total cost function, or both are non-linear. Let us consider a case, where both are non-linear:

$$TR = 24\,Q - Q^2 \tag{7.8}$$

$$TC = 20 + 8Q + Q^2 \tag{7.9}$$

The demand function corresponding to the above total revenue equation would be the following:

$$P = \frac{TR}{Q} = \frac{24Q - Q^2}{Q}$$

$$= 24 - Q$$

or, $$Q = 24 - P \tag{7.10}$$

The profit function would then be the following:

$$\pi = TR - TC = (24Q - Q^2) - (20 + 8Q + Q^2)$$

or, $$\pi = 16Q - 2Q^2 - 20 \tag{7.11}$$

The break-even output would then be given by the solution of the equation $\pi = 0$;

$$16Q - 2Q^2 - 20 = 0$$

or, $$Q^2 - 8Q + 10 = 0$$

i.e. $$Q = \frac{8 \pm \sqrt{64 - 40}}{2}$$

or, $Q = 6.45$ and $Q = 1.55$

Thus, for such functions, break-even output is not unique. Instead, there are two levels of output at each of which profit equals zero.

Given the *TR* and *TC* functions, one can easily compute the levels of *TR* and *TC* for various levels of output, and thereby a profit schedule. The results could then the plotted on a graph to obtain the break-even graph. For the non-linear functions of this sub-section, the so obtained graph would be as shown in Fig. 7.3.

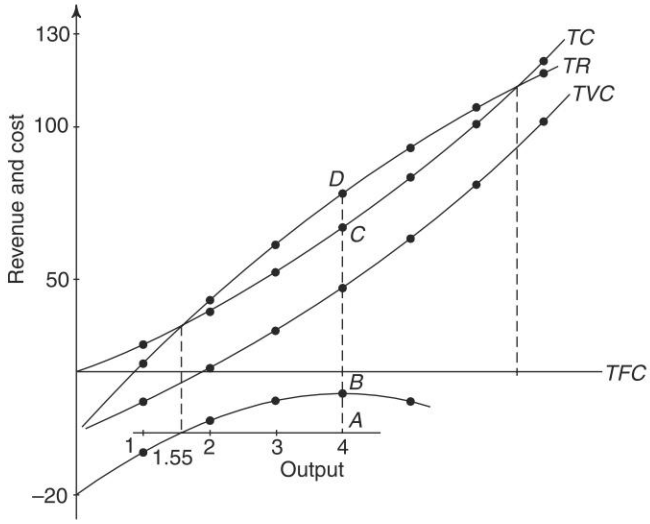



**Fig. 7.3** *Non-linear break-even graph*

In the non-linear analysis, the profit function is also non-linear and accordingly the total profit curve in the Fig. 7.3 is hill-shaped. Profit is maximum at $Q = 4$, which equals *AB,* as well as *CD (AB = CD).* The break-even outputs are $Q = 1.55$ and 6.45. The other useful calculations are given by the following.

$$P = TR/Q = 24 - Q$$

$$TFC = 20$$

$$TVC = 8Q + Q^2$$

$$AVC = TVC/Q = 8 + Q$$

Total contribution $= TR - TVC$

$$= (24Q - Q^2) - (8Q + Q^2)$$

$$= 16Q - 2Q^2$$

Average contribution $= P - AVC$

$$= (24 - Q) - (8 + Q)$$

$$= 16 - 2Q$$

## Applications of Profit and Break-Even Analysis

The profit and break-even analysis is useful for determining price-output combinations under the objectives of profit-maximization, sales-maximization subject to a target profit, and sales-maximization subject to break-even. As a corollary to this, it is useful in finding out the break-even output for a given price, break-even price for a given output, target-profit output for a given price and target-profit price for a given output. Both these applications are illustrated below.

**(a) Determination of Price-Output Combinations under Alternative Objectives** As seen above, when total revenue and total cost curves are linear, both *TR* and *TC*, and hence the total profit, are monotonically increasing functions of output. Under such a situation, the profit-maximizing or sales-maximizing output is thus indeterminate. Further, though break-even output is determinate, the objective of sales revenue maximization subject to no-profit-no-loss makes little sense.

In a situation of non-linear revenue and cost functions, neither the *TR* nor the profit function is a monotonically increasing function, and thus, the price-output determination under varying objectives is of great significance. In the above non-linear case example, the profit function (7.11) could be used to determine the **profit-maximizing price**. For profit to be maximum, the first derivative of equation (7.11) must be zero and its second derivative must be negative (and the price should not be less than the shut down price):

$$d\,\pi/dQ = 16 - 4Q = 0, \text{ or } Q = 4$$

$$d^2\,\pi/dQ^2 = -2 < 0$$

Thus, at $Q = 4$, profit is the maximum. Substitution of this value into the corresponding *TR* equation (7.8) would yield $TR = 80$ and into Eq. (7.11), $\pi = 12$. The $TR = 80$ and $Q = 4$ implies $P = 20$ (note that since $AVC = 8 + Q$, the minimum $AVC = 8$, which exceeds the price =20).

Similarly, the **break-even price-output combination** would be given by setting profit equal to zero in Eq. (7.11). This has been done above, where the two break-even output levels were found to be equal to 1.55 and 6.45, which yields *TR* equal to 34.8 and 113.2, and *P* equal to 22.45 and 17.55, respectively. If the objective were maximization of sales revenue subject to no-profit-no-loss, the equilibrium values would be $Q = 6.45$ and $P = 17.55$.

The equilibrium price-output combination under the objective of **sales revenue maximization subject to a given positive amount of profit** could be determined by setting the profit function (7.11) equal to the target profit amount. Thus, if the target profit were 10, we would have

$$16\,Q - 2Q^2 - 20 = 10$$

or, $$Q^2 - 8Q + 15 = 0$$

i.e. $$Q^2 - 5Q - 3\,Q + 15 = 0$$

or, $$Q\,(Q - 5) - 3\,(Q - 5) = 0$$

or, $$(Q - 5\,)\,(Q - 3\,) = 0$$

which gives, $$Q = 5 \text{ or } Q = 3$$

Corresponding to $Q = 5$, the $TR = 95$, and that corresponding to $Q = 3$ equals 63. Thus, the sales-maximizing target profit output = 5, and accordingly that price = (95/5) =19. If the profit target differs, the equilibrium values of price and output would change. Thus, the profit and break-even analysis is useful in determining price-output combinations under certain objectives.

**(b) Determination of Break-Even and Target-Profit Outputs and Prices** The break-even analysis serves yet other useful functions for decision-makers through its aid in the determination of the break-even output, break-even price, target-profit output and target profit price. The formulae are the following:

$$BEQ = \frac{TFC}{P - AVC} \tag{7.12}$$

$$T\pi Q = \frac{TFC + T\pi}{P - AVC} \tag{7.13}$$

where the new notations are

$BEQ$ = Break-even quantity

$T\pi Q$ = Target-profit quantity

$T\pi$ = Target profit

If the revenue and cost functions are linear, $P$ and $AVC$ would be constants; $TFC$ and $T\pi$ are constants by their definitions. Accordingly, definite values of break-even quantity and target-profit quantity could be determined by equations (7.12) and (7.13), respectively. Thus, if $TFC = 12$, $P = 10$, $AVC = 7$, and $T\pi = 15$,

$$BEQ = \frac{12}{10 - 7} = 4$$

and,

$$T\pi Q = \frac{12 + 15}{10 - 7} = 9$$

Incidentally note that $T\pi Q > BEQ$.

Equations (7.12) and (7.13) could also be used to yield break-even price and target-profit price for a given level of output. In that case, the given output must be substituted for $BEQ$ and $T\pi Q$, and then the equations could be solved for the corresponding values of $P$, called break-even price ($BEP$) and target-profit price ($T\pi P$), respectively. Thus, at $Q = 5$, other variables assuming the same values as before, we have

$$5 = \frac{12}{BEP - 7} \text{ or } BEP = 9.4$$

and

$$5 = \frac{12 + 15}{T\pi P - 7} \text{ or } T\pi P = 12.4$$

In the case of non-linear revenue and cost functions, the formulae (7.12) and (7.13) remain the same but they would also be non-linear. For the non-linear revenue and cost functions (7.8) and (7.9), we have

$$P = 24 - Q$$

$$AVC = 8 + Q$$

and

$$TFC = 20$$

The break-even output is given by (Refer Eq. 7.12)

$$Q = \frac{20}{(24 - Q) - (8 + Q)} = \frac{20}{16 - 2Q}$$

or, $16Q - 2Q^2 = 20$

Solution of which yields

$Q = 6.45$ and $Q = 1.55$

And if target-profit is assumed at 10, the target-profit output is given by (Refer Eq. 7.13)

$$Q = \frac{20 + 10}{(24 - Q) - (8 + Q)} = \frac{30}{16 - 2Q}$$

or, $16Q - 2Q^2 = 30$

Solution of which gives

$Q = 5$ and $Q = 3$

It is obvious that once a break-even quantity is available, the break-even sales revenue can easily be obtained by multiplying the break-even quantity by the given price. The same is true with regard to target-profit sales revenue.

The break-even analysis, also called the **profit-contribution analysis**, may also be used to reach a fairly close approximation to profit maximization by considering expected profits under a variety of alternatives. Such alternatives would include various prices, different plant sizes, and different levels of advertising budgets. Nevertheless, if the firm has full information on its revenue and cost functions, the profit maximization principle is preferred to the break-even analysis. But very often a firm possesses only limited information on its costs and revenues, and in such situations, the break-even analysis is highly significant. In particular, for the new projects, firms have rough estimates of the likely fixed costs, unit variable cost and on the expected price they could charge for their products. Given these, such firms could easily compute the break-even quantities they must ensure for themselves to be able to sell to avoid even initial losses. Alternatively, a firm might receive a confirmed order for a new product and it may be asked to fill a tender on its price. Under such a situation, the firm could estimate its total overheads and unit variable cost and resort to the break-even analysis to determine the break-even or even target-profit prices. Thus, there are numerous uses of the break-even analysis in decision-making.

## 7.5 PRICING UNDER DIFFERENT OBJECTIVES

Various theories of firms' behaviour have been reviewed in chapter 1, section 1.3. As seen there, while some of these theories are amenable to empirical determination and testing, others lack that quality. The present section explains the significance of firms' objectives in pricing with regard to those of such hypotheses which are available for quantification.

Needless to say, one needs the demand and cost functions for price determination. Given the values of all non-own-price determinants of demand and non-output determinants of cost, the demand and cost functions could be reduced to, through substitution of numerical values of the 'other' determinants in the corresponding function, simple equations of demand and cost curves, respectively. Suppose the so

obtained functions were the same as Eqs. (7.9) and (7.10). Note that in these functions, the intercepts contain the influence of 'other' determinants in the respective functions. Thus, the number 24 in Eq. (7.10) includes the effect of non-own-price determinants of demand on demand and the value 20 in Eq. (7.9) that of non-output determinants of cost on cost.

Corresponding to the cost Eq. (7.9) and the demand Eq. (7.10), the price-output combinations under the objectives of profit-maximization, sales, revenue-maximization subject to a positive profit level ($\pi = 10$) and sales-revenue maximization subject to break-even have been obtained in the previous section. The only other quantifiable objective is that of sales revenue-maximization subject to no constraint. The equilibrium price-output combination under this objective would require that the first derivative of the total revenue function (7.8) be zero and its second derivative be negative:

$$\frac{d(TR)}{dQ} = 24 - 2Q = 0$$

or, $$Q = 12$$

$$\frac{d^2(TR)}{dQ^2} = -2Q < 0$$

Thus, at $Q = 12$, $TR$ is maximum. If $Q = 12$, $P = 12$ (Refer Eq. 7.10), and $\pi = 116$ (vide Eq. 7.11).

The results on equilibrium prices and outputs under four alternative objectives are summarized in Table 7.2.

**Table 7.2** *Price-Output Determination under Alternative Objectives*

| | *Objective* | *Equilibrium Price* | *Equilibrium Output* | *Profit* |
|---|---|---|---|---|
| (i) | Maximum Profit | 20 | 4 | 12 |
| (ii) | Maximum sales revenue subject to profit = 10 | 19 | 5 | 10 |
| (iii) | Maximum sales revenue subject to break-even | 17.55 | 6.45 | 0 |
| (iv) | Maximum sales revenue | 12 | 12 | –116 |

It is obvious from Table 7.2 that the equilibrium price depends on the objective pursued by the firm. In general, among the four objectives considered here, the price is the highest under profit-maximization principle and the least under sales revenue maximization. The analysis of results under alternative objectives through geometry is presented in Fig. 7.4.

In Fig. 7.4, $TR$ and $TC$ give the total revenue and total cost curves, respectively, and the difference between the two the profit/loss. Thus, for outputs between $Q_1$ and $Q_5$, there is a profit, while there is a loss associated with every other output level. Profit (= $EF$) is the maximum at $Q = Q_3$. The break-even is at two different levels of output ($Q = Q_1$ and $Q = Q_5$), but the revenue-maximizing break-even output is unique at $Q = Q_5$. Similarly, the target positive profit output is not unique, for the condition is met at $Q = Q_2$ and $Q = Q_4$ ($AB = CD$) but that associated with maximum possible sales revenue is at $Q = Q_4$ only. The unconstrained maximum sales revenue output is single valued at $Q = Q_6$.

## 7.6 MARKET STRUCTURE

Michael Porter's model of competition (1980) identifies five forces of competition which an industry faces. These are

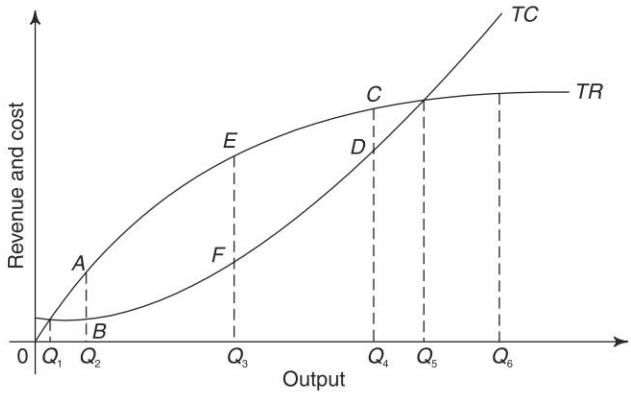



**Fig. 7.4** *Equilibrium under different objectives*

(a) Competition from inputs' suppliers wanting higher prices

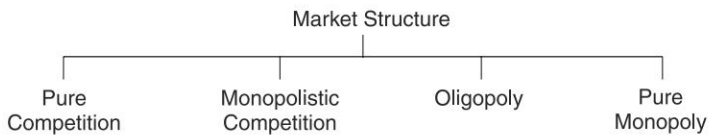
(b) Competition from customers asking for lower prices and better quality

(c) Competition from other industries (firms) supplying substitute products

(d) Competition from potential new entrants

(e) Competition from other firms operating in the industry

The last of these is the intra-industry competition, which is referred to as the market structure a firm faces in the industry in which it falls. Since market structure is an important determinant of the demand function which a firm faces, a clear understanding of it is absolutely essential for learning how a firm prices or ought to price its product.




Traditionally, under this criterion, the markets are classified into four categories:

The major differences among these are in terms of the number of firms, market share, type of the product, and entry/exit conditions:

**Table 7.3** *Basic Features of Various Markets*

| | *Measures* | *Market Structure* | | | |
|---|---|---|---|---|---|
| | | *Pure Competition* | *Monopolistic Competition* | *Oligopoly* | *Pure Monopoly* |
| (i) | Number of firms | Many | Many | Few | One |
| (ii) | Market share | Insignificant | Insignificant | Significant | 100 % |
| (iii) | Type of product | Homogeneous | Heterogeneous | Homogeneous or Heterogeneous | Unique |
| (iv) | Entry/exit conditions | Very easy | Easy | Significant barriers | Blocked |

Thus, purely or perfectly competitive and monopolistically competitive markets are alike in the sense that both have a large number of sellers*, each of whose share in the total supply of the good is insignificant, and the entry in both the kinds of markets is easy. The difference between these two markets is purely in terms of the nature of the product. While the products of all purely competitive firms are identical in all respects (quality, shape, packaging, etc.), those of monopolistically competitive firms are similar but not identical. In the latter case, the differences in products may be real or imaginary. It could be a difference in quality, shape, packaging, or just in brand name. There is perhaps no commodity in any economy whose market strictly satisfies the conditions of pure competition, but the closest to that would include the markets for agricultural goods (like corn, wheat, rice, cotton, barley, oilseeds, eggs, oranges, apples, potatoes, tomatoes, etc.) and financial instruments and services (like bank deposits and loans, shares/stocks of a well market firm-such as Tata Steel, ITC, Reliance Industries, Pfizer, Reliance Capital, Hindalco, Indian Oil Ltd, Bank of Baroda, Axis Bank, etc.). In contrast, there are a number of goods and services that are traded in a monopolistically competitive market. These include all the retail trade items, such as garments, shoes, grocery stores, restaurants, etc. and many services, such as those rendered by barbers, advocates, doctors, chartered accountants, washer men, etc.

The oligopoly markets are characterised by a few sellers, where each one or at least one of them commands a significant portion of the total market supply of the product in question. In fact, this market could have a large number of sellers but in that case there must be some 'big' sellers in the market, which would distinguish it from the purely competitive market, which has all small sellers. The number 'a few' is conceptual and it cannot be quantified exactly. Under this market, the products of various firms in an industry could be standard (homogeneous) as in the case of a purely competitive market or could be differentiated (heterogeneous) as in the case of monopolistically competitive market. Accordingly, this market is further classified into homogeneous oligopoly and heterogeneous oligopoly. Here the entry by new firms is possible but difficult. The examples of this market are plenty. For example, homogeneous oligopoly prevails in the markets for steel, cement and aluminum; and heterogeneous oligopoly exists in almost all capital intensive, expensive and durable goods' markets, like those for cars, two wheelers, heavy farm equipments, kitchen appliances, televisions and TV channel providers, musical instruments, personal computers, internet providers, mobile phones and their service providers, pharmaceuticals, etc.

The purely monopolist market has just one seller, with 100 percent market share, unique product, and total absence of either entry or exit possibility. The examples include the markets for electricity transmission and distribution (not including electricity generation), cooking gas, landline telephone, railway track services, airport services, postal services, etc.

What is the distinguishing feature of each of these kinds of markets? For monopoly, **single firm** happens to be the distinguishing characteristic. In oligopoly, the number of firms is a "few", and hence the firms in a given industry are **interdependent** in terms of their price-output decisions, and this distinguishes it from the other forms of market structure. The products of no two firms in any monopolistically competitive industry are identical, i.e., their **products are differentiated**, and this makes the said feature unique. So far as the perfectly competitive industry is concerned, all of its four features are found in one or the other of the other three types of market structure. Nevertheless, the

*Most goods have a large number of buyers and accordingly all the four types of markets of Table 7.2 are characterised by large number of buyers.

simultaneous occurrence of all these lead to a unique feature, viz. every firm in such an industry is a **price-taker**! The said implication will be pursued further in the next chapter.

## Determinants of Market Structure

It is true that various goods and services have different market structure in all countries. The question is, what determines the market structure, or why different products have different market structures? There are four reasons for this: product characteristics, nature of the production function, number of buyers and government policy.

The degree of competition in the market for a product depends not only on the number of its sellers but also on the closeness of the substitutes available for the product, physical characteristics of the product and the extent of the perishablity of the product, among other factors. The closer the substitute available for a product, the greater is the competition in the market. Thus, for example, the market for text books are more competitive than for specialised or research based books. Similarly, the bulkier is a product, the lesser is the competition, for its geographical movement would be hindered the most by its high ratio of distribution costs to total cost. Thus, the coal market is less competitive than the watch market. The last aspect of the product's nature, which impinges on the market structure, is the degree of perishability of the product. The more perishable the product is, the less is the competition, and vice versa. By this factor, the market for fruits is less competitive than that for furniture.

The nature of the production function is an important determinant of the market structure. Industries whose production function exhibits increasing returns to scale over a large output level in relation to total market demand are characterised by fewer producers and thus by less competition, than are industries where constant or decreasing returns to scale enters at an output level that is small relative to total product demand. The number of buyers, their education and mobility also affect the extent of the competition in the market. The larger the number of buyers, the greater is the competition in the market. When there is only one buyer, the market is called the *monopsony*, in case of a few buyers, it is called the *oligopsony* and when there are many buyers, it is called the purely competitive market from the side of the buyers. Both the education and mobility of consumers positively affect the competition in the market. Government economic policies, such as licensing, competition policy, reservations of products for public vs. private sector, small vs. medium vs. large sector, nationalization, disinvestment, etc. have obvious bearings on the market structure prevailing in a given industry. For example, Indian airlines enjoyed the monopoly in air travel services until the government of India decided to permit the potential firms to enter the sector, and thereby turning the industry into an oligopoly market. Note that though airports, electricity transmission and distribution, etc. have been privatized, they still continue to enjoying monopolies.

Of the four distinguishing features of market structure, the number of firms and the type of products are easy to understand. However, that is not true for the other two features, viz. market share (also called the degree of concentration in the industry) and exit/entry conditions. Accordingly, these are discussed in what follows.

## Measures of Industry Concentration

Industry concentration is usually measured by either or both of the two methods, viz. Four firm concentration ($C_4$) method and Herfindahl-Hirschman index (HHI) method. The latter is so named in the honour of its founders. These measures are quantified through the following formula:

$$C_4 = S_1 + S_2 + S_3 + S_4$$

or, $$C_4 = \sum S_i \text{ , where } i = 1, 2, 3,4 \qquad (7.\ 14\ )$$

$$HHI = 10{,}000\ [S_1^2 + S_2^2 + S_3^2 + S_4^2 + \ \ldots\ldots\ldots + S_n^2]$$

or, $$HHI = 10{,}000\ [\ \sum S_j^2\ ]\ \text{, where } j = 1, 2, 3, 4, \ldots..., n \qquad (7.\ 15)$$

In the above formulas, notations $S_i$ and $S_j$ stand for the shares of firm $i$ and $j$ in the respective industry, respectively. One may wonder why the *HHI* measure has the number 10,000 as a multiplier. This is because it happens to be the square of number 100 (i.e., $10{,}000 = 100^2$), with the latter standing for the full market share in percent. Also, note that in $C_4$ measure, the shares of the **largest four firms** are used. As these formulas would indicate, these measures have lower and upper limits, as follows:

$$0 \le C_4 \le 1 \quad \text{and} \quad 0 \le HHI \le 10{,}000$$

The closer they are to zero value, the less is the concentration and so more is the competition in the corresponding industry. Accordingly, an industry characterized as perfectly competitive would have close to zero values for each of these two measures. In contrast, a value closer to unity for $C_4$ and closer to 10,000 for *HHI*, would suggest maximum concentration and so the least competition (or monopoly) in the industry under analysis. For example, Indian Railways enjoys the monopoly in rail transportation, meaning there is just one form with 100 market share. Thus, when calculated, $C_4$ would be unity and *HHI* would equal 10,000.

$$C_4 = 1 + 0 + 0 + 0 = 1 = 100\ \%$$

$$HHI = 10{,}000\ [1^2 + 0^2 + 0^2 + \ldots\ldots\ldots.. + 0^2] = 10{,}000$$

To illustrate their computations and uses, let an industry have 10 firms with their market shares (arranged in descending order) equal to 0.30, 0.20, 0.20, 0.10, 0.08, 0.05, 0.03, 0.02 0.01,and 0.01, respectively. The two measures are then given by

$$C_4 = 0.30 + 0.20 + 0.0.20 + 0.10 = 0.80 = 80\ \%$$

and

$$HHI = 10{,}000[0.30^2 + 0.20^2 + 0.20^2 + 0.10^2 + 0.08^2 + 0.05^2 + 0.03^2 + 0.02^2 + 0.01^2 + 0.01^2]$$

$$= 10{,}000\ [0.09 + 0.04 + 0.04 + 0.01 + 0.0064 + 0.0025 + 0.0009 + 0.0004 + 0.0001 + 0.0001]$$

$$= 10{,}000\ [0.1904] = 1{,}904$$

According to the number of firms' criterion, a 10 firm industry may not qualify for oligopoly (as 10 firms may not mean "a few" firms requirement). However, when measured for the above hypothetical example, a $C_4$ value of 80 per cent or *HHI* value of 1,904 could surely lead one to conclude that the industry has oligopoly. Thus, the two measures help one to quantify and better conclude on the matter than otherwise. Accordingly, they are often used by the policy makers for considering the approvals for new firms entry and/or mergers of two or more units in the same industry.

The question now is how the two measures differ and if the difference is significant, which is a better measure of industry concentration? There are two differences in them: (a) While $C_4$ considers just the shares of the top four firms only, *HHI* incorporates those of all firms in the industry. Accordingly, the latter is a more comprehensive measure than the former, and (b) The HHI measure squares the

shares ($S^2$) of each firm, while $C_4$ takes their face values ($S$). Thus, under HHI, firms with larger shares get greater weight than those in $C_4$, and quite the opposite holds for firms with smaller weights. To understand this, consider an industry having just three firms with shares of 0.6, 0.2 and 0.2. The measure $C_4$ would attach 0.6, 0.2 and 0.2 weights to the three firms. In contrast, *HHI* would attach weight = $0.6^2 = 0.36$ to the first firm, $0.2^2 = 0.04$ to second firm, and $0.2^2 = 0.04$ to the third firm. A comparison of these numbers would indicate that the difference in weights of first and second firms in measure $C_4$ is just three times (= 0.6/0.2) that in measure *HHI* is five times (= 0.2/0.04). Between the two measures, *HHI* is considered better than $C_4$. The reasons are found in their differences. Nevertheless, $C_4$ is useful, for it requires lesser data than *HHI* and is thus quicker to calculate. Further, while big companies publish their data regularly in easily accessible documents, the same is not generally true for small companies.

What purpose these measures serve? The **HHI value = 1,800** is often used as the **benchmark** for this purpose. By this, if an industry has its *HHI* value more than 1,800, the exits and mergers are denied the permission on this criterion, while the others with less than this number are permitted to go ahead. Needless to say, several other criteria are used by the authorities for the purpose. The latter would include the significances in terms of employment, insolvency, dominance by some, etc. Thus, consider an industry having, say, *HHI* = 1,500, where two firms are seeking mergers. If not allowing the merger would mean the weaker of the two firms would go insolvent and its workers would get fired, the government might grant the approval for the merger. In contrast, consider an example of another industry with *HHI* = 2,500 to which there is a request from a potential firm for entry. On scrutiny, if the government finds that the new entrant would pouch workers from the incumbent firms and thereby create unhealthy competition, it could well refuse the permission. A *HHI* value = 1, 000 is sometime used as a cut off rate below which no industry is allowed to move through mergers/acquisitions/take-overs.

As no measure is perfect, the above two measures are no exception to the rule. Their limitations come from three sources, viz.

(a) *Global markets*: The measures exclude imports of the good that the industry produces. The limitation is getting more and more binding with the increasing globalization.

(b) *Local vs. Regional vs. National Market*: While applying a measure, should the researcher consider the country wide market, regional or simply the local market? For an appropriate calculation, the local market alone must be taken. For example, in case of public utilities like, electricity, cooking gas and telephone, while the corresponding firms really enjoy monopolies, there are many of them at the national level and several even at regional level. Thus, while Torrent Power Company has monopoly in electricity distribution in Ahmedabad, there are regional companies like Paschim Gujarat Vij Company Ltd. (PGVCL), North Gujarat VIJ Company Ltd. (NGVCL), etc. which supply power to the other (their demarked) parts of Gujarat.

(c) *Industry Classification/definition*: While the international standard/definition is used in classifying various products into different industries, they are not panacea. Such anomalies would obviously lead to wrong signals from the two measures under discussion.

Needless to say, in spite of the above limitations, they are used for measuring the level of competition in all industries and thereby while taking decisions pertaining to entry/exit and mergers/acquisitions in those sectors. Before closing this part, it may not be out of place to briefly mention the **measures of firm's conduct** and even **performance**, as they are related in the **Bains' theory of**

**structure-conduct–performance** (1956). While industry structure is practically given to a firm, its decision variables include price (or output), advertisement budget and media option, research and development (R&D), merger activity, etc. Accordingly, the latter list of variables constitutes the firm's conduct variables. Firms' performance is prominently measured by its profits but lately its contribution to the society (social responsibility) in terms of employment generation, foreign exchange, poverty alleviation, and such other social works (like education, health, community services, etc.) is considered. While Bain argued in terms of the one way relationship between structure-conduct and performance (i.e. structure affects conduct, which affects performance, but not the other way round), currently the three are deemed to be mutually interrelated (i.e. each affects the other two, implying the feedback system).

## Entry/Exit Barriers

While some industries pose no or little barriers to firms' entry or exit, many industries do have such constraints throughout the world. To appreciate this, let us look at those barriers. Those pertaining to **entry barriers** could be classified in to three categories, viz.

(a) Legal barriers
(b) Economic barriers, and
(c) Deliberate obstacles

The legal barriers consist of the patents, copy rights, exclusive access to natural resources, licensing requirements, and trademarks that have been awarded to certain firms or instituted in certain industries by the governments. For example, a patented pharmaceutical gives the patent holder exclusive right for a certain maximum period to manufacture and sell that pharmaceutical product within a specified market. Similarly, copyright on a certain intellectual material (e.g. a book) prohibits everyone else to print or copy that material, particularly in large numbers and/or for profiting. These are legal for they are imposed by authorities and are eligible for protection through courts. Needless to say, patents and copy rights are awarded to encourage R&D and scholarly activities, lest they become loss-taking and thereby unattractive. Licensing is practiced to direct firms where the country needs, and exclusive rights to natural resources are given so as not to duplicate efforts and thereby to avoid waste. Needless to say, the pity is that sometimes they are given on political/personal considerations.

Economic barriers include prohibitive capital expenditure requirements, entrepreneurial skills needs, technology, controls on inputs (plants and/or raw materials, intermediate goods, skilled personnel); and the presence of natural monopoly, economies of scale/scope/transaction cost, distribution network, product brand or customers' loyalty, etc. Though these are basically self explanatory, a couple of examples are in order. To put up an automobile firm, one needs huge funds, entrepreneurial caliber, a network of suppliers of spare parts and intermediary goods, etc., a large scale production and potential customers, etc. to be able to establish and profit from such a venture. In an industry where monopoly is by the nature of the product (and not by legal action or otherwise), economic prudence would always limit the number of firms by just one. Railroads/airports come under such a category. In these fixed costs are not only prohibitive, but having more than one incurring them separately would lead to duplication of facilities, thereby increasing the overheads, rendering the project unviable. This is pursued further later when pricing under monopoly is discussed (Chapter 8).

The deliberate barriers to entry are created consciously by the incumbent firms. These are practiced

through setting unattractive (non/un-profitable) prices, known as limit prices and predatory prices, political or social intimidation, lobbing with government officers and so on. In addition, expectations about incumbents' reactions, incumbent's advantages (e.g. learning curve effect) and exit costs (see below) serve as entry barriers. Setting of the lower prices by the incumbents' firms is often used for the purpose, for their erstwhile (current) prices may well be above the free-market level due to the barriers, and these could easily be reversed as soon as the entry threat disappears.

**Barriers to exit** make it harder for a firm to leave a particular industry than it would otherwise be. These include sunk cost (Refer Chapter 6, Section 6.1), cost of laying-off the workers, and backing out from the contractual obligations to suppliers, customers, government, etc.

This concludes the back ground material for the discussion of the price-output decisions by firms. To give some brief on this topic, recall that the industry structure exerts influence on the demand and cost functions that a firm faces, and even on the firms' objective(s), it plays a significant role in price-output determination. Further, as just analysed, different product groups fall into different market structures. In view of these considerations, pricing is usually discussed separately for each groups of products falling into a given market structure. Following this tradition, in what follows, we shall first discuss separately, price-output determination under perfect competition (i.e., for agricultural and popular financial products), under monopoly (i.e., for utilities and products sold by single producers), under monopolistic competition (i.e., for inexpensive, low capital and skill intensive manufacturing products and general services), and under oligopoly (i.e., for expensive and high capital and skill intensive manufacturing products and specialised services), in that order. Under these categories, the discussion will proceed on some usual assumptions, which will be relaxed subsequently in the last chapter on pricing.

## REFERENCES

1. Bain, Joe Staten (1956): *Barriers to New Competition*, Harvard University Press.
2. Baye, Michael (2009): Managerial Economics, McGraw-Hill.
3. Baumol, W.J. (1982): *Economic Theory and Operations Analysis*, 4th edition, New Delhi, Prentice-Hall.
4. Dean, J. (1976): *Managerial Economics*, New Delhi, Prentice-Hall.
5. Gupta, G.S. (1977): "Pricing and Price Controls: Impact on Profitability and Growth of Industry," *Economics Times*, XVI (Jan. 21, 22 and 24).
6. Gupta, G.S. (1981): "Agricultural Prices Policy and Farm Incomes," *Economic and Political Weekly*, XV (September 27).
7. Gupta, G.S. (1981): "Price Policy and the Small Farmer," *Commerce*, 142 (March 14).
8. Gupta, G.S. and Ruhani Ali (2000): Corporate Takeovers in Malaysia; Discriminant Analysis for Bidder and Target Firms, *Asian Academy of Management Journal*, V, 1 (January), 1-14.
9. Henderson, J.M. and R.E. Quant (1980): *Microeconomic Theory*: *A Mathematical Approach*, 3rd edition, New York, McGraw-Hill.
10. Marshall, A. (1920): *Principle of Economics*, 8th edition, London, Macmillan.
11. Porter, Michael (1980): *Competitive Strategy*, The Free Press.
12. Stigler, G.J. (1966): *The Theory of Price*, 3rd edition, New York, Macmillan.

## CASELETS

1. Draw the appropriate demand and supply curves to depict the following observed (or equilibrium) situations over time. Is demand or supply the dominant factor in each of situations (a) and (b)? Why?

(a) Product *X*

| Period | Price | Quantity |
|---|---|---|
| 1 | 100 | 2,500 |
| 2 | 140 | 2,825 |

(b) Product *Y*

| Period | Price | Quantity |
|---|---|---|
| 1 | 120 | 2,250 |
| 2 | 160 | 1,750 |

2. Pintoo Garments (PG) has an investment in the amount of Rs. 5 lakhs in its garments' trading business. Its imputed cost of this capital is 15 percent. Besides this fixed cost, its average variable cost (including the opportunity cost of his own time) per garment is Rs. 250. Given this, answer the following questions:

   (a) If PG prices its garment at 10% above its average variable cost, how many pieces of garments it must sell in a year to break-even? Suppose, its actual sales in 2009 stood at 15,000 garments, how much profit did the company make in that year? (*Hint*: Total fixed cost = Rs. 75,000)

   (b) Suppose the marketing person at PG estimated the demand facing the company as

   $Q = 14{,}000 - 40\,P$

   Then attempt the following questions:

   (i) What is the break-even output? Compare your answer with that in part (a) above and comment.

   (ii) Determine the Company's profit maximizing output, and its corresponding profit.

   (iii) If the company chooses to price its product so as to maximize its sales (in rupees), what price it must set? Determine the company's output and profit under this objective.

   (iv) If the company chooses to set the price such that its total revenue is maximum subject to the constraint that profit = 12,500, determine its equilibrium output and price.

   (v) Compare your answers in all the four parts above and comment.

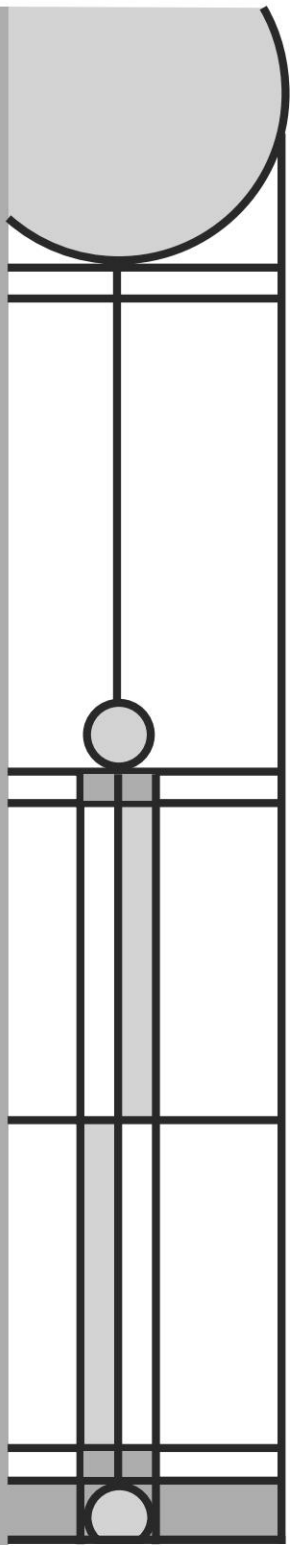
3. The automobiles industry is composed of nine firms of varying sizes. Firm 1 has 35 percent of the market, Firm 2 has 25 percent, Firm 3 has 10, and the remaining firms have 5 percent each. Based on this data, attempt the following questions:

   (a) Compute the four-firm concentration ratio and Herfindahl-Hirschman Index (HHI) for the automobiles industry. What do these numbers indicate about the market structure in the said industry?

(b) Suppose Firms 7, 8 and 9 contemplate a merger and request for the government approval. Given the benchmark value of 1,800 for the HHI, should the merger be approved? Why or why not?

(c) Suppose Firm 1 attempts to take-over all the six firms with market share of 5 percent each, how should the authorities in the country's competition policy react? Justify your answer.

The McGraw-Hill Companies



CHAPTER

8

# PRICE-OUTPUT DETERMINATION–I

## (Pricing under perfect competition, monopoly and monopolistic competition)

Recall from the previous chapter that the market price of a product is governed by the demand for and supply of the good in question. Also, recall that these demand and supply are influenced by a number of general factors as explained in Chapters 2–4 and 5–6, respectively; and by product specific factors like the practice of the law of one price *vis-a-vis* pricing strategies, the access to the symmetric or asymmetric information set, externalities, if any, the degree of risk and uncertainty, and the governments interventions, if applicable, as briefly presented in the previous chapter. In this and the next two chapters, product pricing is analysed through the standard approach. Accordingly, in these chapters, we first present pricing under different market structures, assuming

(a) All firms aim at maximum profits
(b) Law of one price holds
(c) Buyers and sellers have full information about the market and thus the situation of asymmetric information does not apply
(d) There are no externalities and no risk/uncertainty in the market

The inevitability of the profit maximization principle has been explained in Chapter 1 and elsewhere, and the effects of its non-perusal have been analysed in the previous chapter. The last chapter on pricing (Chapter 10) will relax the last three assumptions, one by one. We start with the extreme case of perfect competition, and then go on first to the other extreme case of (pure) monopoly, and then monopolistic competition, and finally to oligopoly market. This traditional order is adhered to here, for though the extreme cases are only rare in today's world, they do provide the necessary base and standards for comparison. The present chapter will delineate pricing under the first three kinds of market structures and the next chapter will deal with the oligopoly market.

### 8.1 PRICING UNDER PERFECT COMPETITION

A perfectly competitive market is characterised by a large number of buyers and sellers, no market concentration, a homogeneous/standard product, and easy entry and exit. Besides, in this kind of a

market, each seller and buyer has insignificant market share, none of them has any preference for buying/selling from a particular seller/buyer, and each one of them possesses full information about the market. In view of these factors (particularly of the insignificant market share of each and every buyer as well as seller, and the standard product), no buyer or seller has any influence on the price and consequently each one of them is a *price taker,* which also means *absence of market power*. Thus, a farmer who produces a few tonnes of wheat and a household who consumes a few quintals of wheat has no say in the price of wheat, and they could sell and buy any quantity of this product at the ruling market price. Who then determines the price of wheat? No individual buyer or seller but all of them taken together, called the market or the *invisible hands*, determine the price. How this is achieved is presented below.

There are many firms, selling an identical good. They together constitute the industry for that good. By summing the supply schedules/curves of individual firms, one can get the aggregate or industry supply schedule/curve. Similarly, there are many buyers, looking for an identical good. They together form the aggregate market (demand) for the good. Summation of individual demand schedules/curves would give the aggregate demand schedule/curve for the industry's product. The intersection of the industry supply and industry demand curves determines the equilibrium price of the industry's product and the level of industry output. Figure 8.1 has a graph for industry which illustrates price determination in a perfectly competitive industry. The price is determined at *OP* in the industry by the intersection of industry demand (ID) and industry supply (IS) curves. Thus, price under perfect competition is determined by the market, the so called, invisible hands. While all firms together have role (through industry supply curve) in pricing their product, each one of them is too small to have any impact on industry supply on the price of its product. The said price is given to every firm and every buyer, and accordingly all of them are price takers.

What role does an individual firm play in its price-output decisions? As just noted, individual firm has no role in pricing its product under perfect competition. What about output decisions? Yes, every firm decides its own output level, given the price of its product. How is this achieved? Simply through the firm's optimization procedure. Recall from the previous chapter that there are three conditions for profit maximization, viz. $MR = MC$, $MC$ cutting $MR$ from below, and price no less than the shut down price (which equals minimum $AVC$ in the short run and minimum $AC$ in the long run). Note that at the industry's determined price, individual firms may sell any quantity and individual buyers could purchase any quantity of the product in the market. Accordingly, each firm faces a horizontal (infinitely elastic) demand curve at the price determined in the industry. If so, each firm's total revenue ($TR$, which equals $PQ$) increases linearly as its output expands, resulting into $P = AR = MR$ (Refer Chapter 2, Fig. 2.8). Accordingly, the horizontal demand curve is also the firms' $AR$ as well as $MR$ curve. Since every firm faces the same price, all of them would have identical $AR$ and $MR$ curves. Thus, $AR_i = MR_i$ for $i = 1, 2, ..., n,$ where $i$ stands for the firm. Moving to the cost curves next, since all firms do not have homogeneous factors of production, their cost functions would be different. Many of them would have relatively low costs, many somewhat medium costs and still many more relatively high costs. For convenience, various firms in the industry, in the short-run, could be classified into four categories:

(a) Efficient (low-costs) and profit-making firms (call this as Firm I)
(b) Efficient but breaking-even firms (Firm II)
(c) Inefficient but operating firms (Firm III)
(d) Inefficient (high costs) and closed-down firms (Firm IV)

Firm I is the most efficient one, as its costs are the lowest, Firm II comes next in terms of efficiency, Firm III still next and Firm IV is the least efficient of the four firms. The various cost curves take their normal shapes. Fig. 8.1 includes all the useful graphs of each of the four firms in different parts. Applying the above mentioned optimizing conditions to the firms in Fig. 8.1, it is seen that the equilibrium outputs of firms I, II, III, and IV stand at $0q_1$, $0q_2$, $0q_3$ and 0 (zero), respectively. This is how firms under perfectly competitive markets decide their output levels. In the example, while firm I make profit, equivalent to the area of the rectangular *PABC*, firm II just breaks even, firm III and firm IV run into losses, equivalent to the area of rectangular *PDEF* and total fixed cost, respectively. Thus, under pure competition in the **short-run**, there could be firms making (economic) profit (called abnormal profit by economists, for normal profit is included in costs), firms just breaking-even, firms operating but running into losses and firms just kept shut-offs.

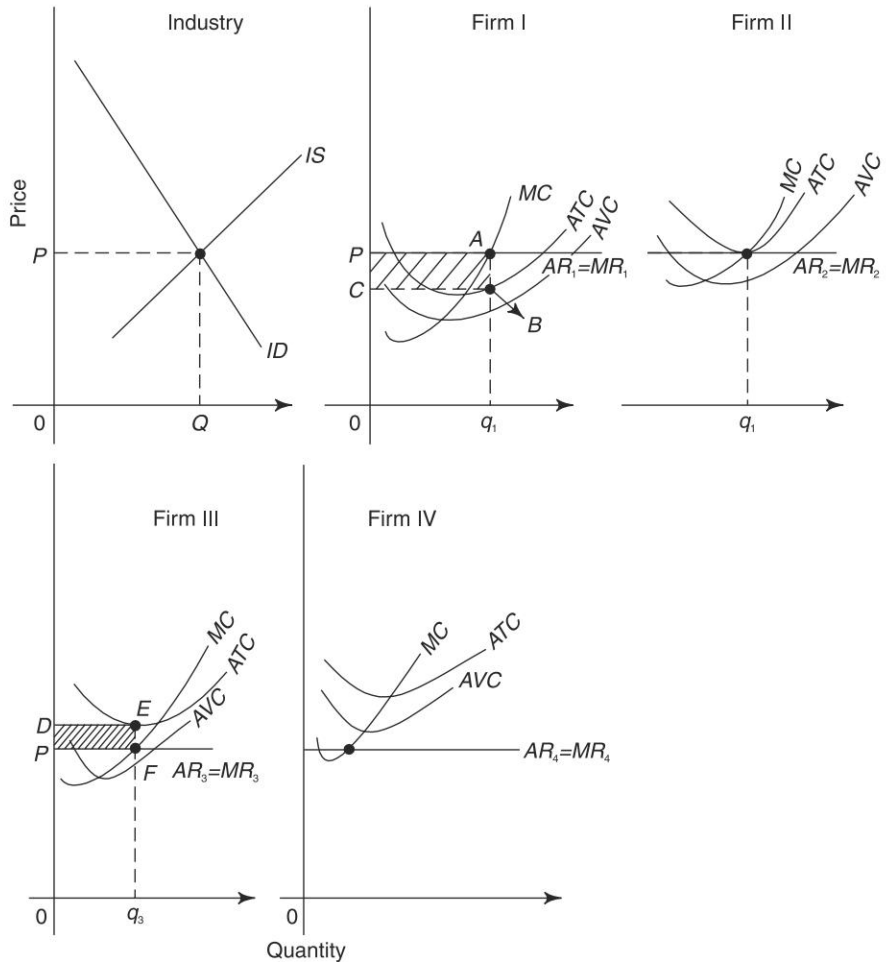



**Fig. 8.1** *Price determination and firms' equilibrium under pure competition: short-run*

Remember that every firm charges the same price (and thus faces identical *AR* and *MR* curves) and differences among them are in terms of factors of production or cost functions only. However, the

optimum size of the firm in terms of output levels varies across firms. Thus, the firm size in terms of output is determined solely by its cost conditions.

What would happen in the **long-run**? Under perfect competition, new firms are free to enter the industry, existing firms are free to leave the industry, and there is perfect information about the market. In view of these features, in long-run, the market would realize that the secret of firm I in profit-making lies in its relatively efficient factors of production. Firms III and IV would either leave the industry or try to attract the efficient factors of firm I to themselves. If firm I allows its efficient factors to leave it in favour of, say, firm III, firm I's cost would rise and firm III's cost would go down, and ultimately both these firms might reach to a break-even situation like that of firms II. Alternatively, if firm I attempt to retain its efficient factors through increases in their remuneration, its cost curves will still shift upward and thereby wipe-off the profit. Thus, in the long-run, no firm would either make profit or loss, all would just break-even like firm II. The short-run profits would disappear and the remuneration of efficient resources would rise. In other words, what was profit in the short-run would become excess remuneration to the factors of production, called the *Quasi Rent*, in the long-run. In fact, some cost differences among firms could be due to inconsistent imputation of implicit costs (particularly normal profits) and if so the excess profits of firms in short run could be due to under valuation of the cost of the owners' time.

It needs to be emphasized that in the long-run, each firm will have, $AR = MR = AC = MC$. Further, although all firms will sell at a uniform price, their equilibrium outputs need not be the same. Each firm would operate at its corresponding minimum average cost, but this output would generally be different for different firms. Only in an extreme case of all firms having homogeneous factors of production, and hence homogeneous cost curves, would the equilibrium outputs of various firms be the same.

The price-output determination and equilibrium of the firm under the conditions of pure competition may now be explained through a numerical example. Suppose the demand and supply conditions of a good are represented by the following equations:

Industry Demand : $Q = 150 - 2P$ (8.1)

Industry Supply : $Q = -25 + 5P$ (8.2)

The equilibrium price would then be the one at which industry demand equals industry supply:

$$150 - 2P = -25 + 5P$$

or, $$P = 25$$

Industry output at $P = 25$ is obtained by substituting this price into either the demand or supply function:

$$Q = 150 - 2(25)$$
$$= 100$$

Thus, equilibrium $P = 25$ and equilibrium industry $Q = 100$.

To determine firm's equilibrium output, we need the firm's cost function. Suppose, the total cost ($TC$) function of a particular firm (firm I) is represented by ($q_1$= firm I's output)

$$TC = 10 + 2q_1 + 0.5\, q_1^2 \quad (8.3)$$

Then, profit-maximizing firm I's supply function would be given by

$$MC = MR = P$$

i.e., $$2 + q_1 = P = 25$$

or, $q_1 = 23$

Firm's total revenue ($TR$), $TC$ and profit ($\pi$) are given by

$$TR = PQ = 25 \times 23 = 575$$
$$TC = 10 + 2 \times 23 + 0.5 \times (23)^2 = 320.5$$
$$\pi = 575 - 320.5 = 254.5$$

Since profits are positive, the profit maximizing condition, that the price is not less than the shut down price, is met. Since $MC$ increases while $MR$ decreases as output expands, the sufficient condition (that the rate of change in $MR$ is less than that of $MC$) is also met. Further, since the firm makes profit, it gives incentives for other firms to enter the business (industry). The supply curve of the firm is given by

$$MC = P$$

or, $2 + q_1 = P$

or, $q_1 = P - 2$, if $P \geq P_S$

$= 0$, if $P < P_S$

To determine the firm's shut down price ($P_S$), we need to check its minimum $AVC$ and minimum $ATC$. Considering $AVC$ first, we have

$$TVC = 2q_1 + 0.5\, q_1^2$$

Thus, $AVC = 2 + 0.5\, q_1$

The $AVC$ is a linear function. Therefore, $AVC$ has no technical minimum value. However, it would equal 2 if output is zero. Thus, the short-run supply of firm I ($q_1^s$) would be given by

$$q_1^s = \frac{1}{8}(P - 2), \text{ if } P \geq 2$$
$$= 0, \text{ if } P < 2 \qquad (8.4)$$

To determine firm I's long run supply function, we need to find out minimium $ATC$. The $ATC$ is given by

$$ATC = \frac{10}{q_1} + 2 + 0.5\, q_1$$

For minimum, the first derivative is set equal to zero. Thus,

$$-\frac{10}{(q_1)2} + 0.5 = 0 \qquad \text{or,} \qquad q_1 = 4.5$$

At $q_1 = 4.5$, $ATC = 10/4.5 + 2 + 0.5\,(4.5) = 6.45$

The second derivative of $ATC$ would be positive and thus the sufficient condition for $AC$ to be minimum at $ATC = 6.45$ is also met.

$$\frac{d(ATC)}{dq_1^2} = \frac{-25}{q_1^2} + 4 = 0$$
$$\frac{d^2(ATC)}{dq_1^2} = \frac{-50}{q_1^3} > 0$$

Accordingly, the firm's long run supply function is given by

$$q_{1s} = P - 2, \text{ if } P \geq 6.45$$
$$= 0, \text{ if } P < 6.45 \qquad (8.5)$$

This means, firm I would supply positive output in the long-run if and only if the price of its product is no less than 6.45. Since $P = 25$, it would supply 23 units of its output, as mentioned above. Thus, the numerical example is of a firm of the type of firm I in Fig. 8.1. Similarly, the equilibrium outputs of other firms could be determined, given their cost functions.

Before we proceed further, we must emphasize four basic factors under perfect competition, viz.

(a) The only **strategic decisions** for a firm are with respect to its **cost competitiveness and output**. The only way available to firms to make profit is through minimizing cost or maximizing efficiency. Accordingly, the firm must concentrate on cost saving avenues to excel in the market.

(b) A perfectly competitive firm has no choice but to maximize its profit, for doing anything else would mean making less than normal profit (meaning loss) in the long run. In consequence, owners will liquidate their investments and firms would go out of business.

(c) Only under perfect competition, profit maximization leads to $P = MC$, which ensures socially desirable output. This is so as price denotes the marginal utility that the society gets by consuming the last unit and $MC$ the marginal cost that the society bears on producing the last unit. Accordingly, under perfect competition, economic surplus (see below under the sub-section "consequences of government interventions" for its meaning) created by the product is the maximum.

(d) The concept of competition is a misnomer here. There is no overt competition, for each firm is basically independent of all other firms and accordingly it can ignore other firms' behaviour without peril.

## Changes in Industry's Product's Demand and Supply, and Product Price

The consequences of changes in industry product's demand and/or supply (henceforth called just as demand and supply) on the product price and equilibrium outputs of individual firms could now be analysed. Recall from Chapters 2–4 that demand for a consumers' good undergoes a change as and when any one or more of the demand determinants (like consumers' income, consumers' taste, prices of related goods, consumers' expectations about their incomes and prices, number and distribution of consumers) changes. Similarly, as elaborated in Chapter 6, supply of goods change as and when any one or more of the supply determinants (like inputs' prices, technology, prices of goods related in supply, sellers' price expectations, number of sellers, taxes and subsidies, weather) change. How such changes affect product price and industry output are discussed here. The following four combinations of such changes are worth pursuing here.

**(a) Change in Demand, Supply Constant** If demand ($D$) increases, supply ($S$) remaining constant, equilibrium price goes up and so does quantity bought and sold in the market. Quite the opposite holds in the event of a decrease in demand, supply remaining the same. These situations are explained in Fig. 8.2.

In Fig. 8.2(a), increase in demand from $D_1$ to $D_2$ is shown to lead to an increase in price from $P_1$ to $P_2$ and an increase in output from $Q_1$ to $Q_2$. Figure 8.2(b), shows exactly the reverse case.

**(b) Change in Supply, Demand Constant** Increase in supply, *ceteris paribus*, leads to a decrease in price and increase in industry output and vice versa. This is illustrated in Fig. 8.3.

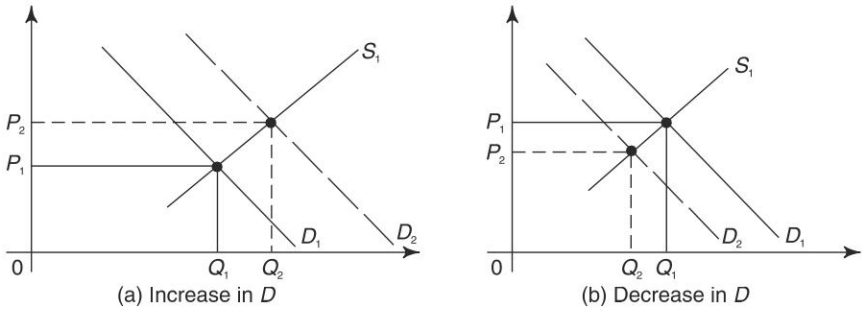



**Fig. 8.2** *Effects of a change in supply on price and quantity*

Figure 8.3(a) shows that an increase in supply from $S_1$ to $S_2$, *ceteris paribus*, causes product price to fall from $P_1$ to $P_2$ and quantity to increase from $Q_1$ to $Q_2$. Figure 8.3(b) shows just the reverse.

**(c) Change in Demand and Supply—Uni-Directional** The effect of an uni-directional change in both demand and supply is ambiguous on price but a clear positive on quantity. This is explained in Fig. 8.4.

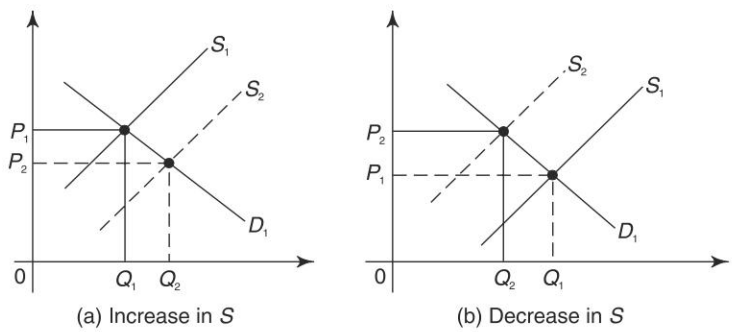



**Fig. 8.3** *Effect of a change in supply on price and quantity*

Figure 8.4(a) shows the effect of an increase in both the demand and supply on price and output. If the said demand increases from $D_1$ to $D_2$ and supply from $S_1$ to $S_2$, price falls from $P_1$ to $P_2$ and quantity increases from $Q_1$ to $Q_2$. However, if the increase in supply remains the same while that in demand becomes from $D_1$ to $D_3$, price remain unaltered at $P_1$ though quantity increases from $Q_1$ to $Q_3$. In still another situation where demand increases from $D_1$ to $D_4$, increase in supply remaining the same, price goes up from $P_1$ to $P_4$ and quantity increases from $Q_1$ to $Q_4$. Thus, it is clear that the effect of an increase in both industry's demand and supply is ambiguous on price but a clear positive on quantity traded in the market. Fig. 8.4(b) shows just the reverse situation.

**(d) Change in Demand and Supply—Opposite Direction** In contrast to the effect of a uni-directional change in demand and supply on price and quantity, the effect of an opposite directional change in these is unambiguous on price and ambiguous on quantity. This is illustrated in Fig. 8.5.

It is clear from Fig. 8.5(a) that an increase in demand and a simultaneous decrease in supply

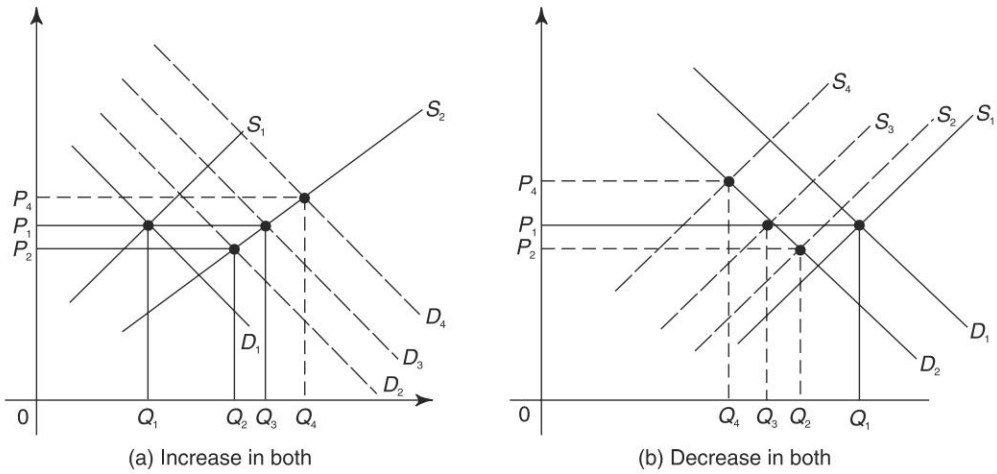



**Fig. 8.4** *Effects of a uni-directional change in demand and supply on price and quantity*

invariably leads to an increase in price but its effect on quantity depends on the relative changes in the two variables. If the increase in demand is relatively more than the decrease in supply, quantity sold and bought would increase, and vice versa. Figure 8.5(b) shows just the reverse of Fig. 8.5(a).

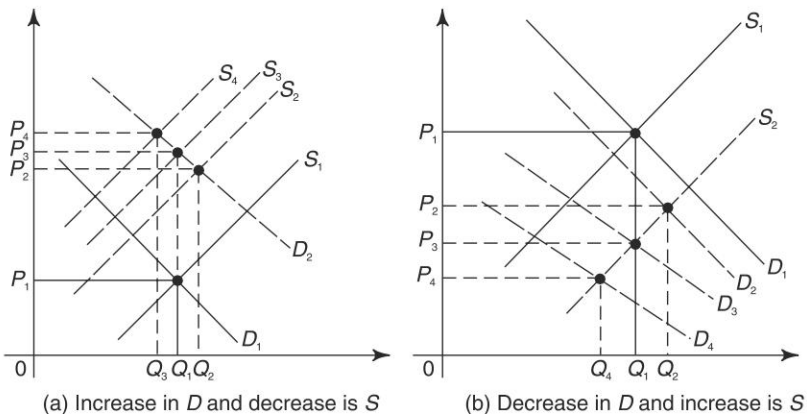



**Fig. 8.5** *Effects of an opposite directional change in demand and supply on price of quantity*

The ones explained above are the normal shapes of demand and supply schedules or curves. However, their actual shapes depend on the structures of the product market and factors' markets. Further, the shape of the supply function is influenced by the objectives of the firms producing the product. Thus, market structure and firms' objectives also have bearings on price.

## Consequences of Government Interventions

Governments' interventions have varying influences on the prices of goods and services. When a

government fixes the price of a good on all its output, the impact is obvious and the **price is pegged**. Under the dual pricing, the price is pegged for a part of the output and that exerts its effect on the free market price for the rest of the output. Generally speaking, the levy price is lower than, and the free market price is higher than, the price that would have prevailed in the market had there been no government intervention of this sort. Consequently, through the dual pricing system, government subsidises the consumption of the quantity available through ration shops, and taxes the consumption of the quantity bought in the open market.

**Price floors** and **price ceilings** have repercussions both on the price as well as the availability of the good in question. This is illustrated in Figs. 8.6 and 8.7, respectively.

In both figures, *OP* is the equilibrium price and *OQ* the equilibrium quantity sold and bought in the absence of price floor/ceiling. Incidentally note that if the floor price is set below the equilibrium

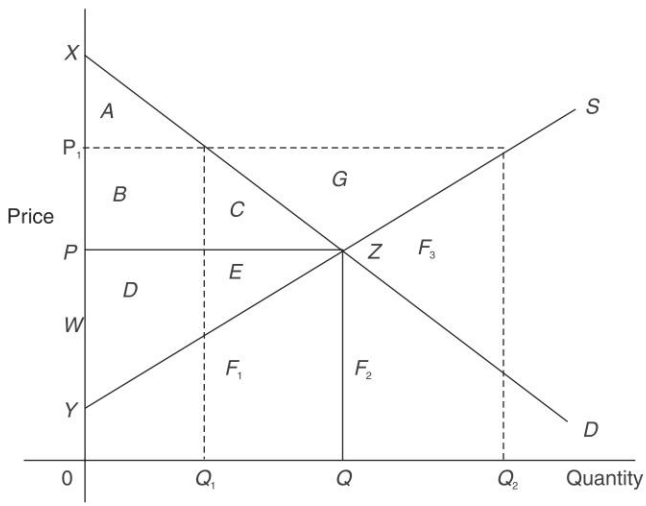



**Fig 8.6** *Consequences of price floor*

price, it would not be effective. Thus, if the floor price is less than *OP*, its consequences would be zero. In such circumstances, the floor price is an unbinding constraint. Since producers are able to sell at a price higher than the floor price, they would be no complaint and there would be no transaction at the floor price. Similarly, if the ceiling price is above the equilibrium price, it is not effective. Consumers would not complain, for they are able to buy the product below the ceiling imposed by the government. Consider the cases where floors and ceilings are the binding constraints.

In Fig. 8.6, *OP* is the equilibrium price and let the floor price equals $OP_1$. At $OP_1$ sellers would like to sell $OQ_2$ quantity while buyers would like to buy $OQ_1$ quantity only, giving rise to excess supply in the amount of $Q_1$ $Q_2$. Thus, effective *price floors lead to excess supply* or *glut* of the commodity in question. To avoid further damage, the government may purchase this excess quantity and store it in godowns, sell it through the public distributiion system (PDS) at a subsidized price to people below the poverty line, or/and export it to the foreign market, if deemed viable. Since storage requires godowns, and it costs in terms of godown space, transport in and out of godowns, loss through spoilage and pests

such as mice, interest cost on the money invested in the purchase of stocks, etc., there is a limit up to which the government can store the goods in excess supply. Selling at a subsidized price through PDS will help poor people but at cost to the governments/tax-payers. Similarly, there is no guarantee about the availability of a foreign market for the goods in excess supply, particularly at economic prices. There are cases where such products are dumped in the foreign market at prices below cost price or those in the domestic market. In view of these repercussions, price floors have their own limitations and they have to be designed rather judiciously.

In addition, price floors, like all other government interventions, have welfare implications. To understand these, we need to go for some new concepts, called **consumer surplus, producer surplus, economic surplus and dead weight loss**. Recall that demand curve reflects the maximum prices that the consumer is willing and able to pay for various quantities. Thus, it is also the **marginal utility (MU) curve**, where the MU is denoted in nominal term (rupees). Accordingly, the area under this curve and the output axis denotes the gain that the consumer receives by consuming the good in question. The cost that the consumer bears for the product is given by the quantity he buys multiplied by the price he pays. The difference between the two is the consumer surplus. Thus, in Fig. 8.6, under no government intervention, consumer buys $OQ$ units at price equals $OP$, which yields consumer gain equal to the area under the demand curve until demand = $OQ$, which equals (area $OXZQ$) = areas $A + B + C + D + E + F_1$ in the said graph. The cost that the consumer bears on the demand $OQ$ equals ($OPZQ$) = areas $D + E + F_1$. The difference between the two = areas $A + B + C$, which denotes the consumer surplus. Similarly, recall that the supply curve denotes the minimum prices that the firm would accept to sell various quantities of its product. Alternatively, it is also (a part of) the industry's $MC$ curve. Accordingly, the area under this curve and the output axis gives the cost that the seller bears. Thus, in Fig. 8.6, the firm sells $OQ$ units of its product and its cost = area $OYZQ$ = area $F_1$ (ignore the dotted line near $F_1$ for this). The gain (revenue) which the firm collects on the said output = output times price = areas $D + E + F_1$. The difference between the producer gain and producer cost measures producer surplus, which = areas $D + E$. Sum of the consumer surplus and producer surplus is known as the economic surplus due to the availability of the good in the market. In this example, this equals = area of triangle $YXZ$ = areas $A + B + C + D + E$. In other words, economic surplus equals the area between the demand and supply curves, which falls between the origin and the equilibrium quantity bought and sold.

To see the welfare effects of the above price floor, recall that $P = OP$ under no government intervention and it is equal to $OP_1$ under price floor, and note the relevant details in Table 8.1.

**Table 8.1** *Welfare Effects of Price Floor*

| *Surplus* | *No govt. intervention* | *Price floor (= $OP_1$)* | *Difference in surplus* |
|---|---|---|---|
| Consumer surplus | $A + B + C$ | $A$ | $B + C$ |
| Producer surplus | $D + E$ | $B + D$ | $E - B$ |
| Economic surplus | $A + B + C + D + E$ | $A + B + D$ | $C + E$ |
| Dead weight loss | – | $C + E$ | – |

**Dead weight loss is the loss due to the intervention to the free market environment** (*laissez faire*) and accordingly it is given by the decrease in economic surplus due to the intervention. In the above example, this equals = areas $C + E$. Incidentally note that the burden is borne heavily by consumers (= $B + C$) and the sellers may bear some burden (if $E > B$) or even gain (if $E < B$) under price floor. Also,

note that there is a transfer of surplus from consumers to producers in the amount equal the area *B*. The pertinent question now is that if government interventions create dead weight loss (DWL), why they are resorted to? Note that government exists for several reasons, two of which are to foster **equity** and **efficiency** in the economy. By setting price floor, government is ensuring fair price to the producers and thereby ensuring the product's availability in future. It the governments have a price support programme, then it would buy the excess supply at its set price, and distribute the same through PDS to help people below the poverty line at a subsidized price. Under such a situation, governments would incur a loss, equal to excess supply times the difference between the support price and the subsidized price, and producer surplus will increase to the areas of triangles B + C + D + E + G (vide Figure 8.6), giving a net increase equal to areas of triangles B + C + G. The dead weight loss would also change accordingly.

We may now illustrate the consequences of price floor through pursuing the above numerical example (See Eqs. 8.1 and 8.2). The equilibrium price was 25 and thus an effective floor price would be above 25, let $P_F = 30$. At this price, demand would equal 150 – 2 (30) = 90, and supply = –25 + 5 (30) = 125. The equilibrium price = 30, equilibrium output would be the lower of the demand and supply, and thus equal to 90, leaving a glut = 35. Demand and supply curve in Fig. 8.6 do not correspond exactly to this numerical example. If they did; the values of prices and outputs corresponding to points along the price and quantity axes; and the values of the various areas of Fig. 8.6 would be as follows:

Points: $X = 75$, $P_1 = 30$, $P = 25$, $W = 23$, $Y = 5$, $Q_1 = 90$, $Q = 100$ and $Q_2 = 125$

Areas: $A = 90 \times 45/2 = 2{,}025$, $B = 90 \times 5 = 450$, $C = 10 \times 5/2 = 25$, $D = (90 \times 2) + (90 \times 18/2) = 990$, $E = 10 \times 2/2 = 10$

Accordingly, DWL = 35, transfer from buyers to sellers = 450

The case of price ceiling is illustrated in Fig. 8.7.

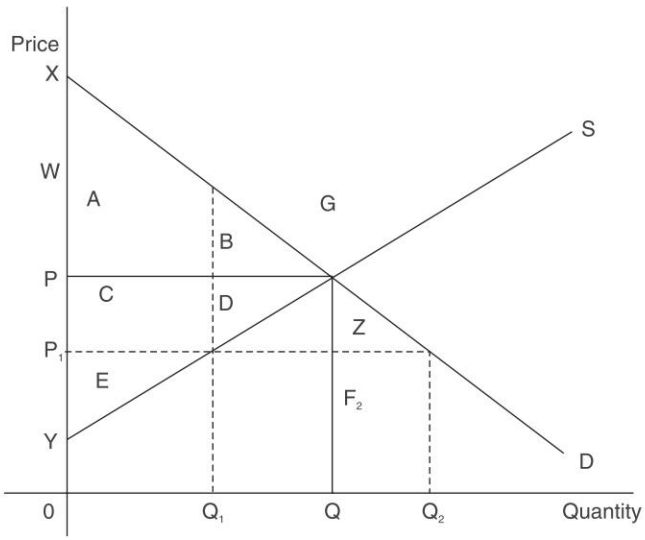



**Fig. 8.7** *Consequences of price ceilings*

If the ceiling price is $OP_2$ (less than the equilibrium price $OP$), demand equals $OQ_2$ while supply equals just $OQ_1$, giving rise to an excess demand or shortage of the commodity in the amount of $Q_1 Q_2$.

Thus, *an effective price ceiling leads to an excess demand*. The excess demand creates its own problems: who would get the product in what quantity or how to apportion its distribution among the consumers. If the distribution is conducted on the basis of a first come first served basis, there would be a queue, which would result in wastage of time, tensions and fights caused by queue breakers, who exist in every society. If it is done through computerised random method, some genuine needs might go unmet. Similarly, if judgemental method of distribution is practised, court proceedings could go unabated. Besides these problems, *price ceilings would give rise to black marketing*. This is because in such a situation, on the one hand, consumers are both willing and able to pay a price higher than the ceiling price and, on the other hand, sellers are both willing and able to sell extra if higher prices are available. Some buyers are eager to resell at prices higher than the ones they paid for the product, to reap some profits in the process. Thus, there are evidences of handsome premiums on many items, such as Bajaj/Vespa scooter in 1970s and Maruti Car in 1980s. To minimize such unfavourable consequences, price ceilings are sometimes accompanied by liberal import policy with regard to the products in question, and they are decided on after serious deliberations. To see the welfare effects of price ceilings, look at the details in Table 8.2.

**Table 8.2** *Welfare effects of price ceiling*

| *Surplus* | *No govt. intervention* | *Price ceiling (at $P = P_2$)* | *Difference* |
|---|---|---|---|
| Consumer surplus | $A + B$ | $A + C$ | $B - C$ |
| Producer surplus | $C + D + E$ | $E$ | $C + D$ |
| Economic surplus | $A + B + C + D + E$ | $A + C + E$ | $B + D$ |
| Dead weight loss | – | $B + D$ | – |

As would be obvious from the Table 8.2, price ceiling causes DWL and transfer of some surplus (= $C$) from producers to consumers. Producers surely suffer a loss, while consumers may gain or lose depending on whether $C$ exceeds $B$ or fall short of it, respectively. Further, in fact, those buyers who are able to buy the good gain but those who are now unable to buy the good suffer. Through this process, government ensures the availability of the product to the vulnerable sections of the society and thus foster equity.

Applying price ceiling to the above numerical example (See Eqs. 8.1 and 8.2), the effective price ceiling ($P_C$) will have to be lower than the free market equilibrium price (=25), thus let $P_C = 20$. At this price, demand = $150 - 2 \times 20 = 110$, supply = $-25 + 5 \times 20 = 75$, and thus equilibrium $Q = 75$, $P = 20$, and shortage = 35. The points along the price and quantity axes, and the various areas of Fig. 8.7 would take the following values:

Points: $X = 75$, $W = 37.5$, $P = 25$, $P_2 = 20$, $Q_1 = 75$, $Q = 100$, and $Q_2 = 110$

Areas: $A = 75 \times 50 - 75 \times 37.5/2 = 2343.75$, $B = 25 \times 12.5/2 = 156.25$, $C = 75 \times 5 = 375$, $D = 25 \times 5/2 = 62.5$, $E = 75 \times 15/2 = 562.5$,

Accordingly, DWL = 156.25 + 62.5 = 218.75 and transfer from producers to consumers = 375

Then what is the effect of price floor and price ceiling on industry output and price under perfect competition? As seen above, equilibrium output falls under both kinds of intervention if government does not buy for selling through PDS; and price goes up under price floor and it goes down under price ceiling. Further, since industry output falls, some firms would leave the industry and move on to new businesses and/or cut their outputs.

Let us now turn to **taxation**. There are three aspects worth considering here, viz.

(a) How taxes affect equilibrium price and output?

(b) Who bears the burden of (indirect) taxes—consumers and/or sellers?

(c) What implications taxes have on economic welfare?

We take up these aspects in this order.

To answer these questions, let us go back to the above hypothetical example, with demand and supply as in Eqs. 8.1 and 8.2; and free market equilibrium price = 25, industry output = 100 and firm output = 23. Now suppose the government imposes a specific (per unit of output) sales (or excise) tax at the rate of Rs. 2 per unit on the product of this industry, what would happen to the equilibrium price and output? The demand equation contains the price that the consumers pay while the supply equation has the price that the firms receive. In the presence of sales tax, the two prices are different, and their difference just equals the tax rate. If $P$ is defined as the price that the consumers pay, then $P - t$ (where $t$ is the tax rate) is the price that sellers receive. Thus, in the event of a specific sales tax of Rs. 2 per unit, the demand equation would continue to remain the same and the supply equation would change, leading to the following.

Demand: $Q = 150 - 2P$

Supply: $Q = 5(P - 2) - 25$

The solution of these two equations yields $P = 26.43$ and $Q = 97.14$. A comparison of this solution with that of the before tax (free market) solution suggests that

(a) Sales tax lowers the output and raises the prices to the consumer. In this specific case, output has declined by 2.86 units and the price to the consumer has increased by Rs. 1.43.

(b) Tax burden falls both on the buyers as well as on the seller. In this case, the former bears a burden equivalent to Rs. 1.43 (= 26.43 – 25) per unit and the latter the residual, i.e., Rs. 0.57 [= (25 – (26.43 – 2)], per unit.

There is a general rule with regard to the distribution of the burden of specific sales tax between the buyers and sellers. The same is the following:

$$\frac{\text{Tax Burden on Buyer}}{\text{Tax Burden on Seller}} = \frac{E_S}{E_D} \tag{8.6}$$

where, $E_S$ = Price elasticity of supply

$E_D$ = Price elasticity of demand

In our example $E_S = 5\ (25/100) = 1.25$, $E_D = (-2)(25/100) = -0.5$, and the ratio of two equals $-(1.25/0.5) = -2.5$, which equals their relative share (absolute value) subject to the rounding error made above, viz. 1.43/0.57 = 2.5. Since **cigarettes demand** is relatively price inelastic while its **supply** relatively highly price elastic, the burden of a tax on it is borne largely by smokers. This explains why when New Jersey (USA) raised cigarettes tax from 40 cents to 80 cents/pack in 1998, the cigarettes price there went up from $ 2.40 to $2.80/pack. An estimate of the price elasticity of demand for this example stands at –0.59.

The proof for the formula (8.6) is provided through geometry in Fig. 8.8.

In Fig. 8.8, $S_1$ denotes the supply curve before tax and $S_2$ that after a specific sales tax at the rate of the length of line $P_1 P_2$. The equilibrium price before tax equals $OP$, and that after tax equals $OP_1$ to

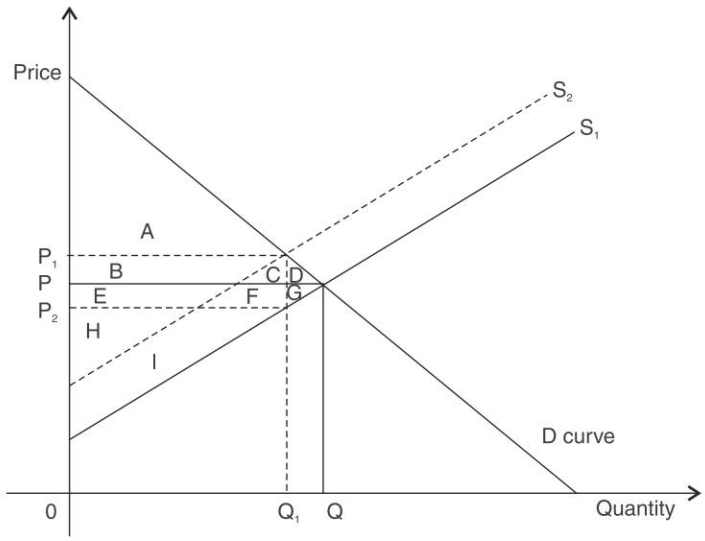



**Fig. 8.8** *Incidence of specific sales tax*

the buyer and $OP_2$ to the seller. Thus, the price to the consumer has increased by $PP_1$ and that received by the seller has declined by $PP_2$. Consequently, the tax burden (= $P_1P_2$) is divided between the buyers and sellers by the amount of $PP_1$ and $PP_2$, respectively.

At the point corresponding to the initial equilibrium point on the demand curve (*D curve*), the price elasticity of demand ($E_D$) is given by

$$E_D = \left[\frac{QQ_1}{PP_1}\right] \cdot \left[\frac{OP}{OQ}\right]$$

and the price elasticity of supply ($E_S$) equals

$$E_S = \left[\frac{QQ_1}{PP_2}\right] \cdot \left[\frac{OP}{OQ}\right]$$

The ratio of the two elasticities is thus given by

$$\frac{E_S}{E_D} = \frac{PP_1}{PP_2}$$

which, as per the previous paragraph, equals the ratio of the tax burden on the buyer to that on the seller. Thus, we have the results of Eq. (8.6) above. In a normal situation, where the demand curve is downward sloping and the supply curve is upward sloping, the tax burden is shared by both the buyers and sellers. However, their burden would be equal only when the two elasticities are equal. Further note the following inferences from the result of Eq. (8.6):

(a) If $E_S = 0$ and $E_D \neq 0$, entire tax burden lies on the seller
(b) If $E_D = 0$ and $E_S \neq 0$, entire tax burden falls on the buyer
(c) If $E_S = \infty$ and $E_D$ lies between zero and infinity, entire tax burden lies on the buyer, and
(d) If $E_D = \infty \neq$ and $E_S$ lies between zero and infinity, entire tax incidence falls on the seller

A careful understanding of the above observations would reveal that there is a premium for high elasticity. That is, the higher the price elasticity of demand, the lesser is the tax incidence on the buyer,

and the higher the price elasticity of supply, the lesser is the tax burden on the seller. In view of these implications, the government will be able to decide the group of goods and services that should be taxed. This decision, of course, depends on the distribution of tax between buyers and sellers. For example, if the government decides to collect tax revenues primarily from the consumers, then it must tax items of necessity (whose price elasticities of demand are low) and if the decision is to collect such revenues mainly from the producers, then the government must tax items of luxuries (whose price elasticities of demand are high). Also by levying taxes on agricultural goods (whose price elasticities of supply are relatively low), the government can collect more revenue from farmers than from households, and by taxing industrial products (whose price elasticities of supply are relatively high), it can collect more revenue from households than industrialists.

Moving on to the welfare effects of taxation, we analyse the losses to consumer surplus and producer surplus, and look at the dead weight loss. The same are presented in Table 8.3.

**Table 8.3** *Welfare effects of specific sales tax*

| *Surplus* | *Free market* | *Specific Tax @ Rs. 2/unit* | *Difference* |
|---|---|---|---|
| Consumer surplus | $A+B+C+D$ | $A$ | $B+C+D$ |
| Producer surplus | $E+F+G+H+I$ | $H+I$ | $E+F+G$ |
| Govt. revenue | Nil | $B+C+E+F$ | $-[B+C+E+F]$ |
| Economic surplus | $A+B+C+D+E+F+G+H+I$ | $A+B+C+E+F+H+I$ | $D+G$ |
| Dead weight loss | – | $D + G$ | – |

In Fig. 8.8, the points along the price and quantity axes, and the various areas of Table 8.3 would take the following values:

Points: $P_1 = 26.43$, $P = 25$, $P_2 = 24.43$, $Q_1 = 97.14$, $Q = 100$ and $Y = 5$

Areas: A $= 97.14 \times 48.5/2 = 2{,}359$, $B + C = 97.14 \times 1.43 = 138.9$, $D = 2.86 \times 1.43/2 = 2.05$, $E + F = 97.14 \times 0.57 = 55.4$, $G = 2.86 \times 0.57/2 = 0.81$ and $H + I = 97.14 \times 19.43/2 = 943.2$

Accordingly, DWL = 2.86, government tax collection = 194.3, loss to consumer surplus = 140.95, and loss to producer surplus = 56.21.

The effects of an *ad valorem* sales tax [tax per unit of revenue = $t$ $(TR) = t$ $(P)(Q)$] could similarly be analysed. Under such a tax, the supply curve would not shift parallel, as it does under a specific tax. In particular, the *TC* equation under *ad valorem* tax would be like

$$TC^* = TC - tQP$$

where $TC^* = TC$ under *ad valorem* tax @ $t$, $TC^* = TC$ under no tax, and $P$ and $Q$ have their usual meanings. Accordingly, the slope of the supply curve would increase as the price increases under the *ad valorem* tax instead of a parallel shift as under the specific tax. Subsidies are negative taxes and their effects would exactly be opposite to those of taxes. To conclude the effects on government interventions, it must be emphasized that while they do **cause dead weight loss, they also foster equity and efficiency** (through appropriate use of tax proceeds and re-distribution effects under price floors/ceilings, (Chapter 7), which constitute their backbone.

## 8.2 PRICING UNDER MONOPOLY

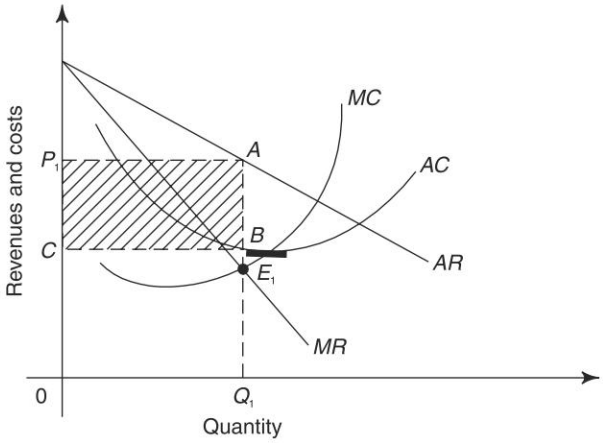
A pure monopolist market is characterised by a single producer and a unique product. Both the entry and exit are totally prohibited, and the product is such that it has no close substitute available in the market. Under such conditions, the firm itself constitutes the industry. The number of buyers in such a market could be large, small or just one. When there is only one buyer, the monopolist market becomes *a bilateral monopoly* market. The bilateral markets are rather rare. For example, for many defence goods, there are single producer and the government is the single buyer. For most other monopolist's products, like public utilities (gas, electricity, landline telephones, etc.), there are a number of consumers. Pricing under bilateral monopoly is basically handled through bargaining between the two parties, and so it is not attempted here. Throughout this section, pricing is discussed for the case of monopolist products, each of which has a large number of consumers.

There are two versions of a purely monopolist market: **simple monopoly** and **discriminating monopoly**. Under the former, the firm charges a uniform price for its product from all its customers, while under the latter, it charges different prices for a given product from different sets of customers or even for different units of its product from the same customer. Since in this and the next chapter, we are assuming the law of one price, we shall discuss here the pricing under simple monopoly only; pricing under discriminating monopoly will be taken up in Chapter 10. Monopoly markets are also classified in terms of **natural monopoly** and **non-natural (or artificially created) monopoly**. As the name implies, the former is due to the nature of the product and the latter is through the human-made or artificial barriers (viz. legal, economic and deliberate barriers, See Chapter 7). Thus, a natural monopoly market is defined as the one where the firm faces a monotonically downward sloping average cost curve for output range no less than the total demand for the product in the market. In other words, the average cost curve facing a natural monopolist slopes downward for all the relevant range of output. Such a market prevails when the fixed costs are exorbitant and could be exploited well only if the output is rather large, such that it may even exceed the total demand for the product. The examples are found in railways, airports, electricity transmission and distribution, landline phones, etc. All such products involve huge initial costs in creating facilities (like railway tracks for railways, take-off and landing facilities at airports for planes, laying down the distribution lines for landline phones and electricity) for any production, which would be highly unrewarding for small productions; and if more than one firm existed, there would be a lot of duplication of facilities. In what follows, we would concentrate at non-natural monopolies, as they are the ones who dominate, and briefly present the case of natural monopoly towards the end.

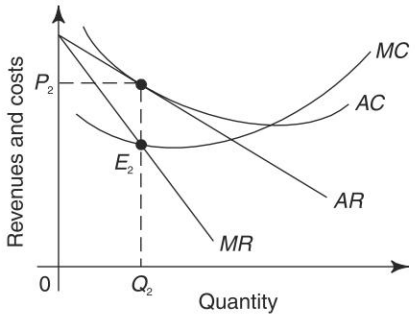
Price under simple monopoly, like that under perfect competition, is determined by demand and supply conditions in the market. Since the number of consumers is large even under monopoly, the monopoly case is similar to the purely competitive case so far as the demand side as a whole (industry demand) is concerned. The difference lies in the demand curve facing a firm. Under monopoly, there is no difference between the industry and the firm, and thus the demand curve facing the monopolist firm is the one faced by the purely competitive industry, which is downward sloping. Further, the downward sloping demand curve implies that more could be sold only at a lower price and vice versa, thus the monopolist firm is not a *price-taker*. Also, the firm is not a price-maker either. This is because the firm and consumers together determine the price. The demand function gives the consumers' side. Thus, if the seller sets the price, buyers would decide the quantity they would buy as per their demand equation. On the other hand, if the monopolist decides the output, the consumers would decide the price at which they would buy that output level, as per their demand equation. As a result, the monopolist has a choice,

i.e., it can either set the price or set the output but not both. Nevertheless, the monopolist is in a better position than his counterpart under perfect competition, who is a price-taker and can merely choose its output level. Further, since the monopolist takes part in pricing its product and the demand for its product varies with the product price, the monopolist does not face a given or infinite demand, as does a purely competitive firm. Accordingly, under monopoly, the relevant curves are revenue (*TR*, *AR*, *MR*) curves instead of demand curve. By similar reasoning, instead of the supply curve, the appropriate curves are cost (*TC*, *AC*, *MC*) curves. As a matter of fact, recall that the concept of supply curve applies only to the perfectly competitive markets. For all other markets, viz. monopoly, monopolist competition and oligopoly, the relevant tools for pricing are *TC, AC* and *MC* curves/equations. The monopolist hires his factors of production from the factors market, just as a purely competitive firm does. Thus, there is no significant difference with regard to the cost curves between the two market structures. Accordingly, the cost curves of the monopolist would be of usual shapes as discussed in the previous chapter.

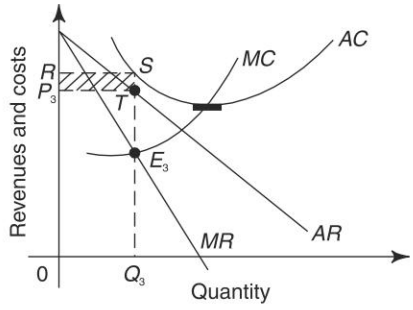
Given the revenue and cost curves, and firms' objective of profit-maximization, price-output determination can easily be explained. The geometric explanation could proceed through the total revenue and total cost curves; or through average and marginal revenue, and average and marginal cost curves. The former approach is less suitable than the latter, for it does not explain price determination




(A): Price-output determination under simple monopoly: normal case




(B): Simple monopolist: breaking-even




(C): Simple monopolist: incurring loss

**Fig. 8.9**

explicitly which the latter does. Thus, the *AR*, *MR*, *AC* and *MC* approach is followed here. Figure 8.9(a) portrays the equilibrium of a monopolist under a normal case.

The monopolist's profit is maximum at a point where $MR = MC$ and the *MC* curve intersects the *MR* curve from below; the option of shut-down is ignored here. Thus, the conditions are met at point $E_1$ in Fig. 8.9. Accordingly, the equilibrium quantity is $OQ_1$ and the equilibrium price, which is given by the *AR* (demand) curve, equals $OP_1$. At this price, the monopolist reaps a profit, equivalent to the area of rectangular $P_1ABC$. A change in either the demand (revenue) curve or in cost curves or in both would cause a change in the equilibrium price and output. It is easy to see that while an increase in demand, which would cause in upward shift in *AR* and *MR* curves, *ceteris paribus*, would lead to an increase both in price and quantity, and in an increase in supply, causing cost to shift downward, *ceteris paribus*, would lead to an increase in quantity but a decline in price.

The monopolist of Fig. 8.9(a) makes some profit. This is both in the short as well as in long-run. This is so, for the entry of new firms in the industry is totally prohibited. In some situations, the monopolist could just break-even or even run into loss, but the loss would not exceed the fixed cost in short run and disappear in the long run. Note that the losses are possible only in short-run, for the monopolist's exit in long run cannot be ruled out. Further, the loss cannot exceed fixed costs, for there is no compulsion on the monopolist to operate the plant and the loss under shut-down condition equals total fixed cost. Such situations are illustrated in Figs. 8.9(b) and 8.9(c), respectively. The monopolist of Fig. 8.9(b) charges price equal to $OP_2$, sells quantity equal to $OQ_2$ and makes no profit or loss. In contrast, the monopolist of Fig. 8.9(c) reaches equilibrium at point $E_3$, charges price equal to $OP_3$, sells quantity equal to $OQ_3$ and runs into loss, equivalent to the area of rectangular $P_3RST$. The break-even equilibrium position, like the one in Fig. 8.9(b), could exist both in the short as well as in the long-run. But no firm, in long-run, would take any loss and thus the equilibrium position shown in Fig. 8.9(c) is merely short run equilibrium. Such a firm would exit from the market in long run.

The equilibrium of a profit-maximizing monopolist may now be explained through a numerical example. Let the equations of the demand curve facing the monopolist and of the total cost curve of the monopolist be the same as in the example under perfect competition above (See Eqs. 8.1 and 8.3):

Demand: $Q = 150 - 2P$

Total cost: $TC = 10 + 2\,Q + 0.5\,Q^2$

It must be noted that since there is just one firm under monopoly, the industry demand equation of perfect competition is taken as the demand equation facing the monopolist and the firm's cost equation of perfect competition is taken as the monopolist cost equation, with a change of $q$ by $Q$ for obvious reason. The demand equation can be transformed as

$$P = 75 - 0.5\,Q$$

The total revenue equation is given by

$$TR = PQ = 75\,Q - 0.5\,Q^2$$

The derivation of which yields the *MR* equation as

$$MR = 75 - Q$$

The derivation of total cost equation yields the *MC* equation as

$$MC = 2 + Q$$

For profit-maximization, $MR = MC$, and so

$$75 - Q = 2 + Q$$

or, $$Q = 36.5$$

Substituting the value of $Q$ in the demand equation yields

$$P = 75 - 0.5\,(36.5) = 56.75$$

At these values of $Q$ and $P$, $TR$ = 2061.375 and $TC$ = 749.125, thus profit = 1312. Since the monopolist is making a profit, the question of its shut-down does not arise. The second-order condition for profit-maximization is also met, for the derivative of $MR$ is negative (= –1) and that of $MC$ is positive (= +1). The shut down condition is also met as the firm is making profit. Thus the profit-maximizing monopolist's equilibrium price = 56.75 and equilibrium output = 36.5. A comparison of these values with those under perfect competition would indicate that the **monopolist charges a higher price** (56.75 > 25) and **produces less** (36.5 < 100) than the **perfectly competitive market**. Why then, monopoly is permitted to exist? The virtues of monopoly are taken up a little later. Let us first look at the three **misconceptions about the monopolist's market**:

(a) Monopolist charges the maximum possible price

(b) Monopolist always makes (economic) profit

(c) Monopolist operates on an inelastic demand curve

The monopolist is often misunderstood by the society and, in fact, none of the above charges holds good. In the above numerical example, demand would be zero if $P = 75$ (to ascertain this, set $Q = 0$ in the demand equation and solve for $P$) and at any price less than 75, demand for monopolist's product is positive. For example, at $P = 70$, demand = 10; at $P = 60$, demand = 30, and so on. Thus, the monopolist could charge any price less than 75 and yet sell some quantity of its product in the market. However, its most profitable price = 56.75 and this is the price the unregulated monopolist would charge in the market. Thus, it is wrong to say that the monopolist charges the maximum possible price. The reason for this wrong impression lies in the fact that, as shown in the previous paragraph, the monopolist price is higher than the price in the perfectly competitive market, given the industry demand and cost conditions.

The second charge against the monopolist also is not true, for it was seen in Figs. 8.9(b) and 8.9(c) above that the profit maximizing monopolist could break-even and even run into losses. However, the charge was conceived as the monopolist generally makes more profit than a purely competitive firm does; and in the long run while a perfectly competitive firm just breaks-even, a monopolist could well reap economic profits.

The last charge is also ill-conceived. It was shown in Chapter 2, section 2.6 that no profit-maximizing firm would ever operate on a price inelastic (absolute value of elasticity < 1) part of the demand curve that it faces. The monopolist is very much a profit-maximizing one and so the question of it operating on the inelastic part does not arise. Once again the charge in the absolute sense is wrong but it is not entirely baseless. The truth is that the demand curve faced by a monopolist is less price-elastic than the one faced by a competitive firm and hence the misconception, though exaggeration.

Thus, all the above three observations about the monopolist's behaviour are wrong. But there is one thing which is quite **undesirable of the monopolist**: the monopolist's equilibrium output is less than the **socially desirable output**. The socially desirable output is one at which $P = MC$. Recall that

this is so because on the one hand, the price represents the money society is willing and able to pay for a unit of the product and thus it measures the benefits the society reaps from the last unit of the good in question. On the other hand, marginal cost measures the cost the society incurs in the production of the last unit. Under pure monopoly, demand curve is falling and hence price is greater than marginal revenue*. But a profit-maximizing monopolist must equate his *MR* to *MC*, and consequently price exceeds marginal cost. Thus, a monopolist produces less than the socially desirable output level. A.P. Lerner has introduced a measure, known as the **Lerner Index**, to measure the degree of monopoly power. This is given by

$$\text{Lerner index} = \frac{(P - MC)}{P} \tag{8.7}$$

The Lerner index takes its minimum value of zero under perfect competition, which indicates no market power. The higher is its magnitude, the greater is the power the firm enjoys. In our numerical example above for monopoly with no government intervention, $P = 56.75$ and $MC = (2 + Q = 2 + 36.5) = 38.5$. Thus, Lerner index equals (56.75 – 38.5) / 56.75 = 0.3216. Recall that monopoly results in socially sub-optimal output. To take this further, one could analyse the dead weight loss due to monopoly. This is explained in Fig. 8.10.

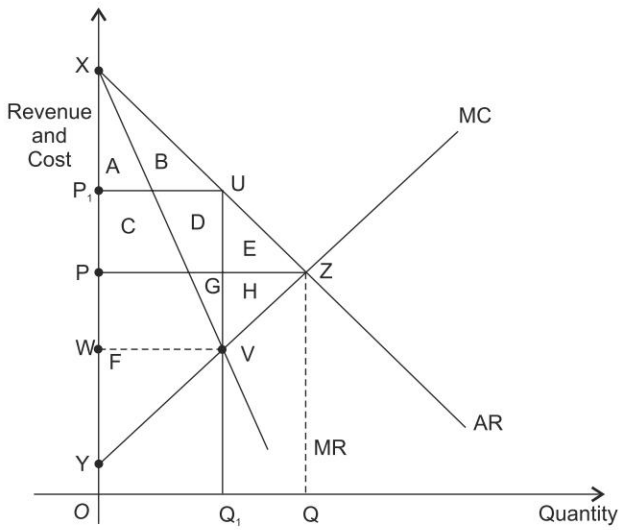



**Fig. 8.10** *Welfare effects of monopoly*

* In the falling demand curve case, $P > MR$, **for** two reasons: One, more can be sold at lower price only. Two, reduction in price is not just for the additional quantity sold but also for the quantity which could have been sold at a higher price. To illustrate this, consider the following example of a falling demand schedule:

| *P* | *Q* | *TR* | *Arc MR* |
|---|---|---|---|
| 10 | 20 | 200 | |
| 8 | 30 | 240 | 4 |
| 6 | 50 | 300 | 3 |

Here $P > MR$, for if sales were to increase from 20 to 30 units, *P* must fall from 10 to 8, and the fall in price applies not only to the additional quantity sold (30 – 20 = 10) but also to the initial quantity (20) which could have been sold at $P = 10$.

Remember that the rising part of the *MC* curve (*MC* curve here is the sum of the *MC* curves of all the firms in the industry) is the industry supply curve under perfect competition. As per the equilibrium conditions, in Fig. 8.10, *OP* and *OQ* are the equilibrium price and output, respectively under perfect competition, and $OP_1$ and $OQ_1$ are those for a monopolist firm. Given these, Table 8.4 gives the various surpluses that are generated under the two types of markets.

**Table 8.4** *Welfare Effects of Monopoly*

| *Surplus* | *Perfect competition* | *Monopoly* | *Difference (loss)* |
|---|---|---|---|
| Consumer surplus | $A + B + C + D + E$ | $A + B$ | $C + D + E$ |
| Producer surplus | $F + G + H$ | $C + D + F + G$ | $H - C - D$ |
| Economic surplus | $A+B+C+D+E+F+G+H$ | $A + B + C + D + F + G$ | $E + H$ |
| Dead weight loss | – | $E + H$ | – |

For our numerical example above, the values for various points on the figure and areas will be as follows:

Points: $X = 75$, $P_1 = 56.75$, $P = 25$, $W = 12.3$, $Y = 5$, $Q_1 = 36.5$, and $Q = 100$

Areas: $C + D = 36.5 \times 31.75 = 1158.875$, $E = 63.5 \times 31.75/2 = 1008.625$,

$H = 63.5 \times 12.7/2 = 403.225$

Thus, DWL = 1411.85 and the transfer of surplus from consumers to producer = $C + D = 1158.875$

As expected, consumers suffer a heavy loss under monopoly. Producers lose due to decrease in their overall output but gain through higher price. The net gain to them is generally positive but could be negative. In the example here, consumers loss (= $C + D + E$) comes to 2167.5 and producers net gain (= $C + D - H$) to 755.65. The society as a whole loses and here it equals 1411.85. It is for this reason that monopoly is considered as bad and accordingly they are often regulated.

Coming to the **virtues of the monopoly market**, there are quite a few good things about monopolist's behaviour, which explain its existence. These include, their profits (which enables them to carry on research and development, so essential for the development of an economy), absence of waste through false advertisements and fancy packaging, etc. (which are inevitable under competitive conditions), reaping the fruits of economies of mass production, abolition of the cost of duplication of facilities (which would arise if competition were permitted in the supply of public utilities, such as electricity, landline phones, water, etc., for competitive firm would need its own wire and pipe connections, etc.), and so on. Due to the above described cons and these pros of a monopolist market, these enterprises are allowed or even encouraged to operate in almost all economies, though their operations are often regulated by the governments. Pricing under regulated monopoly is dealt with next.

## Government Interventions and Monopoly Price

Monopolies are regulated through price ceilings and taxes. The effects of price ceiling on the product of a monopolist are explained in Fig. 8.11. In all three parts of this Fig (8.11)(a), (8.11)(b) and 8.11(c), *OP* marks the equilibrium price and *OQ* the equilibrium quantity of the profit-maximizing unregulated monopolist. In all the three cases, the monopolist reaps economic profit, for the equilibrium price exceeds the average cost. Now, if the government imposes a price ceiling equal to ($OP_1 = MC$), the

effective demand curve becomes $P_1 E_1 D$ and the effective $MR$ curve becomes $P_1 E_1 K_1 L$ (not shown in Fig. 8.11c). The new equilibrium of the monopolist is at point $E_1$, where $OP_1$ is the equilibrium price and $OQ_1$ the equilibrium output. Since $OP_1 < OP$ and $OQ_1 > OQ$, the price ceiling results in a lower price and greater output. As regards its effect on the monopolist's profit, there is some ambiguity. If the monopolist were in a situation like in Fig. 8.11(a) it would continue making profit even under price ceiling, though its profits would certainly fall. However, if the situation were as in Fig. 8.11(b), the monopolist would just break-even under such a price ceiling. The monopolist position would be worse if he were operating in a situation like the one in Fig. 8.11(c). While he was having some profit when unregulated, the price ceiling puts him under a loss, for at $P = MC$, $AC$ exceeds the price by the length

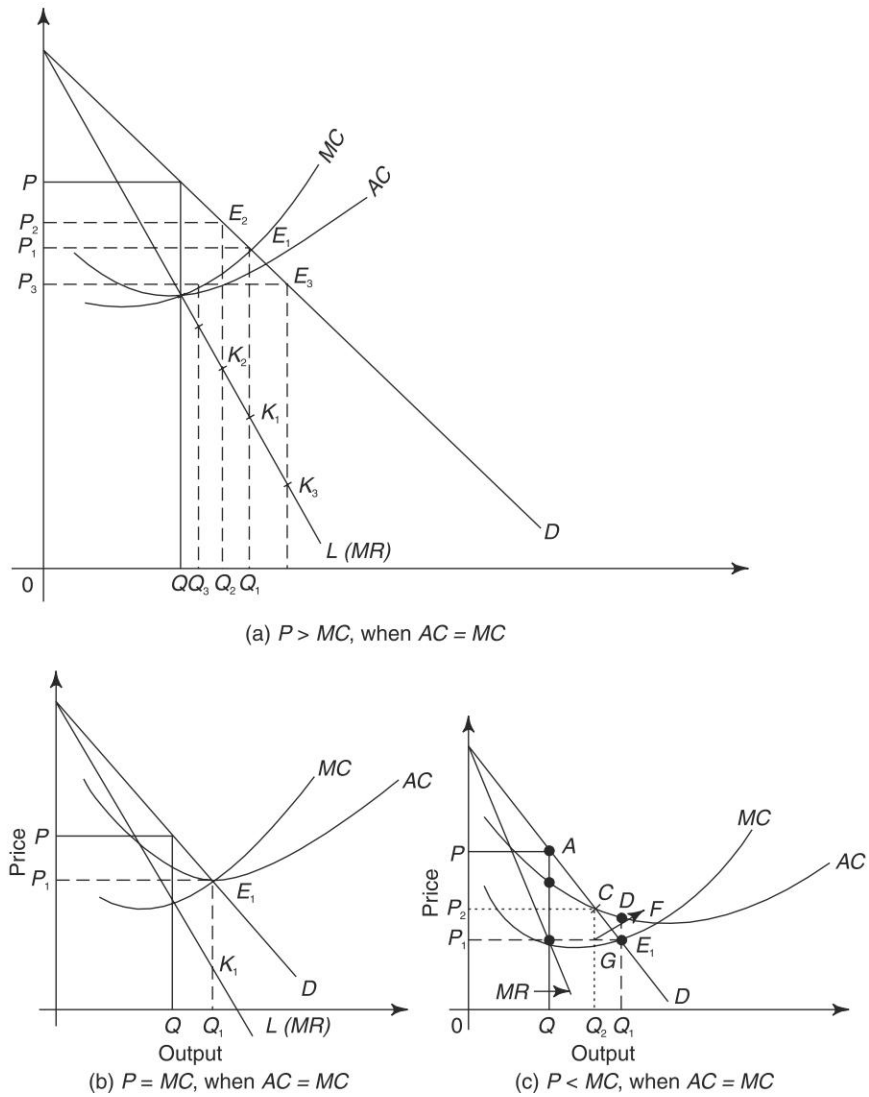



**Fig. 8.11** *Effects of price ceiling on monopolistic equilibrium*

of the line $DE_1$. Thus, if the price ceiling is imposed such that $P = MC$, the monopolist might suffer a loss, which, in turn, argues for the case of a subsidy to a regulated monopolist. If no subsidy is granted, the loss-making monopolist would get out of the business in the long-run.

There is an alternative to subsidy when the price ceiling renders a monopolist in a loss-making situation. The alternative is a sort of price discrimination, not like the one presented in Chapter 10 (different markets), but of charging two prices for the same product in the same market. Under this system, the monopolist could be allowed to charge a price equal to its *AC* (= $OP_2$) for $OQ_2$ and a price equal to *MC* for its remaining output (= $Q_2\ Q_1$). If the *MC* curve is rising in this region as in Fig. 8.11(c), the regulated monopolist would make some profit on the latter ($Q_2\ Q_1$) output (= $\Delta\ E_1\ FG$). Thus, such price discrimination would obviate the need for subsidy and at the same time leave some incentives for the regulated monopolist to operate.

An important point to demonstrate on the consequences of price ceiling on a monopolist's product is that *the ceiling price which would ensure the maximum possible output from the monopolist is the one at which P = MC*, which, as seen above happens to be the socially optimum price and the one that prevails under the conditions of pure competition. This can be proved through Fig. 8.11(a). If the ceiling price were $OP_2$ the effective demand curve would be $P_2E_2D$, the effective *MR* curve would be $P_2E_2K_2L$, and the equilibrium price and output of the regulated profit-maximising monopolist would be $OP_2$ and $OQ_2$, respectively. Similarly, if the ceiling price were $OP_3$, the effective demand curve would be $P_3\ E_3\ D$, the effective *MR* curve would be $P_3\ E_3\ K_3\ L$, and the equilibrium price and output would be $OP_3$ and $OQ_3$, respectively. It was seen above that if the price ceiling were at $OP_1$, the equilibrium price and output would be $OP_1$ and $OQ_1$, respectively. Since $OQ_1$ is greater than both $OQ_2$ and $OQ_3$, and $OP_1$ lies between $OP_2$ and $OP_3$, it is clear that the price ceiling at $OP_1$ extracts the maximum possible output from a regulated profit-maximising monopolist.

This leaves the effects of indirect taxes and subsidies alone to be analysed. While most taxes lead to increase in prices and decrease in outputs, and most subsidies to quite the opposite consequences, the same is not true for all kinds of taxes and subsidies. The consequences depend on the type of tax (subsidy). For this purpose, four kinds of taxes are worth analysing; specific sales tax, *ad valorem* sales tax, profit tax, and tax on real estate (on mere existence).

If $t_1$, $t_2$, $t_3$, and $t_4$ denote the tax rates and $\pi_1$, $\pi_2$, $\pi_3$, and $\pi_4$, the profit after taxes under specific sales tax, *ad valorem* sales tax, profit tax and real estate tax, respectively, the consequent necessary conditions for profit-maximization (*MR = MC*) under various taxes would be as follows.

| *Profit Function* | *Necessary condition for Profit Maximization* |
|---|---|
| $\pi_1 = TR - TC - t_1Q$ | $MR = MC + t_1$ |
| $\pi_2 = TR - TC - t_4(TR)$ | $MR\ (t - t_2) = MC$ |
| $\pi_3 = TR - TC - t_3(TR - TC)$ | $(MR - MC)\ (1 - t_3) = 0$ or $MR = MC$ |
| $\pi_4 = TR - TC - t_4$ | $MR = MC$ |

Note that *TR*, *TC*, *MR* and *MC* stand for the corresponding variables before taxes. Thus, it is seen that while the necessary condition for profit-maximization changes under the first two kinds of taxes, it remains invariant under the last two types of taxes. In consequence, the profit tax and real estate tax are neutral with regard to the equilibrium price and output of the profit-maximizing firm. Under the specific sales tax, *MC* increases by the tax rate for each output, and so for profit to be maximum *MR* must also increase by the same magnitude. Since the *MR* curve is falling, *MR* would increase only if output

declines, which, in turn, would cause price to rise. Thus, specific sales tax leads to decrease in output and increase in price of the profit-maximizing firm. This is illustrated in Fig. 8.12(a).

In Fig. 8.12(a) $AC_0$ and $MC_0$ represent the *AC* and *MC* curves before tax, respectively and $MC_1$ the *MC* curve after tax. Since the tax is specific, $MC_0$ and $MC_1$ are parallel. The equilibrium output and price are $OQ_0$ and $OP_0$ before tax, and $OQ_1$, and $OP_1$, after tax. Thus, specific sales tax causes decrease in output and increase in price.

The case of an *ad valorem* sales tax is illustrated in Fig. 8.12(b). As the necessary condition for

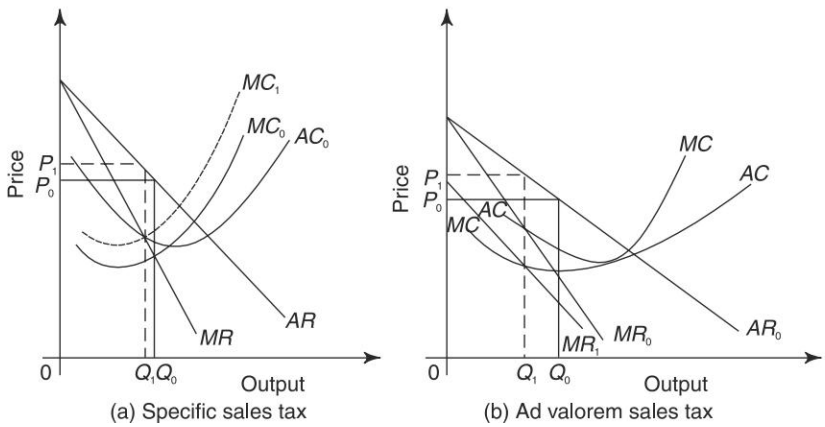



**Fig. 8.12** *Effects of taxes on price and output*

profit-maximization after such a tax would reveal, the tax causes *MR* to decline for all output levels by a fixed proportion, equal to $(1 - t_2)$. Thus, if $MR_0$ is the *MR* curve before tax, $MR_1$, is the *MR* curve after tax. Note that $MR_1$ is not parallel to $MR_0$, for the decline is by a fixed percentage and so the difference between $MR_0$ and $MR_1$ is large for small outputs (when *MR* is high), and small for large outputs (when *MR* is low). The profit-maximizing output and price equal $OQ_0$ and $OP_0$ before tax; and $OQ_1$ and $OP_1$ after tax, respectively. Thus, an *ad valorem* sales tax, like a specific sales tax, leads to a decrease in output and an increase in price of the profit-maximizing firm.

## Natural Monopoly

Recall that under natural monopoly, the firm faces a monotonically falling average cost (*AC*) curve (i.e., decreasing cost industry). The corresponding *MC* curve falls below the *AC* curve for all output levels. The demand (*AR*) curve and *MR* curves take their regular shapes of a monopolist market. The equilibrium under such a situation is illustrated in Fig. 8.13.

In Fig. 8.13, the natural monopolist's equilibrium output and price are *OQ* and *OP*, respectively, and its profit equals the area of rectangle *PABC*. Since *AC* falls monotonically, the firms generally make larger profit than a non-natural monopolist. To curb such huge profitable opportunities, natural monopolists are often regulated through price pegging or price ceilings.

It would be obvious from the Fig. 8.13 that if the price was pegged anywhere between points *A* and *D* on the demand curve, the firm would still make some profit. However, even if all the profits were wiped out by setting the price given by point *D* on the demand curve, the optimum output the firm would

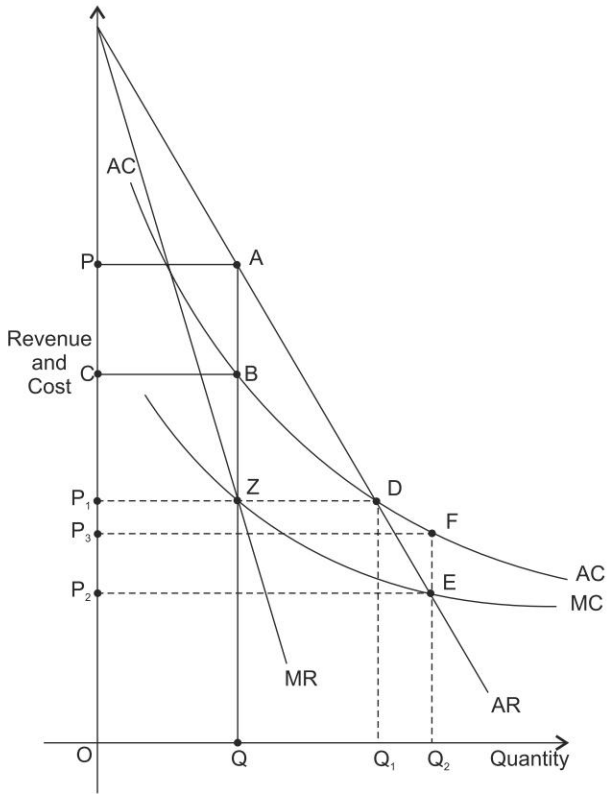



**Fig. 8.13** *Equilibrium under natural monopoly*

produce (which would equal $OQ_1$) would be less than the socially optimum output (which equals $OQ_2$). To induce such a firm to produce socially optimum level of output, government would need to peg the price at *MC*, which is given by the point of intersection (viz. point *E*) of *AR* and *MC* curves. Under the latter situation, the firm would end up running into loss equal to *EF* (*AC* – *MC* at output = $OQ_2$) per unit of output (= $EF \times OQ_2$). Under such an event the natural monopolist would quit the industry/business unless governments grant it a subsidy. This explains an apparent contradiction that sometime governments approve subsidy even to a natural monopolist.

## 8.3 CONTESTABLE MARKETS

The theory of contestable market (*CM*) is of relatively recent origin. It is founded on the premise that a potential competition may be as important, if not more, as the actual competition that firms face in the real market. Such a situation arises when the barriers to entry into a business are either weak or could be managed through economic power, political connections/lobbying, social pressures or terror, etc. To understand it and its implications on price-output decisions, let us first look at the features of a contestable market itself. A market is deemed contestable under the following conditions:

(a) No serious entry/exit barriers

(b) Entry/exit can be executed rapidly

Recall the barriers to entry/exit discussed in Chapter 7. No serious entry barriers would thus mean no legal constraint, no significant economies of scale, no technological issue, etc., and no serious deliberate threat; and no serious exit obstacles mean no significant sunk cost and no serious labour problem if fired, etc. Point (b) is also important, for if these were time consuming, the incumbent firm may manage some barriers in the mean time. It is obvious that the threat of such a potential competition is greater the lower are the entry/exit costs and vice versa. Under the situation where such costs are zero, the market is perfectly contestable. The theory of *CM* suggests that the sheer threat of this happening will ensure that the incumbent firm already in the market will

(a) keep its price down so that it makes no significant abnormal profit and thereby leaves no incentive for the new entrant

(b) produce as efficiently as possible, taking full advantage of economies of scale/scope and new technology

Thus, the message of the *CM* theory is that even when there are no barriers to entry/exit, the incumbent firms may let it be known that any firm that desires to enter will face an all-out war.

In effect, the potential entrants take into sufficient account the possibility of reactions of the incumbent firm while pursuing the entry. To cite an example, the monopolist of Fig. 8.9 was assumed to have no threat of the contested market and accordingly it set the profit maximizing price-output combination. However, if the said firm faces such a potential threat, it may voluntarily opt for a price below its profit maximizing price at $OP_1$, and accordingly produce output in excess of $OQ_1$. Further, such advance actions might be repeated and there by the entry may be blocked altogether. Once the threat disappears, the monopolist may go back to its practice of profit maximization. Furthermore, to avoid such threats, such firms often do lobbying with government officers (which, of course, adds to its cost) to hinder threats/entries. In view of these factors, the *CM* theory exercises downward pressure on product prices and upward pressure on outputs. This leads one to recommend that if a monopoly is encouraged to operate in a contestable market, it might bring *best of both the worlds* for the consumer and the economy. Not only it will be able to achieve low costs through fuller exploitation of economies of scale/scope but also will induce the incumbent to keep prices and profits down. We must here emphasize that a **natural monopoly does not face a contestable market**, for it has significant entry barrier through huge economies of scale.

The *CM* theory goes beyond the monopoly market and there by encourages firms to maintain efficiency, and reasonable prices, outputs and profit margins. Needless to say, the presence of *CM* is good for an economy and governments may be attempting or consider attempting to creating such conditions. Through the policy of liberalization, privatization, and globalization, the scope of contestable markets has grown tremendously in recent years. Such markets are currently found in businesses like airlines, mobile phones, electricity generation, courier services, cable providers, and so on. American Railroads in the early 20$^{th}$ century were weakened not by competing railroads but by airlines, buses and trucks. Full-servicing airlines have been threatened by the low-cost, no frill airlines. Postal services have been contested by courier services. Hindustan Levers has been challenged by Nirma.

There are new pricing theories for contestable markets. These are known as **limit pricing and predatory pricing.** The limit price is the one which is below the monopolist optimum price (i.e., below $OP_1$ in Fig. 8.9) while the predatory price stands for a price below the firm's marginal cost (i.e., below the point of intersection of *AR* and *MC* curves in Fig. 8.9). By practicing these prices, the firm may prevent a

potential entry and thus they serve as means to curb the competition. Needless to say, practicing of either of these prices, the firms' profits are reduced, and may even turn to losses depending on as to whether at the so set price-output combinations, firms' price/*MC* exceeds or fall short of its *AC*. However, as suggested above, the limit/predatory price may be a temporary recourse until the threat disappears, and thus over long run, the incumbent firm should make profits.

## 8.4 PRICING IN MULTI-PLANT FIRMS

The foregoing discussion on price-output decisions had implicitly assumed that a firm has just one plant. However, in real life, many firms own more than one plant to produce the same good and sell all its outputs of all the plants through a common channel at a uniform price. The examples would include electricity, gas and telephone companies, and many companies engaged in the supply of automobiles, heavy plants and machines, televisions, refrigerators, cement, steel, pharmaceuticals, etc. Further such companies have plants not only in different parts of a given country but also in a number of countries (multi-nationals). Obviously, these companies have different production technologies, and they face different inputs' prices and efficiencies, taxes, weather conditions, industrial environment, etc. in different plants. In the circumstances, such a firm incurs different costs in different plants for the same product, though its product of all the plants is sold at a uniform price within the constraint of the consumers' demand function. Under such a situation, how does the firm decide on its price, total production and the allocation of production to different plants?

The decision problem of a multi-plant firm is akin to that of a standard (undifferentiated) oligopoly. The firm has one demand function for its output but multiple cost functions, one for each plant. As a profit maximizing entity, it sets its total output in such a way that its aggregate profit is maximum with regard to its outputs in each of its plants. Thus, the optimization problem for a multi-plant firm is

Maximize, $\pi = TR - (TC_1 + TC_2 + TC_3 + \ldots\ldots + TC_k)$ (where, $k$ = number of plants)

The necessary condition for maximization would mean,

$$d\pi/dQ = MR - CMC = 0$$

or, $$MR = CMC \text{ (where, } CMC \text{ = combined marginal cost)} \tag{8.8}$$

Both *MR* and *CMC* are functions of total output ($Q$) and thus the solution of the above equation would yield the equilibrium level of output. Substitution of that in the demand function would give the equilibrium price. The so determined values would, of course, have to satisfy the sufficient condition (as well as non-shut down condition), which are usually met when *CMC* is upward sloping but ignored here just to avoid complications. The distribution of the total output among two or more plants would be decided on the basis of the following conditions:

$$MR = MC_1 = MC_2 = MC_3 = \ldots = MC_k \tag{8.9}$$

Equation (8.9) contains $k$ equations in as many variables ($Q_1$, $Q_2$, …, $Q_k$ stand for outputs of plants $i = 1, 2,3,\ldots k$) and thus it would yield an unique solution for the outputs of various plants. By design, total output of the firm would equal the sum total of outputs of all its plants.

The equilibrium of the firm and various plants in the multiple plant case could easily be explained through geometry as well. The graph would be similar to that of a standard (undifferentiated) oligopoly

under the agreed cartel case, given later in Fig. 9.1 in Chapter 9. Here, instead of different firms and the industry of standard oligopoly, we would have different plants and the firm. The horizontal sum of *MC* curves of various plants would give the firms' *CMC* curve. The intersection of the so obtained *CMC* curve and the firm's unique *MR* curve would give the total output of the firm. Once the firm's total output is determined, the unique demand curve facing the firm would yield the equilibrium price at which the firm must sell its product. The distribution of the total output among various plants would be given by forcing the equation system (8.9) in the graph through a horizontal straight line on the *Y* (price)-axis passing through the point of intersection between the *CMC* and *MR* curves. We now take a numerical example of a firm with two plants to illustrate the price-output determination.

Suppose the firm faces the following demand function and the cost equations of its two plants are as follows:

$$P = 500 - 2Q \qquad TC_1 = 25 + 2Q_1^2 \qquad TC_2 = 20 + Q_2^2$$

Then, $TR = PQ = 500Q - 2Q^2$

$$MC_1 = 4Q_1 \quad \text{and} \quad MC_2 = 2Q_2$$

Thus, $MR = 500 - 4Q$

To obtain, *CMC*, we invert *MC* functions, which yield

$$Q_1 = 0.25\,MC_1 \quad \text{and} \quad Q_2 = 0.5\,MC_2$$

Summation of the two, gives $Q = 0.75\,CMC$ (where $Q = Q_1 + Q_2$)

or, $CMC = 1.333\,Q$

Equating *MR* and *CMC* then gives

$$500 - 4Q = 1.333\,Q$$

or, $Q = 93.75$

Substitution of this *Q* value in the above demand function, yields

$$P = 312.5$$

Further setting $MR = MC_1$ and $MR = MC_2$ gives,

$$500 - 4Q = 4Q_1 \quad \text{and} \quad 500 - 4Q = 2Q_2$$

Substitution of $Q = Q_1 + Q_2$ in above equations and solution of the resulting equations in $Q_1$ and $Q_2$ would give

$$Q_1 = 31.25 \text{ and } Q_2 = 62.5$$

Further, profits of individual plants would be given

$$\pi_1 = PQ_1 \quad \text{and} \quad \pi_2 = PQ_2 \text{ which would be } \pi_1 = 4232 \text{ and } \pi_2 = 4324.5$$

This completes the solution of the two plant firm. The same procedure could easily be applied for plants with more than two plants. Before moving to the next section, we must point out that the above approach is good only when the *MC*s across plants are not constants. If they were constant and the same, the division of total output across plants would have to be made arbitrarily. If *MC*s were constants but different across plants, then the optimization would require all output of the firm to be produced

in the plant which has the least *MC*. This is easy to see and hence warrants no further explanation. Nevertheless, interested readers could look at hypothetical examples cited under cartel solution of the standard oligopoly case (Refer Chapter 9, See Fig. 9.1).

## 8.4 PRICING IN MULTIPLE PRODUCTS FIRMS

The foregoing sections have assumed that firm produces a single product only. However, most of the firms today are engaged in the production of two or more products simultaneously. If the products are separable both in terms of their costs and demands, price-output determination can proceed in the same manner as in the case of a single product. However, generally speaking, a multiproduct firm faces separate demand functions for individual products but many of its costs, particularly the overheads, are hardly divisible product-wise. For example, a livestock firm which rears sheep produces both wool and meat simultaneously, faces separate demand functions for both its products, but its cost of production of the two goods is totally mixed-up. Similarly, an educational institute like the Indian Institute of Management, is engaged in the production of multiple products (viz., graduates in postgraduate programme in management, fellows of the Institute, trained management teachers through faculty development programme, trained executives through executive *PGPX* and short duration management development programmes, research in various fields, consulting services, institutional buildings, etc.), faces distinct demands for each of its products, has some costs which are divisible product-wise (cost of teaching material, temporary staff, brochure printing, etc.) but its overheads costs (e.g. cost of faculty, permanent staff, buildings, library stationery, electricity, etc.) are just inseparable by product. Thus, under joint products cases, while demands are separable, many significant costs are joint or common. Under such a situation, the determination of prices and outputs could be quite complicated. However, if the various products are produced in a fixed proportion, the determination of price and output would be similar to that under price discrimination. In addition to joint products, there are many firms today who have gone for vertical integration (Refer Chapter 6), where they sell and buy, partly or fully, their own products under different departments. They too set prices and outputs for each department, and such prices are called as the **Transfer Prices**.

In joint products, there is a demand function for each product just as there is a demand function for each segmented market for a product under price discrimination. Thus, there will be as many *AR* curves and *MR* curves as the number of products the firm is producing. From these, one could derive a combined *MR* (*CMR*) curve for all the products through a vertical summation (not horizontal, for the products, are different) of various *MR* curves. Under the fixed proportion of outputs case, the various products are assumed to be produced in a fixed proportion, outputs of various products could be added together in a well defined unit and then the total cost function for all the products could be deemed in terms of the standardized joint product. Given this, the unique *MC* curve could be obtained. The intersection of this *MC* curve with the *CMR* curve just explained, would determine the level of the standardized product. Given this and the fixed proportion of various products in the total output in standard term, the equilibrium output levels of each product could easily be inferred. The equilibrium prices of various products would then be given by their corresponding demand functions.

In the circumstances where the firm does not produce its various products in a fixed proportion, pricing could be achieved either through the allocation of common costs into various products through their individual shares in the firm's total activity and then proceeding just as in the case of a simple

monopoly with regard to pricing of each product or through some advanced method of pricing. The former method is a straight forward one and there is no point in taking the space here. The latter method is too complicated to be included in this text. Transfer pricing in case of vertically integrated firm is taken up in Chapter 10.

## 8.5 MONOPOLISTIC COMPETITION

The price-output determination under the two limiting market structures has been covered. The two special features of these two limiting cases are worth emphasizing:

(a) Neither a purely competitive firm nor a purely monopolist firm has any reason to differentiate its product from the other sellers through advertising or other means.

(b) Neither of the two limiting types of a firm needs to be concerned about the effects of rival firms' activities on its own operations or of the reactions of other firms to its own decisions.

The first feature is the result of the homogeneity of the product of all the firms under pure competition and the uniqueness of the product of the pure monopolist. The second feature is the outcome of the large number of firms, each having an insignificant share of the industry output under pure competition and the absence of any rival firm under pure monopoly.

In the world today, the above two extremes are rarely present. Most businesses today operate either in monopolistic competition or oligopoly markets. These two kinds have features of both the extreme kinds of markets. For example, monopolistic competition has monopoly in the sense that the product of each firm is differentiated from its rivals and at has the feature of perfect competition as the number of firms in the industry is large and no firm has perceptible market share. Also, in oligopoly, of standard oligopoly one, products of two or more firms are identical and thus they resemble perfect competition on that count. However, the entry/exit barriers are at times quite serious and the products of differentiated oligopolists could well be quite different, making the oligopoly market resembles monopolies on these counts. Between monopolistic competition and oligopoly there is one basic difference, viz. while firms in the former are largely independent as under perfect competition, they in the latter are highly mutually dependent. Thus, the distinctive feature of a monopolist competition is **product differentiation** while that of oligopoly is **mutual interdependence**. Incidentally note that product differentiation could be real or just pretension! This unique difference makes the price-output decisions in the two markets very distinct and accordingly they are discussed separately in the text.

Due to product differentiation and/or mutual interdependence, both the markets under discussion here are characterized by vigorous advertisement drives through the media (e.g. television, radio, newspapers, magazines, wall posters, loud speakers, etc.) and firms spend enormous amounts of money on this activity. These are designed to emphasize the qualities and prices of their products in relation to the competitors' products, so as to persuade the potential consumers to buy their own products. Also, there are a large number of products, which are available under different brands, in different qualities and at different prices. Consequently, there are close (though not perfect) substitutes for most products. Under monopolistic competition, the number of firms is large and thus their fortunes may not depend mutually, at least significantly. However, under oligopoly, the number of firms is small and or some firms enjoy significant market share, the fortunes of one seller does depend significantly on the actions and reactions of all the other sellers engaged in the production of similar products. Pricing decisions

under such situations are rather significant and a discussion of this must constitute an important part of a text on managerial economics. Under the present section, we take up the subject for monopolistically competitive markets and oligopoly pricing will be presented in Chapter 9.

As analysed in the previous chapter, there are a large number of goods and services where monopolistic competition prevails. To recall briefly, these include all small items traded through retail shops, such as gasoline, groceries, toiletries, aspirin, fast food restaurants, mouth wash, soap, shampoo, shaving cream, perfumes, hair-cuts, shoe-repairing, laundry services, general physician services, general legal services, etc. While there are a large number of barbers in a given town, the hair-cuts including the reception, etc. which a particular barber offers to his clients is not available in an identical form with all other barbers. The same is true with regard to all the goods and services traded in such a market. How a barber sets his product's price under such a situation? The particular barber determines the price of his services on the basis of the demand for and cost of his services, given his objective. The demand curve facing a monopolistically competitive firm slopes downward from left to right for two reasons, viz.

(a) Sellers' product (e.g. barber's hair cut, food at a fast food restaurant) is unique, and no one else can supply an identical product.

(b) Customers are often found to develop some attachment to the product of a particular firm.

We do witness people going to the same barber, the same fast food restaurant, the same physician, the same chartered accountants, the same teacher, the same advocate, the same laundry shop, the same grocery shop, the same mechanic for car repairing, the same brand of tooth paste, the same brand of cigarettes, the same brand of razor blades, the same brand of a ball pen, etc. year after year. This is so largely because differences among rival products are either unreal, insignificant or not worth the trouble of identifying an alternative brand/firm. Also, sometimes the seller happens to be either a relative or a friend, and human beings are known for their patronage. In view of this, when a monopolistically competitive firm raises its price, *ceteris paribus*, it does not lose all its buyers; while some are lost, some who are attached to it for any reason still shop with it. Similarly, when a firm under such a market structure lowers its product price, *ceteris paribus*, it does get new customers but the rival firms' do not lose all their customers. Therefore, the demand for the product of a monopolistically competitive firm is negatively price elastic. Further, the said curve is highly, though not perfectly, elastic. This is so because the products of competing firms are close but not the perfect substitutes for its products. When the price of a Pizza Hut pizza goes up, other fast foods' (like Subway's sandwich, McDonald's burger, Honest's pau-bhaji, dosa, etc.) prices remaining the same, many fast food consumers will switch over from pizzas to other foods, but some will still continue eating pizza. In effect, there would be a significant decrease in the sales of pizza. Quite the opposite would happen if Pizza Hut alone were to reduce the price of its product. In contrast to this, if electricity tariff goes up, *ceteris paribus*, electricity consumption would fall but this fall would be much less in relation to that in pizzas' consumption in the face of Pizza Hut's price increase to a similar extent. This is because there is no close substitute for electricity. The demand curve of a monopolistically competitive firm would thus be more elastic than that of a purely monopolist firm. However, it would be less elastic than that of a competitive firm, for if a competitive firm like a farmer producing rice, raises its price, other competitors' price remaining unchanged, he would lose all his customers, but as seen above, a monopolistically competitive firm does not lose all its customers under a similar situation.

So far as the cost function of a firm is concerned, there is no significant difference across different types of the structure in the product market. A firm irrespective of its product's market structure, hires

its factors of production from the factor market, which may have different degrees of competition. Thus, the cost function of a cigarette manufacturer would be similar to that of a farmer or that of an electricity company. As seen in Chapter 6, the usual shapes of the *AC* and *MC* curves would be close to the U-shaped.

Given the demand function, and the corresponding *AR* and *MR* curves, and the cost function, and the corresponding *AC* and *MC* curves, the price-output determination of a profit-maximizing monopolistically competitive firm could easily be explained. Figure 8.14 illustrates the same.

The firm in Fig. 8.14 (a) makes a profit equivalent to the area of rectangular $P_1ABC$, while that in Fig. 8.14 (b) just breaks-even. The firm could even run into losses. Incidentally note that these graphs are similar to Fig. 8.9, which illustrates pricing under simple monopoly. The only difference between these is in terms of the slopes (flatness) of the *AR* and *MR* curves. As just explained, these curves are flatter under monopolistic competition than under pure monopoly.

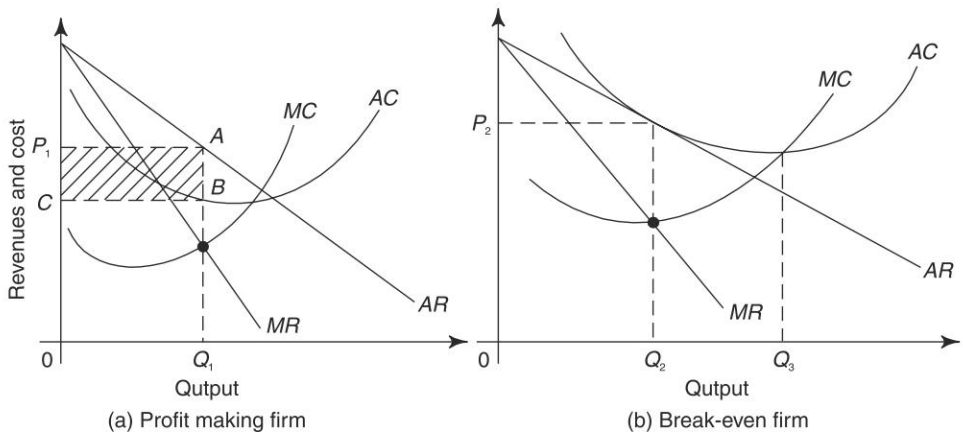



**Fig. 8.14** *Pricing under monopolistic competition*

In the short-run, a firm under monopolistic competition could reap economic profits, just breaks-even or runs into losses. However, as is true in all other market structures, the losses would never exceed total fixed cost. It would necessarily break-even in the long-run. This is because if it makes a profit, other firms would enter the industry, thereby they would impinge on this firm's market, which, in turn, would lower its demand curves and/or raise its cost curves. This would continue until all profits are wiped-out. Similarly if the firm was running into losses, it would get out of the industry in the long-run to avoid losses. Thus, the easy entry-easy exit feature of monopolistic competition, firms under such a market necessarily break-even in long run. Thus, the long-run equilibrium of such a firm is depicted by Fig. 8.14(b), where the *AC* curve is tangent to the falling demand curve. We may now take notes of certain significant implications of such equilibrium.

(a) A monopolistically competitive firm always produces **socially sub-optimal output** both in the short as well as in long-run. This is evident from the definition of socially optimal output (where $P = MC$) and the nature of the *AR* and *MR* curves (Chapter 6, section 6.6).

(b) A monopolistically competitive firm **never operates at its minimum *AC*** (or always has an excess capacity) in the long-run. This is obvious from the firm's long-run equilibrium position as presented above in Fig. 8.14(b). The *AC* curve is U-shaped, *AR* curve is falling, *MC* curve

intersects the *AC* curve at the latter's minimum point, and the firm merely breaks-even in the long-run. For break-even, *AC* must equal *P* (= *AR*), since the *AR* curve is falling, the *AC* curve must be tangent to the *AR* curve at the point of equilibrium when *AC* is falling, and since falling *AC* precedes the minimum *AC*, the equilibrium output must be less than the output at which *AC* is minimum. In Fig. 8.14 (b), while $OQ_2$ is the long-run equilibrium output, $OQ_3$, is the output at which *AC* is the least, and $OQ_2$ is less than $OQ_3$. The full capacity output (the level at which the fixed inputs are optimally used), by definition, is the one at which *AC* is the minimum. Thus $OQ_3$ is also the full capacity output. Since $OQ_2$ denotes the equilibrium output, $Q_2Q_3$ measures the excess capacity, which is always positive, as just explained.

(c) As noted above, this type of market structure would indicate that the market is characterised by **large varieties** of a product, all of which are close substitutes To cite some striking examples, we have icecreams with varying flavours, coke in diet and other varieties, milk in varying fat levels, MBA in regular two year vs. one year executive MBA, and so on. This puts the customers in a quandary as to what to buy. The monopolistically competitive firms, through their ingenious advertisements, try to impress on the customers that their product is superior to all other similar products with regard to the quality vis-a-vis the price. Recall that the difference among rivals' products may be real or just imaginary. Further, no consumer can afford to examine all the alternative varieties (brands) and he/she is often confused as to which one to buy. Quite often consumers judge the quality by price and in the process they get cheated by clever salesmen. Thus, this kind of a market structure leads to a lot of expenditure on advertisements, fancy packaging, etc. (which increases cost and hence price), and confusion to customers. Yet such a market structure prevails in many products' markets; for competition is a way of life. It encourages efficiency and innovations of new products. In such markets, various **firms compete less on price and more through non-price factors, like location, hours of operation, advertisements, packaging,** etc. Thus, one finds retail banks, grocery stores, petrol pumps, laundries, barbers, physicians, restaurants, etc. having different work hours, different locations, and so on. Incidentally note that monopolistic competition leads to avoidable expenditures and sometime unhealthy competition, and accordingly it is avoided in case of those products which are prone to high degree of wastage like plants and machinery, automobiles, trucks, railways, telephones, electricity, kitchen appliances, etc.

(d) Under this kind of market structure, the fortune of a firm does depend upon the actions and reactions of rival firms, but their **interdependence is not very significant**. Thus, the business of a general physician does depend upon the fees and qualities of services rendered by other physicians in the markets, but since each physician is an insignificant part of all physicians in the market, their interdependence is of little substance. In practice, firms under monopolistically competitive market hardly bother about such minor interdependence. Figures 8.14(a) and 8.14(b) are drawn ignoring firms' interdependence. Nevertheless, the prices of the differentiated versions of the product do tend to move together. This is partly because the demand and cost functions of rival firms are similar, and partly because the competition among them ensures a fair price in relation to the product quality.

To explain the price-output decisions through algebra, we consider a numerical illustration. We do so by making the required modifications to the example given for perfect competition and monopoly cases. Recall that the basic difference between monopoly and monopolistic competition is in terms of the slope (or price elasticity) of the demand curve, which is steeper (less price elastic) in the former than in

the latter kind of market. Accordingly, we merely modify the equation of the monopolist demand curve in the example cited above. Under monopoly the demand curve facing the firm (which is the also the demand curve facing the industry under perfect competition, viz. Eq. 8.1) was denoted by

$$Q = 150 - 2P \qquad (8.1)$$

We now change it to

$$Q = 150 - 3P \qquad (8.10)$$

Equation (8.10) is our demand equation facing the monopolistically competitive firm. It can be verified that the price elasticity of demand is higher in equation (8.10) than in Eq. (8.1). Under Eq. (8.10), the said elasticity equal $(-3)(P/Q)$, which exceeds (in absolute value terms) the elasticity in Eq. 8.1, which equals $(-2)(P/Q)$. The cost equation under the two market structure need not be different and thus we assume that it is same as the one under monopoly, viz.

$$TC = 10 + 2Q + 0.5Q^2 \quad (8.3)$$

To find the profit maximizing price-output for the monopolistically competitive form, we set $MR = MC$. To get $MR$, we must invert the demand function, giving price in terms of output. Thus,

Equation (8.10) gives

$$P = 50 - 0.333Q, \quad \text{and} \quad TR = 50Q - 0.333Q_2$$

and $$MR = 50 - 0.67Q$$

From $TC$ equation, we can get $MC$ and then setting $MR = MC$ gives

$$50 - 0.67Q = 2 + Q$$

or, $Q = 28.7$, Substitution of this value of $Q$ in demand equation yields

$$P = 40.3$$

The above solution would give

$TR = PQ = 1159.5$, $TC = 10 + 2(28.7) + 0.5(28.7)^2 = 891.1$ and profit = 268.4

Thus, the firm's optimum output = 28.7, optimum price = 40.3 and it makes profit = 268.4 in short run. In comparison to monopolist, the monopolistically competitive firms' price is lower, profit is lower and the output is lower, as well. The first two are lower due to the competition this firm faces which is absent for the monopolist, and the last (i.e., output) is lower as there are many firms in this market compared to a single firm in monopoly.

Three additional points are worth taking up here.

(a) As in monopoly, the monopolistically competitive firm has the **option** to set its output or price but not the both (as it faces a given consumers demand function).

(b) The above results throw some inputs useful to foresee some future scenario. Since the firm in our example makes a positive profit in short-run, which must vanish in long run, it leaves incentives for other firms to enter the industry. Accordingly, the industry under question is likely to see new firms' entries.

(c) Since $P > MC$, firms in monopolist competition enjoy some **market power**. It is derived largely through the falling demand curve. Its degree could be measured by Lerner index. For example, $P = 40.4$ and $MC = 2 + Q = 30.7$, and thus

$$\text{Lerner index} = (40.4 - 30.7)/\ 40.4 = 0.24$$

Which, as expected, is lower than that (0.3216) under monopoly.

The effects of governments' regulations in such markets can be analysed on the lines we did for other kinds of markets above. This completes the discussion on the three kinds of market structures presented in this chapter. The next two chapters would cover pricing in oligopoly and the pricing strategies practiced by firms enjoying some market power.

## REFERENCES

1. Baumol, W.J. (1982): *Economic Theory and Operations Analysis*, 4th edition, New Delhi, Prentice-Hall.
2. Baumol, J.C., Panzar J.C. and Robert D. Wilig (1982): Contestable Markets and the Theory of Industrial Structure, Amazon.
3. Chamberlin, E.H. (1962): *The Theory of Monopolistic Competition*, 8th edition, Cambridge, Harvard University Press.
4. Cohen, K.J. and R.M. Cyert (1965): *Theory of the Firm*, Englewood Cliffs, Prentice-Hall.
5. Dean, J. (1976): *Managerial Economics*, New Delhi, Prentice-Hall.
6. Henderson, J.M. and R.E. Quant (1980): *Microeconomic Theory* : *A Mathematical Approach*, 3rd edition, New York, McGraw-Hill.
7. Marshall, A. (1920): *Principle of Economics*, 8th edition, London, Macmillan.
8. Pindyck, Robert S., Daniel L. Rubinfeld (2009): *Microeconomics*, 7th edition, Pearson.
9. Robinson, J. (1933): *The Economics of Imperfect Competition,* London, Macmillan
10. Samuelson, P.A. (1948): *Foundations of Economic Analysis*, Cambridge, Harvard University Press.
11. Stigler, G.J. (1966): *The Theory of Price*, 3rd edition, New York, Macmillan.

## CASELETS

1. One of the items of the Jagdish Farm Products (JFP) is sold in a purely competitive market. Its marketing manager, Vishnu Bansal, has estimated the equations of the demand and supply curves of this product in the country as well as the cost equation of the JFP for the production of this item. The same are given below:

   Industry Demand Curve: $Q = 360 - 2P$

   Industry Supply Curve: $Q = 10 + 3P$

   JFP's Cost equation $TC_1 = 200 + 50\ q_1 - 6\ q_1^2 + \frac{1}{3} q_1^3$

   where, $Q$ = Quantity demanded and supplied

$P$ = Price

$TC_1$ = Total cost of JFP

$q_1$ = Output of JFP

(a) Determine the equilibrium price and quantity of the product

(b) Determine the profit-maximizing output of JFP. Does JFP make any profit in the short-run? If yes, how much?

(c) Suppose the JFP succeeds in monopolizing the market without hampering the demand for the product, i.e., all the competitors leave the industry while the industry demand curve remains the same. Determine the new equilibrium price and quantity and the profit of JFP.

(d) Compare your results under the situations of pure competition and pure monopoly and comment.

2. A farm product has the following demand and supply functions:

Demand: $Q = 360 - 2P$

Supply: $Q = 10 + 3P$

where, $Q$ = Quantity demanded and supplied

$P$ = Price

(a) Determine the equilibrium price and quantity under each of the following situations:
   (i) Government imposes a specific sales tax at the rate of Rs. 10 per unit
   (ii) Government imposes a lump sum tax in the amount of Rs. 1000.

(b) Compare your above answers with that of Caselet 1 above and comment. How is the burden of specific sales tax shared by the sellers and buyers? Is it proportional to the reciprocal of their price elasticities? Explain.

(c) Compute the losses to consumer and producers surpluses, tax revenue, and dead weight loss under the specific tax of (a) (i) above.

3. A monopolist uses one input, which he purchased at the fixed price, $P_L = 5$, to produce his output. His demand and production functions, respectively are

$P = 85 - 3\,Q$ and $Q = 2\,L^{0.5}$

Determine the values of $P$, $Q$ and $L$ at which the monopolist maximizes his profit.

4. Radhe Cellular Company (RCC) sells its product to a set of homogeneous retail customers. Each customers' demand for cell phone minutes was estimated as follows:

$$P = 20 - 0.2\,Q$$

RCC's cost function was

$$TC = 1.0\,Q$$

where, $P$ = price per minute, $Q$ = quantity in minutes sold per customer and $TC$ = total cost

(a) Suppose there is perfect competition in the cellular business. If RCC aims at profit maximization, determine its optimum output.

(b) If the cellular business was monopolized by RCC, what output level the company must generate and at what price it must sell the product?

(c) Compare the above two results and comment.

5. The Bharat Power Corporation (BPC) has the monopoly in power supply in its area of operation. Its estimated cost and demand functions are the following:

$TC = 50 + 19\,Q - 0.25\,Q^2$

$P = 75 - 4\,Q$

Notations have the usual meanings.

(a) Do the above data suggest that BPC could be enjoying natural monopoly in the market? How?

(b) Determine the profit maximizing output and price for BPC. How much profit the company makes?

(c) If the government forces BPC to operate at the socially optimum output, how much the company would produce and what price it would charge from the customers? Would the company make profit or loss under the situation, and how much would that amount be? Should the company end up in loss, what would you suggest for its long run survival?

6. Ashok Patel is the city planner in a medium sized city. The city is considering a proposal to award an exclusive contract to ABC Inc., a cable television carrier. Mr. Patel has discovered that an economic planner hired a year before has generated the following demand and total cost functions for cable:

$P = 28 - 0.0008\,Q$

$TC = 120{,}000 + 0.0006\,Q^2$

where, $P$ = monthly price of cable service, $Q$ = number of cable subscribers and $TC$ = total cost

If you were the consultant to Mr. Patel, what would be your answers to his following questions.

(a) If the government were to leave the cable service business unregulated, what price and quantity could be expected to prevail in the market?

(b) If the cable service was to be regulated such that it is socially efficient, what price and quantity it would generate?

(c) Compare the economic efficiency implications of (a) and (b) above. Your answer should include numerical calculations as well as the relevant diagrams to demonstrate dead weight loss under (a) vs. (b).

7. Mohan Brothers has the estimated the following cost schedule for its product:

| Output | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| TFC | 300 | 300 | 300 | 300 | 300 | 300 | 300 | 300 | 300 |
| TC | 300 | 400 | 450 | 510 | 590 | 700 | 840 | 1020 | 1250 |

Mohan Brothers is a profit maximizing enterprise. Based on these data, attempt to answer the following questions to Mohan Brothers:

(a) If the firm faced perfect competition and the price happened to be 50, what quantity the firm must produce? At price = 70? At price =100?

(b) If the firm had monopoly and the demand was zero at $P = 150$, 1 unit at $P = 140$, 2 units at $P = 130$, and falling in this pattern (i.e., demand = 8 at $P = 70$), what price and output the firm would operate at?

(c) If the government had imposed a specific tax @ 10 per unit of output, what would be the equilibrium output and price under each of perfect competition and monopoly situations?

(d) If the tax was *ad valorem* @ 10 per cent of total revenue, determine the equilibrium output-price combinations in each of the two kinds of market structure.

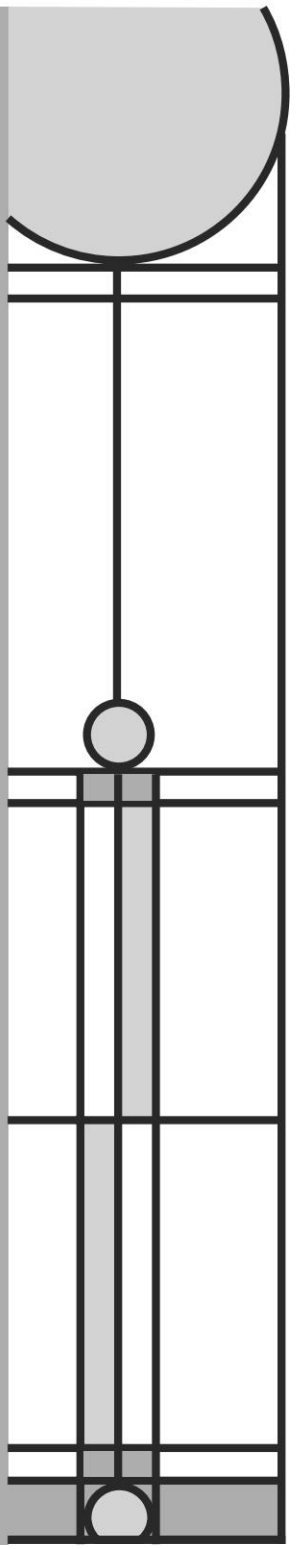
8. Murari Entriprises Ltd. (MEL) runs a business in Happyland. It has three plants at Talgarh, Talstreet and Talgaun. The demand facing the firm and total cost functions at each of the three plants have been estimated as follows:

$P = 2500 - 5\,Q$ $\quad TC_1 = 100 + 4\,Q_1 + 2\,Q_1^2$

$TC_2 = 200 + 3\,Q_2 + 2\,Q_2^2$ and $TC_3 = 250 + 4\,Q_3 + 0.5\,Q_3^2$

If you were the chief economist at the firm, what price you will advice the firm to set and how much sales MEL is likely to have in each market? Would the firm end up with profits at each plant? How much?

The McGraw-Hill Companies



CHAPTER

9

# PRICE–OUTPUT DETERMINATION-II
## (PRICING UNDER OLIGOPOLY)

Recall that an oligopoly market is characterized by the following features:

(a) A few sellers or a few 'big' seller if there are many sellers
(b) Standard or differentiated product, and
(c) Difficult but possible entry/exit

Thus, unlike the other three market structures dealt with in the previous chapter, the oligopoly market is somewhat ambiguous. It could have a large number of sellers as under pure and monopolistic competitions, but the distinguishing factor is that some of them must be dominant sellers; the term 'dominant' or 'big' is again not defined in terms of the market share unambiguously. The product of rival firms could be identical as under pure competition or it could be differentiated as under monopolistic competition. The entry or exist is possible but difficult, where the term 'difficult' is not specified exactly in terms of any measurable variable. Thus, oligopoly is a market which contains the features of the most other market structures.

Although the oligopoly market is not so clearly defined, it is found to prevail for sure in the case of many products today. Thus, we have the examples of homogeneous or standard oligopoly markets in products like cement, steel, aluminum, copper, sugar, petroleum, etc. In each of these commodities, the number of producers is either small, or if large, they have a few 'big' sellers, and the products of all the firms in each of these goods is homogeneous or very nearly so. For example, the number of firms in cement production is by no means 'a few' but the Associated Cement Companies (ACC) commands a very respectable share in the total supply of cement in India to qualify for the 'big' firm. The heterogeneous oligopoly is found to exist in many products like aircrafts, automobiles, two wheelers, trucks, kitchen appliances, air conditioners, televisions, computers, mobile phones, cable service providers, soft drinks, tennis balls, detergents, cigarettes, pharmaceuticals, and in specialized services, such as management education, management consulting, cancer specialists, heart specialists, five star hotels, etc. In each of these products, the number of firms is limited; one, two or a few of the firms are 'big, and the products of

various (rival) firms are similar but not identical. It may be mentioned here that the distinction between differentiated oligopoly and monopolistic competition is more in degree (number of firms—small vs. large, size of firms–significant vs. insignificant of firms, and entry/exit conditions—hard vs. easy) and less in kind; and hence, the exact classification of an industry into one of the two types is sometimes difficult to make.

An important implication of the characteristics of such a market is that various firms in a given oligopoly industry are **significantly mutually interdependent**, which happens to be the **unique feature** of oligopoly industry. That is, the actions of a firm affect the fortunes of other firms in the industry directly, immediately and adversely, and to a significant extent. Thus, if Kelvinator reduces the price of its refrigerators, it would immediately and significantly hamper the sales of Whirlpool, L.G., and the other brands of refrigerators. Similarly, if Maruti Udyog cuts the prices of its cars, the markets for Tata Motors, Hyundai cars, Honda cars, Mahindra and Mahindra jeeps, etc., would shrink immediately and to a large extent. Further, it is not just the price of rivals' which would affect the demand for a firm's product, but even the rivals' advertisement budgets, the quality and the style of the rivals' products would have negative repercussions on the demand for a firm's product. This is why the firms operating in an oligopoly industry are referred to as "**rivals**" to each other firm. In the circumstances, the demand function for a firm's product would have the following determining variables, in addition to the ones explained in Chapters 2–4.

(a) Prices of rival firms' goods
(b) Advertisement budgets of rival firms
(c) Styles and qualities of rival firms' goods

Since rivals' products are close substitutes, prices of rivals' goods would have positive (same direction) effects while the advertisement budgets would have the negative (opposite direction) effects on the demands for a firm's product. To make it more explicit, let us take an example. The demand for, say, Coke, would fall as the price of Pepsi falls, *ceteris paribus*, and *vice versa*. Also, the demand for Coke will fall as the advertisement budget of Pepsi goes up, *ceteris paribus*, and *vice versa*.

Since rivals' actions affect one's fortune rather significantly, the oligopoly firms have no alternative but to react to the actions of their rivals. Thus, if Maruti Udyog reduces the price of Maruti car, prices of Hyundai cars and other cars would be cut almost immediately. On such an event, Maruti Udyog might get upset and it, in turn, might react by further lowering the price of its own car, which may again be reacted by other car companies in a similar fashion, and the process might go on unabated until everyone's profits vanish. The interesting question to answer is, how are the prices of products sold in oligopoly markets determined in such a situation? Also, it is said that prices in oligopoly markets are relatively inflexible/rigid. In this chapter, we shall attempt to answer both these questions. While attempting these questions, we would continue the standard assumptions of firms' seeking **maximum profit, law of one price, no asymmetric information, no externalities, no risk and uncertainty**.

## 9.1 PRICE-OUTPUT DETERMINATION MODELS IN OLIGOPOLY MARKETS

It is said that oligopoly prices are often **indeterminate**. This is true in the sense that the profit-maximizing price, or even the one consistent with any other objective of the firm, is not determined

unless the reactions of rival firms' on the actions of a firm (called the **reaction functions**) are known. This is obvious, for the demand for a firms' product in such a market depends on the prices and quality of its rival firms' products, and rivals' advertisements, besides other standard factors. The demand function thus occupies a rather critical role in the price determination mechanism in oligopoly markets. Several theories of oligopoly prices have been developed and each one of them is based on a particular assumption about the reactions of rival firms on a firm's actions. The popular models under appropriate classification are described below:

(i) Perfect Collusion (Cartel) model

(ii) Competition Oriented models
   (a) Cournot model
   (b) Bertrand model

(iii) Tacit Collusion (Leader-Follower) models
   (a) Dominant Firm (Residual Market) model
   (b) Fixed Market share model
   (c) Stackelberg models
      - Stackelberg-Cournot model
      - Stackelberg-Bertrand model

(iv) Game theory

(v) Non-optimizing models
   (a) Cost-Based models
   (b) Going Rate model
   (c) Price penetration and price skimming model

To give a brief on these alternative methods of pricing under oligopoly, their distinguishing factors may be kept in mind. The so called standard models, which include the Cartel, Cournot, Bertrand, Dominant Firm, Fixed Market Share and Stackelberg models, assume the availability of all pertinent information about the actions and reactions of all firms in the industry. Accordingly, these models assume that the demand, cost functions and unique strategy which each firm would choose in a given situation are known to every firm in the industry. Thus, in these models, equilibrium (optimum) prices and outputs under oligopoly are uniquely determined, though different models give different results. In contrast, the Game theory approach assumes that the demand and cost functions are known, the alternative strategies available to each and every firm are known, but what strategy a firm would choose in a given situation is not known to other (rival) firms. While each of the standard models give unique results for oligopoly price-output decisions, Game theory may or may not even yield conclusive outcomes. Accordingly, optimum price-output may be determined or may remain indeterminate. Under the situations where the above mentioned data are inadequate, neither of the just mentioned two groups of models is available and then the optimum price is in-determinate. Under the last eventuality, and the situations where Game theory leads to inconclusive outcomes, firms price their products on some non-optimizing ways, including the cost-based pricing, going rate pricing, etc. A detailed perusal of all these pricing strategies follows.

## 9.2 PERFECT (EXPLICIT) COLLUSION (CARTEL) MODEL

Under the cartel (**perfect or explicit collusion**) model of oligopoly, all the firms in an industry recognize their interdependence and decide to collaborate in the form of a cartel in the matter of pricing their products. This is what the Organization of Petroleum Exporting Countries (OPEC) has been trying in the matter of petroleum pricing since 1974–75. If the product of various firms in the industry is **homogeneous (standard)**, as is the case of petroleum, the pricing would be similar to that under pure monopoly with multi-plant operations (vide Chapter 8, Section 8.5). There will be one market demand function (*AR* and *MR*) and as many cost functions (*AC*s and *MC*s) as the number of competing firms in the industry. The various *MC* curves could then be summed up horizontally to get the combined *MC* curve (*CMC*). The joint profit-maximizing output for the industry as a whole would then be given by the intersection of the *CMC* and *MR* curve. Given the equilibrium level of industry output, the industry *AR* curve would give the equilibrium price. The distribution of industry output among firms would be obtained by equating *MR* to each of the *MC* through a horizontal straight line passing through the point of intersection between *CMC* and *MR* curves. The procedure is described in Fig. 9.1 for the case of an industry with three firms.

In Figure 9.1, the industry output and price are determined in part (d), where *CMC* was obtained as the horizontal summation of $MC_1$, $MC_2$ and $MC_3$ (e.g. $oq_1 + oq_2 + oq_3 = oQ$), and *AR* and *MR* represent the demand and corresponding marginal revenue curves of the industry as a whole. The industry output *OQ* is divided among the three firms as $oq_1$ to firm 1, $oq_2$ to firm 2 and $oq_3$ to firm 3, the sum of these three outputs equal industry output *OQ* by construction. Each of the firm sells its product at a uniform price equal to *OP*, which is determined in part (*d*)*.

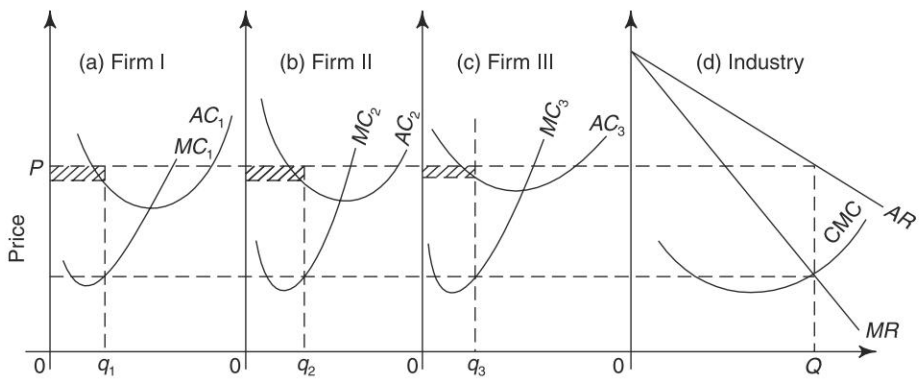



**Fig. 9.1** *Standard (or Homogeneous) oligopoly pricing under a cartel*

* Pricing under homogeneous oligopoly could easily be explained through calculus-algebra. Given the industry demand function and the various cost functions, one of each firm in the industry, one could find the industry profit function in terms of output levels of all firms. Maximization of this profit with respect to each firm's output would yield the equilibrium outputs for various firms. The sum of individual firm's outputs would yield the industry output, which when plugged into the industry demand function would yield the price of the product, which each firm would charge. Alternatively, it could be done through the process as explained in Chapter 8, Section 8.5.

It may be noted that under this method of pricing, while all firms in an industry sell at the same price, their outputs and profits will not be equal. This is because their cost functions will be different as the firms have different factors of production. Consequently, the cost efficient firms will make more profits than others. But some firms might even loose under this system of pricing. The difference in their profits would threaten the continuity of the cartel. In order to avoid the problem of break-down of the cartel, the more efficient firms undertake to transfer a part of their profits to the less efficient firms as a clause of the terms of the cartel.

For a numerical example, we can go back to the example cited in Chapter 8, Section 8.5 on pricing under multi-plants monopolist. The demand curve remains the industry demand curve and the plant-wise cost functions now become the firm-wise cost functions. Under the said example, industry output comes to 93.75 units, industry price = 312.5, outputs of firms = 31.25 and 62.5, and firms' profits = 7,787 and 15,605, respectively. Firm-2 is more efficient (in terms of cost) and thus sells more and makes larger profits than Firm-1. To cite another example, to make a new point (and prove the assertion made in Chapter 8, Section 8.5), consider a two-firm standard oligopoly industry, having the following demand and cost functions:

$$P = 50 - 5\,Q \qquad TC_1 = 20 + 10\,Q_1 \qquad \text{and} \quad TC_2 = 10 + 12\,Q_2$$

Given the above, $TR = PQ = 50\,Q - 5\,Q^2$ $\quad MC_1 = 10$ and $MC_2 = 12$

Recall the necessary conditions for profit maximization under the two-plant monopolist, which apply equally to the two-firm oligopoly cartel (vide Chapter 8, Section 8.5), viz.

$$MR = MC_1 = MC_2$$

Applying these, we get

$$50 - 10\,Q = 10 \quad \text{and} \quad 50 - 10\,Q = 12$$

Since the left hand sides of the above two equations are same and the right hand sides have different values, these equations are inconsistent and give no solution. Thus, note that since *MC*s of the two firms are constants, and different, there is no solution under a cartel. Even if the two *MC*s were equal, there would be no unique solution. Thus, as stated earlier in Chapter 8, Section 8.5, for cartel (or multi-plant monopolist) to yield a solution, *MC*s of various firms (plants) should be different and not constants. Coming to the basic issue in the above hypothetical example, where, *MC*s of two firms under cartel are constants and different, how the equilibrium price-output mix is obtained? The answer is easy to see. Firm-1 has a lower *MC* than Firm-2, and accordingly, a cartel between the two would mean Firm-1 alone supplies the whole market and Firm-2 just remains shut. Under such an event, industry output would be given by

$$MR = MC_1, \quad \text{i.e.,} \quad 50 - 10\,Q = 10, \qquad \text{or,} \qquad Q = 4.$$

The price would then equal $50 - (5)(4) = 30$. Industry profit $= PQ - TC_1 - TC_2 = (30)(4) - [20 + (10)(4)] - 10 = 50$. Profit of Firm-1 = 60 and of Firm-2 equals = –10. Accordingly, to save the cartel, Firm-1 must compensate Firm-2 not only against its loss but also to reward it with some profit. Alternatively, the two firms can split the output through some agreement and resolve the dilemma. For example, if they split equally, each would produce 2 units and make profits as $\pi_1 = 20$ and $\pi_2 = 26$.

The cartel system of pricing can prevail even under **differentiated oligopoly**. But the odds in its continuity are stronger than those in homogeneous oligopoly, for the product quality apparently varies

between rivals and the differences among them are at times vague. The geometry under this case would be complicated and hence it is illustrated through a numerical example, in what follows. For the sake of simplicity, assume that there are just **two firms in the industry** (a two firm oligopoly market it called as ***duopoly***), and that the demand functions for their differentiated products and their cost functions are as follows:

| *Duopolist* | *Demand Functions* | | *Cost Functions* | |
|---|---|---|---|---|
| 1 | $P_1 = 100 - 2Q_1 - Q_2$ | (9.1) | $TC_1 = 2.5\ Q_1^2$ | (9.3) |
| 2 | $P_2 = 95 - Q_1 - 3Q_2$ | (9.2) | $TC_2 = 25Q_2$ | (9.4) |

where $P_1$ and $P_2$ are the prices, $Q_1$ and $Q_2$ are the outputs, and $TC_1$ and $TC_2$ are total costs of duopolists 1 and 2, respectively. Incidentally note that the above demand functions do hypothesize interdependence of the duopolists as $Q_2$ appears in the demand function facing duopolist 1 and $Q_1$ in the demand function facing duopolist 2. There is no such dependence from the cost side. Recall that, we have argued above that demands of each oligopolist's product depends (adversely) on prices and advertisement budgets of rival firms', among other determinants. In the above demand functions, advertisement budgets are ignored for simplicity, the constant terms are assumed to contain the effects of all but price variables, and the functions are presented in their inverse forms. We must emphasize that the case of duopoly is considered for all numerical examples here not just because it is simpler to handle but also because there are several industries in the real life where duopoly prevails. To cite some examples, we have Airbus and Boeing in aircrafts industry, Coke and Pepsi in soft drinks, Xerox and Canon in copying machines, Herschey and Mars in candies, Kodak and Fuji in films.

If the two duopolists are able to get together under the current competition policy and form a cartel, they would determine their quantities (or prices) in such a way that their joint profit is at maximum with respect to the output levels (prices) of each of them. The total profit ($\pi$) function for the above hypothetical example would be

$$\pi = (TR_1 + TR_2) - (TC_1 + TC_2)$$
$$= P_1 Q_1 + P_2 Q_2 - TC_1 - TC_2$$

Substituting the values for $P_1$ and $P_2$ from the above demand functions and of $TC_1$ and $TC_2$ from the above cost functions, we have

$$\pi = 100\ Q_1 - 2\ Q_1^2 - Q_1\ Q_2 + 95\ Q_2 - Q_1\ Q_2 - 3\ Q_2^2 - 2.5\ Q_1^2 - 25\ Q_2$$

or,
$$\pi = 100\ Q_1 + 70\ Q_2 - 4.5\ Q_1^2 - 3\ Q_2^2 - 2Q_1\ Q_2 \qquad (9.5)$$

The necessary conditions for profit maximization with respect to $Q_1$ and $Q_2$ require that each of the two partial derivatives be zero:

$$\frac{\partial \pi}{\partial Q_1} = 100 - 9\ Q_1 - 2\ Q_2 = 0;$$

or,
$$9\ Q_1 + 2Q_2 = 100$$

and,
$$\frac{\partial \pi}{\partial Q_2} = 70 - 6\ Q_2 - 2\ Q_1 = 0;$$

or,
$$2\ Q_1 + 6\ Q_2 = 70$$

The solution of these two equation yields

$$Q_1 = 9.2 \text{ and } Q_2 = 8.6$$

It can easily be verified that this solution meets with the second order conditions for profit-maximization. Substitution of these values into the two demand functions yields

$$P_1 = 73 \text{ and } P_2 = 60$$

The profit of each of the two duopolists are given by

$$\pi_1 = P_1 Q_1 - TC_1$$

$$\pi_2 = P_2 Q_2 - TC_2$$

Substitution of the equilibrium values for $P_1$, $P_2$, $Q_1$ and $Q_2$, and of $TC_1$ and $TC_2$ from the two cost functions would give

$$\pi_1 = 460 \text{ and } \pi_2 = 301.$$

In this case, both the firms are making profits and therefore chances of their breaking-off the Cartel are less. However, since the two firms do not have uniform profits, it is conceivable that the (a) price-differences between the two products may be more or less than their quality differences, and (b) two firms may have more or less differences in terms of their relative social and political powers than reflected in their activities at the above equilibrium, etc. Under such an eventuality, there is still a possibility of threat to the cartel arrangements. To the extent the cartel could survive, the above would be the optimum solution for oligopolists.

The advantages of perfect collusion are obvious. It avoids price wars among rivals and thereby each firm gains at the cost of the consumers. Further, since they operate like a pure monopolist, their joint profit under perfect collusion can never be less than, and would generally be more than, the total profit all of them together would make if they had acted independently or through any partial collusion. The profit of an individual oligopolist as obtained through the above solution procedure could, of course, be less than it could make under no collusion. However, if so, the oligopolists who make extra profits due to the cartel would compensate the oligopolists (through a simple transfer) who suffer due to the cartel, lest the cartel agreement be broken.

The difficulties which such a cartel faces are enormous.

(a) The consumers are adversely affected by it in the form of higher prices and restricted quantities, so they would raise their voice against it to the extent possible.

(b) If the number of firms under oligopoly is large, the chances of arriving at a common understanding would be difficult. The more the number, the more difficult it is to form a cartel.

(c) Cartels are more difficult in the case of differentiated oligopoly than in the case of homogeneous oligopoly, for the differences in products' qualities would be difficult to rationalize and sort out.

(d) In the world of rare honesty, oligopolists would be attracted to enter into secret dealings in order to raise their own profits and in the process they would tend to cheat and thereby threaten the continuity of the cartel.

(e) During the period of recession or bumper crops, all the oligopolists would suffer and a suffering unit can hardly afford to be honest.

In view of these difficulties, cartels are rare today and, in fact, they are regulated under anti-trust laws, the Monopolies and Restrictive Trade Practices (MRTP), Competition Policy, etc. In any case, the feasibility of a cartel in an industry depends negatively on the number of firms in the industry; the degree of product differentiation, as well as size differences across rivals firms; and positively on the quality of the past histories/records of rival firms; practice of the law of one price or otherwise, etc.

## 9.3 COMPETITION ORIENTED MODELS

Since cartels are difficult to enforce, a number of alternative models for pricing under oligopoly have been developed. Some of these alternative models presume a particular type of competition among firms in the industry, while some others postulate some kind of tacit (implicit) collusion among them. We have two models in this group, viz. Cournot and Bertrand models. While the former postulates rival firms to compete on outputs, the latter takes them to compete on price. The two models are now analysed.

**Cournot Model**: The Cournot model happens to be the one which takes **firms to compete among themselves through their output levels**. Augustin Cournot, an early 19th century French economist, developed the said model for price-output determination for oligopoly markets in 1838 under the assumption that the oligopolists compete through outputs by ignoring their interdependence on the changes in outputs of their rivals. In other words, the competing firms assume

$$dQ_i / dQ_j = dQ_j / dQ_i = 0 \tag{9.6}$$

where $Q_i$ is the output of the $i^{th}$ oligopolist and $Q_j$ is the output of the $j^{th}$ oligopolist. This means that when oligopolist $i$ changes its output, oligopolist $j$ does not change its output, and vice versa. On this assumption, each oligopolist endeavours to maximize its own profit with respect to its own output and thereby the oligopoly prices and outputs get uniquely determined. Note that, as in monopoly and monopolistic competition's markets, the firms under oligopoly can determine either the prices or outputs but not the both, for they also face the given demand functions for their products. The model applies to both the standard and differentiated oligopoly markets. Since standard oligopoly is a special case of differentiated oligopoly (where the products are identical and so $Q_1 + Q_2 = Q$ and $P_1 = P_2 = P$), and the latter is more common than the former, we illustrate the application of the Cournot model through an example of differentiated oligopoly. For the purpose, we take the above example of two firms, so that the results under alternative models could be compared as well. For the said example, the duopolist *I*'s total profit ($\pi$) function is given by

$$\pi_1 = TR_1 - TC_1 = P_1Q_1 - TC_1$$

$$= 100\, Q_1 - 2\, Q_1^2 - Q_1Q_2 - 2.5\, Q_1^2$$

or,
$$\pi_1 = 100\, Q_1 - 4.5\, Q_1^2 - Q_1Q_2 \tag{9.7}$$

Maximization of $\pi_1$ with respect to $Q_1$ requires

$$\frac{d\pi_1}{dQ_1} = 100 - 9Q_1 - Q_2 - Q_1 \left[\frac{dQ_2}{dQ_1}\right] = 0$$

But $dQ_2/dQ_1 = 0$ by Cournot assumption (9.6) above. Thus,

$$100 - 9Q_1 - Q_2 = 0$$

or, $$9Q_1 + Q_2 = 100 \tag{9.8}$$

Duopolist 2's total profit function is given by

$$\pi_2 = TR_2 - TC_2 = P_2 Q_2 - TC_2$$

$$= 95\ Q_2 - Q_1 Q_2 - 3\ Q^2_2 - 25\ Q_2$$

or, $$\pi_2 = 70Q_2 - Q_1 Q_2 - 3\ Q_2^2 \tag{9.9}$$

Oligopolist would maximize its profit with respect to its own output, which requires

$$\frac{d\pi_2}{dQ_2} = 70 - Q_1 - Q_2 \frac{dQ_1}{dQ_2} - 6\ Q_2 = 0$$

Again by the Cournot assumption, $\frac{dQ_1}{dQ_2} = 0$, so

$$70 - Q_1 - 6\ Q_2 = 0$$

or, $$Q_1 + 6Q_2 = 70 \tag{9.10}$$

Equations (9.8) and (9.10) are known as the Reaction Functions of firms' 1 and 2, respectively. A graph of the two equations on the same diagram would give a point of intersection, which would mark the equilibrium values of output levels of the two firms. Algebraically, the solution of Eqs. (9.8) and (9.10) would give

$$Q_1 = 10 \text{ and } Q_2 = 10$$

Substitution of these values into the two demand and two profit functions would yield.

$$P_1 = 70 \text{ and } P_2 = 55$$

$$\pi_1 = 450 \text{ and } \pi_2 = 300$$

This is the solution of the duopoly problem. This solution procedure will hold good for any number of firms in an oligopolistic industry. Suffice it to say here that each firm would attempt to maximize its own profit with respect to its own output, assuming the output of all rival firms as parameters; which would yield as many necessary conditions as the number of firms, a solution of which together with the demand functions for each oligopolist's product would yield the profit-maximising outputs and prices, and the corresponding profits of all the oligopolists. Incidentally note that the equilibrium output under oligopoly is never less than under monopoly, and in standard oligopoly, the relationship between the two is given by

$$Q_N = 2\ Q_M\ [N/(N + 1)]$$

Where $Q_N$ = equilibrium output under oligopoly having $N$ firms and $Q_M$ = equilibrium output under monopoly. Thus, under standard duopoly ($N = 2$), equilibrium output would be 1.33 times that of under monopoly.

**Bertrand Model**: Another competition based model for oligopoly pricing is due to Joseph Bertrand, a French economist, and accordingly it is known as the Bertrand model. The model developed in 1883

postulates that the **duopolists compete through price** and accordingly assume that they do not change their prices in the face of a change in the price of their rival firms. Thus, the model is based on the premise

$$\frac{dP_i}{dP_j} = \frac{dP_j}{dP_i} = 0 \qquad (9.11)$$

If this were true, each oligopolist could simply attempt to maximize its profit with respect to its own price and that would yield a unique solution to the oligopoly market. For a numerical illustration of this model, we must have demand functions in prices rather than in outputs. We could, of course, transform the demand functions (9.1) and (9.2) in terms of prices but the so derived functions would be rather complicated to apply the Bertrand model. We thus consider a different example. Suppose the relevant functions were the following:

$$Q_1 = 100 - 3P_1 + 2P_2 \quad Q_2 = 200 - 4P_2 + 3P_1 \quad TC_1 = 500 \quad \text{and} \quad TC_2 = 1500$$

The profit function of duopolist-1 will then be given by

$$\pi_1 = 100P_1 - 3P_1^2 + 2P_1P_2 - 500$$

For this to be maximum, the first derivative must be zero and the second derivative negative. Thus,

$d\pi_1/dP_1 = 100 - 6P_1 + 2P_2 = 0$ [using the Bertrand's assumption in (9.11) above]

or, $6P_1 - 2P_2 = 100$, which gives, $P_1 = (100 + 2P_2)/6$ [reaction function of Firm-1]

and, $d\pi_1/dP_1^2 = -6 < 0$

Similarly, profit function of duopolist-2 and the optimization conditions for the same would give

$$\pi_2 = 200P_2 - 4P_2^2 + 3P_1P_2 - 1{,}500$$

$$d\pi_2/dP_2 = 200 - 8P_2 + 3P_1 = 0$$

or, $3P_1 - 8P_2 = -200$, which gives, $P_2 = (3P_1 + 200)/8$ [reaction function of Firm-2]

and, $d\pi_2/dP_2^2 = -8 < 0$

The solution of the two reaction functions would yield

$$P_1 = 28.6 \quad \text{and} \quad P_2 = 35.7$$

Substitution of these values in the two demand functions and profit functions would give

$Q_1 = 85.6$ $Q_2 = 143$, $\pi_1 = 1948$ and $\pi_2 = 3605$, and $\pi_1 + \pi_2 = 5553$

Which mark the solution of the duopoly price-output problem under the Bertrand's model.

If the two rival firms form a cartel, they would maximize the joint profit with respect to each price. Thus,

$$\Pi = \pi_1 + \pi_2 = 100P_1 = (100P_1 - 3P_1^2 + 2P_1P_2 - 500) + (200P_2 - 4P_2^2 + 3P_1P_2 - 1500)$$
$$= -2000 + 100P_1 + 200P_2 - 3P_1^2 - 4P_2^2 + 5P_1P_2$$

i.e., $\partial\Pi / \partial P_1 = 100 - 6P_1 + 5P_2 = 0$, by first derivative and necessary condition

or, $6P_1 - 5P_2 = 100$

and $\partial \Pi / \partial P_2 = 200 - 8 P_2 + 5 P_1 = 0$, by first derivative and necessary condition

or, $5 P_1 - 8 P_2 = -200$

Solution of the two above equations will give

$P_1 = 42.02$ and $P_2 = 30.43$

Substitution of the values of prices in demand equations and calculation of profits will yield

$Q_1 = 31.8 \quad Q_2 = 204.34 \quad \pi_1 = 962$ and $\pi_2 = 4718$, and $\pi_1 + \pi_2 = 5680$

As expected the total (industry) profit is greater under cartel than under Bertrand model. The validity of the Bertrand's assumption is hard to believe and therefore the Bertrand model could not become a popular solution to the oligopoly pricing problem.

## 9.4 TACIT (IMPLICIT) COLLUSION MODELS

Due to anti-trust regulations and opposition from consumers' organizations, oligopolists find it rather hard to form cartels. They, therefore, sometimes reach an agreement (Memorandum of Understandings: MOU) among themselves to "live and let live", and such agreements in the oligopoly literature are known as tacit collusions. Under this, all the oligopolists in an industry agree on something, known as the Leader-Follower relationship among them. There are alternative versions of such agreements, and thus they are known under different names, viz. Dominant Firm (Residual Market Share) model, Fixed Market Share model and Stackelberg model. Together these models are also knows as the **Leader-Follower models**. Each of these models provides a unique solution to the oligopoly pricing problem. There is a distinct difference between the three models discussed in the last two sections (viz. Cartel, Cournot and Bertrand) and the three models under discussion here and this is with regard to their timings of moves. In the former group, various firms in an oligopoly industry **move simultaneously** with regard to their price-output decisions; while in the latter group, the **rivals move in sequence**—leader moves first and the followers follow. While it is difficult to cite examples of such tacit collusions among firms in a given industry, one can surely venture to suggest the possibility of their existences, if not already exist. These would include Wal-Mart in general merchandise, State Bank of India in banking services (in India), US Dollar-Euro rate in exchange rates, Harvard Business School in management education, Toyota in automobiles industry, etc. How each of the tacit collusion models work is explained now.

**Dominant Firm Model** As the name implies, the **dominant firm model** is applicable when one of the competing firms happen to command a dominant share in the industry and when the industry falls under standard oligopoly market. Under this version of the leadership model, the most dominant firm in the industry assumes the market leadership and the rest follow it in price-output determination. For example, in the cement industry in India, the Associated Cement Companies (ACC) may become the leader and all the other manufacturers of cement agree to follow the ACC. In such a situation, the leader moves first and sets the price of the commodity or service, and it lets all the followers to sell as much as they wish at its determined price and takes for itself the residual market only. Thus, the followers face a horizontal demand curve at the price set by their leader, like the one under perfect

competition, and they are mere **price-takers**. Such a strategy is feasible in case of homogeneous oligopoly only, for otherwise the leader would face enormous difficulties in setting the price for differentiated products of all the oligopolists. To explain price-output determination in such a situation, consider an example of a homogeneous oligopoly market. Let the equations of the relevant functions be the following

Industry ($I$) Demand Function: $Q_I = 40{,}000 - 0.2P$

Leader's ($L$) Total Cost Function: $TC_L = 1{,}000 - 20{,}000\ Q_L + 3\ Q_L^2$

Each follower's ($f$) Total Cost Function: $TC_f = 50 + 44{,}000\ Q_f + 20\ Q_f^2$

(For simplicity, all the follower firms are assumed to have identical cost functions)

Since all followers are price takers, each one of them has a supply function; and their summation would give all the follower's supply curve. Given the above relationships, we can find out the supply function of all the followers. This could be attempted through first finding out the individual follower's supply function (vide Chapter 6, Section 6.7) and then summing (horizontally) over the firms to get their aggregate supply function. To follow this procedure, we first need to determine the marginal cost function of a follower, which is given by the derivative of its total cost function. Thus,

$$MC_f = 44{,}000 + 40\ Q_f$$

Since the follower firms are price takers and they could sell as much as they wish at the ruling price, $MC_f = P$ and hence

$$P = 44{,}000 + 40\ Q_f$$

or, $$Q_f = -1{,}100 + 0.025\ P$$

This marks the supply function of the individual follower, assuming the price is no less than the shut-down price. Assuming that there are 10 such firms (each having the same supply function, for their cost functions are assumed to be identical), the total supply ($Q_F$) function of all the followers becomes

$$Q_F = 10Q_f = -11{,}000 + 0.25\ P$$

The demand function facing the leader ($Q_L$) is then given by the difference between the industry demand and all the followers' supply:

$$\begin{aligned} Q_L &= Q_I - Q_F \\ &= (40{,}000 - 0.2\ P) - (-11{,}000 + 0.25\ P) \\ &= 51{,}000 - 0.45\ P \end{aligned}$$

or, $$P = 113{,}333 - 2.22\ Q_L$$

The leader would set the price in such a way that its profit is maximum with respect to its own output. The profit function of the leader ($\pi_L$) is given by

$$\begin{aligned} \pi_L &= TR_L - TC_L \\ &= PQ_L - TC_L \end{aligned}$$

$$= (113{,}333 - 2.22\ Q_L)\ Q_L - (1{,}000 - 20{,}000\ Q_L + 3Q_L^2)$$

$$= -\ 1{,}000 + 133{,}333\ Q_L - 5.22$$

Maximization of $\pi_L$ with respect to $Q_L$ requires that

$$\frac{d\pi_L}{dQ_L} = 133{,}333 - 10.44\ Q_L = 0$$

or, $$Q_L = 12{,}771$$

and, $$\frac{d^2\pi_L}{dQ_L^2} = -\ 10.44 < 0$$

Thus, at $Q_L = 12{,}771$, $\pi_L$ is maximum. Substitution of this into the leader's demand function yields

$$P = 113{,}333 - 2.22\ (12{,}771)$$

$$= 84{,}981$$

Substitution of the above value of $P$ into the supply function of all the followers, gives

$$Q_F = -\ 11{,}000 + 0.25\ (84{,}981)$$

$$= 10{,}245$$

Since each follower has the same supply function and there are 10 of them, the supply of each follower equals 1024.5. Thus, the price in the industry would be Rs. 84,981, each follower would sell a quantity equal to 1024.5, all followers together would sell 10245 units of output, the leader would sell a quantity equal to 12,771, and the total output sold by all oligopolists would equal 23,016 (=10,245 + 12,771) units. It can now be checked that this solution is consistent with the industry demand curve i.e.,

$$Q_I = 40{,}000 - 0.2\ P$$

$$= 40{,}000 - 0.2\ (84{,}981)$$

$$= 23{,}004$$

This result is different from the above total output of 23,016. The difference is due to the rounding of numbers in the above calculations only. Thus, the model yields unique solutions for the oligopoly industry. The above solution procedure is described through geometry in Fig. 9.2.

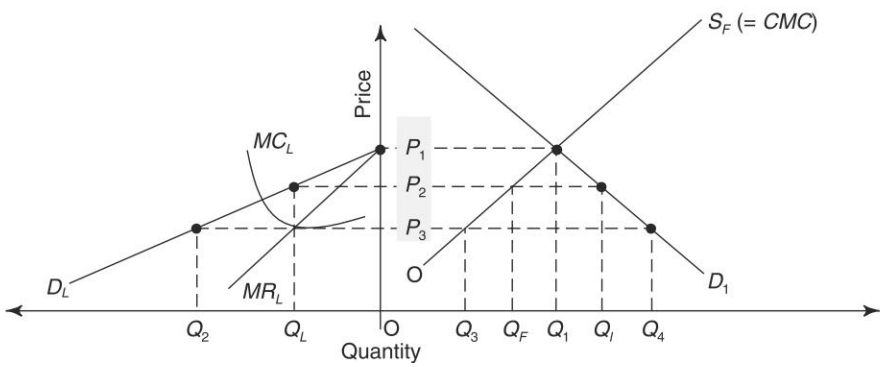



**Fig. 9.2.** *Homogeneous oligopoly pricing under dominant firm model*

In Fig. 9.2, $D_I$ represents the industry demand curve, $S_F$ the supply curve of all the followers (the part of their combined marginal cost curve), $D_L$ the residual demand curve facing the leader (dominant firm) and $MR_L$ the marginal revenue curve of the leader corresponding to its $D_L$ (= $AR$) curve. The $D_L$ curve is constructed as the difference between $D_I$ and $S_F$ at various prices. Thus, if price were $P_1$, the demand for industry product (= $OQ_1$) would just equal the supply of all the small firms (followers) and there would be zero demand for the dominant firms output. However, at lower prices, such as $P_2$ and $P_3$, the industry demand exceeds the all followers' supply and residual represents the demand for the dominant firms output. Thus, at $P = OP_2$, $OQ_L$ (= $Q_F\ Q_1$) is the demand for the dominant firms output, and at $P = OP_3$, $OQ_2$ (= $Q_3\ Q_4$) is the demand for the dominant firms output, and so on. Once the $D_L$ is constructed, $MR_L$ is obtained through the definite relationship between $AR$ and $MR$. The price is determined by the dominant firm such that its profit takes its maximum value. This occurs at $Q = OQ_L$, where $MR_L = MC_L$ and $MC_L$ intersects the $MR_L$ from below. Corresponding to this output, the equilibrium price equals $P_2$ at which all the followers sell an output equal to $OQ_F$. By construction, $OQ_L + OQ_F = OQ_I$ which represents the industry demand at the equilibrium price of $OP_2$. This is how the price and quantity are determined under the dominant firm (residual market share) version of the leader-follower model. Such a pricing model is quite relevant in India for various goods and services produced under homogeneous oligopoly situation. These would include cement, steel, aluminum, sugar, petrol, etc.

**Fixed Market Share Model** Under the fixed market solution system, one firm becomes the leader and all the other firms take the pre-set fixed market shares. The leader is decided jointly by all the firms in the industry and it generally goes to the most efficient firm. Accordingly, the approach is also called the **efficient firm model**. However, the leader could be named on the basis of some other criteria as well, like social or political clout or simply by election and/or rotation. To explain the working of the model, let there be just two oligopolists, their individual demand and cost functions be the same as the ones given above in Eqs. (9.1) through (9.4), the oligopolist-1 be the leader and oligopolist-2 the follower, and the leader and follower's market share be pre-decided at two-thirds and one-third, respectively. Thus,

$$\frac{Q_2}{Q_1 + Q_2} = \frac{1}{3}$$

or,
$$Q_2 = 0.5\ Q_1 \tag{9.10}$$

Given the market share of the follower, the leader would make the **first move** and sets its output level such that its own profit is at the maximum. Recall that the profit function of oligopolist-1 is the following:

$$\pi_1 = 100\ Q_1 - 4.5\ Q_1^2 - Q_1\ Q_2$$

Substitution for $Q_2$ in terms of $Q_1$ from (9.10), gives

$$\pi_1 = 100\ Q_1 - 4.5\ Q_1^2 - Q_1 (0.5\ Q_1)$$

$$= 100\ Q_1 - 5\ Q_1^2$$

For $\pi_1$ to be maximum with respect to $Q_1$

$$\frac{d\pi_1}{dQ_1} = 100 - 10\ Q_1 = 0$$

or, $Q_1 = 10$

and, $\frac{d^2\pi}{dQ_1^2} = -10 < 0$

Thus, at $Q_1 = 10$, $\pi_1$ is maximum. Substitution of this into Eq. (9.10) yields $Q_2 = 5$. Further substitution of these values into the two demand and two profit functions would give

$P_1 = 75$ and $P_2 = 70$, and

$\pi_1 = 500$ and $\pi_2 = 225$

The above results mark the solution to the duopoly pricing problem. It must be emphasized here that the solution depends not only on the market share of the oligopolists but also on the factor as to who becomes the leader. To prove this, let us keep the market shares as in (9.10) above and change the leadership. Thus, if oligopolist 2 takes up the leadership, giving 2/3 of the market to oligopolist-1, it would set its output such that its profit is at its maximum. Oligopolist-2's profit function is given by

$$\pi_2 = 70\,Q_2 - Q_1 Q_2 - 3\,Q_2^2$$

Substituting for $Q_1$ from (9.10) above, we have

$$\pi_2 = 70\,Q_2 - (2Q_2)\,Q_2 - 3\,Q_2^2$$

$$= 70\,Q_2 - 5\,Q_2^2$$

Maximization of $\pi_2$ with respect to $Q_2$ requires that

$$\frac{d\pi_2}{dQ_2} = 70 - 10\,Q_2 = 0$$

or, $Q_2 = 7$

and, $\frac{d_2\pi_2}{dQ_2} = -10 < 0$

Thus, at $Q_2 = 7$, $\pi_2$ is maximum. At this value of $Q_2$, $Q_1 = 14$ by the market share Eq. (9.10). Substitution of these into the demand and cost functions of the duopolists yields

$P_1 = 65$ and $P_2 = 60$

$\pi_1 = 420$ and $\pi_2 = 245$

A comparison of this solution with the previous one clearly demonstrates that the pricing under the fixed market share model depends upon the choice of the leader, among other things. Further, when oligopolist-1 was the leader, its profit stood at 500, which fell to 420 when he turned as the follower. Similarly, under the leadership of oligopolist-1, oligopolist-2 made a profit equals 225, which goes up to 245 when the leadership goes to oligopolist-2. Such changes always occur in this leader-follower model. Thus, it may be concluded that there exists the **First Mover Advantage** in the leader-follower model. A corollary of this is, viz. **each rival firm would try to grab the leadership**! May be such gains are available in political fields as well and that is why the coalition governments fight in who would be the first Chief Minister in an agreed rotation ships?

**Stackelberg Model** The last model under the group of tacit collusion models is due to the German economist, H.V. Stackelberg, and accordingly is known as the **Stackelberg model**. As in other leader-

follower models, under this approach to oligopoly pricing, one of the oligopolists becomes the leader on the criteria similar to the ones in the other leader-follower models, and the rest agree to be the followers. Also, the **leader makes the first move**, which is given (known) to the followers when they make their moves. The similarity ends here. The difference is in terms of the assumption, made by the leader, about the actions of the followers in its optimization behaviour. There are two versions of the Stackelberg model in this regard. One is the Cournot version and the other is the Bertrand version. In the former, the leader assumes that the followers maximize their respective profits with respect to their corresponding **output levels**, taking the leader's output as given; while in the latter case the leader's assumption is that the followers maximize their respective profits with respect to their corresponding prices, taking the leader's price as given. In other words, the leader takes the followers' Cournot model's reaction functions for granted in its decision making in the **Stackelberg–Cournot** model, and it takes the followers' Bertrand model's reactions functions for granted in the **Stackelberg–Bertrand** model. Further, in the Stackelberg-Cournot version, the leader sets its output such that its profit is maximum with respect to its output, while in the Stackelberg-Bertrand version, the leader sets its price such that its profit is maximum with respect to its price. The solution procedure for the Stackelberg-Cournot model is illustrated below for the above differentiated duopoly example.

Recall the given demand and cost functions of Eq. (9.10) through (9.4). Also, recall the corresponding profit Eqs. (9.7) and (9.9) of the two duopolists. Further, suppose duopolist-1 becomes the leader and duopolist-2 the follower. Under the Stackelberg–Cournot model, the leader sets its output such that its profit is maximum subject to the reaction functions of the followers. Consider the duopolist-1's profit function:

$$\pi_1 = 100\, Q_1 - 4.5\, Q_1^2 - Q_1 Q_2 \tag{9.7}$$

Substitute in (9.7) the value of $Q_2$ from the Cournot reaction function of duopolist-2, i.e., Eq. (9.10), viz. $Q_2 = (70 - Q_1)/ 6$. Thus,

$$\pi_1 = 100\, Q_1 - 4.5\, Q_1^2 - Q_1 (70 - Q_1)/6$$

$$= 100\, Q_1 - 4.5\, Q_1^2 - 11.67\, Q_1 + 0.167\, Q_1^2$$

or, $$\pi_1 = 88.33\, Q_1 - 4.333\, Q_1^2$$

Maximisation of profit with respect to output requires the first derivative to be zero and the second derivative to be negative. Thus,

$$\frac{d\pi_1}{dQ_1} = 88.33 - 8.\ 666\, Q_1$$

or, $$Q_1 = 10.19$$

and $$\frac{d^2\pi}{dQ_1^2} = -4.333 < 0$$

Substitution of the value of $Q_1$ in the duopolist-2's reaction function gives

$$Q_2 = (70 - 10.19)/6 = 9.97$$

Substitution of the above equilibrium values of the two outputs in the demand functions (9.1) and (9.2) yield the equilibrium prices:

$$P_1 = 100 - 2(10.19) - 9.97 \quad \text{and} \quad P_2 = 95 - 10.19 - 3\,(9.97)$$

or, $$P_1 = 69.65 \quad \text{and} \quad P_2 = 54.90$$

Further substitution of the values of $Q_1$ and $Q_2$ in two profit functions, give

$$\pi_1 = P_1 Q_1 - TC_1 = 69.65\,(10.19) - 2.5\,(10.19)^2$$

$$= 450.34$$

$$\pi_2 = P_2\,Q_2 - TC_2 = 54.90\,(9.97) - 25\,(9.97)$$

$$= 298.10$$

The above function provides the unique solution of the duopoly price–output decision problem under the Stackelberg–Cournot model. Like in the fixed market share model, if the leader changes the solution would change and one would discover that there is the **first mover advantage** in this model as well. The detailed working of the Stackelberg–Bertrand model is not illustrated here. Suffice to say here, the example will need demand functions in terms of the prices; and the procedure would be similar to the Stackelberg–Cournot version, though the leader and followers will maximize their respective profits with respect to their own corresponding prices; and one would discover that in this case there would be **first mover disadvantage.**

Note that, if the rival firms fail to resolve the leadership issue and decide to act as if all were the followers (no leader), the Cournot solution would emerge. Alternatively, if each one tries to act as the leader (no follower), no solution would emerge and we would have the so called "**Stackelberg disequilibrium**". This completes the explanation of all the three tacit collusion models. However, before we proceed to the next model, it would be interesting to provide a comparative picture on the solutions of the duopoly problem under various hypotheses. The results are summarized in Table 9.1.

**Table 9.1** *Price Output Determination under Different Models of Oligopoly*

| | *Model* | *Prices* | | *Outputs* | | | *Profits* | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $P_1$ | $P_2$ | $Q_1$ | $Q_2$ | *Total* | $\Pi_1$ | $\Pi_2$ | *Total* |
| (i) | Cartel | 73 | 60 | 9.2 | 8.6 | 17.8 | 460 | 301 | 761 |
| (ii) | Cournot | 70 | 55 | 10 | 10 | 20 | 450 | 300 | 750 |
| (iii) | Fixed Market Share | | | | | | | | |
| | Leader: 1 | 75 | 70 | 10 | 5 | 15 | 500 | 225 | 725 |
| | Leader: 2 | 65 | 60 | 14 | 7 | 21 | 420 | 245 | 665 |
| (iv) | Stackelberg-Cournot | | | | | | | | |
| | Leader -1 | 69.65 | 54.90 | 10.19 | 9.97 | 20.16 | 450.34 | 298.10 | 748.44 |

A scrutiny of the results in Table 9.1 would reveal that the profit-maximizing prices, outputs and corresponding profits are different under different models of oligopoly market. Although there are several differences in the results, a couple of differences, which hold in all cases, are worth noting.

(a) Aggregate profit would never be less and would generally be more under the cartel model than in any other model of oligopoly pricing.

(b) If an inefficient or weak firm becomes the leader, the outcome would be unfavourable from the point of view of all the oligopolists taken as a group. It is seen in the table that when firm II, which is otherwise inefficient, becomes the leader, the aggregate profit is the least of all other models. For this reason, the fixed market share model and Stackelberg–Cournot model are generally associated with the leadership of the most efficient firm.

The models of price-output determination under oligopoly could easily be extended to the cases of multiple differentiated oligopolists as well as homogeneous oligopolist situations. In the former, the number of market share equations would equal $n - 1$, where $n$ is the number of oligopolists; on substitution of the values for all but one (leader) oligopolist's output in terms of the leader's output into the leader's profit function would express the leader's profit in terms of its output only; maximization of which would yield the equilibrium output of the leader; which, on substitution into the demand and profit functions, would yield the equilibrium values for the remaining variables. The homogeneous oligopoly situation would be simpler than the heterogeneous one. Instead of multiple demand functions, there would be just one (industry) demand function, the sum of all firms output would equal the industry output, and the solution could then be easily obtained on the same procedure as for the case of heterogeneous oligopoly.

## 9.5 GAME THEORY

In all the foregoing oligopoly models, there was no uncertainty by design as each model had a fixed assumption with regard to the actions/reactions of all the firms which are mutually dependent. However, in the real world, there are examples where none of those assumptions hold. The Game theory provides a possible solution under such situations. There are several versions of the game theory. The one with which we are concerned here is known as the **non-zero sum, non-cooperative game theory**. Under this, just to highlight this part only each competing firm in a given industry has a set of alternative strategies, known to all, and it is not known which strategy it will use under which situation. Thus, there is uncertainty in the industry. Under the circumstances, firms play non-cooperative games with their rivals in the industry. How such games are played or how a given uncertainty is handled, is the subject matter of the game theory? Before explaining its working in the oligopoly market, we must understand the ingredients of the Game theory.

**What is a game?** A game describes how people/firms behave in strategic situations. In every game, more than one decision-maker is involved and the actions of each of them depend not only on the choices available to self but also on the choices available to others. Thus, a game has three ingredients, viz. (a) List of decision-makers/players, (b) List of choices/strategies available to each player, and (c) Pay-offs associated with each and every possible set of alternative strategies. The alternative choices available to each player are called **strategies**; for while choosing an option, each player must consider how its rivals may react to its actions and thereby affect its own pay-off. Obviously, the number of players and number of strategies have to be finite for a game to be feasible. The pay-offs would be positive for good outcomes and negative for bad outcomes. For example, if the reward is in terms of profits or satisfaction, it is positive; and if it is loss or prison term, it would be negative. Obviously, if the outcome is desirable, the players would like to maximize that and if it is undesirable (bad good), they would like to minimize that. For the game to have full details, its particular type must be mentioned. There are two types of games, viz. **simultaneous** (or **one shot**) **and sequential,** also known as **static and dynamic**,

respectively. Further, whether the **game is played just once or more times; and if repeated, is it repeated finite times or infinite times**. In a game everyone tries to predict what everyone else will do and then chooses the best available alternative. The issue to settle in a game theory is, "**Is the outcome of the game predictable**"? If yes, the game theory offers a solution, if not, the game theory is of little use.

The game theory, like all others, is based on some usual **assumptions**. These are

(a) All players act rationally

(b) All players have the full information like those contained in each of above games' tables below.

**Concepts/Theories** Game theory has some **new concepts** which must be understood well before the theory could be applied to resolve the decision issue, even if the same could be predicted. These include

(a) Dominant strategy

(b) Secure strategy

(c) Nash equilibrium

A **strategy is the dominant for a player if it guarantees the best outcome pay-off to the player irrespective of his/her rivals' choices of their strategies**. Thus, the dominant strategy **caps the player's reward no matter what his/her rival choose to play**. Accordingly, this offers the best choice that a player could make when he/she faces the uncertainty with regard to his/her rivals' choices. Under the situation, if any player could find such a strategy, he would grab it and forget about seeking guidance from any other game theory. As a corollary to the dominant strategy, if a strategy is dominated by any other strategy, then it (the former) is a dominated strategy. Obviously, all dominated strategies must be rejected in favour of their respective dominant strategies. Neither of these two strategies is bothered about the rivals' options. Of course, rivals' choices would ultimately determine his/her ultimate outcome, within his/her (chosen) dominant strategy. The outcome of the game, of course, depends on the choices all players make. Thus, if each player has a dominant strategy, the game has an equilibrium (the term "equilibrium" in game theory context means the game has a solution), which would be stable as well. Further, if all but one player (i.e., $n$–1 players out of $n$ of them) has dominant strategies, the game would still have unique and stable equilibrium; the last player (the one who does not have a dominant strategy) would know the dominant strategies of all his rivals and given those he would choose the best from the available choices. As will be seen later, unfortunately, **dominant strategy equilibriums** are hard to find in the real life games.

In contrast, the secure strategy aims to do the best under the worst situation. Accordingly, the **secure strategy of a player is the one which guarantees him/her the best pay-off under the worst situation**. Thus, the secure strategy **caps the loss** to its player. The said strategy suffers from two **limitations**. One, it is conservative and may be considered only if the player has good reason to be absolutely averse to risk. Two, it ignores the optimal decisions of rivals and thus prevent the player from earning a significantly higher pay-off. In consequence, the secure strategy is inferior to the dominant strategy which focuses on maximum profit to the player. Accordingly, it could be opted to only when a dominant strategy does not exist. Further, the secure strategy is useful for a game to have equilibrium, but it does not guarantee the existence of equilibrium. Only when the secure strategy of one player coincides with those of all the other players, the game has equilibrium, and the equilibrium point is

referred to as the "**Saddle Point**". The question is how to find a secure strategy and the saddle point? The method is known as the maximum minimorum (or just maximin) if the outcome is a desirable one and minimum maximorum (or just maximin) if the outcome is a bad good. As the names imply, the former is the maximum among the lot of minimums and the latter represents the minimum among the lot of maximum values. When the maximin or minimax, as the case may be, for all the players in a game coincides, we get the saddle point. How to calculate each of those is explained in examples below. Again, like the dominant strategy equilibrium, **saddle point equilibrium** is rare to find in the real life cases.

John Nash, in his doctoral thesis (27 pages) at Princeton university, USA (at the age of 22 years) and subsequently in his articles (vide Nash 1951) has advanced a famous game theory, now known as the "Nash Equilibrium", which is more often available than the above mentioned two theories to solve the games and the oligopoly decision issues. It must be mentioned here that John Nash was awarded the **NOBEL PRIZE,** which he shared with two others, on this contribution (cited as "most trivial work") in 1994. According to this theory, the **Nash equilibrium is the situation which guarantees the best pay-off to each player, given the strategic choices of all the rivals**. In other words, Nash equilibrium is a **set of strategies** such that none of the players in the game can improve their pay-offs given the strategies of all the players. In Nash equilibrium, the strategy played by each individual is a best response to the strategies played by everyone else, and thus everyone is content with its outcome in the sense that no player has any incentive to move out of such equilibrium. Further, each player desires all the rivals to choose the Nash equilibrium. In view of these characteristics, Nash equilibrium is **stable**. No wonder why this hypothesis is considered as the backbone of the game theory. However, the theory suffers from some **limitations** as well. One, a game could have more than one Nash equilibrium and thereby renders the outcome of the game unpredictable. Two, like other theories, it could give misleading result if any of the players, acted irrationally. Three, it can never give results better than under a dominant strategy. Due to these reasons, **oligopolists must choose this strategy only when a dominant strategy is not available, and may even follow a secure strategy when the probability of irrational behaviour is high**. Recall that since both the Cournot and Bertrand models are based on the premise that "given the output (Cournot) or price (Bertrand) of rivals", the solution of each of them represents a Nash equilibrium. In addition to the above game theories, there is a less popular strategy, called "**tit-for-tat strategy**". Under this strategy, a player cooperates as long as opponents cooperate and switches to a non-cooperating strategy if his opponents switch strategies. This strategy is rational for the infinitely repeated prisoners' dilemma. We now apply each of the three theories to some famous games and to oligopoly (duopoly) issues.

## Examples and Applications

A classical example of games is the so called "**Prisoners' Dilemma**", which is presented in Table 9.2. In this example, there are just two players (Prisoners 1 and 2), each with just two alternatives/strategies (Confess and does not confess), and accordingly there are just four possible outcomes (5,5; 0,20; 20,0; and 1,1). Note that in this and other tables (matrices) of game theory, columns have player-1 and rows player-2; and their pay-offs are separated by a comma, the first number denotes the pay-off of player-1 and the second number that of player-2. In Table 9.2, the outcome is the number of years of prison term (a bad good), so each player would aim at its minimum value.

**Table 9.2** *Prisoners' Dilemma (Pay-offs: Prison term)*

| *Prisoner-1/Prisoner-2* | *Confess* | *Not Confess* |
|---|---|---|
| Confess | 5, 5 | 0, 20 |
| Not Confess | 20, 0 | 1, 1 |

To give a brief background of the game, the two persons involved are currently in prison in separate cells, not allowed to communicate with each other, have been caught together with weapons and are suspected to have jointly committed a murder. The authority has declared the rewards (years of prison term) as per their responses listed in the table. The rewards are in terms of years of prison term and the game is to be played simultaneously and only once. Each player will like to minimize his jail term. Is the outcome predictable? Before we answer such questions, let us understand how the pay-offs are generated. Obviously, the pay-offs are derived through the pay-off equations which denote each player's pay-off as a function of its strategies and strategies of its rivals. There is no set mathematical model by which the authorities decide the prison terms exactly and even if there is one, we are unable to explain the above pay-off matrix. Thus, the above could be taken as hypothetical. However, for our games in an oligopoly industry, we can attempt to even explain their pay-offs. Thus, let us put the results of our one such hypothetical example in the chapter above in the game form. This is given in Table 9.3.

**Table 9.3** *Differentiated Oligopoly: Cartel Model vs. Cournot Model—Fair and Cheating Game (Pay-offs: Profits)*

| *Firm-1/Firm -2* | $P_1 = 55$ *(Cournot)* | $P_2 = 60$ *(Cartel)* |
|---|---|---|
| $P_1 = 70$ (Cournot) | 450, 300 (Cournot) | 467, 280 (Firm-1 cheats on Cartel agreement) |
| $P_2 = 73$ (Cartel) | 431, 318 (Firm-2 cheats on Cartel agreement) | 460, 301 (Cartel) |

Recall from the previous section that if the two duopolists (vide Eqs. 9.1, 9.2, 9.3 and 9.4) form a Cartel, their prices would be $P_1 = 73$ and $P_2 = 60$, and their profits (460, 301) would be as given in the East-South column in Table 9.4. Alternatively, if they move in the Cournot way, their prices would be 70 and 55, and profits 450 and 300, respectively. Further, if they agree to form a cartel but Firm-1 cheats (by setting its Cournot price-output mix), they would end up with profits at 467 and 280. Lastly, they agree for the Cartel but Firm-2 cheats (by setting Cournot values for its price-output mix), then profits would be 431 and 318. Note that the results under cheating cases could be verified by inserting the cheat (Cournot model) prices in the corresponding demand and profit functions. Thus, there are four possible outcomes and thus the outcome is uncertain. Could the outcome be predicted through the game theory? The answer would be yes if, and only if, at least any one of the game theories had a unique equilibrium.

Consider the dominant strategy first and let us start with the prisoners dilemma game. Starting with Prisoner-1; if he confesses the crime, he would end up with either 5 or 0 year of jail; but if he declines, he receives 20 or 1 year of jail term. Since each of the former decision's jail term in confessing are less than the corresponding ones in declining (i.e. $5 < 20$ and $0 < 1$, and jail is a bad good), his confess strategy is dominating. Similarly, confess strategy is dominating strategy for prisoner-2 (the pay-offs are symmetric in this game). Accordingly, both would confess and receive jail term of 5 years each. This outcome is thus predictable through the dominant strategy rule. No matter whether they have committed the crime or not, it is viable for both to confess the crime. The said (equilibrium) outcome is, of course, inferior to the

outcome under a cooperative game (i.e., sum of the outcomes of all players available under "not confess-not confess" option, where each receives a mere one year of jail term)! The jointly (socially) optimum will not happen because the one who chooses "not confess" runs the risk of other choosing "confess" and end with a larger jail term (20 years instead of 5 years). However, in repeated games, cooperation among players could be enforced through the threat of penalty. Thus, if player-1 cheats in first shot, player -2 could cheat in second shot, and so on. Nevertheless, the result will hold good even if the game is repeated but finite times, for in the last occasion the one who cheats will surely gain (the cheat may not gain in all other but last chance to play because the other player will have no chance to reciprocate). However, if the game is played infinite times, the jointly optimum strategy of "not confess-not confess" will prevail.

Consider the oligopoly game next. As one would expect, the joint profits are maximum (not just jointly but for each as well) under the Cartel with $P_1 = 73$ and $P_2 = 60$. Look for the dominant strategy of each oligopolist. Profit figure 450 > 431, and 467 > 460, and hence $P_1 = 70$ is the dominant strategy of Firm-1. Similarly, profit of 300 > 280 and 318 > 301, and thus strategy $P_2 = 55$ is the dominant strategy for Firm-2. Both the players have dominant strategies, which cross at the North-West corner. Accordingly, the outcome of the game is predictable at these prices giving profits of 450 and 300 to firms 1 and 2, respectively. Recall that this was the result in the Cournot model also.

Recall that if dominant equilibrium exists, it is the best to choose and thus leaves no room for any other theory. However, just for the purpose of explanation, let us apply the other two theories. What is the Saddle point equilibrium, if any, in each of the above two games? For this, let us look first for the secure strategy for each player. In prisoners' dilemma game, outcomes are a bad good and thus the secure strategy would be the minimax strategy. For player-1, the maximum jail term that he would get would be 5 years if he opts for confess and it would be 20 years if he chooses don't confess. The minimum of these two outcomes is 5 years and thus his secure strategy would be to confess. Similarly are the outcomes for player-2 (since the matrix is symmetrical) and so his secure would also be to confess. The two secures cross at confess-confess, which, thus, represents the saddle point equilibrium. In oligopoly pricing game, the pay-offs are profits, a desirable good, and thus the secure strategy would be the maximin strategy for each player. If Firm-1 prices at 70, its minimum profit would be 450 and if it prices at 73, the minimum profit would equal 430. Accordingly, its maximin is at price =70, which is its secure strategy. For Firm-2, if it prices at 55, it would end up with a minimum profit of 300 and if prices at 73, the minimum profit would be 280. Accordingly, its maximin or secure strategy would be at price = 55. The two secure strategies cross at North-West corner in the matrix and that represents the saddle point equilibrium. Once again, the Cournot equilibrium is hailed!

We move to apply the Nash equilibrium to the two games above. In the Prisoners dilemma's game; if player-2 chooses to play "Confess", Player-1's best pay-off (minimum jail term) is in him playing "Confess", and vice versa. Thus, "confess-confess" is a Nash equilibrium. Further; if Player-2 chooses "not confess", player-1's best pay-off is in "confess". However, if Player-1 chooses "confess", Player-2's best strategy would be to "confess", and not "not confess". Thus, "not confess"-"not confess" is not Nash equilibrium. Accordingly, there is just one Nash equilibrium in the Prisoners' dilemma game, viz. "confess-confess". This equilibrium is stable, for it leaves no incentive for any player to move out of this position. For example, when player-1 is at the North-West corner in the matrix; his only other option is to move down in the matrix and by doing so his jail term only increases from 5 to 20 years. So is the case for player-2 at the North-West corner, who could move to right lengthening his jail term from 5 to 20 years. The game has unique equilibrium and thus its outcome is predictable by this theory. Let us now apply the theory to our duopoly game of Table 9.3. Given that Firm-2 chooses strategy $P_2 = 55$, Firm-1's

all options are in column-1 of the matrix. Of the two options in it, his profits are the most under $P_1$=70. Taking the reverse side, if Firm-1 chooses $P_1 = 70$, the best that player-2 could do is to opt for strategy $P_2 = 55$. Thus, $P_1$=70 and $P_2 = 55$, giving the pay-offs 450, 300 in North-West corner happens to be a Nash equilibrium. Further, if Firm-2 opts for strategy $P_2 = 60$, the best available strategy for Firm-1 is to choose strategy $P_1 = 70$ (for 467 > 460). However, if Firm-1 chooses $P_1 = 70$, firm-2's best strategy would be $P_2 = 55$ (for 300 > 280). Thus, North-East corner is not a Nash equilibrium. Similarly, it can be verified that neither of the remaining two corners in the matrix represents a Nash equilibrium. It can therefore be concluded that the outcome of the duopoly game in Table 9.3 can be predicted and it would be profits of 450 and 300 to firm-1 and firm-2, respectively. This outcome is same as in Cournot model and this confirms that the Cournot solution is Nash equilibrium. It can easily be verified that this Nash predicted solution is stable as well.

Many more meaningful games can be designed for oligopoly markets. For example, firms in such an industry compete on price, output, advertisement budgets, research and development (R & D), etc. Also, firms face competition between the owners (Share-holders: Masters) and management (agents), management and trade unions, etc. Even countries play games through foreign exchange rates, interest rates, tariff rates, etc. which affects their relative trades, capital flows, etc. Furthermore, students as well as faculty at times play games within their own groups while carrying group projects. For example, if two students work on a joint project where their individual grades depend on the quality of their jointly submitted work; then no matter who puts on more or less efforts, each gets the same grade. Under the situation, both may work hard and receive best (like *A*) grade; if neither work hard, each will end up with the worst grade (like *C* or *D*); if one works hard and the other soft, both would end up with an average (like *B* or *C*) grade. This is an example of **free-rider problem**. Professors' could play a similar game in their joint research/consulting projects, and this offers a reason why joint projects are so rare these days.

Now we consider a pricing game which has no dominant equilibrium but has both the saddle point and Nash equilibrium. Such a game is provided in Tables 9.4.

**Table 9.4:** *Pricing Game (Pay-offs: Profits)*

| *Firm-1/Firm-2* | *Hold Price Constant* | *Increase Price* |
|---|---|---|
| Hold Price Constant | 10, 10 | 100, – 30 |
| Increase Price | –20, 30 | 140, 35 |

In the game of Table 9.4, neither firm has a dominant strategy. This is easy to see. For Firm-1, under the rivals' strategy of constant price; under its own constant price its pay-off of 10 exceeds its pay-off of –20 under its own strategy of price increase; but under the rivals' strategy of increase price, its pay-off of 100 under its former strategy is less than under its latter strategy. Thus, there is no dominant strategy for Firm-1. Similarly, for the rivals, while 10 > –30, 30 < 35, implying no dominant strategy. Looking for the secure equilibrium, we have to check if maximin of two matches. For firm-1, the minimum pay-off under constant price is 10 and that under price increase is –20, thus maximin strategy for it is to hold price constant. For Firm-2, the minimum pay-off under constant price is 10 and that under increase price strategy is –30, thus its maximin strategy is hold price constant. The two match at the North-West corner with pay-offs of 10, 10. Accordingly, the saddle point equilibrium is "price constant-price constant". However, recall that saddle point does not always represent the best option and it is surely not a socially optimum strategy either. So let us check if Nash equilibrium exists. If Firm-2 opts for constant price strategy, the best strategy for Firm-1 is to hold price constant. Similarly, if Firm-1 chooses to hold price

constant, Firm-2 would prefer to go for price constant strategy. The two strategies match, and hence North-West corner is a Nash equilibrium. Also, if Firm-2 chooses to increase price, the best for Firm-1 is to increase price; and if Firm-1 opts for price increase, Firm-2 would also prefer to go for price increase strategy. Thus, South-East corner with pay-offs of 140, 35 also represent Nash equilibrium. Accordingly, the game has two Nash equilibriums and one saddle point equilibrium, and no dominant strategy. Of these three equilibriums, two are the same with pay-offs of 10, 10; and thus this outcome is likely to be the predicted one.

Another interesting game worth discussing here is on competitive advertisements given in Table 9.5.

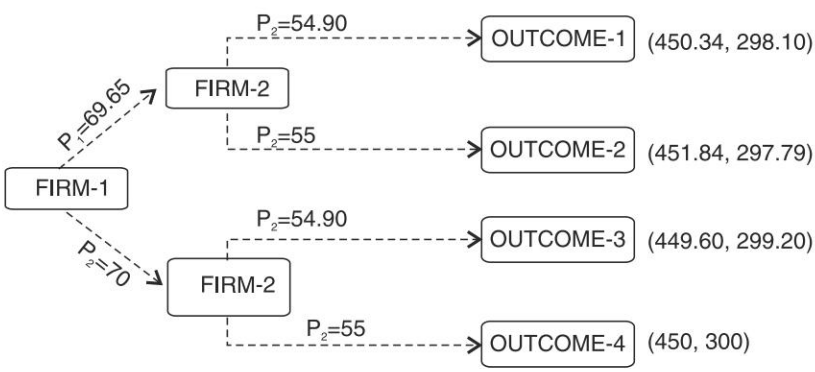
**Table 9.5** *Competitive Advertisements; Cigarette Industry (Pay-offs: Profits)*

| *Firm-1/Firm-2* | *Low* | *High* |
|---|---|---|
| Low | 300, 300 | 100, 400 |
| High | 400, 100 | 200, 200 |

In the advertisement game, the readers must verify that the strategy high is the dominant one for each of the two firms; the set of strategy "high-high" is the dominant equilibrium; strategy high is the maximin strategy for each of the two firms; the set of strategy "high-high" is the saddle point equilibrium; and the game has unique Nash equilibrium, viz. "high-high". Obviously, the prediction would be for "high-high" set of strategies. The outcome is poorer (having a total pay-off of 400) than the socially best (which happens to be "low-low" with a total pay-off of 600). The latter conclusion suggests that spending money on competitive advertisements, particularly in an industry like Cigarette industry, is twice harmful. This provides a clear case for government intervention! This completes our discussion of simultaneous (or, static) games.

**Sequential (or dynamic) Games** Under the dynamic games, one of the players (called the leader) makes the first move, which is followed by all his rivals (called followers). Thus, while the leader is unable to react to the actions/strategies of his rivals, his rivals have the opportunity to react to his action/strategy. This is the situation like the one we discussed under the section on tacit collusion (vide section 9.4) above. To illustrate this, let us consider the case of the Stackelberg–Cournot model vs. Cournot model's results from the hypothetical example above.

**Table 9.6** ***(Network diagram)*** *Dynamic Game under Differentiated Oligopoly: Stackelberg-Cournot Model vs. Cournot Model—Fair and Cheating Game (Pay-offs: Profits)*



In the dynamic game in Table 9.6, Firm-1 is the leader having two options, viz. to set its product's price at 69.65 (as per the tacit collusion) or at 70 (as per the Cournot's non-collusion). If it chooses the first option, Firm-2 could either price its product at 54.90 or at 55; giving pay-offs set of (a) 450.34, 298.10 or (b) 451.84, 297.79, respectively. In case Firm-1 opts for $P_1 = 70$ (as per the Cournot non-collusion), Firm-2 could opt for either $P_2 = 54.94$ (as per the tacit collusion) or $P_2 = 55$ (as per the Cournot non-collusion model); yielding pay-offs set of (c) 449.60, 299.20 or (d) 450, 300, respectively. Incidentally note that while the first and the last pay-offs are the ones explained above under the respective models, the other two were calculated through putting the values of the chosen values of $P_1$ and $P_2$ in the two demand functions, giving values for $Q_1$ and $Q_2$, which together with prices and cost equations yielded the levels of profits*.

The question in the game is, which of the four possible sets of pay-offs is likely to result? Note that, Firm-1's best pay-off is in option (b) while that of Firm-2 is in option (d) and of two together (i.e., socially the best) is in option (d). If it were the cooperative game, the two would settle at their joint best, viz. option (d). In the non-cooperative dynamic game that we are considering right now, how could Firm-2 ensure to secure the best result (i.e., option d) for itself? For this to happen, Firm-2 must **threaten** Firm-1 (the leader) to play $P_1 = 70$! If so, **would the threat be credible**? It is obvious that a threat is credible if it is in the player's (threatens') own interest to carry out the threat when given the option. To give an analogy; suppose a teenage girl is used to return home late in the night almost every day. Her father threatens her that if she does not return by 9 p.m. any day, he would commit suicide. Is the father's threat credible? The girl will probably keep returning late! In the game, the threat to Firm-1 is not credible either. This is simply because Firm-1's both the alternative pay-offs under $P_1 = 69.65$ are better than its pay-off under $P_1 = 70$. Accordingly, Firm-1 would surely opt for $P_1 = 69.65$ and in that case Firm-2 would opt for $P_2 = 54.90$; certifying the superiority of the Stackelberg–Cournot model. Thus, the game has a predictable solution at the pay-offs set given by (a) above. It may be verified that if the game was played simultaneously (static game), it would have the same expected solution. Students must verify that the pay-offs 450.34, 298.10 represent the dominant equilibrium, saddle point equilibrium as well as the unique Nash equilibrium. While in the case of the game in Table 9.6 above, the static and dynamic games have the same outcome; this need not be the case in all games. Before we conclude this section, note that, for simplicity, we have assumed that players have **pure strategies** only, i.e., they can use either of the alternative strategies available to them and not a **mix of (pure) strategies**, based on probabilities. Obviously, the outcome under mixed strategy could be different than under pure strategies.

## 9.6 NON-OPTIMIZING MODELS

Recall that all the above models, including the game theory, assume that the information on demand

*To some readers, the calculations in four sets of pay-offs may sound inconsistent, but that is not the case. Profits do not necessarily go up when price is either raised or lowered, as profit depends on price elasticity of demand as well. To clear the doubt, note that under the option (b), $Q_1 = 10.21$ and $Q_2 = 9.93$, and under option (c), $Q_1 = 9.98$ and $Q_2 = 10.04$; the values of $Q$s under the other two options, as noted in the previous section, are, under option (a), $Q_1 = 10.19$ and $Q_2 = 9.97$, and under option (d), $Q_1 = 10$ and $Q_2 = 10$. Also, it is true that the differences in pay-offs under various options appear small, but that should never be bothered in any optimization exercise, where more (no matter how much more) is always better than the less. Besides, the profits could be in millions of rupees/dollars, and apparently looking small may not really be that small.

and cost functions facing all the firms is available to every firm and that the game theory has a unique equilibrium. In the absence of such information, none of the above model is available to generate optimum values for prices and outputs of various firms in an industry. In the real life, quite often, firms have fairly accurate data on cost functions but little information on demand functions facing them. Under such a situation, the above pricing methods are not available and, in fact, there is no technique by which firms could determine their optimum prices. As a result they have to be content with sub-optimal results which, of course, they could keep revising as they gain more and more experience/data. If so, what pricing methods are available to the firms which could assist them to get the best possible, though still non-optimal results? Three prevailing methods under this are worth discussing here. They are (a) Cost-based pricing, (b) Competition-based or Going Rate pricing, and (c) Strategy-based or Price Penetration and Price Skimming method.

## Cost-Based Pricing Model

When the oligopolists discover that they are not able to determine their optimum prices for want of sufficient data and/or the possibility of unique outcome through game theory, they often opt to go for the cost-based pricing methods. Even firms under non-oligopoly markets use the cost based pricing when they do not have sufficient data; and/or to get quick figures, which could serve some kind of benchmark while actually deciding the prices for their products. Under the cost-based pricing approach, each oligopolist determines its floor price on the basis of its average cost and then experiments with alternative margins over that cost. Depending on the competition it faces from the rivals, it continues revising (increasing or reducing) prices, keeping in mind its own fixed costs, break-even output, installed capacity, etc. Thus in this method, there is no equilibrium price for the oligopolist's product. There are three versions of the cost-based pricing approach, viz.

(a) Full-cost or break-even pricing

(b) Marginal cost pricing, and

(c) Cost plus pricing

Under the first version, price just equals the average (total) cost. In the second version, price is set equal to the marginal cost. In the last version, some mark-up is added to the average cost in arriving at the price. The mark-up is decided on the basis of the price elasticity of demand; the higher is the elasticity, the lower is the mark-up. For example, under perfect competition, the mark-up is zero, and it increases as one move to oligopoly, monopolistic competition and monopoly. To appreciate this better, let us understand the relationship between the **cost-based prices and economic (optimum) price**. From the demand analysis of Chapter 2, we know that

$$MR = P\left(1 + \frac{1}{e}\right)$$

where, $MR$ = marginal revenue, $P$ = price (= $AR$), and $e$ = price elasticity of demand. The necessary condition for profit-maximization is

$$MR = MC$$

Combining the two relations, we get

$$MC = P\left(1 + \frac{1}{e}\right)$$

or, $$P = MC\left(\frac{e}{1+e}\right) \tag{9.11}$$

Equation (9.11) describes the economic price. By this equation, economic price varies positively with the firm's marginal cost and negatively with the price elasticity of demand for the firm's product (Remember, $e$ takes a negative value usually). If the price elasticity of demand and *MC* are known, the optimum price can be determined simply by using equation (9.11). In particular, the equation indicates the following:

(a) If $e = \infty$, $P = MC$

(b) If $AC = MC$ (i.e. $AC$ = constant), $P > AC$ for any value of $|e| > 1$; recall that $e$ is normally negative and no profit-maximizing firm operates at $|e| < 1$.

Under the conditions of perfect competition, $e = \infty$ and (vide Chapter 8, section 8.1) $P = MC$. Thus, the marginal cost pricing leads to economic (profit-maximizing) price under the conditions of pure competition. Also, recall from section 8.2 of Chapter 8 that the profit-maximizing price of the regulated monopolist, where the regulation is such that the ceiling price equals *MC* (which generates the maximum possible output from the monopolist), is the one equal to *MC*. Furthermore, in the case of goods produced exclusively in the public sector, the ones whose production requires huge capital expenditure, and the ones which are deemed essential for human consumption, price is often found to be equal to the marginal cost. Railway, postal and telephone services; and electricity tariffs may very well belong to this last category of goods and services. Incidentally note that the *MC* pricing would lead to profits if *MC* is rising, break-even if *MC* is constant, and to losses if *MC* is falling.

When *AC* is constant (constant cost industry), which implies $AC = MC$, economic price is tantamount to cost-plus pricing; the magnitude of 'plus' depends on the value of $e$. Thus, if $e = -2$ and $AC = MC = 1.5$, $P = 3$ (vide Eq. 9.11). Here, the 'plus' or the mark-up is 100 percent. But if $e = -3$ and $AC = MC = 1.5$ as before, $P = 2.25$, which gives a mark up in price over the full cost (*AC*) of only 50 per cent. Further, if $e = -1.5$, and $AC = MC = 1.5$ as before, $P = 4.5$, which implies a mark-up of 200 per cent. From these examples, it is clear that under the constant cost conditions, the profit-maximizing price equals the cost plus pricing, where the 'plus' is inversely related to the price elasticity of demand.

There is another concept worthwhile here, viz. **Rothschild Index** (RI), which is defined as below.

$$\text{Rothschild Index} = \frac{E_i}{E_f} \tag{9.12}$$

where, $E_i$ = elasticity of demand for the industry's product

$E_f$ = elasticity of demand for the firm's product

Under perfect competition; $E_f$ is (negative) infinity while $E_i$, though also takes a negative value, its magnitude depends on a number of factors (vide Chapter 2) and could vary within its usual values. Thus, RI takes a value of zero under perfect competition. Under monopoly, the two elasticities take the same value, giving RI = 1. In the other two market structures, $E_i < E_f$, for, a firm's product has a larger price elasticity of demand than an industry's product due to the difference in the availability of close substitutes; and accordingly the RI takes a value somewhere between zero and unity. The higher is the competition, the lower is the value of RI. For the Cournot standard duopoly model, the *RI* takes a value equal the inverse of the number of firms in the industry, i.e. RI = 1/*k*, where $k$ = number of firms in the industry. Thus, if there are 5 firms in a standard oligopoly industry (like steel industry), $RI = 0.20$ or $E_f = 5\,E_i$. Given this, if a firm knows that the price elasticity of demand for steel in the country is, say –1.2,

then the price elasticity for the firm's product equals $-1.2 \times 5 = -6$. Further for $e_f = -6$, formula (9.11) gives $P = 1.2\ MC$, which suggests that the firm must price its product at 20 per cent above its *MC*.

From this analysis it follows that the magnitude of "plus" in the cost plus pricing would vary negatively with the price elasticity or alternatively positively with the magnitude of RI. In view of the above relationship between the cost-based pricing and economic pricing, it may be concluded that the former could surely serve a useful tool, particularly for pricing under oligopoly markets.

There is a relevant useful rule here, called the **Pass Through Rule** (PTR). This denotes the sensitiveness of price to a unit change in *MC* which is given by

$$\frac{dP}{d(MC)} = \frac{e}{1+e} \tag{9.12}$$

Under perfect competition $e = \infty$ (infinity) and thus PTR $\leq 1$. In imperfectly competitive markets, PTR value depends on the shape of the demand curve. If the price elasticity of demand were a constant, PTR would $\geq 1$. In a study on the cigarette industry in USA, Summer and Sullivan (*Journal of Political economy*, 1981) found that during 1950s through 1980s, the PTR in that industry was 1.07, which exceeds unity and thus the authors concluded that the US cigarette industry was not perfectly competitive.

While cost based pricing methods appear to be easy and straight forward, they are, in fact, associated with a number of difficulties, as described in Chapter 7. Nevertheless, in the case of an individual firm, these difficulties do not usually pose serious problems; and accordingly the cost-oriented pricing is quite popular today. The method has several strengths as well as limitations. Among the former are, its simplicity, acceptability and consistency with a target rate of return on investment and the price stability in general. Among the latter are the difficulties in getting accurate estimates of cost (particularly of the future cost rather than the historic cost), volatile nature of the variable cost, and its mere implicit consideration of the demand side of the market.

## Going Rate Pricing Model

The going rate, as the name connotes, is the price which is currently prevailing in the market for an identical or similar product. Such a price is often employed, either exactly or with some small plus or minus mark-up, by new firms entering the industry for the first time. Under this model,

$$\text{Price} = \text{Going Rate plus or minus Small Margin} \tag{9.13}$$

The going rate would be the price prevailing in the country if the good is produced domestically and it would equal the effective import price in the home country if the product is available through imports only. The model is practiced because either the producer/seller does know its own costs and/or demand functions reasonably well or it merely initially, at least, desires to induce customers to experience its product. The firm might add a plus to the going rate if it considers its product really superior to the then existing competing products, and it might attach a minus to the going rate if it feels that its product is either very similar to the existing competing goods or because it wishes to provide an incentive to its potential customers. To the method's credibility, we may mention that we do come across cases where firms distribute their new products even free to the select target group of their potential customers. The method is particularly useful under oligopoly markets due to intense competition and mutual interdependence of firms. In monopolistically competitive markets, the competition among

differentiated products is tough and demands are highly volatile and thus the going rate price has a place in product pricing techniques. Since the method focuses on competition, it is also known as a competition-based pricing method. To cite some examples, when television was first manufactured in India, its import cost must have been a guiding force in its price determination. Similarly, when Maruti Car was first manufactured in India, its price must have taken into account the prices of existing cars, price of petrol, price of car accessories, etc. Recall that the going rate price could be below or above the average cost, and it could even turn out to be an economic price. Another competition based price is the **sealed bid pricing**, where the seller invites bids from the prospective buyers in sealed covers and the one who bids the highest get the product. This is practiced particularly in old goods trades, vehicles damaged in accidents, stock markets, agriculture crops and products, etc. In some cases, the bids are even open ones.

### Price Penetration and Price Skimming

Another pricing method goes in the name of price penetration and price-skimming. Under the former, firms sell their new products at a low price in the beginning in order to catch the attention of consumers. Once the product image and credibility is established, the seller slowly starts jacking up the price to reap good profits in future. Under this strategy, a firm might well sell its product below the cost of production and thus runs into losses to start with but eventually it recovers all its losses and even makes good overall profits. The Hindustan Levers' RIN washing soap perhaps falls into this category. This soap was sold at a rather low price in the beginning and the firm even distributed free samples. Today, it is quite an expensive brand and yet it is selling very well. Under the price-skimming strategy, the new product is priced high in the beginning, and its price is reduced gradually as it faces a dearth of buyers. Such a strategy may be beneficial for products which have attractive features and the prospective buyers have relatively price inelastic demands. Once the market for the product is saturated, adequate capacity has been built up and economies of scale become available, the price is reduced. Many electronic goods including calculators, laptop computers, televisions, iPods, iPads, iPhones, etc. may have been subjected to such pricing methods.

A prudent producer follows a good mix of the various pricing methods rather than adapting anyone of them. This is because no method is perfect and every method has certain good features. Further, a firm might adopt one method at one time and another method at some other occasion. To conclude this long section on pricing in oligopoly markets; recall that profit-maximizing prices are not determinate under oligopoly unless some understanding, formal or informal, is reached among the competing firms, or when the non-cooperative games among oligopolists yield definite outcomes. Also, we have examined the difficulties which crop up in reaching and maintaining such understanding, and these are surmountable, particularly in the case of differentiated oligopoly. Further, today there are many more cases of differentiated oligopoly than the standard (homogeneous) oligopoly. In view of these, pricing in oligopoly markets is indeed a difficult job. The oligopolists know rather well that if they compete among themselves through price, the price would go no falling unabated and all of them would make small profit, lose or just break-even in the process. Thus, instead of competing through price, they compete through non-price means, like the firms under monopolistic competition. These include product quality, service to the customers, fancy packaging, selling on credit, and guarantees against their products breakdown until a stipulated period, and so on. In view of this, today we find varieties in most oligopolists products and there is a lot of promotional activities for such products all around the globe. In consequence, there is a lot of obsolescence in these products and shopping has become quite an art.

## 9.7 PRICE RIGIDITY

The salient feature of oligopoly prices is that, they are rigid/inflexible/stable in relation to prices under pure competition, pure monopoly and monopolistic competition. The basic rationale for this is found in the firms' desire to avoid mutually destructive price competition and the 'kinky' demand curve approach to the oligopoly market.

Paul Sweezy (1939) argued that the demand curve facing an oligopolist has a kink at the current price. This, according to him, is because of the asymmetric behaviour of firms in oligopolist markets. He hypothesized that when a firm under oligopolist market lowers its product price, the rivals also lower their product prices; but when it raises its product price, the rivals do not raise their products prices. Such behaviour exists because if firms do not react to price reductions by their rivals, they lose a significant part of their market to their rivals who have lowered their prices; but if they do not react to their rivals' price increases, they gain significantly the market from their rivals and thereby enjoy an enlarged market for their own products. Since a cut in price is followed, the price cutting firm is unable to secure a significant increase in the demand for its product through its price reduction moves. In contrast, since an increase in price is not followed; the price increasing firm faces a significant decrease in the demand for its product when it raises its price. Thus, in either case, firms have little incentive to change their prices and accordingly they just avoid what is called the mutually destructive price wars. Under the situation, every oligopolist firm faces two demand curves, viz. one, for price decreases, which is less price elastic; and the other for price increases, which is more price elastic. Corresponding to these two demand curves, there are two marginal revenue curves. These curves are drawn in Fig. 9.3.

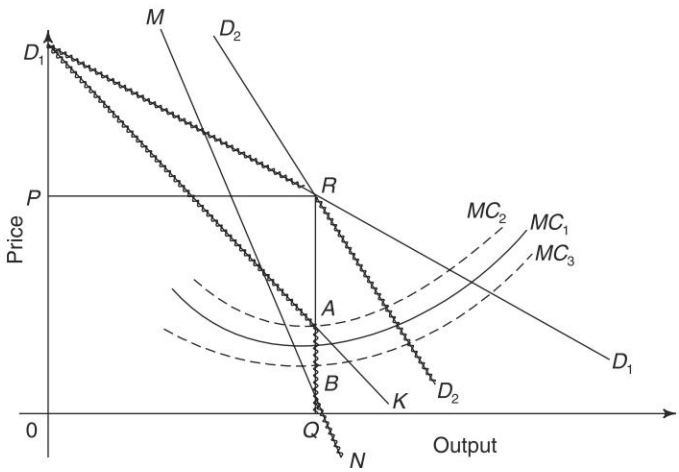



**Fig. 9.3** *Kinky demand curve and equilibrium under oligopoly*

In Fig. 9.3, *OP* is the current optimum price and *OQ* is the corresponding quantity currently produced by the oligopolist. The curve $D_1 D_1$ is the demand curve valid for price increases and $D_2 D_2$ that relevant for price decreases. Thus, the effective demand curve facing the oligopolist is $D_1 R D_2$, which has a kink at point *R*, i.e., at current price *OP*. Corresponding to demand curve $D_1$, $D_1$, the $D_1 K$ line is the marginal revenue curve, while corresponding to $D_2 D_2$, the *MN* line is the marginal revenue curve; and the effective *MR* curve corresponding to the effective demand curve $D_1 RD_2$ is the curve marked $D_1 ABN$. Given these effective demand and *MR* curves price rigidity is easy to explain.

There are two reasons for price rigidity under oligopoly in relation to price under other market conditions.

(a) Under oligopoly, the demand curve is kinky while in other market structures it is continuous. Thus, an oligopolist would avoid reducing its price, for if it did, its rivals would follow suit and so it would not be able to gain much of the new market. Similarly, it would avoid a raise in price, for if it did its rivals would not follow suit and so it would lose its existing market significantly. Since the demand curve facing non-oligopolist firms is continuous, there is no such asymmetry and consequently no such aversion against price changes. This phenomenon result could be explained through the oligopolists' *MR* curve as well. As explained above, the kinky demand curve implies a discontinuous *MR* curve under oligopoly, in contrast to a continuous *MR* curve in other market structures. Thus, in Fig. 9.4, the effective *MR* curve has a break (gap) in the region *AB*. Due to this break, even if there is a change in the cost within a given range, the profit-maximizing price would remain invariant. Thus if $MC_1$ is the relevant *MC* curve, the profit-maximizing price equals *OP*. Now if the cost rises such that the new *MC* becomes $MC_2$, the equilibrium price remains the same at *OP*. Similarly, if *MC* curve shifts downward to $MC_3$ due to a fall in the cost, there is no change in the equilibrium price. In contrast, in non-oligopoly markets, since *MR* curve is continuous, any change in *MC* curve would cause a change in the profit-maximizing price. Incidentally note that the equilibrium price would change even under oligopoly if the change in *MC* is significant, such that the new *MC* curve cuts the effective *MR* curve at some point away from its gap (=*AB*).

(b) Unlike non-oligopoly markets, under oligopoly, firms' fortune depends mutually on their rivals actions and reactions. Thus, as was obvious in prisoners dilemma above, each firm tries to avoid mutually destructive price competition.

To illustrate price rigidity, we consider a hypothetical example. Suppose the two demand curves facing a firm in an oligopoly industry and its cost functions are as follows:

Demand equation for price increase (vide $D_1 D_1$ curve): $Q_1 = 40 - 2P$

Demand equation for price decrease (vide $D_2 D_2$ curve): $Q_2 = 25 - P$

Cost equation: $TC = 10 + 25Q - 1.5Q^2 + 0.04Q^3$

The corresponding *TR*, *MC* and *MR* equations will be the following:

$TR_1 = 20Q_1 - 0.5Q_1^2$ $\quad TR_2 = 25Q_2 - Q_2^2$ $\quad$ and $\quad MC = 25 - 3Q + 0.12Q^2$

$MR_1 = 20 - Q_1$ $\quad$ and $\quad MR_2 = 25 - 2Q_2$

The equilibrium price and output are given by the point of intersection between the two demand curves. Thus,

$$40 - 2P = 25 - P$$

or, $\quad P = 15 \quad$ and $\quad Q = 10$

Note that the demand curve corresponding to price increase is more elastic than that to price decrease. The effective (kinky) demand curve is one part of each of the two demand curves:

$Q = 40 - 2P$, if $P \geq 15$ $\quad$ and $\quad Q = 25 - P$, if $P \leq 15$

And the effective (discontinuous) *MR* curve is also one part of the two *MR* curves:

$$MR = 20 - Q_1, \text{ if } Q \leq 10 \qquad \text{and} \qquad MR = 25 - 2Q, \quad \text{if } Q \geq 10$$

And the points *A* and *B* on the discontinuous part of *MR* curve (vide *AB* in Fig. 9.3), where $Q = 10$, have the values for marginal revenues equal 10 and 5, respectively. Thus, the *MR* curve is discontinuous in the price range 5 to 10. Further, the value of marginal cost at $Q = 10$ (equilibrium value) = 7. As $MC = 7$ lies between $MR = 5$ and $MR = 10$, the *MC* curve intersects the effective *MR* at the latter's discontinuous part. Thus, if *MC* moves within the range of 5 to 10 at $Q = 10$, the equilibrium points stay at $Q = 10$ and $P = 15$. However, if *MC* increases beyond 10 or decreases below 5 at $Q = 10$, the equilibrium price and output would undergo a change. In view of this, in this example of an oligopolist, the firm's optimum price is rigid but not fixed. The same conclusions can be demonstrated through tabulation of various results as well. The same are provided in the table below.

| $P$ | $Q_1$ | $Q_2$ | $TR_1$ | $TR_2$ | $MC$ | $MR_1$ | $MR_2$ |
|---|---|---|---|---|---|---|---|
| 25 | 0 | 0 | 0 | 0 | | – | – |
| 20 | 0 | 5 | 0 | 100 | 13 ($Q = 5$) | – | 15 |
| 15 | 10 | 10 | 150 | 150 | 7 ($Q = 10$) | 10 | 5 |
| 10 | 20 | 15 | 200 | 150 | 13 ($Q = 20$) | 0 | –5 |
| 5 | 30 | 20 | 150 | 100 | 43 ($Q = 30$) | 15 | –15 |
| 0 | 40 | 25 | 0 | 0 | | –20 | –25 |

It will be seen that the effective *MR* curve is discontinuous at $Q = 10$ and between price/*MR* equals 5 (on $MR_2$) and 10 (on $MR_1$). At $Q = 10$, $MC = 7$. Thus, *MC* between the range of 5 and 10 would always equal the effective *MR* curve. Accordingly, any change in *MC* between its values 5 and 10, the optimum price-output mix would remain invariant at $P = 15$ and $Q = 10$. If *MC* changes beyond the range 5 to 10, the optimum price and output would undergo a change. This shows that the kinky demand curve causes price rigidity but not price stability.

This concludes the long discussion on oligopoly prices, both as to how they are determined and why they are relatively rigid. Before we proceed to the next section, a brief commentary on the virtues and shortcomings of the oligopoly market is in order. The **pros of oligopoly markets** include the availability of varieties of products, their quality, relative price stability, exploitation of economies of scale, and both the willingness and ability of the oligopolists to carry on research and development (*R* and *D*) activities and to introduce new products. The first three virtues have been argued for above, the fourth comes from their large size, and the last virtue is simply because the oligopolists generally make profits so they have money to spend on *R* and *D*, and since there is a tough competition among them in the form of non-price factors, including product quality, they have enough incentives to keep up the quality of their products and to introduce new products. The **cons of the oligopoly market** are similar to the ones about the monopolist and monopolistically competitive markets. These include market inefficiency or the socially sub-optimal output (for $P > MC$), the so called dead weight loss, the usual existence of the excess capacity ($P >$ Min. *AC*), social wastes through promotional activities and mutually destructive competition and confusion among customers about the product quality (not found under pure monopoly).

## REFERENCES

1. Baye, Michael (2009): *Managerial Economics*, McGraw Publishers.
2. Cohen, K.J. and R.M. Cyert (1965): *Theory of the Firm*, Englewood Cliffs, Prentice-Hall.
3. Fellner, W. (1949): *Competition among the Few*, New York, Knopf.
4. Friedman, J.W. (1986): *Game Theory with Applications to Economics,* New York, Oxford University Press.
5. Gibbons, Robert (1992): Game Theory for Applied Economists, New Jersey, Princeton University Press.
6. Nash, John (1951): Non-Cooperative Games, *Annals of Mathematics*.
7. Pindyck, Roberts and Daniel L.Rubinfeld (2009): *Microeconomics*, 7th edition, Pearson.
8. Rothschild, K.W. (1947): "*Price Theory and Oligopoly*", Economic Journal, XLII.
9. Sweezy, Paul (1939): Demand under Conditions of Oligopoly, *Journal of Political Economy*.

## CASELETS

1. There are just two firms in shampoo business, known as Batliwala ($B$) and Talchiriwala ($T$). They sell shampoo under their own trade-marks and aim at maximum profits. The demand functions for their products are as follows:

$$P_B = 65 - Q_B - Q_T$$
$$P_T = 120 - 2Q_B - 3Q_T$$

where, $P_i$ = price of firm i and $Q_i$ = Output/demand of firm $i$

The total cost functions of the two firms are as follows:

$$TC_B = 40 + 4Q_B + Q^2{}_B$$
$$TC_T = 60 + 3Q_T + 2\,Q_T{}^2$$

Where, $TC_i$ = total cost of firm $i$

(a) Determine the equilibrium prices and quantities of the two firms under the Cournot assumption.

(b) Would a collusion arrangement between the two firms be advantageous to the firms? If so, determine the equilibrium prices and quantities under the cartel.

(c) Suppose the two firms reach an understanding and decide that Batliwala would be the leader in setting the price but ensure 1/3 market share for Talchiriwala. Determine the equilibrium prices and quantities in such a market. Verify if the first-mover advantage exists in this solution.

(d) Suppose the two firms decide to work on the Stackelberg–Cournot model under the leadership of Talchiriwala, what would be their optimum prices and outputs? Check if the first-mover advantage is available to Talchiriwala.

(e) Compare your results in all the above alternative models and comment on the differences.

2. Consider the example under the Bertrand model in the text (vide after Eq. 9.11) and attempt to answer the following questions.

(a) If the two duopolists agree to form a Cartel, what prices and quantities they would set for their products?

(b) If the two duopolists enter into an implicit collusion, decide that Firm-1 would be the leader and Firm-2 the follower, and follow the Stackelberg-Bertrand model of pricing, determine the equilibrium prices and quantities for their products.

(c) Check for the presence of the first mover advantage, or disadvantage, or neither under the model.

(d) Compare the above results under different hypotheses and comment.

3. The steel industry is characterised by homogenous oligopoly with 11 firms. Of these, 10 are small and one is large. The large firm (Aditya Steel) acts as the leader and all the rest as followers in price setting. The leader sets the price, allows the followers to sell their outputs first and thus taking the residual market. Aditya Steel's total cost ($TC_A$) function is

$$TC_A = 500 + 0.1Q_A^2$$

and all the small firms (followers) have the identical total cost function:

$$TC_B = 25 + 0.5Q_B^2$$

The market demand for the product is given by

$$P = 210 - Q$$

If each firm aims at profit-maximization, determine the following.

(a) The price of steel.

(b) Total steel output that all the 11 firms would sell.

(c) The distribution of total sales between the leader and the followers.

4. There are just two firms, called Rohan Enterprises Ltd. (REL) and Varun Enterprises Ltd. (VEL) in the cable service industry. They compete through advertisement budgets via a simultaneous game of pure strategies. If neither firm advertises, REL would make Rs. 5 millions of profits and VEL would make Rs.10 million of profits. If both pursue advertisements, they would each make Rs. 15 million of profits. If VEL does not advertise and REL does, VEL would earn Rs. 8 million and REL would earn Rs. 25 million. Under the last event when REL does not advertise and VEL does, REL would earn 3 million and VEL would earn Rs. 9 million.

(a) Express the game in normal (matrix) form

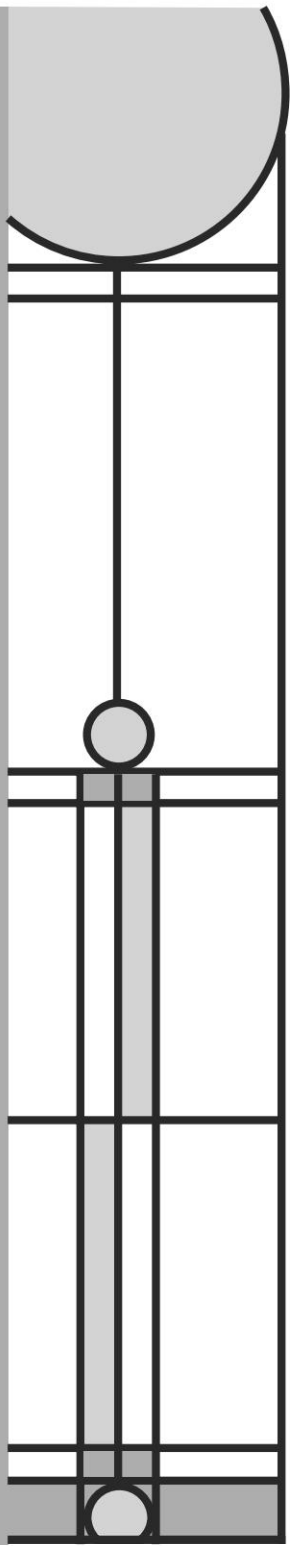
(b) Does REL have a dominant strategy? Does the game have a dominant equilibrium? If yes, which one?

(c) Does VEL have a secure strategy? Check if the game has the saddle point equilibrium. If yes, which one?

(d) Determine the Nash equilibrium of the game. Suppose you were the CEO of REL. Does the Nash equilibrium give you the best available result? If not, how much would you be willing to bribe your rival so that you maximize your pay-off?

5. If the game of Caselet 4 above was conducted in its dynamic (sequential) form in which REL were to act as the leader, write the game in the decision tree format. In the new form, does the game have equilibrium? If yes, which one? Examine if the game has the first-mover advantage.

The McGraw-Hill Companies



CHAPTER

10

# PRICE–OUTPUT DETERMINATION-III

## (PRICING UNDER MARKET POWER, ASYMMETRIC INFORMATION, EXTERNALITIES AND RISK)

In the previous two chapters, pricing was discussed under various forms of market structure under the usual assumptions, viz.

(a) The law of one price is practiced

(b) Information about the market is available uniformly to all buyers and sellers

(c) Costs and benefits associated with a good are available to producers and consumers only

(d) Demand and cost functions pertaining to the good are known with certainty

In addition, the foregoing analysis implicitly assumed that the product is priced as a whole and not in parts (like entry/membership fee and user charge), and that each product is priced separately and not in a given bundle of products (tying or bundle of products). In the real world, none of these assumptions hold in pricing a number of goods and services. Accordingly, it is imperative for us to understand where and why each of these assumptions is violated; and if violated, how the optimum prices are determined in such cases; and what consequences they lead to for the firms, consumers and the society. This is the subject to which this chapter is devoted to. We would take each of the above assumptions one by one in an appropriate order and examine each of the relevant issues in this regard.

### 10.1 MARKET POWER

The assumptions of the law of one price, and products being priced as a whole and individually only do not hold under the situations in which the firm enjoys, what is called, the **market power**? It means the power to set price! Recall that under perfect competition, an individual firm is merely a price taker as it faces a given but horizontal demand curve for its product, and thus has no power to set the price of its own product. In all other market structures, the firm has some power in this regard. The power is not absolute, for it faces a given but downward demand curve for its product, which implies that the firm can

set either the price or output of its product but not the both. It is obvious that the market power (or the **power to set price** of own product) is tantamount to a **falling demand curve** facing the producer. Why uniform pricing is unsuitable under market power can best be explained through the graphs in Fig. 10.1.

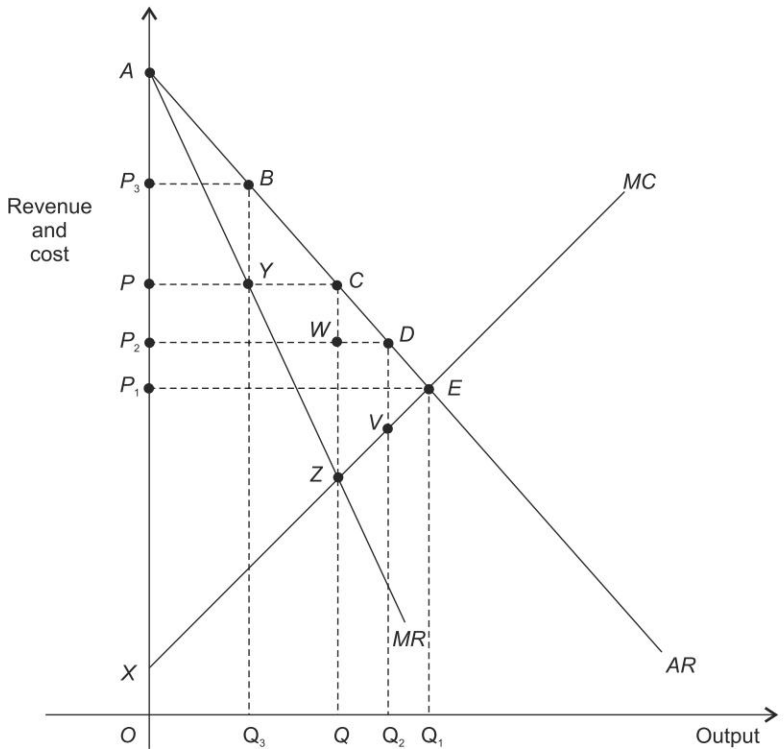



**Fig. 10.1** *Incompatibility of Uniform Pricing and Market Power*

Figure 10.1 corresponds to price-output decisions under all non-perfectly competitive markets. For simplicity, it assumes linear demand and *MC* curves and ignores the kink in the demand curve which exists under oligopoly. As explained in previous chapters, under (simple) monopoly, the optimum price and output would be *OP* and *OQ*, and resulting consumer surplus and producer surplus would equal the area of triangle *APC* and the area within *PCZX*, respectively. As the demand curve facing the firm is the falling curve, the firm has market power. If this power is used, it could collect its potential profit which is otherwise lying on the table! In other words, the firm could add to its profits if it departs from uniform pricing rule and/or does not restrict its output to *OQ* level. How? For example, it could sell $OQ_3$ output at price equals $OP_3$, $Q_3$ *Q* output at price equals *OP*, and $QQ_2$ output at price equals $OP_2$. If it did so, producer surplus (*PS*) would increase and consumer surplus (*CS*) would decline, and thereby the firm would be able to capture a part of *CS* which was otherwise lying on the table. What would be these changes in *CS* and *PS* as per Fig. 10.1? Let us look at these first in terms of consumer surplus. For the first output level of $OQ_3$, the *CS* would decline from area within *PABY* to area of triangle within $P_3BA$, a net fall equivalent to area within rectangle $P_3BYP$. This measures the captured part of *CS* which is now transferred to *PS*. Furthermore, on the additional output being sold under non-uniform pricing over

uniform pricing, which equals $QQ_2$ at price equals $OP_2$, the firm would have additional *PS* equivalent to area within *WDVZ* and consumers would have additional *CS* equivalent to area of triangle *CDW*. Thus, as the result of non-uniform pricing in comparison to uniform pricing, *PS* goes up by areas within $P_3\ BYP + WDVZ$ and *CS* changes by areas $CDW - P_3\ BYP$. Accordingly, producers gain for sure and consumers may gain or lose depending on whether area of *CDW* exceeds or fall short of area $P_3\ BYP$. In other words, producer is able to capture a part of *CS* through violating the law of one price. Further, the social welfare (sum of *CS* and *PS*) may, in fact, be more under discriminating than simple monopoly, due to the additional sales under the former. If this were so, price discrimination could even reduce market inefficiency (dead weight loss) of monopoly market.

Our above illustration still leaves scope for additional capture of *CS* and more of *PS*, as we have restricted to just three prices and outputs up to $OQ_2$ even though scope exists for more number of prices as well as additional outputs until output reaches close to the level of optimum output under perfect competition, i.e. $OQ_1$. We now turn to this common practice of non-uniform pricing called price discrimination which, among other practices like two-part pricing and commodity bundling, prevails under market power.

## 10.2 PRICE DISCRIMINATION

Economists do not believe that money was available so easily and business people are ever ready to exploit any such opportunity, like money lying on table. Accordingly the former recommends and the latter follows price discrimination and other practices, such as two-part pricing and commodity bundling, when a firm enjoys some market power. The law of one price states that the price of an item is the same no matter who buys it, for what purpose and where, subject to differences in transaction cost, if any. This is due to the otherwise availability of arbitrage opportunities. For example, suppose petrol is available at Rs. 50/liter at location Vastrapur to Mr. *X* (and anyone else) and at Rs. 60/liter at location Ambli to Mr. *Y* (and others), then some arbitrator *Z* would buy petrol at Vastrapur and sell the same at Ambli, making a profit of Rs. 10/liter minus the transaction cost. In consequence, demand for petrol at Vastrapur would increase and the supply of petrol at Ambli would increase, causing petrol price to rise at Vastrapur and fall at Ambli. The process will continue until the two prices net of transaction cost equalize. If this were true, the assumption of the law of one price, called uniform pricing, would hold. If it were not true, we would have the opposite phenomenon, called, price discrimination. Thus, price discrimination is the situation under which an **identical product is sold at different (more than one) prices, not distinguished by cost differences**. In the real life, we do come across cases where price discrimination is practiced. This is so, firstly because practicing firm possesses **market power** and secondly because **arbitrage is not permitted** due to demarcation of markets and no resale across markets. In particular, three conditions must be met for price discrimination to be **feasible**. These are

(i) Firm enjoys some market power
(ii) Market is clearly separable into two or more parts through some criterion
(iii) No resale of the product in other market segments must be possible

Only under perfect competition, firm has no market power and thus no price discrimination is possible in such a market. The power is measured by the Lerner's index (vide Chapter 8, Eq. 8.7). It is at maximum under monopoly and so price discrimination is most common in such markets. By this criterion, discrimination is also available both under monopolistic competition and oligopoly. If the

market was not divisible into distinct categories, the firm would not be able to decide as to whom to charge what price, which, in turn, would lead to confusion and defeat the practice of price discrimination. Similarly, if resale of the product was possible, then, as explained above under the law of one price, the arbitrators could make profits by buying the product in the market where the price was low and then selling the same in the other market where the price is high. This would lead to the movements of the product from the low price market segment to the high price market segment, affecting the supply and demand of the good in the two market segments in the direction to correct the price differences. The process would go on until the price difference was eliminated and would thus defeat the price discrimination itself. Due to these three pre-requisites, price discrimination is practiced largely between domestic and foreign markets, for the two markets are distinctly separable; and by the suppliers of services, for services cannot be resold. Like the urban and rural areas are distinctly separable, males and females are separable, students and non-students are separable, rich and poor are separable, children, adults and senior citizens are separable, agricultural and non-agricultural use of a product are separable, household and commercial use of a product are separable, days and nights are separable, week-days and week-ends are separable, and so on, and correspondingly separate prices for them are also feasible.

The market is divided into two or more parts in two different ways, viz. (i) on the criteria such as geographical location, product usage, and demographic characteristics of the customers; and (b) on the basis of the time the customer buys the product. These distinctions are neat and the market powered firm, assuming government rules permit, could force no resale across the part so divided. Two kinds of price discriminations are, thus, distinguished, viz.

(a) Price discrimination by market segments

(b) Inter-temporal price discrimination

Under the first type, discrimination is at the same time across **market segments**; and under the second category, discrimination is at the same market segment across (over) **time.** This distinction is alike the distinction between the cross section and time series data. Accordingly, the pricing through price discrimination of the first category could be treated as the **static price discrimination**, and that under the second category, along with the peak-load pricing (to be explained later) as the **Dynamic Pricing**.

Feasibility of price discrimination would transfer to the practice of price discrimination only if it is found profitable by the firm. Thus, we need to examine the additional requirements, if any, for the profitability of this practice. Profit from production/sales depends on demand (or revenue) and cost functions the firm faces for the product. In view of this, either the cost or demand function must differ across the markets for price discrimination to be profitable. Under price discrimination by market segments, product is sold at the same time in various market segments. Thus, if the product is homogeneous, it would have the **same cost** no matter where it is sold. Further, if the product is differentiated, cost would, of course, differ but for discrimination to occur, the **price difference must exceed the cost difference**. Similarly, under inter-temporal price discrimination, cost could surely vary even for homogeneous good over time (through inflation/deflation in cost items) but once again for price discrimination to happen, the **price difference must exceed the cost difference**. Thus, for price discrimination to exist, mere presence of cost difference is not enough; differences must be found in demand equation. Demand across market segments and/or time could differ in two ways, viz.

(a) It could be less price elastic at one market segment/time and more price sensitive at other market segment/time, and

(b) It could be low in one market segment/time and high at other market segment/time. Recall that by demand more (less) we mean, demand is more (less) at each price (Chapter 2); i.e., the more or less is due to difference in any non-own price determinants of the demand for the good in question. Further, low demand implies more price elastic demand, and high demand implies less price elastic. This can be demonstrated through an example. Consider two demand equations for a product, viz.

$$Q = 25 - 2P \quad \text{(i)}$$

and

$$Q = 40 - 2P \quad \text{(ii)}$$

Equation (i) has low demand than Eq. (ii), for at any price the demand in the former exceeds that in the latter by 15. Further, the price elasticity of demand in (i) equals $(-2)[P/(25 - 2P)]$ while that in (ii) equals $(-2)[P/(40 - 2P)]$. Obviously, in absolute term, the said elasticity is larger in Eq. (i) than in Eq. (ii).

In view that high demand means low price elastic and vice versa, the above condition (b) is contained in condition (a). Thus, it can be concluded that the pre-requisite for price discrimination across market segments or time period to be **profitable**, the additional condition (over the above three feasibility conditions) is that

(iv) Price elasticity of demand must be different in different markets

A more concrete proof of this condition would be provided later. We now turn to the two types of price discriminations individually and in detail.

## 10.3 PRICE DISCRIMINATION BY MARKET SEGMENTS

Recall that under price discrimination by market segments, markets are classified on the criteria such as geographical location, product usage and demographic characteristics of the customers. The geographical location-wise classification consists of domestic vs. foreign market, urban vs. rural areas, near the factory site vs. other places, and so on. Of course, due to liberalization and globalization, discrimination across countries has come down but it still exists at least to some degree. Also, one offer hears that so and so country is selling dearly in foreign markets than in domestic market, and/or dumping its certain goods in foreign markets. Product uses are classified by household vs. commercial use, agricultural vs. industrial use, charity vs. education vs. other uses, etc. Demographic factors include sex, age, caste, religion, income group, student vs. non-student, etc.

Under price discrimination by market segments, a firm endowed with market power charges different prices for its unique quality product in different segments of its market which are not distinguished by the cost differences to the firm. Such discrimination could be practiced in different degrees. A.C. Pigou has classified them in three degrees as follows:

(a) First degree (perfect) price discrimination

(b) Second degree price discrimination

(c) Third degree price discrimination

Under the first degree price discrimination, also called the perfect price discrimination, the firm charges different price for (from) each unit (different customer) of its product. Thus, it takes the highest price for the first unit and it goes on reducing the price for every successive unit that a customer buys from it. This is so because, by the law of diminishing marginal utility, a consumer derives less and less

satisfaction from the successive units of a good as he/she goes on consuming more and more units of that good, and thus unless he/she is offered the successive units at falling prices, he/she would not purchase the additional unit of a good. In other words, the maximum price that a customer would pay for a unit of the good equals the marginal utility (MU) that he/she derives from its consumption. The maximum price referred to here is called, the **customer's reserve price** for the corresponding unit of the good. If MU is expressed in terms of money, then the points on the demand curve denote the customer's reserve prices for the successive units of the good. If the firm knows this reserve price and enjoys the market power, it may be able to set, **price = reserve price**, and thereby practice perfect price discrimination, leaving no consumer surplus. By doing so, the firm (the monopolist only) takes full advantage of its monopolist position and thus exploits the consumers to the maximum possible extent. This is the extreme form of discrimination and it is rarely found to prevail anywhere today. When reserve price cannot be assessed, prices are set by trial and error. We do sometime come across cases where the seller offers the first unit at a higher price and the subsequent units at lower and lower prices; and surely there are many cases where price varies across customer, i.e., **personalized pricing,** particularly in services. Consultants and experts in critical fields like medicine, law, management, etc. are often blamed to be practicing such tactics. Universities charge personalized tuition fees through the use of financial aid packages. By practicing such price discriminations, these firms enable the otherwise non-affording customers to avail their facilities and even to earn larger their profits through larger sales than otherwise.

Under the intermediate form of price discrimination, the monopolist offers quantity discount which varies positively with the size of the purchase/order. This form of discrimination is quite prevalent today. We have wholesale vs. retail prices for a number of commodities and even price variations across regular and ad hoc customers. One often comes across of lots of advertisements, stating "buy one gets another free". Electricity boards have a minimum charge for the first certain units of electricity consumed and then a reduced/enhanced rate for the extra units used. The railways' fare per kilometre gets reduced as the passenger/goods travel distance increases. Thus, one train ticket of a particular class between Ahmedabad and Mumbai is less than the two train tickets, one between Ahmedabad and Vadodara and another between Vadodara and Mumbai; Vadodara falls en route from Ahmedabad to Mumbai. Although, Indian Railways might incur a little extra cost (in terms of vouchers or tickets) in case of selling in smaller quantities vis-a-vis larger quantities, the price differences are much more than accounted for by the cost difference, and hence this is still the case of price discrimination. In addition to such examples, there are cases of **block pricing,** which fall in the group of second degree price discrimination across market segments. Under one form of block pricing, up to a certain number of units, say, 1–5, are sold at the highest price, 6–15 units at a little lower price, and units in excess of 15 are sold at the lowest price, and so on. Prices per unit are lower at Costco and Sam club than at Wal-Mart and other stores. Through the practice of this kind of price discrimination, firms reduce their transaction costs, experience larger sales and thereby earn larger profits than otherwise.

The price discrimination of the third degree is the **most popular** form of price discrimination prevalent in most societies. Under this form, different prices are charged for the same homogeneous product from different customers depending upon various factors such as geographical location, product usage, sex, age, caste, religion and income group of the customer. Thus, there are several goods, whose prices depend upon the place they are bought. Cigarette manufacturers have historically charged much higher prices in US than in foreign markets, particularly for their premium brands. The price of Maruti Zen 2010 model car, for example, in India is lower than its price in the foreign market. Of course, there are differences in distribution (transport, insurance, customs, etc.) cost but very often the price

differences are much more than the cost differences. Also, there are cases, called dumping, where a product is sold at a high price in the home country and at a relatively low price in a foreign country. For example, during the late Seventies India had a bumper sugar cane crop, there was a glut in the domestic market and the Indian sugar was sold at a lower price in foreign markets than in India.

Prices are also found to vary with product use. For example, if electricity is used for illumination, its tariff could be higher than if it is to be used for energizing kitchen appliances. Similarly, electricity charges are higher for its uses in industry/commerce than for its uses in agriculture or household. Paper price for news papers use is lower than for its use in luxury hotel's napkins, cement price for house construction use is lower than its price for commercial/factory building uses, land price for charity/education/hospital uses are lower than for household/commercial uses, LPG price for household cooking use is lower as compared to its uses in commercial/industrial activities; and so on.

Price is also found to vary with age, sex, income group, student-non-student category, etc. to which the customer belongs. Air, train and bus fare depend upon whether the passenger is under or above a certain age (like 12 years). Senior citizens get a discount in income tax, train journeys, banking services, movies, etc.; females get a discount in income tax, college fees in some places, sports' tickets, etc.; students get a discount in travelling to homes during vacation, subscriptions to journals/magazines, etc.; poor customers get discounts in ration shops, hospital charges, college fees, consulting/advocate fees, and so on. In Gujarat, girls pay no tuition fees for school and college education. Government hospitals often charge from patients on the basis of the income category to which he or she belongs. Indian railways have different rates for second class and first class, and non-air conditioned and air conditioned accommodation. Airlines have different rates for first class, business class and economy class. In all these, though costs differ, the price differences are much more than the ones accounted for by the cost differences. Similarly, Indian railways and airlines offer concessions to students for some pre-specified travelling, and professionals like physicians, advocates, chartered accountants, engineers and professors often offer their services at lower rates to poor and middle income people than they do to rich people. The IIMs and IITs are practicing such tactics to some extent in fixing their charges which are negatively related to the family incomes. Many companies offer discount/rebate coupons on their merchandises. Customers who take trouble to clip and carry/mail such coupons are able to get the discounts while others pay the regular price, which tantamount to price discrimination. Studies show that only about 20–30 percent of all customers regularly bother to clip, save and use coupons. Also, salespersons do entertain bargaining on price and depending on the bargaining ability; different customers buy the same product at different prices. Such bargains are particularly popular in white goods, like cars, televisions, computers, mobile phones, etc. In real life, firms offer even **price matching**, under which they guarantee the customer that the price charged by them is the minimum in the market and in the event the customer finds a lower price anywhere now or until some fixed period (one month or so), the selling firm would refund the difference. By this tactic, firms tend to increase their sales and only those customers who search and succeed finding lower price elsewhere are able to collect the price difference, if any. In consequence, firms get the information about prices elsewhere and succeed in practicing price discrimination.

Thus, there are plenty of examples of price discrimination of the third degree. The rationale for its existence lies in the principles of social justice and/or profit-maximization. Although it is rather hard to identify the theory for a particular case of price discrimination, it is generally believed that government hospitals, government-run schools and colleges, railways, electricity boards (for discrimination between agriculture and industrial uses), etc. practice price discrimination on the premise of social justice. In

contrast, professionals, sport organizations, multinational organizations, banks, airlines, etc., are believed to be guided by profit motives behind their discriminations.

**Optimization** Under price discrimination across market segments, the decision issues confronting the practicing firm are

(i) How much to produce?

(ii) How to distribute the total production in various market segments?

(iii) What prices must be charged in different market segments?

Questions (ii) and (iii) are interdependent, i.e., if one is decided, the other is known. This is so because the firm faces given demand schedules in its various markets and it could set either the quantities of its product or the prices of its product in various markets. Thus, if the firm sets its outputs, the prices are given by the demand schedules; and if it sets prices, outputs are known from the respective demand schedule. The firm faces as many demand functions as the number of markets it serves, one for each market. However, since all the markets are supplied by the single firm, its total production for all the markets may take place at a single plant located in just one market. Further, if the product sold in different market segments is homogeneous, as it is under monopoly, the firm would have just one cost function. For a discriminating monopolist, the problem of optimization can thus be pursued through **calculus** as follows:

$$\text{Maximize} \qquad \pi = TR - TC = [TR_1 + TR_2 + TR_3 + \ldots\ldots + TR_k] - TC \tag{10.1}$$

With respect to $Q_1, Q_2, Q_3, \ldots\ldots\ldots, Q_k$ , if $TR$ and $TC$ 's are expressed in outputs ($Q_1, Q_2, Q_3, \ldots\ldots\ldots, Q_k$); or with respect to $P_1, P_2, \ldots\ldots\ldots, P_k$ , if they were expressed in prices ($P_1, P_2, \ldots\ldots\ldots, P_k$). In the above equation, $TR_i$ = total revenue in market-$i$, $Q_i$ = equilibrium output in market-$i$, $P_i$ = equilibrium price in market-$i$, $TC$ = firm's total cost, which depends on its total output ($Q = Q_1 + Q_2 + Q_3 + \ldots\ldots\ldots + Q_k$), and $k$ = number of market segments.

It is immaterial whether the firm chooses outputs or prices, as both lead to identical results. Usually, functions like (10.1) are in outputs and we follow the same procedure. To obtain the optimum results, the first derivatives with respect to total output as well as each output level must be zero (besides the sufficient conditions, which are ignored here for simplicity). The one with respect to total output would yield,

$$\frac{d(TR)}{dQ} = \frac{d(TC)}{d(Q)}$$

or, $$CMR = MC \tag{10.2}$$

And the others with respect to each of output levels will give,

$$MR_1 = MC \qquad MR_2 = MC \qquad MR_3 = MC\ , \ldots\ldots\ldots, \text{and} \qquad MR_k = MC$$

or, $$MR_1 = MR_2 = MR_3 = \ldots\ldots\ldots = MR_k = MC \tag{10.3}$$

In these equations $CMR$ = combined marginal revenue, $MR_i = MR$ in market-$i$, and $MC$ = marginal cost.

Equation (10.2) and $k$ number of Eq. (10.3) give the equilibrium conditions for optimization. The first equation would yield the total output the firm must produce and the $k$ number of equations will determine the distribution of the total output among its different market segments. The two solutions would always yield consistent results, for the solution uses the definition, viz. $Q = Q_1 + Q_2 + Q_3 +$

…… + $Q_k$. Since the price discrimination involves one firm and multiple markets, pricing under such a condition is also referred to as **pricing under multiple markets**.

It may now be shown that the price would be higher in the less elastic demand market than that in the more elastic demand market. For simplicity, let us explain this for two market segments case. By the necessary condition for profit-maximization under price discrimination (vide Eq. 10.3) we have

$$MR_1 = MR_2$$

Recall the definition of *MR* in terms of price and price elasticity of demand (vide Chapters 2 and 9),

$$MR = P\left(1 + \frac{1}{e}\right)$$

Combining the two, we have

$$P_1\left(1 + \frac{1}{e_1}\right) = P_2\left(1 + \frac{1}{e_2}\right)$$

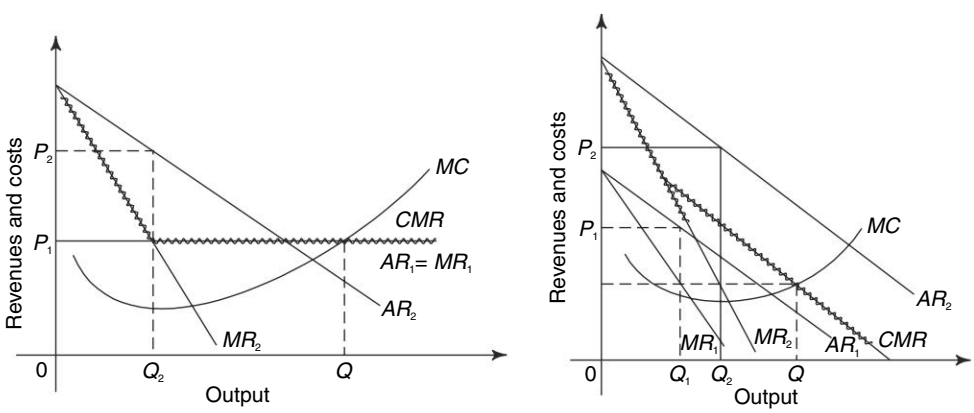
or,
$$\frac{P_1}{P_2} = \frac{1 + \frac{1}{e_2}}{1 + \frac{1}{e_1}} \tag{10.4}$$

where, $e_1$ and $e_2$ are price elasticities of demand (negative values) in markets 1 and 2, respectively. From Eq. (10.4), it is clear that

(i) $P_1 > P_2$, if $|e_1| < |e_2|$
(ii) $P_1 = P_2$, if $|e_1| = |e_2|$
(iii) $P_1 < P_2$, if $|e_1| > |e_2|$

This proves our above assertion mentioned under prerequisites for price discrimination to be profitable viz., that price discrimination across market segments would be profitable if and only if the price elasticities of demand are different in different market segments. This further substantiates the significance of the price elasticity of demand in the price of the product.

To explain the decision mechanism through **geometry**, let us consider a simple case where a firm divides its market in two segments, viz. foreign market and domestic market. Further, we know that




(a) Pure competition in one and monopoly in other market

(b) Monopoly in both markets

**Fig. 10.2.** *Pricing under Price Discrimination across Market Segments*

the foreign market is likely to have perfect competition while the domestic market may be a monopoly market for the firm. The equilibrium of the firm under such a situation is explained in Fig. 10.2(a).

The foreign market (market-1) has a perfectly elastic demand curve and its *AR* and *MR* curves are represented by $AR_1$ and $MR_1$, respectively. The firm has a monopoly in the domestic market (market-2), and accordingly there are $AR_2$ and $MR_2$. The *CMR* curve represents the combined *MR* curve of both the markets. It is obtained through a horizontal summation of $MR_1$ and $MR_2$. The firm maximizes its profit at an output at which *CMR* = *MC*, and the *MC* curve intersects the *CMR* curve from below; the third condition for profit maximization, viz. price ≥ shut-down price is ignored for simplicity. Thus, the profit-maximizing output equals *OQ*. The distribution of this output in two markets is optimum when $MR_1 = MR_2 = MC$, as required by conditions (10.3) above. Also, note that if these equalities did not hold, the firm could add to its profit by moving some output from the market where *MR* was low to the market where *MR* was high. Thus, to determine the allocation of output between the two markets geometrically, one draws a horizontal straight line through the price axis to the point of intersection between *CMR* and *MC*. In this particular case, this line coincides with $AR_1 = MR_1$. The point at which this horizontal line intersects the $MR_2$ curve gives the equilibrium level of output ($OQ_2$) and the equilibrium price ($OP_2$) in the domestic market. The remainder of the total output ($Q_2$ $Q$) is sold in the foreign market at a price equal to $P_1$. In consequences, $P_2 > P_1$, that is, the firm would charge a higher price in market-2 than in market-1. Note that the demand is more price elastic in market-1 than in market-2. Thus, a profit-maximizing discriminating monopolist would charge a higher price in the less elastic (domestic) market than in the more elastic (foreign market) market. This conclusion is always true and it is because of this that firms prefer to dump their products in foreign markets. Further, for this reason alone, dumping is often banned. Nevertheless, dumping is hard to prove as the true cost data are rarely known to outsiders and thus the price discrimination of this variety often goes unabated.

The case where the price discriminating firm enjoys monopoly in both the markets is illustrated in Fig. 10.2(b). Here again, the demand is assumed to be more (though not perfectly) elastic in market-1 than in market-2. Going through the similar procedure as above, the firm's equilibrium position could be understood. The results suggest that $P_2 > P_1$, which confirms that profit maximizing price is lower in the more price elastic market than in the less price elastic market. Incidentally note that the firm is selling $OQ_1$ quantity in market-1 and $OQ_2$ quantity in market-2, and their sum is necessarily equal to the total quantity *OQ* by the construction of the *CMR* curve.

**Hypothetical Example** The price-output determination under discriminating monopoly across market segments may now be explained through a numerical example. Let the Anu Power Company (*APC*) which sells power to households and industries, face the following total cost function:

$$TC = 22{,}000 + 100\,Q$$

Further, assume that the demands for power by the two clearly separable market segments are given by the following demand functions:

Industries (1): $P_1 = 300 - 0.5Q_1$

Households (2): $P_2 = 200 - 0.5Q_2$

If the *APC* was allowed to discriminate, the profit-maximizing firm would set its quantities (or prices) in the two markets such that its total profit is at its maximum. The total profit ($\pi$) function would be

$$\pi = P_1Q_1 + P_2 Q_2 - TC$$
$$= (300\ Q_1 - 0.5\ Q_1^2) + (200\ Q_2 - 0.5\ Q_2^2) - (22{,}000 + 100\ Q)$$

But $Q = Q_1 + Q_2$ by definition. Substitution of this into the profit function yields.

$$\pi = 200\ Q_1 + 100\ Q_2 - 0.5\ Q_1^2 - 0.5\ Q_2^2 - 22{,}000$$

The necessary conditions for profit to be at its maximum with respect to $Q_1$ and $Q_2$ are that each of the two partial derivatives must equal zero.

$$\frac{\partial \pi}{\partial Q_1} = 200 - Q_1 = 0$$

or, $$Q_1 = 200$$

$$\frac{\partial \pi}{\partial Q_2} = 100 - Q_2 = 0$$

or, $$Q_2 = 100$$

It could be verified that $MR_1 = MC$ would also yield $Q_1 = 200$, and $MR_2 = MC$, $Q_2 = 100$. The second order conditions, though not detailed here, are also met. Given the quantities to be sold in the two markets, the equilibrium prices in the two markets can be determined through the corresponding demand function. Thus, setting $Q_1 = 200$ in industries' demand function, one gets $P_1 = 200$ and substituting $Q_2 = 100$ in households' demand function, one gets $P_2 = 150$. Substitution of these four values in the profit function yields $\pi = 3{,}000$.

**Simple versus Discrimination Monopoly** It would be interesting to examine the consequences of price discrimination vis-a-vis no such discrimination, as it prevails under simple monopoly. To do so, let us consider a situation in which the Anu Power Company of the above example does not practice price discrimination either because it is not permitted by law or it decides to act like a simple monopolist voluntarily. In this situation, the *APC* could set any price but its constraint is that the price in two markets must be the same. Thus, this is a case of constrained profit-maximization, where the constraint is

$$P_1 = P_2$$

Substituting the values for $P_1$ and $P_2$ from the two demand functions, we have

$$300 - 0.5Q_1 = 200 - 0.5Q_2$$

or, $$0.5Q_1 - 0.5Q_2 - 100 = 0$$

Combining this constraint with the profit function above in the Lagrangian multiplier ($\lambda$) technique of constrained optimization, we have the following Lagrangian expression ($L$):

$$L = 200\ Q_1 + 100\ Q_2 - 0.5\ Q_1^2 - 0.5\ Q_2^2 - 22{,}000 + \lambda\ (0.5\ Q_1 - 0.5\ Q_2 - 100)$$

Maximization of $L$ with respect to $Q_1$, and $Q_2$ and $\lambda$, requires that each of the three partial derivatives are zero

$$\frac{\partial L}{\partial Q_1} = 200 - Q_1 + 0.5\lambda = 0$$
$$\frac{\partial L}{\partial Q_2} = 100 - Q_2 - 0.5\lambda = 0$$
$$\frac{\partial L}{\partial \lambda} = 0.5Q_1 - 0.5Q_2 - 100 = 0$$

Solution of these three equations for $Q_1$, $Q_2$ and $\lambda$ yields

$$Q_1 = 250,\ Q_2 = 50 \text{ and } \lambda = 100$$

Substitution of the values for $Q_1$ and $Q_2$ into the two demand functions gives,

$$P_1 = P_2 = 175$$

And substitution of the values for two outputs and two prices into the profit function yields,

$$\pi = 500$$

**Alternative Procedure** The above problem could be solved alternatively through direct application of the necessary conditions for optimization, viz. conditions given in Eqs. (10.2) and (10.3). To obtain *CMR*, we must first convert the two demand functions in to a combined demand function in its inverse form. Thus, the two functions in their inverse form are

$$Q_1 = 600 - 2P_1 \qquad \text{and} \qquad Q_2 = 400 - 2P_2$$

Imposing the simple monopoly condition, viz. $P_1 = P_2$ , and using $Q_1 + Q_2 = Q$, and summing the two above functions, we get the combined demand function:

$$Q = 1{,}000 - 4\,P$$

Inverse of this gives, $P = 250 - 0.25\,Q$

And combined *TR* (*CTR*) is then given by $CTR = P\,Q = 250\,Q - 0.25\,Q^2$

And combined *MR* is thus given by $CMR = 250 - 0.50\,Q$

Using the equilibrium condition $CMR = MC$, we get

$$250 - 0.50\,Q = 100 \qquad \text{or,} \qquad Q = 300$$

Substitution of the above value of $Q$ in the demand function gives,

$$P = 250 - 0.25\,(300) \qquad \text{Or,} \qquad P = 175$$

Substitution of the value of $P$ in each of the two demand functions, and noting that $P_1 = P_2$ under simple monopoly, we get

$$Q_1 = 600 - 2\,(175) = 250 \qquad \text{and} \qquad Q_2 = 400 - 2(175) = 50$$

It may be verified that these equilibrium values for two outputs and the price are same as obtained above and so the profit would also be the same at $\pi = 500$.

**Perfect Competition vs. Discriminating Monopoly** To evaluate the effect of price discrimination on the society, we need the outcome that would have resulted had there been perfect competition in the power industry. Under perfect competition, $P = MC$. Putting this equilibrium condition as the constraint in the above numerical example, we get

Industries (1): $P_1 = 300 - 0.5\,Q_1 = MC = 100$

The solution gives $P_1 = 100$ and $Q_1 = 400$

Households (2): $P_2 = 200 - 0.5\,Q_2 = MC = 100$

The solution yields $P_2 = 100$ and $Q_2 = 200$ and thus $Q_1 + Q_2 = 600$

And, $\pi = 100\,(400) + 100\,(200) - [22,000 - 100\,(400 + 200)]$

$= -\,22{,}000$

**Dead Weight Loss** To analyze the effects of price discrimination on the society, we need to compute the dead weight loss (*DWL*) under such a pricing rule. Recall from Chapter 8, Fig. 8.10 that the *DWL* is given by the area of the triangle $UVZ = E + H$. Further, if *MC* is constant as it is in our numerical example, $H = 0$ and $DWL = E$, which equals the area of the triangle falling between the equilibrium position of the monopolist and that of the perfectly competitive outcome. Accordingly it is given by

$$DWL = (0.5)(Q_C - Q_M)\,(P_M - P_C) \tag{10.5}$$

where, *C* stands for perfect competition and *M* for monopoly. We already have results under both discriminating and simple monopoly, as well as under perfect competition for our numerical example. Substituting the corresponding values in formula (10.5), we get

*DWL* under simple monopoly vs. Perfect competition = (0.5) (600 – 300)(175–100) = 11,250

*DWL* under discriminating monopoly vs. perfect competition = *DWL* in market-1 + *DWL* in market-2

= (0.5) (400 – 200) (200 – 100) + (0.5)(200 – 100)(150 – 100)

= 10,000 + 7,500 = 17,500

As one would expect, *DWL* is more under price discrimination than under simple monopoly.

The results under price discrimination, simple monopoly and perfect competition may now be compared. They are tabulated below.

| *Variable* | *Simple Monopoly* | *Discrimination Monopoly* | *Perfect Competition* |
|---|---|---|---|
| $P_1$ | 175 | 200 | 100 |
| $P_2$ | 175 | 150 | 100 |
| $Q_1$ | 250 | 200 | 400 |
| $Q_2$ | 50 | 100 | 200 |
| $\pi$ | 500 | 3000 | – 22, 000 |
| *DWL* | 11, 250 | 17, 500 | – |

In this example, it so happens that (i) simple monopoly price is just the simple average of the two discriminating prices, (ii) total output is the same under both the situations, (iii) profit under simple monopoly is less than that under discriminating monopoly, and (iv) *DWL* is more under price discrimination than under simple monopoly. In general, while the first of these findings may or may hold, the other three relationships are always true. The conclusion that the total output under simple and discriminating monopoly are always equal comes from the equilibrium condition $CMR = MC$, which is applicable under both the situations. The conclusion that profit under discriminating monopoly is generally larger (but never smaller) than that under simple monopoly comes from the very foundation of price discrimination. If discrimination does not fetch larger profit, then the firm has no reason to pursue discrimination at all. The conclusion that the *DWL* is larger under discrimination than otherwise emerges, for discriminations are always socially more harmful than otherwise. An additional conclusion that follows from price discrimination of this variety is that price is more in less elastic market than in more elastic market, implying that there is a **premium on high elasticity and penalty on low elasticity**. This confirms what we discovered under the division of the incidence of tax burden between producers and consumers. Further, the theory of price discrimination suggests that it leads to **cross-subsidization**

in which the customers of the low price elastic market subsidize the customers of the high price elastic market, and in the process the **discriminating firm gains**.

## 10.4 INTER-TEMPORAL PRICE DISCRIMINATION

In contrast to price discrimination by market segments; under the inter-temporal price discrimination, the discriminating firm divides its potential customers on the **basis of the time they buy its product**. Time may be classified by clock hours, days of the week, weeks of the month, months or the year, quarters of the year, years of the decade, and so on. Thus, a firm may have different prices for its product during, say, 8–10 am, 10 am – 5 pm, 5 pm – 8 pm, and 8 pm – 8 am. Also, the firm could have separate prices during week-days and week-ends. In addition; a pricing strategy in which a firm charges low (high) price in the beginning and raises (lowers) it after a few months or even after one or more years, in one or more installments, also falls under the inter-temporal price discrimination. The latter tactics are called the price penetration and price skimming strategies of pricing, as explained in Chapter 9. Further, it must be emphasized here that the peak-load pricing theory (pricing differently over seasons, viz. peak and slack seasons), which of course is very much inter-temporal pricing, is not a part of the inter-temporal price discrimination theory; for, as will be shown later, the peak-load pricing may or may not involve discrimination, and it is practiced under distinctly different situations and under different objective. Suffice at this stage, under the peak-load pricing, the firm sells its product at a high price during the peak season and at low price during the slack season; and the season varies from product to product and is thus subjective.

Under inter-temporal price discrimination, the discriminating firm divides its market by the pre-set time zones the prospective customers buy its product, and charges different prices for an identical good (or differentiated good but not distinguished by cost differences) in the so divided different markets. The examples of such discriminations generally include tariffs by clock hours and days of the week, such as airfare (day time rate, night rate; week day rate and week-end rate), movie rates (matinee show rate and afternoon and evening rate; week-day rate and week-end rate), college fees for morning session vs. afternoon/evening session, specially announced discount sales of merchandizes during fixed hours/ days of the week, and so on. As stated above, it also includes sales during initial launch of a product and afterwards, called pricing penetration and price skimming (vide Chapter 9). Under the former, the product is sold at a lower price on its launch until some period, and price is revised upward after that in one or more installments. Quite the reverse strategy is followed under price skimming. For example, electronic (white) goods are often priced through skimming route while cosmetic and general merchandize items follow the penetration route. Under price skimming, the more enthused customers buy the product in the earlier period, and less enthused wait and buy when the product gets cheaper. Novels, textbooks, and even movies are often priced through such discrimination. In addition, quite often books are initially brought out in hard bound editions only and sold at high prices. Subsequently as the dedicated and enthused people's market gets saturated, the paper bound edition is brought out and sold at a lower price to the rest of the customers. Many popular textbooks published in Western countries are first sold only in those countries and subsequently they are brought out in internationals editions and sold in developing countries at much reduced rate. The paper bound and international editions are surely a little poor in quality (not in content but in paper, print, design, etc.), but the price difference is much more than the cost difference, qualifying for inclusion under price discrimination. Some firms

are seen to practice **randomized pricing**, under which they keep changing their select products' prices randomly. The strategy helps firms to attract customers by low price of select goods on a particular day as well as in hiding the information on the prices charged by them.

How the optimum price and output get determined under such a pricing practice? As discussed in Section 10.2, inter-temporal price discrimination can be caused only by differences in price elasticity of demand in different periods. For such a case, the above numerical example could be cited as the inter-temporal price discrimination case, where the power demand instead of being classified as that for industrial and household uses, could be reclassified as the demand in periods 1 and 2, respectively, while treating total cost equation as invariant over time. As the industries' demand is higher than the households' demand (with higher intercept and same slope), the former is less elastic than the latter, and thus electric tariff for industries ($P_1$) is higher than that for households ($P_2$). If cost varies over time, then the optimization condition, viz. $MR = MC$ in each time period would determine the equilibrium prices in two periods. If the price elasticity differs from period to period, the optimum price in two periods would differ by more than $MC$ difference, and thus it would be a case of inter-temporal price discrimination. Under inter-temporal price discrimination, price may fall or rise in over time beyond the cost difference. If it falls, it is like price skimming and if it increases, it will be like price penetration. Figure 10.3 explain inter-temporal price discrimination for price skimming.

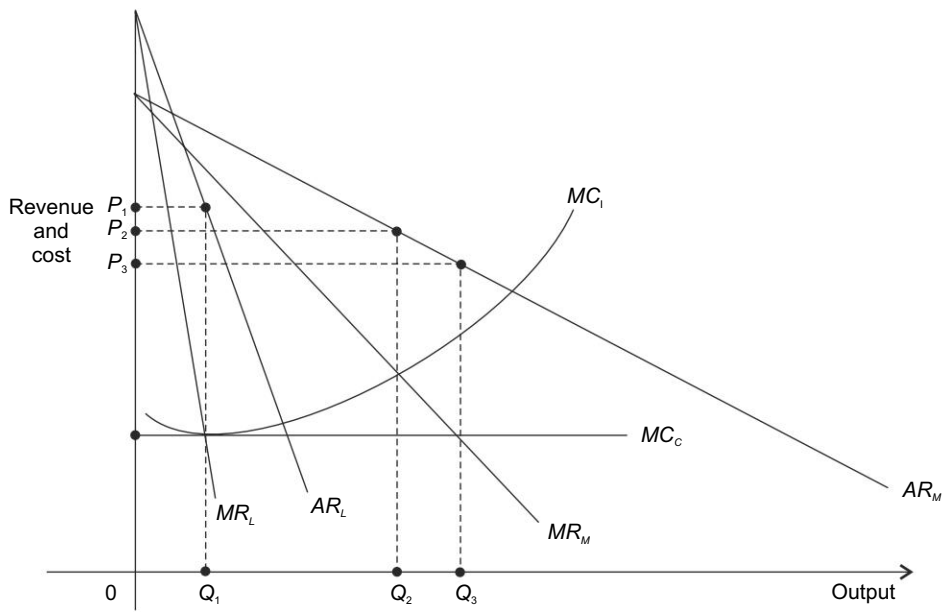



**Fig. 10.3.** *Inter-temporal Price Discrimination*

There are two $MC$ curves, $MC_C$ (constant $MC$) and $MC_I$ (rising $MC$) and two demand curves, $AR_L$ (less elastic $AR$) and $AR_M$ (more elastic $AR$), and the corresponding $MR$ curves. While the $MC$ curve is the same in both the periods under each of the two (constant and increasing) cost situations, the less elastic demand curve is for period 1 and more elastic for period 2. The optimum prices under constant $MC$ curve are given by $P_1$ and $P_3$, and those under increasing $MC$ by $P_1$ and $P_2$. It would be seen that the optimum price ($MR = MC$) of period 1 (where demand is relatively less price elastic) is higher than

that for period 2 (where demand is relatively more elastic) under each of the two cost conditions. This explains price discrimination over time. Further, note that, if *MC* is higher in period 2 than period 1, then inter-temporal price discrimination would be there even if the two prices were equal or the increase in price in period 2 over period 1 is less than the increase in cost during the same period. The case of price penetration could similarly be analyzed.

Under Inter-temporal price discrimination, the practicing firm would earn larger profit than if it were to follow uniform pricing. Accordingly, its pros and cons would be similar to the one under price discrimination across market segments.

## 10.5 PEAK-LOAD PRICING

Demands for various goods and services have seasonal fluctuations. For example, jewellery, sweets white goods, etc. experience higher demands during festivals and marriage seasons than at other times. Hotels at hill stations have higher demands during summer times than at other times. Ski, resorts and amusement parks are more in demand during week-ends than week-days. Phone calls are more during day time than in the evenings and nights, movie shows are more in demand during afternoons and evenings (week-ends) than in mornings (week-days), and so on. Demand for train and air travels is more during summer months and winter break (December) than other periods. Road and tunnel bridges are used more during daylights than nights. To cope with such fluctuating demands, firms charge higher price for an identical product in peak period than in slack period. Such a practice is called the peak-load pricing. This practice is for three reasons, viz.

(a) To meet high peak demand, firm needs to create high capacity, resulting in high capital cost. While the capacity may be used optimally during the peak season, a part of it would remain idle during the slack period, leading to heavy overheads during slack period. Thus, the high capacity provides no benefit to customers in slack period. To alleviate this issue, firms are likely to budget the cost of the additional capacity (to meet peak demand over and above slack demand) exclusively or largely to the peak demand. Further, if this cost is distributed to marginal cost, the latter would rise steeply beyond the slack period output/sales.

(b) If the industry is subject to increasing cost, the higher is the output, the higher is the marginal cost.

(c) High demand, *ceteris paribus*, means low price elasticity of demand (vide section 10.2). Further, even otherwise, demand during peak period is generally less price elastic than slack season because the former is more urgent or essential than the latter. By these factors, demand is less price elastic during peak period than slack period. Since price is negatively related to price elasticity, price is higher during period than in slack period.

Every firm keeps above factors in mind while deciding on its capacity. Accordingly, to minimize the initial cost as also under-utilization of capacity, it is prudent for every firm to attempt to shift some of the demands for its product during the peak period to the slack period. By the law of demand, customers are price sensitive. Thus, by charging high price during peak period and low in slack season, firm could shift some demand from the former period to the latter period and thereby reduce its required capacity and save on capital cost. It must be emphasized that **peak-load pricing does not mean price discrimination over time**; for while the latter aims at capturing some part of consumer surplus, the former attempts to bring price closer to marginal cost or at **improving of economic efficiency**. The

two differ more in **concept** than in substance. The confusion between the two, if any, can be sorted out by looking at the differences in their optimal values of marginal costs and outputs. Under peak-load pricing, the difference between the marginal costs as well as in outputs in two periods will be substantial, while that under inter-temporal price discrimination will be small. Another way of justifying the non-price discrimination feature of peak-load pricing is that for a homogeneous good (which is what we have under peak-load pricing), *MC* is same in different market segments under price discrimination by market segments, which is not the case under peak-load pricing. For example, for an amusement park, *MC* is the same whether it is used by a male or female, or an adult or a senior citizen, etc., but its *MC* on week-ends is independent of (higher than) that on week-days.

It must be emphasized here that **peak-slack division of time is product subjective**. For goods like, amusements parks, movies, eating out, etc., week-ends are peak period while week-days are slack period. For banking services, particularly in India, October–March is considered as the peak period and the rest as slack season. For transport services, March–June and Dec. 15-Jan 10 or so are deemed as the peak season; for rain coats, umbrella, etc., monsoon months (June–August) are treated as the peak period; for movies, afternoons and evenings are the peak and mornings are slack; for resort hotels, summer months are peak period; and so on. The peak-load pricing is illustrated in Fig. 10.4.

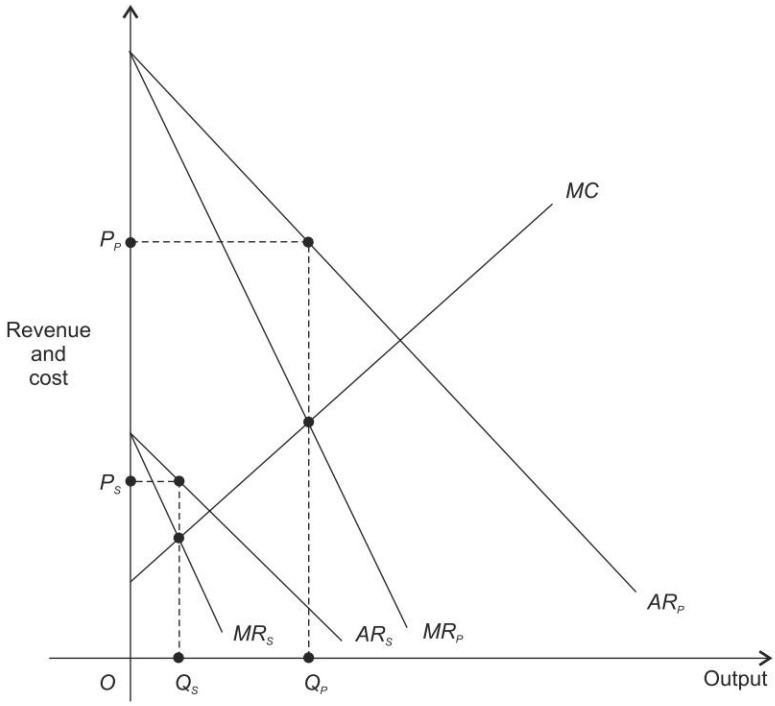



**Fig. 10.4** *Peak-Load pricing*

In Fig. 10.4, *MC* curve is same for both peak and slack periods but it gets steeper (to account for increasing cost and allocation of fixed cost to peak demand exclusively or largely) as output increases. $AR_S$ and $MR_S$ are *AR* and *MR* curve for slack period, and those for peak period are $AR_P$ and $MR_P$, respectively; where the former two are more price elastic than the corresponding latter two. The

corresponding optimum prices and quantities are $P_S$ and $P_P$, and $Q_S$ and $Q_P$, respectively. It is seen that both outputs and prices are larger during peak than slack period. Profit and economic surplus (= sum of consumer and producer surpluses) are not shown in the figure for simplicity but it would be easy to understand that each of these would be more with peak-load pricing than they would be if the two prices were equal. The reason for this virtue of peak-load pricing is seen in the fact that this pricing theory brings price closer to *MC*, and the closer the two are the more is the economic surplus. Before moving to the next topic, it must be emphasized that for peak-load pricing to be prudent, three conditions must be met. They are

(a) Product is not storable and thus demands in slack and peak periods are non-competing.

(b) Same facility is used to provide the service during both the slack and peak periods.

(c) Demand for the product varies significantly across slack and peak periods.

It is obvious to see that in the absence of any of these conditions, peak-load pricing is unlikely to be profitable.

## 10.6 BLOCK PRICING

Block pricing is another strategy by which firms attempt to capture some consumer surplus. One form of block pricing was discussed above under second degree price discrimination. The block pricing relevant here is of a different kind. Under the block pricing form relevant here, multiple units of a product are packed together, called the blocks, and sold either in blocks only or both in blocks as well as in individual pieces; and price per unit is lower in block purchases than in individual unit buys. Such examples are found in soft drinks (Coke and Pepsi), bath and washing soaps, ball pens, notebooks, many more retail items, meals at restaurants, movies, etc. For example, Gap *t*-shirts may be sold in packs of 3 pieces, Coke may be sold in cartons of 10 pieces/cans/bottles, meals at Pizza Hut or Subway outlet may be sold in, say, 3/weeks, movies in a multiplex may be sold in, say, 5/month, and so on. This pricing tactic brings more sales as well as more profit to the firm. For example, suppose a customer values (called the reserve price) the first *t*-shirt at Rs. 500, and the second *t*-shirt at Rs. 400. If the firm sells the t-shirt in individual pieces, it has to price it at Rs. 400 or less to get that customer to buy two pieces. However, if the firm sells in packs of two pieces only, the customer would buy one pack at Rs. 900, giving the effective price at Rs. 450 apiece; and yielding an additional profit of Rs. 100 to the firm.

To illustrate this form of block pricing, assume the demand and cost equations facing the multiplex firm are as follows:

$$Q = 12 - P \qquad \text{and} \qquad TC = 5 + 2\,Q + 0.25\,Q^2$$

For simplicity, we assume that the above demand equation is of an individual consumer and that all consumers have identical demand equation. Accordingly, the market demand equation would equal the number of consumers times the individual demand equation, i.e., market demand = $nQ = n[12 - P]$, where $n$ = number of consumers.

The profit-maximizing firm, would set price such that $MR = MC$, assuming other conditions are met. Equations for *MR* and *MC* could be obtained as follows. Transform the demand equation in its inverse form and then obtain *TR*, whose differentiation would give *MR*. Thus,

$$P = 12 - Q \qquad \text{and} \qquad TR = 12\,Q - Q^2\,, \quad \text{and} \quad MR = 12 - 2\,Q$$

Differentiation of *TC* gives *MC*, which may then be equated to *MR* for profit maximization,

$$MC = 2 + 0.5\,Q = MR = 12 - 2\,Q$$

or, $Q = 4$, and using demand equation, we get $P = 8$

For better understanding, Fig. 10.5 gives the various curves corresponding to the demand and cost functions, and indicates the optimum values of price and output under both monopoly, as well as under other pricing systems.

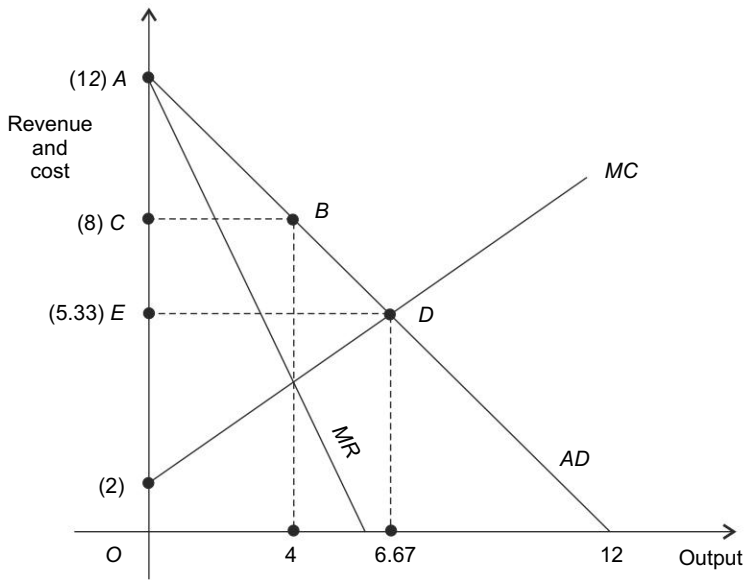



**Fig. 10.5** *Block pricing*

At the equilibrium values of *P* and *Q*, firm's profit = $TR - TC = 4 \times 8 - [5 + 2 \times 4 + 0.25\,(4)^2] = 15$. The consumer surplus is given by the area of triangle *ABC* in Fig. 10.5. The consumer surplus thus equals $[4 \times (12 - 8) \times 0.5 = 8$.

Thus, under uniform pricing by each unit, monopoly firm would price its product at Rs. 8, sell 4 units to each customer and makes a profit of Rs. 15 per customer, and generate consumer surplus equal to 8. Alternatively, if the firm were operating under perfect competition, its optimum price-output would be given by $P = MC$. Applying this, we get,

$$P = MC \qquad \text{or, } 12 - Q = 2 + 0.50\,Q, \text{ which yields } Q = 6.67 \text{ and } P = 5.33$$

and, $$\text{Profit} = TR - TC = 6.67 \times 5.33 - [5 + 2 \times 6.67 + 0.25\,(6.67)^2] = 6.09$$

Thus, under perfect competition, the firm would sell 6.67 units to each customer at price = 5.33 and make profit = 6.09. The consumer surplus would be given by the area of triangle *ADE*, which equals $[6.67 \times (12 - 5.33) \times 0.5 = 22.24$.

Under block pricing, the firm can capture the whole of consumer surplus by selling its product in blocks (in number of units), where the **number of units per block is given by the optimum output under perfect competition, and the price per block is given by consumer surplus plus the price**

**per block under perfect competition**. Accordingly for our example, block would consist of 6.67 units and price per block would equal 6.67 × 5.33 + consumer surplus. The consumer surplus here equals that under perfect competition, which is given by the area of triangle *ADE* in Fig. 10.5. This, as above, is given by 6.67 × (12 – 5.33) × 0.5 = 22.24. This gives block price = 6.67 × 5.33 + 22.24 = 57.79. The price per unit under the block price thus comes equal to 57.79/6.67 = 8.66. The firm's profit ($\Pi$) is then given by

$$\Pi = TR - TC = PQ - TC = 8.66 \times 6.67 - [5 + 2 \times 6.67 + 0.25\ (6.67)^2] = 28.25$$

The above results are summarized in the table below.

| *Market structure* | *Price* | *Output/customer* | *Consumer surplus/customer* | *Profit/customer* |
|---|---|---|---|---|
| Monopoly | 8 | 4 | 8 | 15 |
| Perfect competition | 5.33 | 6.67 | 22.24 | 6.09 |
| Block pricing | 8.66 | 6.67 | 0 | 28.25 |

Thus, through block pricing in this particular fashion, the firm is able to generate competitive results and transfer the entire consumer surplus to itself. Incidentally note that consumer surplus under block pricing equals the sum of consumer surplus and profit under perfect competition, the minor difference in our results in the table are due to rounding errors in prices and quantities. As mentioned earlier, the above results assume same demand function across customers. In real life, demand functions do vary from customer to customer and if that was allowed, setting of block size and price will not be that easy and the capturing of the entire consumer surplus may not be possible for the firm.

## 10.7 TWO-PART PRICING

The two-part pricing theory is yet another tactic which if adopted could transfer the entire consumer surplus to the practicing firm. Under this, the product is split into two parts and each part is priced separately. This practice is obviously not available in case of goods which cannot be split this way. For example, car, camera, computer, air conditioner, house, textbook, shirt, shoes, hair cut, etc. are not subject to divisions into parts and two-part pricing is out in these cases. However, several other goods could be split into two parts. For example, a video (movie DVDs) renting firm could decide to rent out movie DVDs to only those who are its members. In that case, every potential customer must buy its membership first, and then rent out as many movies as desired. The customer is charged separately the membership fee initially and then on each occasion she rents a movie, she pays the rent per movie. Thus, there is two-part pricing. Such a practice is possible in buying goods which a club house might offer, like swimming, jogging, eating, lodging, partying, conferencing, attending special sports, movies, talks, etc. One will have to pay the **membership fee** (also known as **entry fee**) as well as the price for the use of a particular facilities/good (called the **user charge**). The other goods where two-part pricing is feasible will include library of books, journals/periodicals, news papers; sports' stadiums, academic associations like American Economic Association, only members may be allowed to attend conferences or members get a discount on participation, and so on.

Thus, several goods and services could be subjected to two-part pricing theory. The question is, how the optimum price is determined under such situations? The procedure is similar to the block pricing

discussed above. Different prospective customers would again have different demand functions and the firm would have just one cost function, if we restrict to a single product firm. On the assumption that all customers have identical demand functions, the above example of block pricing could be used to illustrate two-part pricing. Under the situation, the two prices would be given by

(a) Entry fee = Consumer surplus under perfect competition, and

(b) User charge per unit of the product = Price under perfect competition

Thus, suppose the firm runs swimming pool services, whose demand by a consumer and its own cost of production are given (same as for the multiplex firm above) by

$$Q = 12 - P \qquad \text{and} \qquad TC = 5 + 2\,Q + 0.25\,Q^2$$

Then, the user charge is given by the price that would prevail if the firm was operating under perfect competition. This price would thus equal *MC*. Calculation of that would give,

User charge per swim = 5.33.

The output/customer (number of swimming /customer) would then equal 6.67. The entry/membership fee for swimming pool area is given by the consumer surplus that would occur if perfectly competitive conditions had prevailed. This surplus is given by output under perfect competition multiplied by the difference in price at which demand is zero and the price under perfect competition and multiplied by 0.5 (vide Fig. 10.5, area of triangle *ADE*). Thus,

$$\text{Entry/membership fee/customer} = 6.67 \times [12 - 5.33]\,(0.5) = 22.24$$

The swimming pool firm's profit from each customer equal to

$$\Pi = TR - TC = [\text{Entry fee} + PQ] - [TC]$$

$$= [22.24 + 5.33 \times 6.67] - [5 + 2 \times 6.67 + 0.25\,(6.67)^2] = 28.25$$

The consumer surplus would be nil. Thus, the firm is able to capture consumers' entire surplus through two-part pricing and generate price, and output equals their perfectly competitive levels.

However, if the firm were to have a single price, the outcomes would be the same as under simple monopoly. In that case, $MR = MC$ rule will apply, and price and output/customer would equal 8 and 4, respectively, and firm's profit/customer would equal 15, and consumer surplus would equal 15.

## 10.8 COMMODITY BUNDLING

Commodity bundling is a pricing technique through which the practicing firm captures a part of consumers' surplus. Under this, the firm offers more than one good or service in a bundle for a single price. The bundled goods may be related, either as substitutes or compliments, or unrelated. The case where the bundled goods are compliments is a special case of commodity bundling and it is known as **tying**. The examples for tying would include a bundle containing, for instance, Polaroid camera and its films, a car and its spare parts and insurance policy, television and DVD player, Digital camera and battery charger, MBA education and room and boards, and so on. The non-tying but commodity bundling examples would include petrol and meal, movie ticket and popcorn, car and hotel accommodation, camera and fountain pen, and so on.

How the price is set and a part of consumer surplus is captured under commodity bundling? This can be illustrated via a hypothetical example. Suppose the firm is able to find out its potential customers'

reserved prices (maximum that the customer would offer instead of going without the product) of the two goods which it sells, and let them be as given below. Demands for camera and DVD player (player) are independent and the marginal cost to firm for each is Rs. 6,000.

| *Customer* | *Reserved price for Camera* | *Reserved price for Fountain pen* |
|---|---|---|
| Anu | 20,000 | 2,000 |
| Ritu | 12,000 | 4,000 |
| Ishika | 4,000 | 12,000 |
| Arushi | 2,000 | 20,000 |

The firm could sell the two goods separately, in bundles only, and both separately as well as in bundles. When the sales are exclusively in bundles only, it is called **pure bundling**; and when the sales are administered both through bundling as well as separately, it is called **mixed bundling**. Suppose the firm operates in mixed bundling; and in separate sales, it charges Rs. 14,000 for each product; in bundles (of one each) only, the firm prices a bundle at Rs. 22,000; and both in separate and in bundle sales, it charges Rs. 11,500 for each and Rs. 22,000 for the bundle. In this case, the sales, profit and consumer surplus would be as shown in table below.

| | *Pricing method* | *Sales* | *Total revenue (Rs.)* | *Total cost (Rs.)* | *Profit (Rs.)* | *Consumer surplus (Rs.)* |
|---|---|---|---|---|---|---|
| i. | Separately only | Camera - 1 | Camera - 14,000 | Camera - 6,000 | 16,000 | Anu - 6,000 |
| | | Player - 1 | Player - 14,000 | Player - 6,000 | | Arushi - 6,000 |
| ii. | Bundles only | Bundles - 2 | 44,000 | 24,000 | 20,000 | 0 |
| iii. | Both ways | Bundles - 2 | Bundles - 44,000 | Bundles - 24,000 | 31,000 | Ritu - 500 |
| | | Camera - 1 | Camera - 11,500 | Camera - 6,000 | | Ishika - 500 |
| | | Player - 1 | Player - 11,500 | Player - 6,000 | | |

From the results under three pricing methods, it is clear that selling both individually as well as in bundles fetch the maximum profit to the firm. Thus, given the choice, the firm would choose selling its two products under mixed bundling. Though the example is simple, it does demonstrate a procedure through which a firm could transfer a part of the consumer surplus to itself through commodity bundling.

## 10.9 TRANSFER PRICING

Transfer price is the price of a product which one division of a particular firm charges on its sales to another division of the same firm. It is thus an **internal price**. Accordingly, transfer price is relevant only in cases where a firm has more than one division and they are linked in such a way that one or more divisions produce one or more goods which are used as inputs in the production of the final good in the last division. The supplier divisions are called the up-stream divisions and the buying division as the down-stream division. In the language of industrial organization, such firms are known as the **vertically integrated firms**. By integrating vertically, firms save on transaction costs (vide Chapter 7) and therefore the real world offers many cases of vertically integrated firms. Thus, we have ITC, in which one up-stream division supply tobacco to the cigarette division, which supplies cigarettes to hotel division. Also, there are automobile firms having multiple divisions, in which engine, tire, etc. divisions

sell their outputs to the car assembly division, which sell cars to the outside market; and many more industries offering scope for vertical integration. In such cases, firms may (or may not) have options of its up-stream division selling not only to its down-stream division but also to the outside market, and the down-stream division acquiring its input both from its up-stream as well as from outside sources. The optimum price and output of each of the divisions would usually depend on the availability or otherwise of such an option. This is so because in the absence of outside market, the firm may aim at maximum profit to the firm as a whole, thus ignoring division-wise issues, but that cannot be a case in the presence of outside market. However, a word of caution is needed here. To monitor/evaluate performance, integrated firms usually looks at the profit of each division and there are government regulations under competition policy which prohibit even internal manipulations, including transfer prices.

How the optimum price and output are set under such situations? In our discussion of pricing so far, we had **implicitly assumed that all output was sold to the outside market only**. Transfer pricing involve complicated mathematics and thus it is pursued thoroughly in advance texts only. Suffice here; we present briefly the logic behind this theory and that too just for the case where the up-stream divisions (*USD*) sells its entire output to the down-stream division (*DSD*) and the latter buys all its requirements from the former only; rendering the transfer price merely an internal price. Also, we ignore government regulations, if any.

Under such a scenario, we would have the demand function facing the firm (i.e. *DSD*), cost functions for each of the divisions, and the fixed relationships between the outputs of each *USD* and that of *DSD* (i.e., input-output ratios). To illustrate, suppose a firm just two divisions, up-stream and down-stream divisions. Further, assume that each output unit of *DSD* requires just one output unit of *USD*, like one engine for each car. Under these assumptions, firm's profit would be given by

$$\Pi = TR_D - TC_D - TC_U$$

The total revenue (*TR*) and total cost (*TC*) functions would be functions of the respective divisions' output ($D = DSD$, $U = USD$). But since the outputs of two divisions have to be in a fixed proportion (1 to 1 under our assumption), they all could be expressed as functions of the output of *DSD*. For profit to take maximum value, the first derivative must equal zero (and second derivative must be negative). Thus,

$$MR_D - MC_D - MC_U = 0$$

or,
$$MR_D = MC_D + MC_U$$

This equilibrium condition would give the equilibrium output of the down-stream division. Given the demand function for the firm's output in *DSD*, the price of the final product could then be obtained. Since we are assuming that one unit of output of down-stream division requires exactly one unit of output from the up-stream division, the two output levels would be equal. To determine the transfer price (*TP*), we have the constraint that the output of *USD* must be in accordance to the optimum output of *DSD*. Thus, *USD* is **output-taker**, and accordingly it has no role in pricing its product, or has no market power. In other words, the demand facing the *USD* is vertical at a given output (i.e., its price elasticity of demand is zero). Under the situation, for *USD*, $AR = MR = P$. Further, *USD* would maximize profit when its $MR = MC$. The two relationships imply, $P = MC$, which is the same as experienced by a firm under perfect competition. Accordingly, price of the output of *USD*, called the **transfer price** (*TP*), is given by its *MC* at the output determined by the requirements of *DSD*. We may now give a numerical example to illustrate the working of the model. Once prices of both the divisions are determined, their outputs would be given by their respective demand functions. Given prices and outputs, total revenues

could be found and then using the cost functions, profits could be determined. To explain the working of this model, we consider a numerical example.

Suppose the firm sells wall clocks. The up-stream division supplies the clocks, and down-stream division assembles and sells wall clocks. The estimated demand and cost functions are as below.

$$Q_D = 493 - P_D \qquad TC_D = 1500 + 5\,Q_D \quad \text{and} \quad TC_U = 5000 + 8\,Q_U + 0.5\,Q_U^2$$

The firm's profit is then given by

$$\Pi = TR_D - TC_D - TC_U$$
$$= [493 - Q_D](Q_D) - [1500 + 5\,Q_D] - [5000 + 8\,Q_U + 0.5\,Q_U^2]$$

One wall clock requires one clock, so $Q_D = Q_U$. Assuming this and setting the first derivative equal to zero gives,

$$493 - 2\,Q - 5 - 8 - Q = 0, \qquad \text{or,} \quad Q = 160$$

The second derivative would equal –3, which confirms profit is at maximum at $Q = 160$. Substitution of this value in demand function will give, $P = 333$

To determine the transfer price, we set $MC_U = P_U = TP$, which yields

$$8 + Q_U = TP$$

We know that the output of *USD* must equal the requirement of *DSD*. Thus, $Q_u = Q = 160$. Substitution of this, gives

$$TP = 168$$

The profit of *DSD* would be given by

$$\Pi_D = P_D\,Q_D - TC_D - P_U\,Q_U = 333 \times 160 - [1500 + 5 \times 160] - 168 \times 160$$
$$= 53{,}280 - 2300 - 26{,}880 = 24{,}100$$

And profit of *USD* would be given by

$$\Pi_U = P_U\,Q_U - TC_U = 168 \times 160 - [5{,}000 + 8 \times 160 + 0.5\,(160)^2]$$
$$= 26{,}880 - 18{,}600 = 8{,}280$$

The firm's total profit will be given by

$$\Pi = \Pi_D + \Pi_U = 24{,}100 + 8{,}280$$
$$= 32{,}380$$

Under this way of setting the transfer price, the firm as whole earns the maximum possible profit and the up-stream division earns the maximum possible profit subject to the constraint that its output must exactly equal the requirements of the down-stream division.

For situations when the *USD* is allowed to sell partly in outside market and *DSD* to partly buy in the outside market, mathematics gets complicated and so we do not go into those details in this text. This completes our discussion on pricing under market power and also under usual assumptions. We now proceed to pricing under special cases.

## 10.10 PRICING UNDER ASYMMETRIC INFORMATION

In our discussion on pricing so far, we had implicitly assumed that both the firm and prospective customers have identical information on demand and cost functions. With the developments in information technology, the assumption is valid in most cases but not universally. The cases where the assumption is not valid include

(a) Trades in old goods, like old (prior owned) cars, old two wheelers, old plants and equipment, etc. In these, the seller (owner) knows more about the product than the prospective customers.

(b) Transactions in (new) goods like vehicle insurance, health/life insurance, etc. In these, the buyer of the policy (product) knows more about the product (condition of vehicle, health/life) than the seller of the product (insurance company).

(c) In recruitments, prospective employee knows more about his/her ability and sincerity than the employer/experts selecting the person.

(d) In business decisions, we are well aware of the agency problem where the agents (managers) know more about their firms' costs, competitive positions and investment opportunities than their masters (owners/equity holders).

(e) In credit/loan markets, loan providers (banks) know less about the credit worthiness of the borrower than the borrower.

(f) In rental business, the firm renting out the house/office space/equipment knows less about the care, etc. the tenant/renter will take of the asset than the renter.

Thus, there are many cases where the **information set is asymmetric**, i.e., either the firm or the customer has better information about the product they are planning to trade than the other. Under such a situation, how the product price/insurance premium/wage/interest rate/rent is determined and what problems it creates is the subject matter addressed in this section. In the literature, this is known as the **lemon theory of pricing**. It was first advanced by **George Akerlof** (1970), for which the founder shared the economics' **Nobel prize** on the subject in 2001 with Michael Spence and Joseph Stiglitz.

To appreciate the significance of asymmetric information in pricing, let us consider an example first. Suppose, a customer is interested in buying an old Maruti Zen cars, 2009 model, and he goes to a dealer in prior-owned cars. The dealer has large inventory of all kinds of cars, including Maruti Zen 2009 model. However, the quality of those cars varies across cars. Suppose, the dealer has just two quality Maruti Zen 2009 model cars, viz. good quality (plum) and poor quality (lemon), and his inventory had many in each category, roughly in about 50-50 percent. Further, assume their true reserve prices to the seller/dealer and prospective customer were as follows:

| *Quality* | *Dealer's Reserve Price (Minimum Price) (Rs.'000)* | *Customer's Reserve Price (Maximum) (Rs. '000)* |
|---|---|---|
| Plum | 300 | 350 |
| Lemon | 200 | 250 |

If both the dealer and the prospective customers had complete information on the quality of the cars, the two kinds of cars will be sold and bought in separate markets for plums and lemons. Those who wanted plums would get plum and those who wanted lemons would get lemons, and the price for plums will settle somewhere between Rs. 300 thousands and Rs. 350 thousands and that for lemons between Rs. 200 to 250 thousands, depending on the relative bargaining ability of dealer and customers.

As a result, the two markets will be cleared and able to allocate the product efficiently. However, in real life; while the dealer, being well informed, knows which cars are plums and which are lemons; the customer, being uninformed, may not be able to locate a plum; though the customer knows that the dealers' inventory has both plums and lemons. Under such a real life case of asymmetric information, how would exchange take place? There are two possibilities, viz.

(a) The buyer is ignorant (or uninformed) and ignores adverse selection, and

(b) The buyer is uninformed but considers adverse selection

Under the first possibility, if the customer is risk neutral (See Section 10.12), he/she would offer Rs. 300 thousands (expected value = 0.5 × 350 + 0.5 × 250) for the car and the dealer would happily sell him a lemon for that price. What is the consequence? Dealer is happy as his producer surplus equals Rs. 100 thousands [price (300) – his reserve price (200)] and the customer is cheated as his consumer surplus is negative Rs. 50 thousands [his reserve price (250) – price (300)]. If the same process was repeated with other customers, all lemons will be sold out and all plums will remain unsold. If the customer was risk- seeker, he would offer somewhere between Rs. 300 to Rs. 350 thousands, and the dealer will be happier than before to sell lemons to him. In case the customer was risk averse, he would offer somewhere in the range of Rs. 250 to Rs. 300 thousands, and dealer will still let him have lemon, though with a lower producer surplus.

Under the second scenario, the customer will offer a maximum of Rs. 250 thousand for a car and the deal will settle between Rs. 200 to 250 thousands and lemon would be traded. While the uninformed party would like to know, the informed party has incentive to hide the characteristics of the lemon. Thus, in all cases, because of hidden characteristics of lemon, only lemons would be sold. This is called,

- **Adverse Selection**

It is called adverse selection of goods, as the selection is "adverse" from the standpoint of the uninformed party. In other words, given plums and lemons, the buyers are likely to get lemons due to asymmetric information. This has two consequences, viz.

(a) The quality products (cars) are driven out of the market. Same would happen with all exchanges of the kind mentioned above in points (a) through (f) above. Thus, due to asymmetric information, low quality goods drive high quality goods out of the market. This is a case of **market failure**, for the system has failed to efficiently clear the market, as the sellers have not been able to sell plums and buyers have failed to buy plums. This phenomenon is known as the **lemon problem.**

(b) Customers get cheated by the dealer through dishonesty in trade. Accordingly, buyers may avoid buying such goods and the product may even go out of the market.

The cost of asymmetric is thus not limited just to the loss incurred by the customers but it also includes the loss caused by driving the legitimate business out of the market. If the process is allowed to continue, it would eventually drive out the market for all old and similar goods as described above. To see how this would happen, let us continue with the old cars sales. Since only lemons are sold, each buyer would discover that the car needs frequent repairs, and he/she has, in fact, been cheated in the purchase. Subsequently, the word would get spread and in subsequent purchases, the prospective customers would offer lower price, the sales will happen until the lower price still exceeds the dealer's reserve price; the latter would also have downward trend due to poor sales and poor reputation of the dealer. The process would go on and the difference will get lower and lower, and would eventually turn negative, and the market would get terminated. Thus, eventually, the market for old cars would just

disappear. To take another example, consider Medicare insurance. It is subscribed more by sick prone people than healthier. Thus, the insured people get sick and claims on the insurance company go up. The company, in turn, raises insurance premium, which reduces company's sales. Due to enhanced premium, insurance is rendered viable to lesser people and to those who are even more prone to sickness. This leads to further increase in sickness and claims from Medicare. Once again premiums get hiked, still more sick people enter the market, more get sick and claims shoot up. The process goes on and health insurance market goes on shirking and eventually the business of health disappears from the market.

Under the theory of asymmetric information, uninformed party feels that the informed party may be, for example, getting rid of the car quickly because of the car's hidden problems, which may or may not be true, and that is why a used car which is just 3 months old and driven for just 500 miles is often sold for around 20 per cent less than its original cost. Similarly, it explains why people with poor health records dominate the insured people under life/medical insurances, why banks have queues of prospective borrowers with poor credit worthiness, why organizations find hard to recruit quality personnel, and so on. To alleviate such issues under adverse selection, market has innovated new products/avenues, like

- **Signalling**
- **Screening/testing to judge the quality**
- **Guarantee/warranty to quality**
- **Exclusions in insurances**
- **Group insurance**

Quality certificates and degrees are sought and examined by the buyers before making the deals. For example, recruiters look for degree certificates; insurance companies look into the age and require the prospective clients to go through prior medical tests; old car dealers offer guarantees/warranties for fixed periods against unexpected breakdowns; insurance companies stipulates some minimum payments/hospitalization by the clients, insert exclusions of certain diseases from coverage, and offer group insurances at reduced rates; and so on. Such requirements and provisions help retaining the affected goods in the market even under adverse selection. However, there is another manifestation of asymmetric information, which leads to customers' /clients' cheating and thereby aggravates the market. This is known as

- **Moral hazard**

By moral hazard here means the hidden actions of the clients. For example, the person with car insurance against accidents/thefts may drive recklessly and/or leave the car unlocked, particularly because he/she is covered under such losses, and end up in accident/theft causing loss to the insurance company. Similarly, people with health insurance may start eating junk food/drinks/smoking, stop exercises, stop taking common medicines against minor sickness, etc. and consequently suffer from serious illness, requiring hospitalization and claims from the insurance company. Also, people having insurance against fire, may not care to install fire extinguishers in the house. Thus, moral hazard is the risk of inappropriate or immoral behaviour of the client when the firm cannot perfectly monitor the former's behaviour. In consequence, the costs of insurance companies (uninformed party) go up, which lead to enhanced premium rates in future, which in turn, would drive out some of the clients who are price sensitive. The process, if allowed to continue, like in lemon cars case, would eventually drive insurance out of the market. Since insurance companies would like to sustain their business, they would take precautions like

- **Deductions/minimum hospitalization**
- **Supervision**
- **Incentives (like no claim rebates in future premiums)**

Under the first head, insurance companies stipulate some minimum payments by the clients for every claim and require some minimum period of hospitalization for the eligibility of the claim. Under the second factor, insurance companies have supervisors who oversee the clients, treatments, and try to minimize moral hazards. Under the third factor, insurers provide rebates on future premiums when there are no claims. Companies train their workers, have supervisors to ensure that the workers are faithfully working and promote (fire), give stock options, commissions to good (poor) employees. In addition, organizations offer temporary employments, particularly if they are not convinced about the suitability of a potential recruit.

In contrast to the problems faced by the poorly informed party under asymmetric information, there are issues faced by the better informed party as well. These are known as

- **Enforcement of contracts' issues**

It is quite common to find cases where the clients/employees/customers face difficulties in getting their claims honoured by their insurance companies/employers/sellers. Such parties often insist for all possible kinds of documents/proofs, delays in processing of claims, and pretends/gives excuses against their admissions, etc. Thus, there are issues faced by both sides in such transactions and accordingly solutions are designed, and the transactions are kept alive. Nevertheless, the trade is surely discouraged and is much less than what one would like to see in these days of uncertainty. In consequence, useful services like health/accident/theft insurance, and trades in old goods have fallen to a rather low level, and regrettable goods like court cases against cheating, etc. have increased to alarming level. Fortunately, group insurance, out-sourcing, etc. have been designed which are surely of great help in coping with asymmetric information.

Three related theories are worth covering here. These are

- **Insider-outsider model**
- **Efficiency Wage theory**
- **Labour hoarding practice**

Under the insider-outsider theory, asymmetric information inspires organizations to prefer the existing employees over the prospective employees, for the organizations have better information on the former than the latter. As a result, the value of the incumbent employees goes up and they reap what is called as **economic rent**, i.e. receive wages is excess of what they must to stay in the job. Due to this bias in favour of incumbent employees, firms tend to grant wage increases even when they have applicants from potential candidates with good looking qualifications. Thus, the insider-outsider theory provides a rationale for the apparent paradox of rising wages in the midst of unemployment. By the efficiency wage theory, to save on supervision cost (required due to asymmetric information) and yet ensure better performance from employees, firms tend to reward its employees better than their market values. For, firms believe that if the employees are well paid, they and their families will have good standard of living and thereby better health; and they will be well motivated, shirk less from sincere work, and avoid looking for potential employment opportunities; and so on. Through such results, the efficiency wage theory offers another rationale for the paradox of rising wages in the midst of unemployment. It is not uncommon to find firms avoiding retrenching of employees during recessions, which is explained by

the labour hoarding practices. Many firms consider this a good business practice because by doing this they illicit faithfulness, commitments, respects, stability, etc. from the employees and thereby minimize the damage caused by asymmetric information. Further, firms use such surplus labour for carrying out the pending works, like arranging files, preparing documents, cleaning the office and even getting help in their personal/family works, etc. Such works do not get reflected in firms' outputs, though retention of employees help keeping unemployment low, and the economy experiences recession with no or little increase in unemployment, reflecting fall in labour productivity.

These three theories, besides others discussed under this section, fully endorse the significance of asymmetric information in the management of firms, and in product pricing and output decisions in particular.

## 10.11 PRICING UNDER EXTERNALITIES

In the foregoing discussion, we had implicitly assumed that any transaction between a firm and its customers brings benefits and costs that accrue to the participants alone and none others. Thus, when you buy a new or old car, or pursue MBA course at a business school, etc., the benefits and costs of that activity are restricted to you and the provider alone, called the two parties in the trade. However, in real life, there are many (surely not all) goods where the benefits and/or cost of such transactions accrue, besides the two parties in the transaction, to some outsiders as well, called the "third party". For example, your MBA education costs and generates benefits not just to you and the business school, but also to some others in the society (economy as well as the world), such as the cost of having denied the admission to someone else in the business school, and benefits to others through your enhanced qualifications and improved performance and interactions/inspirations with/from you, etc. Similarly, when you buy a new car, your neighbours may get free rides sometime but face difficulties in parking around the house as parking area gets congested, and the people living around the car factory may suffer through traffic hazard, pollution, etc. In this regard, the cost incurred by Bhopal (India) people under the world's worst industrial accident at Union Carbide factory in December 1984, and that suffered very recently by Mexican fishermen and others under the BP (British Petroleum)'s gas spill during April–July 2010 are well known. Air pollution suffered by people living around coal/oil based power generating plants, noise pollution suffered by people living around railway tracks and airports, nuisance caused to neighbours by people keeping pet dogs offer other familiar examples of (negative) externalities. Benefits (positive externalities) which people receive through others education, research and development, nurseries, gardens, etc. and from having good neighbours are well known as well. One can go on citing many more such examples but the point to drive home is that the benefits and costs do get spread to the third party in many cases. In the literature and practice, this is known as **externalities**, or the so called **third party effects, or side effects**. It would be obvious that externalities could be both **positive** as well as **negative**, former meaning benefits and the latter costs to the outsiders. For example, neighbours getting free rides at some occasions from you having a car denotes the positive externality and people living around the factory manufacturing car suffering pollution, accidents and traffic hazards denote negative externality.

How do externalities affect price-output decisions? They do through total revenues and total costs, which determine the optimum price-output combination. An individual firm considers the revenues and costs to it only and thereby ignores externalities associated with the production and consumption of the

good in question. Since externalities are ignored from revenues and costs, and they may not cancel out, the optimization results based on such data, though yields optimal results for the firm/industry (**private** party), give sub-optimal decisions for the society (**Social welfare**). Let us take a hypothetical example to illustrate the contention. Note that there could be either net negative or net positive externality from the production and consumption of a good. Under net negative externality, the cost to firm/industry (**internal cost**) would be one cost component and the externality cost (**external cost**) would be the other component of cost to the society. Thus, the social cost would exceed the private cost, and benefit to both the private as well society would be the same. In contrast, when there is net positive externality; the social cost and private cost would be equal but social benefit would exceed private benefit. Handling the two situations to determine the optimum price-output would be similar. We take the case of net negative externality.

## Negative Externalities

Consider a perfectly competitive cement industry, having 200 firms, all firms having identical cost functions, and the demand facing the industry as follows;

$$TC_F = 40{,}000 + 500\, Q_F + Q_F^2$$

$$Q = 150{,}000 - 100\, P$$

In addition, the production of cement generates net negative externality to the society, which is given by

$$TEC = 100\, Q + Q^2$$

where, $TC_F$ = total cost to a typical firm (or total internal cost), $Q_F$ = output of the typical firm, $Q$ = cement industry demand/supply, $P$ = cement price, and $TEC$ = total externality (or external) cost (negative externalities = total cost of cement industry incurred by the third party.

To evaluate the effect of net external externalities on the optimum price-output decision, we must perform a comparative statics exercise. Thus, we need to determine the optimum solutions under two situations, viz. ignoring externalities (private or market optimum) and considering externalities (social optimum). Taking the first exercise first, we need to find out the optimum results under usual perfectly competitive market.

**Market Optimization (Ignoring Externalities)** For this, we find the supply function first for the single firm and then for the industry. Recall that the firm's supply function under perfect competition is given by, $P = MC$. Thus,

$$P = MC_F = 500 + 2\, Q_F \text{ (the RHS comes from first derivative of the firm's cost function)}$$

or,
$$Q_F = -250 + 0.5\, P$$

Since all firms have identical cost functions, each would have the same supply function. The sum of all firms' supplies gives the industry supply curve, and since all firms have identical supply, the industry supply is given by the number of firms multiplied by one firm's supply. Thus, industry supply ($Q_I$) is given by

$$Q_I = 200 \times Q_F$$

or, $Q_I = -50{,}000 + 100\,P$

Equilibrium in the industry requires, industry demand = industry supply. Equating the two yields

$$150{,}000 - 100\,P = -50{,}000 + 100\,P$$

or, $P = 1{,}000$

Substitution of $P$ in either the demand or supply function, gives

$$Q_I = 50{,}000$$

Thus, the optimum price-output under usual perfect competition would give $P = 1{,}000$ and $Q_I = 50{,}000$.

Incidentally, note that each firm's output would equal 50,000/200 = 250 and, of course, all firms are price-takers and thus $P = 1{,}000$ for each of them. The profit that each firm would make is then given by

$$\Pi_F = TR_F - TC_F = P_F\,Q_F - [40,\,000 + 500\,Q_F + Q_F^2]$$

$$= 1000 \times 250 - 40{,}000 - 500 \times 250 - (250)^2 = 250{,}000 - [227{,}500]$$

$$= 22{,}500$$

All firms in the industry would make equal profit (recall they all have identical costs). Since firms are making profits, it is a short-run equilibrium. In the long-run, other firms would be attracted to enter the industry and that would tend to wipe out all profits in the long-run. We now move to determine optimum results recognizing externalities.

**Social Optimization (Considering Externalities)** This requires considering the cost to society besides that to the cement firm/industry. Accordingly, the $P = MC$ rule for the supply curve would now give,

$$P = MC_S = MC_F + MEC, \text{ where } MC_S = MC \text{ to the society}$$

or, $P = (500 + 2\,Q_F) + (100 + 2\,Q_F)$ [Note that $TEC$ depends on $Q$, so at firm level $Q = Q_F$ and $MEC$ is the derivative of $TEC$]

or, $P = 600 + 4\,Q_F$

or, $Q_F = -150 + 0.25\,P$

The above function denotes the firm's supply function. To get the industry's supply curve, we multiply by the number of firms, which equals 200. Thus, cement industry supply curve is given by

$$Q_I = 200\,Q_F = -30,\,000 + 50\,P$$

Equating, industry supply with industry demand, yields

$$-30,\,000 + 50\,P = 150,\,000 - 100\,P$$

or, $P = 1{,}200$

Substitution of $P$ value in demand or supply equation, gives

$$Q_I = 30{,}000$$

Thus, the socially optimum outcome would be at $P = 1,200$ and $Q = 30,000$. The comparison of the market and social optimum results for the cement industry indicate that, the social output is lower and social price is higher than their counterpart market or private results.

To get the equilibrium output of each firm, we divide industry output by number of firms, which equals 200, and thus, $Q_F = 150$. At $P = 1,200$ and $Q_F = 150$, each firm's profit is given by,

$$\Pi_F = TR_F - TC_F = P_F Q_F - [40,000 + 500 Q_F + Q_F^2]$$

$$= 1200 \times 150 - [40,000 + 500 \times 150 + (150)^2] = 180, 000 - [137,500]$$

$$= 42,500$$

The industry results under the two models are shown in Fig. 10.6. The downward sloping linear demand curve $AD_1$ denotes the industry demand curve, and $BS_1$ and $CS_2$ curves denote the industry supply curve under private (ignoring negative externality) and social cost (taking account of the negative externality), respectively. Alternatively, $BS_1$ could be called the private supply curve and $CS_2$ the social supply curve. Points $Y$ and $X$ represent equilibriums under private optimization and social optimization, respectively. Equilibrium output and price under market optimality are $OQ_1$ and $OP_1$ and under social optimality are $OQ_2$ and $OP_2$, respectively.

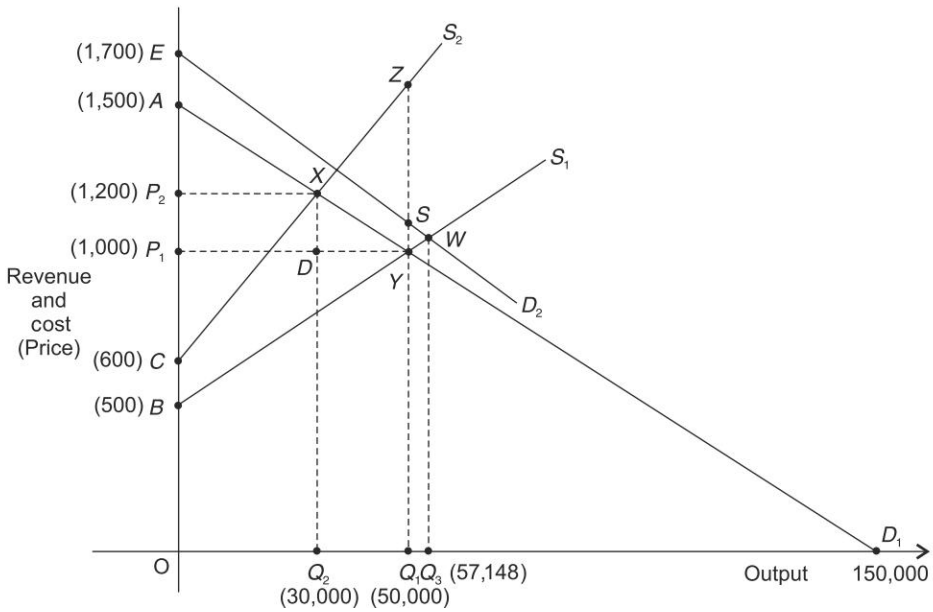



**Fig. 10.6.** *Optimization under net negative (positive) externalities*

The comparison of the two sets of results would confirm the above conclusion, viz. optimal social output is less and price is more than their private counterparts. Why it is so? This happens because the cost is under estimated and hence supply is overestimated under the private optimization, *ceteris paribus*. The market results are optimal for the firm and the industry but sub-optimal socially, and hence represent socially inefficient outcome. In other words, ignoring of (negative) externality prohibits the free market mechanism to allocate resources efficiently and thereby results in **market failure**. In

Fig.10.6, the loss to the society due to negative externality, called the dead weight loss (*DWL*) is given by the area of triangle *XYZ*. This can be explained as follows:

| | *Under Social Optimality (incorporating externalities)* | *Under Market optimality (ignoring externalities)* |
|---|---|---|
| Consumer surplus | Area of triangle $AXP_2$ | Area of triangle $AYP_1$ |
| Producer surplus | Area of triangle $P_2XC$ | Area of triangle $P_1YB$ |
| Externality cost | – | Area *CZYB* |
| Economic surplus | Area of triangle *AXC* | Area of (*AXC* – *XYZ*) |
| Dead weight loss | – | Area of triangle *XYZ* |

Thus, in Fig. 10.6,

$$DWL = YZ \times YD \times 0.5 = [Q_1Z - Q_1Y] \times Q_1Q_2 \times 0.5$$

Note that $Q_1 Z$ equals the height of $S_2$ curve at $Q = OQ_1 = 50{,}000$, which is given by $S_2$ curve. Recall that $S_2$ curve represents the industry supply curve under social optimal situation, whose equation is,

$$Q_I = -30,000 + 50P$$

or, $$P = 600 + 0.02\, Q_I$$

Thus, $$Q_1 Z = 600 + 2\,(50{,}000)$$

$$= 1{,}600$$

And $Q_1 Y = OP_1 = 1{,}000$, and $Q_1Q_2 = OQ_1 - OQ_2 = 50{,}000 - 30{,}000 = 20{,}000$. Thus,

$$DWL = (1{,}600 - 1{,}000) \times 20{,}000 \times 0.5$$

$$= 6{,}000{,}000$$

**Market Optimization under Monopoly** The equilibrium outcomes under the condition of monopoly (ignoring externalities) could also be found by assuming just one firm with the same cost function as above and industry demand curve as the demand facing the lone firm. Pursuing this, and setting *MR* = *MC* (necessary condition for profit maximization), we get

$$TR = PQ = [1500 - 0.01\, Q]\, Q = 1{,}500\, Q - 0.01\, Q^2$$

or, $$MR = 1500 - 0.02\, Q = MC = 500 + 2\, Q$$

or, $$Q = 495 \quad \text{and} \quad P = 1495.05$$

As expected, under monopoly equilibrium output is less and equilibrium price is high relative to perfect competition. The consequences of externality under monopoly for the society could be analyzed similarly. We leave this for students' practice.

## Positive Externalities

Positive externalities case is exactly opposite to the negative externalities case. Positive net externalities are experienced under those goods and services which bring more gains than losses to the third party.

Such goods include education, research and development that bring technical progress, plantation, parks and gardens, etc. Thus, instead of external cost of net negative externality, we would have external benefits under net positive externality. Accordingly, while the market (internal) cost and benefits, and hence the market outcome remains unaltered; the social benefits are enhanced, with no change in social cost, and hence the optimum output goes up under socially optimum solution. The effect on price is ambiguous because the demand curve is downward sloping. Thus, if one analyzes the consequences of positive externalities, he/she would discover that ignoring of those would cause optimum output smaller with price moving to either direction or no change, than it would happen if such externalities were considered. Since the two optimum results differ, the market optimum in relation to social optimum would cause dead weight loss under net positive externalities. To demonstrate this through an example, we could simply change the above numerical example by (i) renaming the company, from cement firm (which cause net negative externality) to, say, Simla Apple Orchard (SAO) (which bring net positive externality), and (ii) replacing the total external cost equation by total external benefit (TEB) equation. Under the changed situation, the market solution would remain unaltered (with equilibrium at point *Y* in Fig. 10.6) and the social optimum would have to be worked out under the new demand curve, which would incorporate the net external benefit in the private benefit as per the market demand curve. To explain the procedure, first get the industry demand curve in its inverse form, as below:

Demand in its normal form is

$$Q = 150{,}000 - 100\,P$$

Its inverse form is given by

$$P = 1{,}500 - 0.01\,Q$$

In this form, the demand curve (recall from Chapter 2 that it is also the marginal utility curve when utility is measured in money terms as the price) reflects the **marginal benefits to the consumers** of the good (private marginal benefits). Suppose, the net total external benefit (*TEB*) is given by

$$TEB = 2{,}000 + 200\,Q - 0.0005\,Q^2$$

Its differential would give the marginal external benefit (*MEB*),

$$MEB = 200 - 0.001\,Q$$

Sum of the private marginal benefit (*PMB*) and *MES* gives the marginal social benefit (*MSB*),

$$MSB = [1,\ 500 - 0.01\,Q] + [200 - 0.001\,Q]$$

or, $$MSB = 1{,}700 - 0.011\,Q$$

or, $$P = 1{,}700 - 0.011\,Q$$

or, $$Q = 154{,}545 - 90.9\,P$$

This equation stands for the **social demand curve, which gives MSB**. It must be noted that, the sum of private demand and external demand is obtained at each output, not at each price as is done when one adds various consumers demand curves to get the industry demand curve. Demand is higher under social demand curve than the private demand curve due to net positive externality. It would be noticed that the social demand equation has a larger intercept and smaller slope ($dQ/dP$) than the private/market demand curve. Accordingly, if it the former were superimposed on Fig. 10.6, the social demand curve would be above the market demand curve (*D*) and the gap between the two would diminish as output

expands. The equilibrium point would be the point of intersection between the social demand curve and social/private supply curve, as shown by point *W* in Fig. 10.6, which is at a point above and to the right of point *Y* in Fig. 10.6. To algebraically determine the social optimum price-output, we must set social demand = social supply = private supply. Thus,

$$154,545 - 90.9P = -50,000 + 100P$$

The solution of this would give, $P = 1,071.48$, which on substitution in demand or supply equation gives $Q = 57,148$. This represents the socially optimum results under positive externality. In comparison with private optimum outcome, viz. $P = 1,000$ and $Q = 50,000$, the socially optimum results have higher output and higher price. The economic surplus to the society under social optimum outcome = area of triangle *EWB* in Figure 10.6 while that under market/private equilibrium, is given by

Economic surplus under market equilibrium = area of triangle *AYB* + marginal external benefit

= area of triangle *AYB* + area *AYSE* (vide Fig. 10.6)

The economic surplus under social optimum exceeds that under market optimum by the area of triangle *SYW*, which represents the dead weight loss under market/private equilibrium versus the socially optimum outcome. We leave it to students to quantify the area of triangle *SYW* and thus *DWL*.

## Countering Externalities

It is because the *DWL* results when we ignore either the net negative externality or net positive externality, it is alleged, that **externalities cause market inefficiency and thereby market failures** under free market economy. The question is, how the *DWL* caused by externalities could be avoided or minimized? There are two alternative ways to do this, viz.

(a) Amend market rules

(b) Amend government rules

Under the first, the moral codes are so designed that no one creates the third party effects or if they do, they compensate (receive compensation from) the third party for its losses (gains). For example, smoking is associated with air pollution. To de-associate smoking from air pollution, either people give up smoking, smoke at private places only, or compensate the people who are suffering from their smoking. Similarly, a cement factory pollutes river/sea when it deposits effluents in them, which cause loss to fishermen who fish in them. To avoid such negative externality, the factory owner must treat those effluents rather than throwing in river/sea or pay compensation to the affected fishermen. If that is done, the negative externalities are internalized, and accordingly the private cost tends to equal the social cost, and thereby the socially optimum results are realized, which are free from *DWL* or misallocation of resources. When externalities are net positive, like in education and *R&D*, the external cost of such activities could be funded entirely through donations, government funding, scholarships, and like and the patent rights are awarded to patent awardees to recover external benefits. If this is done, private benefits would equal social benefits and socially optimum outcomes would be attained. Such are the recommendations of the **Coase theorem**, known behind its inventor, Ronald Coase, to deal with externalities. However, to ascertain its working, the concerned parties must agree to do so, which is easy to say but difficult to practice. It is well-known that negotiations involve valuation of externality and valuation is subjective, and thus is subject to transaction cost. Further, in externalities, the third party is not just one person but usually a large number of affected people/firms. For example, in case of polluting

river/sea, affected people are large number fishermen/fishing firms and even the people who use rivers/sea for swimming, partying, etc. and who live around the river, etc. The Coase theorem merely states that the market rule can solve the externality problem provided the creating and affected parties can bargain successfully, without cost, over the allocation of resources.

The second solution to avoid the social loss of externalities is through what is called corrective taxes and subsidies, also called the **Pigovian taxes/subsidies**, in honour of Arthur Pigou who was the first to recommend the solution. Under this solution, government must impose tax on the product which creates negative externality at the rate at which they create such costs and give subsidy at the rate the product creates positive externality. If this is done, the external cost/benefit would become internal cost/benefit and accordingly socially optimal results would emerge. Economists prefer this solution over the market based solution, for the latter is difficult to practice and the former not only corrects the resource allocation problem but also fetches revenue to the government which can help it to serve other goals like equity and growth. Incidentally, note that, as analyzed in Chapter 8, a tax/subsidy usually creates dead weight loss but the tax/subsidy referred to here and thus called corrective tax/subsidy, is removing DWL. Also, note that **Tobin tax**, which is a tax on international transactions, is like the corrective tax, for it tends to discourage global trade and investments which are alleged to aggravate international inequalities.

## 10.12 PRICING UNDER RISK AND UNCERTAINTY

Recall that all through our discussion so far, we have implicitly assumed that the cost and demand functions known/given to us are true and valid in output and price decisions. The assumption is far from truth for two reasons. One, recall that the said functions are obtained largely (with support from experts' intuitions) on the basis of the past data analyses through multiple regression technique and the data are known to suffer from lies besides the limitations of the estimation technique. Accordingly, at best, the estimated functions may merely be approximately valid for the past average functions. Two, the optimum decisions precede the actions, and thus the results are meant for use in future. Further, future is neither a simple average of the past nor is ever predictable exactly. In view of these concerns, the cost and demand functions, and even the firms' objective are subject to risk and uncertainty.

How to incorporate risk and uncertainty in the determination of optimum price-output decisions is the subject matter for this last section on pricing? Before attempting to answer this question, let us understand the difference between risk and uncertainty and also the various attitudes of decision-makers with regard to risk-taking. It is convenient to start with risk. Risk refers to a situation in which the concerned party has full knowledge about the likelihoods (called probabilities) of alternative outcomes as well as the outcomes or consequences of alternative happenings. Thus, to give an example, suppose an investment firm promises you a guaranteed return of, say 5 percent on your investment and also offers you an option where the firm promises you that there is 50 percent charge that you will make 10 percent return and there is 50 percent chance that your return would be zero. If so, you are in a risky situation, for you know the probabilities attached to all possible rates of return on your investment, and you could measure/quantify your risk. In contrast, uncertainty is a situation in which the decision-maker does not have the complete probability function. Stated alternatively, either all probabilities are not known or/and all possible outcomes are not known fully, rendering the degree of uncertainty non-measurable. In real life, most businesses are confronted with uncertainty. However, since uncertainty still remains non-quantifiable, businesses base their decisions on risk considerations only.

Three kinds of attitudes towards risk are distinguished. They are known as

(a) Risk neutrals

(b) Risk averse

(c) Risk seekers/lovers

A decision maker is described as a risk neutral if he/she takes decisions on the basis of expected outcomes. Thus, in our above example, the investment bank's offer of 5 percent sure return versus 10 percent with 50 percent probability and zero percent with 50 percent probability leaves him indifferent, for the expected ($E$) outcome ($E = 0.5 \times 10 + 0.5 \times 0 = 5$) under the second option exactly equals the sure outcome. Further, a person is risk averse if he prefers a sure outcome over a risky outcome, where the two outcomes have equal expected value. Alternatively, a person is risk averse if he selects an option under which he gets the same expected return with lower risk or larger expected return with same risk than the alternative option. Thus, if our investor in the above example prefers 5 percent sure return over the other option, he is deemed as a risk averse. In contrast, if the investor chooses the risky option over the sure option of our example; or prefers an option with the same expected value but higher risk or lower expected return with same risk; the investor is deemed as a risk seeker or risk lover. In other words, while risk-averse people attach some premium to sure outcome, risk seekers attach a premium to risky situations. The rate of premium varies with the degree of risk-averseness, and hence is subjective. Most decision makers today fall in the category of risk-averse but with varying averseness.

Since uncertainty is not quantifiable, businesses have to let it go in favour of risk consideration. How then the risk is incorporated in arriving at the optimum price-output decisions? This is simple in calculations but difficult in generating the required details. Business must consider all possible scenarios with respect to each crucial parameter (like input prices, input productivities, raw-material prices, prices of related goods both in production and demand, etc.), assign probabilities to each scenario, and assess the likely outcome (demand and costs) under each assumption. Once that is done, calculations could be done to generate expected values and measures of risk (like standard deviation) for each possible scenario. Once this is ready, business must compare and take the so called calculated risk, given its attitude towards risk, while deciding on the final estimates of cost and demand functions, which could then be analyzed just the way we have explained above while ignoring risk and uncertainty. We consider below a simple example.

Suppose a risk neutral firm has to decide its output/price level in the midst of risky situation faced by the industry. The risk is in terms of the demand for the industry output. Two estimates of demand are available and each has a probability of 50 percent. They are,

$$Q_d = 500 - 6P \qquad \text{and} \qquad Q_d = 300 - 4P$$

The firm is sure that the supply function of the industry is,

$$Q_s = -200 + 3P$$

The firm has full confidence that its total cost function is,

$$TC = 10 + 25Q + 5Q^2$$

Under the above data, the expected industry demand curve is given by

$$E(Q_d) = 0.5[500 - 6P] = 0.5[300 - 4P] = 400 - 5P$$

Equilibrium in the industry requires, $E(Q_d) = Q_s$,

or, $400 - 5P = -200 + 3P$

or, $P = 75$, and industry output = 25

Equilibrium at firm is given by, $MC = P$

or, $25 + 10Q = 75$

or, $Q = 5$

Firm's profit = $PQ - TC = 75 \times 5 - [10 + 25 \times 5 + 5(5)^2]$

$= 115$

Thus, the risk-neutral firm must produce 5 units of its output and, expect to earn profit in the amount of 115. However, if the firm were risk-averse, as most are, or risk seeker, it would need to compute standard deviation (or some other measure of risk, like range and coefficient of variation), a measure of risk also to take the appropriate decision. This is easy to calculate and we leave this for students to practice. Similarly, the firm may not be sure about its cost function, as technology, input prices, weather, tax rates etc. are subject to change. However, if the firm could generate probabilities and costs under all alternative scenarios, it could determine its expected cost function. Given the expected cost and expected demand function, the risk-neutral firm would easily determine its optimum output-price decision. Further discussion on risk and uncertainty analysis is provided in the next chapter. This completes our long drawn discussion of product pricing.

## REFERENCES

1. Akerlof, George (1970): The Market for Lemons: Quality Uncertainty and the Market Mechanism, *Quarterly Journal of Economics*.
2. Baye, Michael (2009): *Managerial Economics*, McGraw Publishers.
3. Cohen, K.J. and R.M. Cyert (1965): *Theory of the Firm*, Englewood Cliffs, Prentice-Hall.
4. Fellner, W. (1949): *Competition among the Few*, New York, Knopf.
5. Pindyck, Robert s and Daniel L.Rubinfeld (2009): *Microeconomics*, 7th edition, Pearson.

## CASELETS

1. Singh Brothers (SB) runs a bar and aims at profit maximization. It has a fixed cost of Rs. 50 and its average variable cost is constant at Rs. 20 per drink. The estimated demand function for drinks by men ($M$) is

   $P_M = 180 - 20\,Q_M$

   and for women ($W$) is

   $P_W = 80 - 5Q_W$

   where, $P$ = price, and $Q$ = quantity demanded

   (a) If price discrimination is legal and the Singh Brothers decide to practice the same, what prices would they set in the two markets? Determine the corresponding quantities in the two markets and the profit that Singh Brothers would make under such a situation.

(b) If price discrimination is illegal, determine the aggregate demand function facing the Singh Brothers. Also determine the equilibrium price, quantity and the profit under uniform pricing.

(c) Check on the validity of the proposition, "price is lower in the more elastic market than in the less elastic market under price discrimination".=

(d) Determine the dead weight loss under each of situations in (a) and (b) above.

(e) Compare the results in (a) and (b) above and comment.

2. Radhe Multiplex Ltd. (RML) offers movie shows both as matinee shows ($M$) as well as afternoon-evening shows ($A$). Its demands for each of two shows and the uniform cost function are the following:

$P_A = 400 - 0.5\ Q_A \qquad P_M = 100 - 0.5\ Q_M \qquad TC = 25 + 20\ Q + 0.125\ Q^2$

where, $P$ = movie ticket price, $Q$ = number of seats sold, and $TC$ = total cost

(a) Determine the optimum price and quantity for each show under price discrimination. How much profit *RML* would make from this business?

(b) If *RML* is required to charge uniform price in two shows, what prices, quantities it would set and how much profit the firm would make?

(c) Which of the above two pricing practices is better for the economy? Why?

3. Mohan Pool Services (MPS) provides swimming facilities. The estimated market demand (identical for all customers) and cost functions are as follows;

$Q = 150 - 2\ P \qquad TC = 200 + 2\ Q + 0.5\ Q^2$

Where, $Q$ = number of customers, $P$ = price per swim, and $TC$ = total cost

(a) If MPS operates its pool facilities under uniform pricing, what price it should set for swimming? Determine the number of people who will avail its facilities, the profit that the firm would make and consumer surplus that would occur to the swimmers.

(b) Is there a scope for MPS to practice two-part pricing? If yes, how should it go about doing that and what profit the firm would make under that situation?

(c) Compare the results in parts (a) and (b) above and check if the two-part pricing enables MPS to capture consumer surplus, partly or fully.

(d) Could MPS practice block pricing? If yes, how?

4. The steel industry is subject to the following demand and supply relationships:

$Q_d = 50{,}000 - 10\ P \qquad \text{and} \qquad Q_s = -20{,}000 + 25\ P$

The industry generates net negative externalities, whose marginal cost to the third party is given by

$MEC = 50 + 0.01\ Q$

where, $Q$ = industry output in number of houses, $P$ = price in rupees thousands and $MEC$ = marginal external cost in rupees thousands.

(a) If the industry ignores externality cost, what price and output would prevail in the economy?

(b) Determine the socially optimum price and output for the industry.

(c) Determine the dead weight loss that would occur if externalities were ignored.

5. Mohan Real Estate Ltd. (MRE) builds on sells houses to its prospective customers. It must decide on its expansion plan. Its total cost function is given by

$$TC = 90,000 + 25\,Q + 2.5\,Q^2$$

The economy is currently getting out of slowdown brought about by global recession. The company expects that there is 50 percent chance that the growth rate would be 8 percent, 30 percent chance that growth will be 7 percent and 20 percent chance that growth might be at its lowest rate of 6 per cent. Depending on which of the three scenarios will happen, the company expects its demand functions to be the following under the three scenarios, respectively:

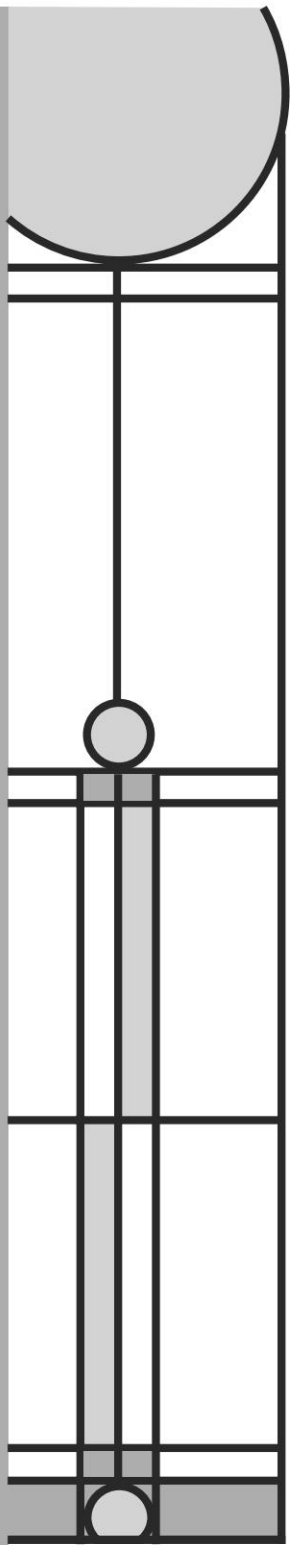
$$Q = 5,000 - 5\,P$$

$$Q = 4,000 - 4\,P$$

$$Q = 3,000 - 3\,P$$

If MRE is risk-neutral, determine its profit-maximizing price and output. How much profit the company will be expected to make if it implements such a decision? Determine the level of risk that MRE is taking in the business.

The McGraw-Hill Companies



CHAPTER

11

# INVESTMENT ANALYSIS

One of the significant decisions a firm takes is "what to produce?" This question is relevant not only for a new enterprise but also for an existing firm. This is so because while a new entrant must decide the good(s) it would like to produce, an existing unit might like to expand its business and in that case it must decide as to whether expand in the old product line or with a new business, i.e., related or unrelated goods. Today, the competition is tough, most firms are going in for diversification, more and more of new products are coming out month after month, integration has proved rewarding, and globalization has become the buzz word. In the circumstances, a firm has to constantly examine the viability of its old ventures vis-a-vis the potential ventures. Since the number of goods a firm could produce is much more than what a firm could possibly undertake and production involves investments, investment analysis is an inevitable activity for every firm. The present chapter provides a full discussion of the various concepts, techniques and steps involved in investment analysis, and useful insights into decision-making with regard to "where to invest" or "what to produce".

## 11.1 MEANING AND SIGNIFICANCE

Investment is an economic activity of employment of funds with the expectation of receiving a stream of benefits in future. Thus, it would include commitments of funds in

(a) Financial assets
(b) Physical assets
(c) Productive activities.

The first group includes bank deposits, company deposits, contributions to provident fund, purchases of national savings certificates, mutual funds, shares and debentures of companies, options and derivatives, etc. The second category consists of purchases of precious metals (gold, silver, diamond, etc.), real estate (land and buildings) and other durable goods, which are purchased not for use but for profit-making. Incidentally, note that if an individual buys a house for own family dwelling, for

renting it out, or even for selling it at a later date, it is considered as an investment in economics. The last category, viz. investment in productive activities, which include renting out residential houses, consists of investments in the production of goods and services. Thus, if I quit my job and start a firm to manufacture computers or television or to supply management consultancy services or to run an educational institute, I would need funds, commitment of which in any one of these and similar activities would mean investment in a productive activity.

In terms of the formal definition, all the three types of above activities refer to investments. However, in the practical sense, investments in productive activities alone are considered as capital expenditures (the phrase synonymous to investments), and these alone are analysed under the subject "capital budgeting". Investments in financial and physical assets are dealt with under the topic "portfolio management". This chapter concentrates on capital budgeting and accordingly investment of the first two kinds would not be analysed explicitly. Incidentally, the concepts and techniques discussed below are quite versatile, and thus an understanding of them would be useful even for portfolio selection and management.

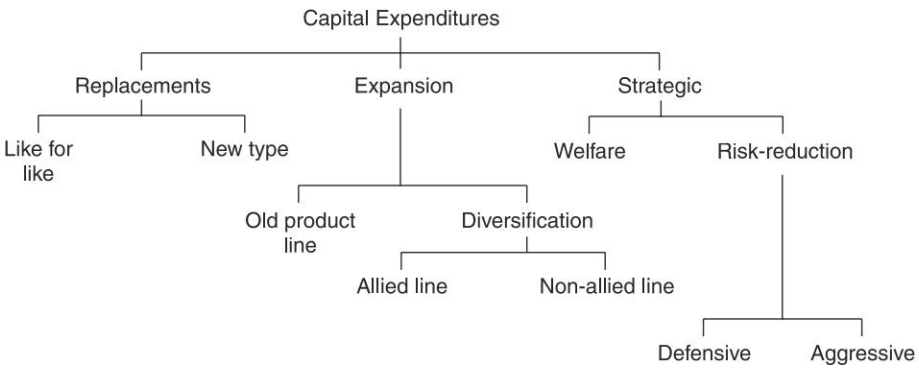



**Chart 11.1** *Type of capital expenditures*

Investments in productive activities are further classified into three categories, each of which is further divided into sub-categories. A detail categorization is provided in Chart 11.1.

Capital expenditures on replacements include those on replacements of worn-out machines by identical new machines, replacements of obsolete plants/machines by modern plants/machines, replacements of labour by machines, relocation of factory/office, renovation of factory/office, etc. These expenditures are obviously undertaken by the existing firms only.

Capital expenditures on expansion consist of those incurred on enlarging the capacity in the existing lines and new lines of production. For example, if an existing cotton textile firm, having a capacity of, say, 500,000 metres of cloth per month wishes to expand, it could do so in a number of ways. First, it could increase its capacity in the cotton textiles business from 500,000 metres per month, to say, 800,000 metres per month. Second, it could launch on a new but allied production, say, on woolen and synthetic clothing. Third, it could simultaneously start manufacturing some unrelated good, such as typewriters, televisions, calculators or even running hotels and colleges. The first category belongs to expansion in the old product line group, the second to diversification in allied line, and the last to diversification in non-allied line group. While the existing firms could go for anyone of these three expansions, the new entrants (investors) could think of diversifications in the non-allied line only.

Capital expenditures of strategic nature include those on improving the welfare of the employees and on reducing the risk of the business. The former category consists of expenditures incurred by a firm on providing facilities such as canteen, indoor games, common room with musical instruments, etc. and on controls of air pollution, etc. All such investments are strategic, for these might resolve industrial relations problems, which hamper production and factor efficiency. A product manager has to carefully assess the mood and motives of the union leaders to decide such expenditures. If it works, a little investment of this kind can save a firm from serious disputes and strikes, but if it does not work, it would mean unnecessary capital expenditure.

The latter group of capital expenditures, called the risk-reducing type of strategic investments, is further divided into two classes, viz., defensive and aggressive investments. Examples of the first category would include expenditures on storing inventories for unusually long periods in the fear of their shortages or inflation in future and purchases of power generators as safeguards against power breakdowns. These expenditures are strategic, for in case the fear turns out to be true and significant, the firm is able to operate its plant regularly just because of advance precautionary planning, but if the fear proves wrong, those expenditures become redundant and thus uneconomic. Expenditures on research and development fall in the category of aggressive expenditures to avoid risk. Thus, if a cotton textile firm fears shortage of cotton in the coming years, it could commission a research/consulting study to look into the matter and discover an alternative raw-material for clothing. This perhaps could have been one of the forces behind the discovery of woolen and synthetic materials for clothing. Expenditures of this kind are strategic in the sense that research and development activities could produce miracles or just nothing. Further, these are aggressive, for through these a firm tries to do without a thing, which is either not available or is too expensive.

A clear understanding of the various kinds of capital expenditures would indicate that, replacement investments are made generally to reduce the unit cost of production, while expansionary investments are undertaken to boost up sales. However, each category could have some impact on the other variable. Thus, if an old machine is replaced by a new machine, either of the earlier type or of a new variety, it would lead not only to a reduction in unit cost but also in some expansion of output, for usually the new machine would be more efficient than the old one. Similarly, expansionary investments could very well result in reduction in unit cost, for there are economies of scale in mass production.

It is important to note that only replacements and expansionary investments are amenable to analysis. This is because analysis requires identification and quantification of costs, and benefits due to projects, and this is possible only in the case of replacement and expansion projects. In case of strategic projects, it is difficult to foresee all the possible benefits and quite often decisions have to be made instantaneously and/or the expected net benefits are so much and so obvious that the analysis is redundant.

From the above discussion it would be clear that the investment decisions of the type discussed here are concerned with the managerial question "what to produce?" This decision is significant for four reasons:

(a) Investment funds are limited and versatile

(b) Investment opportunities are plenty and varied in terms of return on investment (ROI) and risk

(c) Investment decisions have long term impact on the well-being of the investor and the society

(d) Investments are practically irreversible

(e) Investments are self-financing

The first two reasons apply even for consumption decisions, for consumers' income is limited and

versatile in the sense that it can be used to purchase any good or service, and there are numerous candidates for inclusion in a consumption basket, all of which cannot be accommodated within a limited income. The last three factors are specific to investment decisions. A good investment decision can bring fortune not only to the investor but also to coming generations. Today we have the Tatas and Birlas, not necessarily because their present generations are making good investment decisions but probably because their forefathers had made excellent decisions in such matters. Investment decisions are irreversible in the sense that once a decision is made and the project implemented, undoing all that might be possible only at a significant loss. This is because such decisions involve buying land, constructing buildings, buying and installing plants and machines, recruiting managerial and other essential staff etc., and undoing all this could very well lead to significant loss and embarrassment. For these reasons, all investment decisions need to be made after a thorough and careful analysis. Lastly, capital expenditures (or investments), unlike consumption expenditure, if taken after proper analysis, usually recover not only all the costs behind them but also bring considerable profits to the investors.

## 11.2 TIME VALUE OF MONEY

Capital expenditures involve costs and benefits over time. For example, if an investor wants to go in for a sugar factory, he would incur capital expenditure in the purchase of land, construction of buildings, buying of plant and machinery, purchasing of sugarcane, payment of wages and salaries to the employees, etc. While some of these expenditures are non-recurring, others are recurring expenditures. Further, even all non-recurring expenditures may not be incurred in a particular month or a year but rather spread over a number of years. Once the project is commissioned, it would generate output in the form of sugar and molasses, which would be forthcoming year after year until the project terminates. Thus, investment projects are associated with costs and benefits which are spread over a period of time.

Analysis of an investment project involves comparison of costs and benefits associated with it. It is not only the amount of costs and benefits that are relevant for this purpose, but also the timings of their occurrences. This is because money has time value for the following reasons:

(a) Earning power of money
(b) Inflation/deflation
(c) Uncertainty

Whenever there are investment opportunities, money has the earning power. The earning power is represented by the opportunity cost of money, the least of which would be the rate at which banks accept deposits. By this virtue, today's sum of money is equivalent to a larger sum in the future.

Money is needed not for its own sake but for its purchasing power, which varies inversely with the price level. Thus, when one considers the use or exchange value of money, he/she would discover that it changes with inflation and deflation for a given sum. In particular, during inflation, today's sum of money is equivalent to a larger sum in the future, and quite the reverse holds good during deflation. This is easy to see. Thus, for example, if the price of, say, rice of a particular quality, is Rs. 40 per kilogram today and Rs. 50 per kilogram after one year, then Rs. 100 equals 2.5 kilogram of rice this year and 2 kilogram next year. Incidentally, note that the price of rice is taken just for illustration purpose. The relevant price is the weighted average price of goods and services on which the investor spends his/her money income.

Investment as a rule deals with the future and the future is uncertain. Investment is concerned with commitment of funds today with the expectation of receiving a stream of benefits in the future. Thus, it involves a trade-off between a certain sum today and an uncertain series of sums in the future. Uncertainty is undesirable in the sense that given a particular amount, an individual would always prefer receiving it right now instead of receiving the equivalent amount at a future date. This is because "a bird in hand is worth two in the bush". Thus, due to uncertainty, today's sum of money is equivalent to a larger sum in the future.

During the period of inflation, all the three factors representing the time value of money, work in the same direction—today's money is worth more than the equivalent amount in the future and thus money gains in value over time. But during deflation, while money gains in value over time due to its earning power and uncertainty associated with the future, it loses in value due to falling prices. Further, there is no guarantee that the gain and loss either just cancels out or one is always larger than the other, and consequently nothing unambiguously can be stated regarding the direction of change in the value of money over time during deflation.

Since investment involves expenditures and revenues over time, it is imperative to adjust for the time value of money for conducting a meaningful investment analysis. The adjustment for the earning power of money is made through the principles of compounding and discounting, which can, in fact, be also used to adjust for inflation/deflation as well as uncertainty. A discussion on these follows:

## Compounding Principle

Under the compounding principle, the future value of a present sum is found, given the earning power (interest rate) of money and the frequency of compounding. For example the value of Rs. 100 of today after one year, given the rate of interest of 10 per cent per year and the compounding to be done once in a year, equals

$$\text{Rs. } 100 + 0.10\,(100) = \text{Rs. } 100\,(1 + .10) = \text{Rs. } 110$$

The value of this sum after two years equals,

$$\begin{aligned} &\text{Rs. } 110 + .10\,(110) \\ &= \text{Rs. } 110\,[1 + .10] \\ &= \text{Rs. } 100\,(1 + 0.10)\,(1 + .10) \\ &= \text{Rs. } 100\,(1 + .10)^2 \\ &= \text{Rs. } 121 \end{aligned}$$

Similarly, the value of the sum after three years equals Rs. $100\,(1 + .10)^3$ and so on. Thus, the general formula is

$$Y = X\,(1 + i)^T \tag{11.1}$$

where, $Y$ = Final sum

$X$ = Present sum

$i$ = Interest rate per period (year)

$T$ = Number of periods (year)

Thus, the value of Rs. 100 after 7 years, interest rate being 10 per cent equals Rs. 194.87 (vide Appendix B, Table B.1).

If the interest rate is compounded more frequently than once in a period, the formula undergoes a change. For example, if the compounding was done twice a period, then Rs. 100 of today, given the interest rate of 10 per cent per period, would, after one year equal

$$\text{Rs. } 100 + \frac{0.10}{2}(100) + \frac{0.10}{2}\left[100 + \frac{0.10}{2}(100)\right]$$

Where, the second term represents the interest in the first half period and the third term the interest in the second half period. This amount can be rewritten as

$$\text{Rs. } 105 + \frac{0.10}{2}(105)$$

$$= \text{Rs. } 105\left(1 + \frac{0.10}{2}\right)$$

$$= \text{Rs. } 100\left(1 + \frac{0.10}{2}\right)\left(1 + \frac{0.10}{2}\right)$$

$$= \text{Rs. } 100\left(1 + \frac{.10}{2}\right)^2$$

$$= \text{Rs. } 110.25$$

Two year later, the value of the above sum would equal

$$\text{Rs. } 11.25 + \frac{0.10}{2}(110.25) + \frac{0.10}{2}\left[110.25 + \frac{0.10}{2}(110.25)\right]$$

$$= \text{Rs. } 110.25 + 5.5125 + \frac{0.10}{2}(115.7625)$$

$$= \text{Rs. } 115.7625\left(1 + \frac{0.10}{2}\right)$$

$$= \text{Rs. } 110.25\left(1 + \frac{0.10}{2}\right)\left(1 + \frac{0.10}{2}\right)$$

$$= \text{Rs. } 110.25\left(1 + \frac{0.10}{2}\right)^2$$

$$= \text{Rs. } 100\left(1 + \frac{0.10}{2}\right)^2\left(1 + \frac{0.10}{2}\right)^2$$

$$= \text{Rs. } 100\left(1 + \frac{0.10}{2}\right)^4$$

Similarly, three years later the value would be Rs. $100\left(1 + \frac{0.10}{2}\right)^6$ and so on. Thus, the general formula would be

$$Y = X\left(1 + \frac{i}{2}\right)^{2T}$$

Which on further generalization on number of times interest is compounded to $n$ times/period, leads to

$$Y = X\left(1 + \frac{i}{n}\right)^{nT} \tag{11.2}$$

Thus, the value of Rs. 100 after 7 years, given the interest rate of 10 per cent per year and the interest to be compounded quarterly equals

$$\text{Rs. } 100\left(1 + \frac{.10}{4}\right)^{(4)(7)}$$

$= \text{Rs. } 100\ (1 + 0.025)^{28}$

$= \text{Rs. } 199.65$

In this case, money almost doubles in a period of seven years.

If the interest rate is compounded continuously (instantaneously), that is, $n$ tends to infinity, the formula 11.2 changes to

$$Y = X\, e^{iT} \tag{11.3}$$

Where '$e$' stands for the mathematical jargon 'exponent' which equals 2.78. Thus, the value of Rs. 100 after 7 years, interest rate being 10 per cent compounding to be carried out continuously, equals

$\text{Rs. } 100\ e^{(.10)\,(7)}$

$= \text{Rs. } 201.38$

From the formulae (11.1), (11.2) and (11.3), and the numerical example of the value of Rs. 100 after 7 years under the different frequencies of compounding, it is clear that the more frequent is the compounding, the greater is the compounded value. Also, as is obvious from the formulae, compounded value is a positive function of each of the present sum, interest rate and the number of periods.

## Discounting Principle

The discounting principle is just the inverse of the compounding principle. Under this theory, one finds the present value of a future sum, given the rate of interest, the future date and the frequency of discounting. Thus, the present value of Rs. 110 of one year hence, given the interest rate of 10 per cent and discounting frequency as once in a year, equals Rs. 100. Similarly, the present value of Rs. 121 of two years hence under the same interest rate and frequency equals Rs. 100, and so on divide Appendix B, Table B-2). Under different frequencies of discounting, the amount varies. The three discounting formulae corresponding to the three compounding formulas of above are the following:

$$X = Y\left[\frac{1}{1+i}\right]^{T} \tag{11.4}$$

$$X = Y\left[\frac{1}{1+\frac{i}{n}}\right]^{nT} \tag{11.5}$$

$$X = Y\left[\frac{1}{e}\right]^{iT} \tag{11.6}$$

Where, formula (11.4) stands for once in a period discounting, formula (11.5) for $n$ times in a period discounting and formula (11.6) for continuous discounting. Given the future sum, interest rate, discounting frequency and the future value, the present value, also called the **discounted value**, can easily be obtained through the aid of these formulae. It is clear from the formulae that the present value depends positively on the future sum and negatively on each of the interest rate, frequency of discounting and the length of the period between the future data and the present time.

As presented above, formulae (11.1) through (11.6) serve the purpose of adjusting for the earning power of money alone. However, these could easily be modified to serve as the useful tools even for adjusting for the other two factors associated with the time value of money, viz. changing price and

uncertainty. This is done through redefining the variable *i*. In the foregoing discussion, *i* stood for the interest rate or the earning power of money. It can be redefined to include the rate of inflation/deflation and charge for uncertainty, if the latter could be quantified. Thus, if *i* stood for the interest rate as before, *p* for the rate of change in price which could be positive (inflation) or negative (deflation), and *u* for the charge (positive) of uncertainty, the *i* in the above formulae could be replaced by *d* (the discount rate), which is defined as

$$d = i + p + u \tag{11.7}$$

Thus, if $i = 10$ per cent, $p = 5$ per cent and $u = 3$ per cent, $d$ equals 18 per cent; and Rs. 100 of today is worth Rs. 118 after one year or Rs. 118 of next year equals Rs. 100 this year, the time value of money being adjusted once in a year.

It is not difficult to estimate the earning power of money (*i*). Also, the likely rate of change in price in future can be predicted on the basis of price models. Further, fairly reasonable estimates of these two variables are often available for most countries. However, neither are there reasonable models for estimating the degree of uncertainty nor do we have any reliable estimates for it, for uncertainty varies a great deal from project to project Thus, the above method of adjusting for uncertainty is not quite appropriate. In a later section, the alternative methods of dealing with uncertainty will be dealt with in detail.

## 11.3 CASH FLOWS AND MEASURES OF INVESTMENT WORTH

Investment generates cash outflows and cash inflows over time. The former are partly non-recurring and partly recurring, while the latter are of the recurring variety only. The non-recurring cash outflows consist of expenditures on land, buildings, plant and machinery, furniture, technical know-how fees, etc., and these are committed until the project is commissioned. The recurring cash outflows include expenses on wages and salaries, raw-materials, electricity, telephones, promotional activities, etc., and these are incurred year after year from the day the project is commissioned until the day the project is wound up. Besides these costs, there are interest cost, depreciation cost, taxes and subsidies. In addition, there are externalities [the side effects of economic activities] which could be beneficial as well as harmful. The beneficial externalities include improvements in the environment, which are off-shoots of activities such as construction of dams, gardens, swimming pools, etc. The harmful externalities consist of air and water pollution, which results from activities leading to throwing of smokes and waste (e.g. textile mills), traffic congestion (caused by, say, establishing a factory in the heart of a city), accident hazards (such as the one caused by Union Carbide in December 1984 at Bhopal), etc.

The cash inflows essentially comprise of the sales of basic and by-products which take place year after year and of the salvage or scrap value of the project which is realized when the project is sold out/ wound up. Incidentally note that while estimating the cash outflows and cash inflows from an investment project, care must be taken to truly identifying these flows which are the result of the project *per se*. Recall from section 11.1 above that an investment need not be an altogether new activity and if it were in the form of a replacement investment or an investment resulting into multiplication of capacity in the old product line, etc., then there will be cash flows both in the absence of such an investment as well as in the presence of it. In that case, the analyst must attribute only the net cash flows to the project, which consists of the difference between **cash flows with the project and those without the project**, and not the difference between cash flows after the project and before the project. This is because such projects generate cash flows even if the additional investment is not undertaken.

Since money has time value, all cash outflows and cash inflows must be recorded by the time of their occurrence. Of course, time even by day is relevant in this context, but for the sake of simplicity it is customary to treat the year as a unit of time in capital budgeting. Such an assumption does not lead to too much of distortion, for the time value of money for periods less than a year is not usually very significant. In view of this, we shall follow the tradition of treating the year as a unit of time while adjusting for the time value of money. For the similar reasoning, the compounding and discounting would be assumed to be done once in a year.

A project would thus have **expected cash outflows and cash inflows**, recorded on a yearly basis, spread over its life time. In the beginning, until the project is commissioned into production, there would be cash outflows only but later on both outflows and inflows would co-exist, and inflows would normally exceed the outflows. If we treat outflows as negative and inflows as positive, and convert these into net cash inflows (NCI), there would be negative NCI in the beginning and positive NCI later on. Thus, an investment project with two years of gestation period and seven years of productive life would generate NCI of the following type:

| Year: | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 |
|---|---|---|---|---|---|---|---|---|---|
| NCI (Rs. millions): | –100 | –200 | –10 | 20 | 50 | 100 | 100 | 100 | 150 |

In this example, the project implementation starts in 2010, when expenditure is incurred on technical known-how, purchase of land, construction of buildings, purchases of plant and machinery, etc., all of which is not complete in the first year but spills over to the following year as well. In the beginning of the third year, the project is ready for commercial production and starts yielding cash inflows. As one would normally expect, it take a few years for the project to mature in terms of the optimum capacity utilization. In the above example, capacity utilization is considered to be limited in the first three years of operation after which it goes to the optimum level. In the last year of the project life, the project yields cash inflows not only through sales but also through its scrap value, which consists of the market value of its assets and goodwill, if any.

In the above project, the capital cost is Rs. 300 millions, if the time value of money is ignored. It is customary to estimate the project cost as on completion, giving due consideration to the time value of money. Thus, under the compound rate of 20 per cent, the project cost equals rupees

$$100\ (1 + .20) + 200 = 320$$

From 2012 the project generates cash inflows in terms of sales, and cash outflows in terms of recurring costs of wages, raw materials, etc. The NCI during 2012 through 2018 represents the year-wise difference between such cash inflows and cash outflows.

The investor or the analyst would need to examine the viability of this project *per se* and in relation to alternative projects available to him. In other words, there is a need for **accept-reject decision** and **ranking decision** with regard to each alternative project. Obviously, such decisions depend on the objective of the investor, among other things. Most investors aim at maximizing the value of their firm, which is tantamount to the maximization of the return on investment (ROI). Accordingly, in the analysis that follows, unless otherwise stated, we will assume maximization of ROI as the sole objective of the investor.

Given the firm's objective and given the expected NCI from the alternative projects, how does an investor or analyst arrives at the investment decision? Since various projects generate cash flows over

years, net cash inflows of one project may not be either more or less in each year than that of the other project, and money has time value, it may not be possible to arrive at investment decision of the basis of raw data on net cash inflows. For the above project, for example, it will not be possible to decide whether it is a viable project or not because money has time value and the amounts spread over many years cannot be just added and judged on. The situation is worse when one has to rank alternative projects. To aid under such circumstances, economists and finance experts have designed certain techniques which are now known as the techniques for investment decisions or measures of investment worth. In what follows, we shall discuss each of these techniques in terms of their meanings, computation procedures and rules for investment decisions through a common example. Subsequently, we shall examine their salient features (merits and limitations), comment on the relationship among them and provided some tips for their use. For this purpose, let us consider the following four investment projects:

**Table 11.1** *Description of Hypothetical Investments (Rs. millions)*

| *Project* | *Investment cost in year 2010* | *Net cash inflows in year 2011* | *Net cash inflows in year 2012* | *Net cash inflows in year 2013* |
|---|---|---|---|---|
| *A* | 2500 | 2500 | 125 | 125 |
| *B* | 2500 | 1250 | 1250 | 1250 |
| *C* | 2500 | 500 | 1000 | 3000 |
| *D* | 2500 | 2500 | 750 | 750 |

For the sake of simplicity, the above example assumes that all alternative projects have uniform investment cost (Rs. 2500 millions), are commissionable within a year, and have uniform project life (3 years).

## Pay-Back Period

The pay-back (or pay off) period is the number of years a project takes to recover its investment (original) cost. For calculating its value, one simply takes a cumulative sum of NCI until the sum equals (or exceeds) the investment cost. The number of years to which this amount is cumulative, gives the pay-back period. Putting it mathematically, the pay-back period is defined as $P$ where $P$ is the lowest value of $t$ for which the following condition holds:

$$C \leq \sum_{t=1}^{P} R_t \tag{11.8a}$$

Where, $C$ = Project (investment) cost

$R_t$ = net cash inflow in year $t$

If NCIs are uniform, the pay-back period ($P$) is given by

$$P = \frac{C}{R} \tag{11.8b}$$

Where, $R$ = NCI in a year

For the four projects in the above hypothetical example, the pay-back periods are as follows:*

* In calculating the pay-back period we have assumed that NCF are evenly distributed throughout the year.

**Table 11.2** *Pay-Back Period (in years)*

| Projects | *A* | *B* | *C* | *D* |
|---|---|---|---|---|
| Pay-Back period | 1 | 2 | $2\frac{1}{3}$ | 1 |
| Rank | 1 | 3 | 4 | 1 |

On the pay-back period criterion, a project would be acceptable if and only if its pay-back period is no more than the desired pay-back period as stipulated by the investor. The desired pay-back period, in turn, depends on the investor's need for funds and his expectations about the investment opportunities in future vis-à-vis present. If the investor needs funds next year, say, to perform his daughter's marriage, his desired pay-back period is just one year. Similarly, if he feels that investment opportunities are going to be much better after two years than today, he might stipulate a two-year desired pay-back period.

In terms of ranking alternative investments, the lower the pay-back period, the better is the investment. Thus, on this criterion, projects *A* and *D* are equally good, and they are better than projects *B* and *C*, and project *B* is better than projects *C*. Thus, if the investor has Rs. 5,000 millions to invest, he would go in for projects *A* and *D* alone and if he has Rs. 7,500 millions, he would undertake projects *A*, *B* and *D*.

There is a variant of the above pay-back period. It is called the discounted pay-back period. Under this method, instead of using the raw-data on net cash flows, the discounted values of net cash inflows are used. In all other ways, the two versions are identical. Since the discounted values are less than their undiscounted values, the discounted pay-back period is never less than the undiscounted pay-back period. At the discount rate of 10 per cent, the discounted value of NCI from project *B*, for instance, would be as follows:

$$\frac{1250}{1.10} = 1136; \quad \frac{1250}{(1.10)^2} = 1033; \quad \frac{1250}{(1.10)^3} = 939$$

The cumulative sum of the first two numbers is less than the project cost of Rs. 2,500 millions and hence the discounted pay-back period is more than two years $\left(2\frac{1}{3} \text{ years approximately}\right)$ as against the undiscounted pay-back period of exact two years.

## Average Annual Rate of Return

The average annual rate of return (AARR), also called the accounting rate of return, on an investment is defined as the percentage of average annual net returns to the investment costs. For its computation, add all net cash inflows from the project, treating capital cost as the cash outflow, during its life period, divide the sum by the number of years the project will last, and express the result as a percentage of the project's capital cost. Mathematically, the formula for the AARR could be expressed as follows:

$$\text{AARR} = \left[\frac{\sum_{t=1}^{T} R_t - C}{T}\right]\left[\frac{100}{C}\right] \tag{11.9}$$

Where, the new notation *T* stands for the project life.

The AARR from project *A* of Table 11.1, for example, is given by

$$\text{AARR}_A = \frac{(2500 + 125 + 125) - 2500}{3} \times \frac{100}{2500}$$

$$= 3.33\%$$

Similarly, the AARR from other projects could be computed. The result would be as given in Table 8.3.

**Table 11.3** *Average Annual Rate of Return (Per cent)*

| Project: | *A* | *B* | *C* | *D* |
|---|---|---|---|---|
| AARR | 3.3 | 16.7 | 26.7 | 20.0 |
| Rank | 4 | 3 | 1 | 2 |

On the AARR criterion, a project will be acceptable if and only if it yields an AARR which is greater than or equal to the firm's desired rate of return. The desired rate of return, also called the **cut-off rate** or the **hurdle rate of return**, is stipulated by the management and is, of course, not less than the firm's cost of capital. Thus, if the hurdle rate were 20 per cent, project *C* would be accepted, investor would be indifferent with regard to accepting or rejecting project *D*, and projects *B* and *A* would be rejected. Obviously, the higher the AARR, the better is the project. Thus, under this technique, project *C* would top the list and project *A* would be the worst of all.

## Net Present Value

The net present value (NPV) of an investment is defined as the difference between the discounted value of all net cash flows (called the present value) and the capital cost of the project. Symbolically, it can be stated as

$$\text{NPV} = \sum_{t=1}^{T} \frac{R_t}{(1+i)^t} - C \tag{11.10}$$

Where the new notation $i$ stands for the appropriate discount rate. The NPV of project A at $i$ = 10 per cent would thus be given by

$$\text{NPV}_A = \frac{2500}{1 + 0.10} + \frac{125}{(1 + 0.10)^2} + \frac{125}{(1 + 0.10)^3} - 2500$$

$$= (2273 + 103 + 94) - 2500$$

$$= 2470 - 2500$$

$$= -30$$

This means that if project *A* is undertaken, the investor would have a negative NPV of Rs. 30 millions. The NPV from various projects of Table 11.1 are tabulated below.

**Table 11.4** *Net Present Value (i = 10%): (Rs. millions)*

| Project: | *A* | *B* | *C* | *D* |
|---|---|---|---|---|
| NPV | –30 | 608 | 1035 | 951 |
| Rank | 4 | 3 | 1 | 2 |

On the NPV criterion, a project is worth undertaking if its NPV is positive. If NPV is negative, the project *A* must be rejected. When NPV = 0, the investor is indifferent between accepting and rejecting the project. Thus, in this measure, projects *B*, *C* and *D* are acceptable, and project *A* must be rejected. For ranking the alternative projects, the rule is that the higher the NPV the better is the project. Hence, in this method, project *C* tops the list and project *A* goes to the bottom of the list. It would be interesting to note here that the NPV is inversely related to the discount rate. For example, at the discount rate of 10 per cent, the NPV from project *A* stood at minus Rs. 30 millions but at the discount rate of 5 per cent, it would be millions of rupees

$$\frac{2500}{1.05} + \frac{125}{(0.05)^2} + \frac{125}{(1.05)^3} - 2500$$

$$= 103$$

The discount rate equals or is greater than the cost of capital and thus the cost of capital is a determinant of NPV and thereby of the viability of the project.

## Benefit-Cost Ratio

The benefit-cost ratio (BCR) measure is essentially a variant of the NPV measure. Under this method, one takes the ratio of the discounted value of all the net cash flows from the project and the capital cost of the project instead of their difference as under the NPV measure. Thus, mathematically, it is defined as

$$\text{BCR} = \frac{\sum_{t=1}^{T} \frac{R_t}{(1+i)^t}}{C} = \frac{PV}{C} \tag{8.11}$$

The BCR from project *A* at $i = 10$ per cent could be computed as

$$\text{BCR}_A = \frac{\frac{2500}{1.1} + \frac{125}{(1.1)^2} + \frac{125}{(1.1)^3}}{2500}$$

$$= 2470/2500$$

$$= 0.988$$

Similarly, the BCR from other projects could be computed. The results are provided in Table 11.5

**Table 11.5** *Benefit-Cost Ratio (i = 10%)*

| Project: | *A* | *B* | *C* | *D* |
|---|---|---|---|---|
| BCR | 0.988 | 1.2431 | 1.414 | 1.3824 |
| Rank | 4 | 3 | 1 | 2 |

Under this criterion, a project would be acceptable if and only if its BCR is no less than unity. Thus, while projects *B*, *C* and *D* are worth undertaking, project *A* must be rejected. For ranking purpose, the higher the BCR, the better is the project. Thus, project *C* is the best and project *A* the worst among the four projects under analysis. As is the case of the NPV, the BCR depends negatively on the discount rate. It can be verified that the BCR from *A* goes up from 0.988 to 1.0412 as the discount rate changes from 10 per cent to 5 per cent. Thus, project *A* is not acceptable if the discount rate is 10 per cent, it is acceptable if that rate is 5 per cent.

## Internal Rate of Return

The internal rate of return (IRR) is also known by several other names. These include **yield, break-even rate, marginal efficiency of capital** (J. M. Keynes), **internal rate of profit** (K. L. Boulding) and the **discounted cash flow (DCF) rate of return**. It is defined as that rate of return which when used to discount the net cash inflows; the discounted value of all the net cash inflows just equals the capital cost. In other words, it is that rate at which NPV = 0. Mathematically it is given by

$$C = \sum_{t=1}^{T} \frac{R_t}{(1+r)^t} \tag{11.12}$$

Where, $r$ = IRR

Incidentally, note that the IRR formula (Eq. 11.12) is simply the NPV formula (Eq. 11.10) with $i$ replaced by $r$ and the NPV value forced to be equal to zero. Under the former, NPV is known (= 0) and $r$ is computed, while under the latter, $r$ is known (= $i$) and NPV is computed.

Using the IRR formula, the IRR from any project could be found if its capital cost and net cash inflows during the project life are known. Thus, the IRR from project $A$ would be given by

$$2500 = \frac{2500}{1+r} + \frac{125}{(1+r)^2} + \frac{125}{(1+r)^3}$$

There is one equation in one variable ($r$). However, the equation is a polynomial of degree $T$ (project life), which equals three in this case. Thus, it would have three different solutions, which can be obtained either through solving the polynomial algebraically or through the trial and error method. If $T$ is large, the recourse to the computer might be essential if the former method is applied. In the latter method, the analyst starts with some numerical value of $r$ (chosen randomly or on his guess about the true IRR) and see if the equation holds good for the value of $r$. If it does, that itself is a value of $r$. If it does not, he/she has to try with other values of $r$. The following illustrates how one should choose another value of $r$. If the analyst finds that at the first chosen value of $r$, the RHS of the above equation, which is nothing but the present value of the project, exceeds its LHS, which is the capital cost of the project, then he must choose another value of $r$, which is greater than the first chosen value of $r$. In contrast, if the RHS is less than LHS, he must choose a lower value of $r$. Though the procedure looks tedious, a careful analyst should be able to compute the approximate value of IRR in three-four iterations. If one has the convenience of computer and $T$ is large, one must solve the equation on computer only. Although mathematically the equation would have as many values for $r$ as the size of $T$, under normal cases, only one of them would be an acceptable solution. This is proved by the Descarte's Rule of signs, which states that the individual polynomial would have only one real root provided, of course, that the individual terms of the polynomial change signs only once. Equation (11.11) can be rewritten as

$$C - \frac{R_1}{1+r} - \frac{R_2}{(1+r)^2} - \ldots - \frac{R_T}{(1+r)^T} = 0$$

If all R's are positive, the sign is changing only once (from $C$ to $R_1$ term) and so the solution would be unique. Even if some $R$'s are negative (meaning initial net cash inflows are negative) but they are succeeded by all positive $R$'s, the sign would change only once and there would be a unique value of IRR. The problem of multiple roots would arise if NCI changes from positive to negative any time during the currency of the project. In actual practice, NCI is quite often negative in the beginning but once it turns positive it always remain so. Thus, the problem of **multiple root/solution** is only remote in case of investment projects.

Solution of the IRR equation for project *A* by either method would yield $r = 8.8\%$. The IRR calculation for various projects of Table 11.1 would be as shown in Table 11.6 below:

**Table 11.6** *Internal Rate of Return (Per cent)*

| Project: | *A* | *B* | *C* | *D* |
|---|---|---|---|---|
| IRR | 8.8 | 23.4 | 26.5 | 37.6 |
| Rank | 4 | 3 | 2 | 1 |

In this method, a project is acceptable if its IRR exceeds the **hurdle (desired) rate of return**, is to be rejected if the inequality is the other way round, and the investor is indifferent if the two rates are equal. Thus, at the hurdle rate of 20 per cent, while projects *B*, *C* and *D* are worth undertaking, project *A* must be rejected. For ranking the mutually exclusive projects, the higher the IRR, the better is the project. Thus, project *D* is the best, *C* the second best, and so on. If the investor had only Rs. 2500 millions to invest, he would go for project *D* only.

## Net Terminal Value

The Net Terminal Value (NTV) of an investment is the difference between the compounded value of all net cash inflows and the compounded value of the capital cost of the project. Symbolically, it is defined as

$$\text{NTV} = C\sum_{t-1}^{T} R_t (1 + r^*)^{T-t} - (1 + r^*)^T \tag{11.13}$$

Where, $r^*$ = re-investment rate
$i^*$ = cost of capital

If the NCI are negative, $r^*$ would stand for the cost of capital, incurred to finance the negative net cash inflows. To illustrate the computation, NTV from project *A* at $r^* = 20$ per cent and $i^* = 10$ per cent,

$$\begin{aligned}\text{NTV}_A &= 2500\,(1 + .20)^2 + 125\,(1 + .20) + 125\,(1 + .20)^0 - 2500\,(1 + .10)^3 \\ &= 3{,}875 - 3328 \\ &= 547\end{aligned}$$

Similarly, the NTV from other projects were computed, the results are reported in Table 11.7.

It should be noted here that the NTV is a positive function of the re-investment rate and a negative function of the cost of capital. Also, note that formula (11.13) assumes uniform re-investment rate throughout the project life. However, this is not a restrictive assumption. If the re-investment rate changes over time, $r^*$ must be replaced by $r_t^*$ in the above formula.

**Table 11.7** *Net Terminal Value ($r^* = 20\%$, $i^* = 10\%$) (Rs. millions)*

| *Project:* | *A* | *B* | *C* | *D* |
|---|---|---|---|---|
| NTV | 547 | 1222 | 1592 | 1922 |
| Rank | 4 | 3 | 2 | 1 |

According to this criterion, a project is worth undertaking if and only if its NTV is non-negative. Since the NTV from all four investments of Table 11.1 are positive, all of them are acceptable. In a situation of capital constraint, projects would require ranking. Obviously, the higher the NTV, the better is the project. Thus, if the investor has only Rs. 5,000 millions, he would undertake projects *D* and *C*

only. While projects *B* and *A* are also economically viable, they cannot be accepted due to the capital constraint and their inferiority to projects *C* and *D*. Incidentally, note that if projects have different lives, the NTV must be computed as on the terminal year of the longest living project. This concludes our discussion of all techniques for investment decisions. In what follows, we shall present their comparative merits and limitations, and throw some light on the choice of an investment technique for use.

## Compatibility of Various Techniques

The various measures of investment worth discussed above, while usually yield consistent investment decisions, are capable of producing inconsistent accept-reject as well as ranking decisions. To highlight this, the results of the various techniques for four hypothetical investments of Table 11.1 are summarized in Table 11.8.

**Table 11.8** *Measures of Investment Worth (Rs. millions)*

| *Project* | *Pay-Back period* | | *AARR* | | *NPV* | | *BCR* | | *IRR* | | *NTV* | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Year* | *Rank (Undis-counted)* | *%* | *Rank* | *Rs. (i = 10%)* | *Rank* | *Ratio (i = 10%)* | *Rank* | *%* | *Rank* | *Rs. (r * = 20% i * = 10%)* | *Rank* |
| *A* | 1 | 1 | 3.3 | 4 | –30 | 4 | 0.99 | 4 | 8.8 | 4 | 547 | 4 |
| *B* | 2 | 3 | 16.7 | 3 | 608 | 3 | 1.24 | 3 | 23.4 | 3 | 1222 | 3 |
| *C* | $2\frac{1}{3}$ | 4 | 26.7 | 1 | 1035 | 1 | 1.41 | 1 | 26.5 | 2 | 1592 | 2 |
| *D* | 1 | 1 | 20.0 | 2 | 951 | 2 | 1.38 | 2 | 37.6 | 1 | 1922 | 1 |

In this example, there are consistencies as well as inconsistencies. The former are found in the rankings by three methods, viz. AARR, NPV and BCR, and in the rankings by two methods, viz. IRR and NTV, the latter are also witnessed in terms of rejecting project A by the NPV method as well as BCR method and accepting the same by the NTV method. Inconsistencies are also observed in terms of ranking of the four projects by these methods. The ranking by the pay-back period differs from that by the AARR, NPV and BCR methods, and both of them differ from the ranking by the IRR and NTV methods. Such consistencies and inconsistencies are not uni-directional, and this aggravates the decision problem. This necessitates a discussion on the usual consistencies and inconsistencies, and the specific strengths and limitations of various techniques.

The first two methods, together known as the **traditional methods**, are each independent and have no definite relationship with other techniques, which together are known as the **discounted cash flow (DCF) techniques**. The NPV, BCR and IRR techniques necessarily give consistent accept-reject decision but they are capable of producing both consistent as well as inconsistent rankings. This is easy to see. From the three formulae of these methods, it is clear that if NPV > O, BCR is necessarily greater than unity and IRR is necessarily greater than the discount rate, which is tantamount to the hurdle rate. Quite the opposite is true when NPV < O. In the situation when NPV = O, BCR = 1, and IRR = discount rate ($r = i$). These inferences together with the accept-reject decision rules of the three techniques prove that these methods always give identical accept-reject decisions.

The NPV and BCR methods could give consistent as well as inconsistent ranking decisions. Under the situation where all alternative investments have identical capital costs, the two methods necessarily give identical ranking. The four hypothetical investments under analysis have identical investment costs

and accordingly the two methods yield identical ranking of those investments. Even if the capital costs of projects differ, the two methods could yield consistent ranking. This is demonstrated through the following example:

**Table 11.9** *NPV vs. BCR Methods (Discount rate = 10%)*

| *Project* | *Capital Cost (Rs.)* | *Present Value (Rs.)* | *NPV (Rs.)* | *Rank* | *BCR* | |
|---|---|---|---|---|---|---|
| | | | | | *Ratio* | *Rank* |
| *E* | 500 | 1500 | 1000 | 2 | 3 | 1 |
| *F* | 1000 | 1800 | 800 | 4 | 1.8 | 4 |
| *G* | 1000 | 2000 | 1000 | 2 | 2.0 | 3 |
| *H* | 1000 | 2200 | 1200 | 1 | 2.2 | 2 |

Both the NPV and BCR methods rank project *E* superior to that of project *F*. However, when project *E* is compared to project *G*, there is an inconsistency. According to the NPV measure, there is a tie between projects *E* and *G* but project *E* is superior to project *G* by the BCR technique. In fact, the two methods would yield incompatible results with regard to ranking of even projects *E* and *F* at some discount rate other than 10 per cent. The question then is why such inconsistencies arise and if they do which method is superior. In this context, one must note that while the NPV method is absolute in the sense that it ignores the capital requirements of projects, the BCR method is relative in the sense that it finds the worth of a project in relation to its investment cost. In the above example, both projects *E* and *G* give equivalent amount of NPV (Rs. 1000) but the former on an investment of Rs. 500 only while the latter on an investment of Rs. l000. Since the capital base is smaller in project *E* than project *F*, the BCR measure ranks project *E* better than project *F*. In view of this, it is recommended that the BCR technique must be preferred to the NPV method when projects involve different investment costs. When the projects' costs are identical, the two methods converge. There is a caveat in this context. If the capital market is perfect or there is no capital/funds constraint (or no capital rationing), meaning any amount of funds could be raised at a given cost of capital, the NPV method is preferred to the BCR measure. To show this, compare projects *E* and *H*. Project *H* yields an NPV of Rs. 1200 in contrast to an NPV of Rs. 1000 of project *E*. Thus, *H* is preferred to *E*. But by the BCR method, quite the opposite is true, for BCR of project *E* (= 3) is higher than that of project 4 (= 2.2). Although project *H* requires a larger fund (= Rs. 1000) than does project *E* (= Rs. 500), *H* is still better than *E*, for it yields a higher NPV and the cost of the fund is already taken into account in the computation of NPV.

The NPV and IRR methods are capable of producing both consistent as well as inconsistent ranking. This is clear from the results in Table 11.8 itself. While both the techniques rank projects *A* and *B* consistently, they rank projects *C* and *D* differently. It is argued that the two methods could give incompatible ranking under any of the following three situations:

(a) When the capital costs of various projects differ significantly;

(b) When projects have different lives; and

(c) When cash flow pattern of various projects is different, that is, when net cash inflows of some projects' increase over time while those of other projects either stay constant or decrease over time.

The basic reason behind inconsistency in ranking lies in the implicit assumptions of the two techniques with regard to the return on re-investment (and/or cost of borrowings) of intermediate cash

flows. While the NPV method assumes that the intermediate net cash inflows (outflows) are re-invested (financed) so as to yield (cost) a rate equal to the discount rate, the IRR method assumes that the re-investment (finance) rate equals the internal rate of return. These assumptions are comparative only when the discount rate equals the internal rate of return. Projects *C* and *D* are ranked differently by the two methods, for while project *C* gives low intermediate net cash inflows, project *D* generates high intermediate net cash inflows (vide Table 11.1) and in case of both the projects, the IRR is significantly different from the discount rate (vide Table 11.8).

The problem, thus, is which of these two methods should be relied upon in case of inconsistencies. There is no definite answer to this question. However, as a guideline one could say that the choice depends on the reasonableness of the implicit assumptions of the two measures. If the intermediate net cash inflows are used to repay long-term debt, to payout dividends and/or to add to reserves, the return on them would tantamount to the cost of capital, which is the major determinant of the discount rate, and so the NPV method should be preferred to the IRR method. In contrast, if the intermediate net cash inflows are used to plough back into the project, returns on them would approximate the IRR from the project, and hence the IRR method would be the appropriate one. In actual practice, intermediate flows are used for all such purposes and hence it is not possible to argue for or against anyone of these two techniques. Theoretically, both the methods have limitations and therefore we have the net terminal value method.

The NTV method explicitly takes into account the re-investment rate and thus is superior to other methods. This method could even argue to select a project which is rejected on the basis of any other method, and vice versa. This is obvious from the results in Table 11.8. Project *A* is rejected by each of the NPV, BCR and IRR methods, but is accepted by the NTV method. This is because, for this project while the IRR (8.8%) and discount rate (10%) are about the same, they differ significantly with the assumed re-investment rate (20%). The NTV method could, of course, yield both consistent as well as inconsistent ranking in relation to other investment criteria. This is evident from the findings of Table 11.8 again, where projects *A* and *B* are ranked consistently by all the four discounted cash flow techniques, but projects *C* and *D* are ranked differently by NPV (and BCR) and NTV methods. It is only accidental that the ranking by the IRR and NTV methods are in harmony for all the four projects under analysis.

## Evaluation of Various Techniques and the Choice

No decision problem has a unique technique for analysis. Investment decisions are no exception to this fact of life. We have seen in the foregoing section that techniques may not yield consistent decisions. The analyst and the investor must understand the salient features of each technique so that he/she can choose the best for his/her use.

The major strength of the pay-back period method lies in its simplicity and its emphasis on the quick recovery of the capital cost. The method emphasizes liquidity and risk-minimization at the cost of profitability. It may be, thus, be an appropriate method for conservative investors, who tend to minimize risk and care little for profit-maximization. Further, under two assumptions, this method converges to the IRR method. They are (a) uniform net cash inflows (i.e. $R_1 = R_2 \ldots = R_7$) from the project and (b) infinite project life. To verify this, note that under uniform net cash inflows for infinite project life, the IRR formula reduces to

$$C = \sum_{t=1}^{T} \frac{R}{(1+r)^t}$$

$$= R\left[\frac{1}{1+r} + \frac{1}{(1+r)^2} + \ldots + \frac{1}{(1+r)^T}\right]$$

The bracketed term is in geometrical series and its sum could be found as follows:

Let $$S = \frac{1}{1+r} + \frac{1}{(1+r)^2} + \ldots + \frac{1}{(1+r)^T} \quad \text{(a)}$$

Then, $$S\left(\frac{1}{1+r}\right) = \frac{1}{(1+r)^2} + \ldots + \frac{1}{(1+r)^T} + \frac{1}{(1+r)^{T+1}} \quad \text{(b)}$$

Subtracting (b) from (a), we get,

$$S\left(1 - \frac{1}{1+r}\right) = \frac{1}{1+r} - \frac{1}{(1+r)^{T+1}}$$

or, $$S = \frac{1+r}{1}\left[\frac{1}{1+r} - \frac{1}{(1+r)^{T+1}}\right]$$

or, $$S = \frac{1}{r}\left[1 - \frac{1}{(1+r)^T}\right]$$

Substituting this value into the IRR equation, we get

$$C = R\left(\frac{1}{r}\right)\left[1 - \frac{1}{(1+r)^T}\right]$$

or, $$r = \frac{R}{C}\left[1 - \frac{1}{(1+r)^T}\right] \quad (11.14)$$

From Eq. (11.14) it is clear that if $T = \infty$, $r = \frac{R}{C}$, which equals the reciprocal of the pay-back period (vide Eq. (11.8 b)). Thus, it follows that for projects with uniform net cash inflows and infinite life, the IRR is simply the reciprocal of the pay-back period. The decision rules of these two methods (the higher the IRR, the better the project and the lower the pay back, the better the project) would thus suggest that the two methods would yield similar decisions.

For projects with uniform NCI but finite life, Eq. (11.14) provides some clues. The expression for IRR ($r$) consists of two parts; $\frac{R}{C}$ the reciprocal of pay-back, and a correction factor, which depends on $T$, the project life. The larger the $T$, the smaller the correction factor and better is the pay-back reciprocal as an estimate of the IRR. For long-lived projects, reciprocal of the pay-back provides a good approximation of IRR. For short-lived projects, IRR is less than the reciprocal of pay-back period. Thus, if the pay-back period is 5 years, IRR would be less than, or at the most equal to, 20 per cent, depending upon the life of the project.

The major limitations of the pay-back period (undiscounted) measure are two, viz.

(a) It ignores the time value of money. Thus, while computing the pay-back period from four projects (vide Table 11.2), the NCI arising in different years have been taken at their face value and added together. This limitation is overcome by the discounted pay-back period method.

(b) It ignores the cash flows accruing after the pay-back period. Thus, projects *A* and *D* are ranked equally by the payback period, even though the latter is clearly superior to the former, for both the projects' capital costs are equal, promise equal NCI in the first period, and project *D* assures higher levels of NCI than project *C* in each of the future periods thereafter.

The average annual rate of return (accounting rate of return) is fairly simple to compute and thus it offers yet another easy method for investment decisions. Further, it considers all cash flows from the project. However, this method also ignores the time value of money altogether. Besides, it is biased in favour of short-lived investments. This is obvious from its formula (Eq. 11.9) which has project life (*T*) in the denominator. To elaborate this limitation, consider project *B* once again. The AARR from it stands at 16.7 per cent. Now suppose, NCI from it changes slightly, where in the third period NCI of 1250 is split into 1249 in the third period and 1 in the fourth period, then its AARR would change from 16.7 to 12.5%,

$$\frac{(1250 + 1250 + 1249 + 1) - (2500)}{4} \times \frac{100}{2500} = 12.5\%$$

Thus, the consequence of one unit of NCI shifting from the third, to the fourth year is as large as reducing the AARR from 16.7 to 12.5 per cent. Due to this limitation, the method is considered as **deceptive** and **dangerous**.

All the remaining measures, jointly termed as the DCF techniques, are free from the limitations of both the pay-back period and AARR methods. This speaks for their credence and popularity. The choice among them is difficult and somewhat unnecessary. In general, the BCR technique is preferable to the NPV method when the analyst wishes to rank mutually exclusive projects with different capital costs, or faces capital constraint. Under both these techniques, the analyst needs the discount rate, which significantly affects both the accept-reject as well as the ranking decisions. A correct estimate of the discount rate is thus crucial for using these techniques effectively. A discussion on the discount rate is included in the following section.

The computation of IRR does not require the discount rate. However, the IRR method for investment decisions needs the hurdle rate, which is similar, to the discount rate in the NPV and BCR methods. Thus, the data requirements of these three methods are similar. The basic difference between these two groups of methods (NPV and BCR, and IRR), as pointed out in the previous section, is that while the former assumes the reinvestment rate equal to the discount rate, the latter takes that as equal to the IRR.

The NTV method is free from any unreasonable assumption. It explicitly incorporates the re-investment rate and thus is devoid of the limitation of the NPV, BCR and IRR measures. It is thus, theoretically, the best investment criterion. Nevertheless, we must add a rider here. The NTV method requires information on the re-investment rate, which is difficult to estimate at the time of project selection. Since the only significant difference among NPV, IRR and NTV methods is on their use of the re-investment rate which is not known at the time of investment decision, the choice among criteria has to be somewhat arbitrary. As a practical guide, the choice among these investment criteria is not necessary and if made, it can be dangerous. It is not significant for three reasons:

(a) Most of these techniques tell different things, all of which are of utmost significance to the investors. Thus, the pay-back period tells the number of years a project takes to recover the project's capital cost, the NPV and NTV tells the amount of profit in rupees the investor is likely to make if the investment is undertaken and the IRR the gross rate of return in percentages (gross of the interest cost on capital cost) that the project is expected to yield. Surely, all these are useful yardsticks to judge on the variability of an investment

(b) The application of the pay-back period, NPV and IRR methods requires the data on more or less the same variables (over the parameters on which the pay-back period method requires the data, NPV method needs data on the discount rate only and the IRR method on the hurdle rate only) and thus if one is using any one of these methods, he can easily use other methods as well; the computation part is not difficult and, if necessary, it can be handled through easily accessible computers these days.

(c) It is conceivable that a particular method might award the identical ranks to two or more projects. Under such an eventuality, the other methods can be used to distinguish between those projects.

In view of these three factors, the various investment criteria are **complementary rather than substitutes**. All one could say in conclusion is that the AARR and undiscounted pay-back period methods could be ignored, the discounted pay-back period, NPV (or BCR) and IRR must be used simultaneously and if the investor has a fairly good idea about the reinvestment opportunities, he could use the NTV method instead of the NPV method. Incidentally, there is a counterpart of NTV, like BCR (at the beginning year) is to NPV, and that could be named as BCR at the terminal year, and computed as the ratio of the terminal value of all net cash inflows to the compounded value of the capital cost as on project's termination. In terms of the formula, it will be given by the ratio of the first term to the second term of the formula (11.13) above.

## Guidelines for Using Investment Techniques

Techniques provide tools for analysis, which help in decision-making. However, they do not give decisions, which is the prerogative of human judgement. Thus, it must be noted that if a particular criterion argues for the selection of a particular project, the investor must go for it if and if it is satisfied on other grounds, which are not incorporated in that particular technique. The "other" grounds would include the comparative risk involved in alternative projects (which is discussed below), assessment of liquidity positions during the currency of the project, inflationary trends in and outside the economy, any qualitative information that could not be incorporated in measures of investment worth, etc. All that the various investment techniques do is that each one of them reduces all the quantifiable costs and benefits into one number, and by examining that number in relation to the decision rule, the investor is able to conclude on the viability of the project. In the absence of these techniques, the analyst would have difficulties on comprehending the huge data into a meaningful parameter to arrive at a decision.

The saying, "garbage in garbage out" is well known. It is also often said, "There are three lies, viz. lie, damn lie, and data". Thus, it is essential that the data fed in any measure of investment worth is as accurate as possible. To at least try to achieve this accuracy, the parameters of various investment measures must be understood unambiguously lest they misguide the investor. In this connection, five parameters merit attention. They are:

- Capital cost ($C$)
- Project life ($T$)
- Salvage value ($SV$)
- Discount or hurdle rate ($d$), and
- Net cash inflows ($R$'s)

Most investment projects are large and they take years to commission. The capital cost of such projects is accordingly spread over years. While estimating the capital cost of such projects, due

consideration to time value of money must be given. In general, analyst computes the cost as on the year the project is ready for commercial production. Further, for projects which are located in the backward districts, there is often some provision for capital subsidy from the Centre/State, which could be, say, 15 per cent of the total capital cost, subject to a maximum of Rs. 15 millions. The capital cost of the project must be adjusted against this subsidy. Thus, if the capital cost of a project is say, Rs. 200 million, which is spread as Rs. 50 millions in the first year, Rs. 100 millions in the second year and Rs. 50 millions in the third year, when the project will be completed; the cost of capital is 15 per cent; and the capital subsidy of Rs. 1.5 millions becomes available in the third year; then the capital cost of the project as on project's completion stands at Rs. 210. 125 millions:

$$50\ (1 + .15)^2 + 100\ (1 + .15) + (50 - 15) = 210.125$$

Sometimes the financial institutions have commitment charges on the commitments they make to advance loan at a pre-specified interest rate and future date. If so, such charge must be included in the capital cost of the project with proper account of the time value of money.

There are four relevant lives in investment projects, viz. **asset life, activity life, economic life** and **planning horizon**. The asset life is concerned with the life of assets that are inevitable for running the project. A project usually has several assets, and if so, there are several assets' lives. The least of these becomes the binding constraint for the project life unless the project cost has included the replacement of that asset. If so, the life of the least surviving asset whose replacement is not included in the project cost becomes the asset life. The activity life is concerned with the number of years the activity (product) of a project would be in demand in the market. Thus, if one is thinking of a radio manufacturing plant and the plant has assets which could be operative even for 20 years but the radios would be in demand, say, for only 10 more years, then the activity life is 10 years and that would put a constraint on the continuation of the project. The economic life of a project is defined as the number of years until which the project would be worth operating on economic consideration. The last, viz. planning horizon, is the period within which the investor would care to reap the benefits. The least of all these lives is the project life. Incidentally, note that most investment projects last fairly long and, in fact, many continues for ever with or without, of course, additional investments in future. Under the circumstances, the assumption of a project life of about 10 years may sound inappropriate. However, it may not be really so, for two reasons, viz. (a) the more distant is the future, the more uncertain the cash flows are, in general and (b) the discount rate falls as the period of cash flows' accrual moves more into the future. By these two reasons, the future cash flows get less and less valuable as they get more and more far off, and thus ignoring those become less and less distorting the measures of investment worth. Accordingly, most investments are appraised (evaluated) on the assumption of a life of around 10 years.

The salvage value of a project, also called the scrap value, is the price all the assets of the project, including the goodwill, if any, will command in the year the project will be wound up. This need not equal to the written down value (WDV), also called the **book value**, of the project, for there could be capital gain or loss. It is obvious that the salvage value accrues on the expiry of the project life.

The discount rate and the hurdle rate, in the absence of changing prices, uncertainty and capital constraint, is the weighted average cost of capital, weight being the percentage of capital raised through a particular source. To illustrate this, let the capital cost ($C$) of Rs. 100 millions consists of Rs. 40 millions of equity and Rs. 60 millions of debt, and let the cost of equity (dividend + capital gain) be 20 per cent and that of debt 15 per cent. Incidentally note that, as per the corporate taxation rule, interest cost on borrowings is a deductible expense while dividend on equity is not a permissible deduction.

Thus, to compute the effective cost of capital, we need the corporate tax rate as well. Suppose that the tax rate is 30 per cent. The cost capital ($i$) then would be

$$i = \frac{40}{100}(20) + \frac{60}{100}(15)(1 - 0.30)$$

$$= 14.3\ \%$$

Thus, the discount rate to be used in the NPV formula would be 14.3%. The hurdle rate for IRR method as well as the cost of capital for NTV method would also be 14.3%. This rate would need to be adjusted upward for inflation, uncertainty and capital constraint, if any. A discussion on the methods of these adjustments would be presented later in this chapter.

The net cash inflows of a project accrue all through the project life and yet the concept is used erroneously by some analysts. This is partly because it is a difficult concept and partly because it is rarely defined unambiguously. Its definition depends on whether the project is evaluated on the basis of the *total capital cost* of the project or on the basis of the *equity capital* only.

In case of understanding of the relevant net cash inflows, the confusion is merely with regard to the inclusion or exclusion of five items, viz. *depreciation* ($DE$), *interest cost on each of fixed capital* ($I_F$) and *working capital* ($I_w$) *borrowings, dividend* ($DI$), loans repayments ($L$) and taxes ($T$). If the project is evaluated on the basis of total (entire) capital, then net cash inflow (NCI) in a particular year is defined as the profit ($P$) before (market as +) depreciation, interest cost on the borrowings for fixed capital, dividend on the equity capital (both ordinary and preference), and loan repayments ($L$); and after (marked as –) interest cost on borrowings for working capital and taxes ($P + DE + I_F + DI + L – I_w – T$); the terminal year NCI include, in addition, the salvage value of the project (SV). Also note that this concept has nothing to do with the repayment of loan (borrowing). To appreciate the reasons for this, we offer the rationale as follows. The NCI are before (gross of) depreciation, not because depreciation is ignored but because it is incorporated through the recovery of the entire capital cost and the inclusion of salvage value in the terminal value of NCI. For example, in the NPV formula (vide Eq. 11.10), $C$ is deducted and the last NCI (i.e. $R_T$) is inclusive of the salvage value ($SV$), and the cumulative depreciation is equivalent to ($C – SV$). Similarly, the interest and dividend cost of the entire fixed capital are incorporated through the discount rate (= i) in the NPV formula and the hurdle rate in the IRR rule of investment decision. Since interest cost on working capital borrowings and taxes on profits are costs to the investor and they are not incorporated through the investment measures, the NCI are net, of these obligations. Obviously, NCI are net of all other variable costs, such as, wages and salaries, raw-materials' cost, promotional expenses, etc.

If the project is, however, evaluated on the basis of the equity capital alone, then not only the NCI but also the fixed capital and the discount rate undergo a change. In the various measures of investment worth then, equity capital ($E$) replaces the fixed capital ($C$), and the cost of equity capital becomes the discount rate. Further, the relevant NCI then becomes profit before (marked as +) depreciation and dividend; but after (marked as –) interest cost on both fixed and working capital borrowings, repayments of loans ($L$) and taxes ($= P + DE + DI – I_F – I_w – L – T$); and the terminal year NCI include, in addition, the salvage value of the project ($SV$).

If one were evaluating the viability of the entire project, the project must be evaluated on the basis of the total fixed capital (equity + debt) only. Thus, the financial institutions which advance loans to such projects follow this procedure. However, some investors might bother about their own (equity) capital only and accordingly they could examine the worth of an investment on the basis of the equity capital only.

It is instructive to note that the depreciation rule which must be adopted to generate depreciation data should be the one permissible by the corporate tax rules. Currently, the declining balance (called WDV) method is the approved method for this purpose. Under this method, depreciation is charged at a specified rate on the depreciated (written down) value. Thus, if an asset costs Rs. l00, and the permissible depreciation rate is 10 per cent, then depreciation in the first year would be Rs. l0 (100 × .10), in the second period it would be Rs. 9 [(100 – 10) × .10] in the third period Rs. 8.1 [(100 – 10 – 9) × .10], and so on. The other depreciation methods, viz. straight line, sum-of-digit number, double declining method, etc. should not be used in the generation of data on depreciation and thereby on net cash inflows for computing various measure of investment worth.

Another point to note here is with regard to the tax shield (advantage). There are three possible sources of tax saving for the investor. These are

(a) Tax shield accrues through depreciation of capital assets. Since most capital assets are subject to depreciation, which is a permissible cost of production, tax is saved through it as tax is based on profit.

(b) Tax is saved through payments of interest on borrowed capital. This is again due to profit tax laws, which base taxes on profits after interest cost. However, dividend is not permissible deduction for tax purposes. Due to these laws, other things remaining the same, debt financing is cheaper than equity financing.

(c) Tax shield occurs if the project results in capital loss, i.e., if the salvage value falls short of written down value. In the situation of capital gain, there is a tax liability instead of tax shield.

The last thing an analyst should take note of at the stage of evaluating its viability is concerned with the kind of investment analysis one must attempt. In this connection, we talk of **financial analysis**, also called as the **private cost-benefit analysis** and **economic analysis**, also known as the **social-cost-benefit analysis**. The former examines the viability of an investment from the point of view of the private or individual investor while the latter attempts the analysis from the view point of the society as a whole. The two analyses are not the same thing and the difference between them would be examined later in this chapter.

## 11.4 INVESTMENT ANALYSIS

In this section, we shall first discuss the various steps in capital budgeting and then the analysis of investment projects from the view point of an individual investor vis-a-vis the society as a whole.

### Steps in Capital Budgeting

A proper investment analysis must proceed through the following consecutive steps:

(a) Project Identification

(b) Project Formulation

(c) Project Appraisal and Selection

(d) Project Implementation and Monitoring

(e) Project Evaluation or Post-completion Audit

Projects are identified through a search of investment opportunities. The prospective investor could carry on this step either through collecting this information from development organizations, who are engaged in developing full project profiles, through borrowing ideas from well to do investors in the country and abroad, or/and through inventing new project ideas. In this context, it should be noted that there is no dearth of potential projects, particularly in relation to the quantum of investment one desires to undertake. Further, it costs both time and money to analyse an investment. Thus, one must try to generate a limited number of project ideas. However, one has to be careful in limiting such ideas, for if a good project is not included in the list of projects identified for scrutiny, that project can never be selected. In view of this, this step is quite crucial and so it is often entrusted to/guided by the professionals/consultants.

Once a list of projects is ready, each project in it must have a blue print, providing details of the requirements of various assets such as land, buildings plant and machinery, raw materials, labour, etc. and their price tags, together with the expected capacity utilization over time and the products' prices, among other things. After the alternative investment opportunities are well formulated, each must be examined in terms of the feasibility of their implementation. Here one would examine the technical feasibility in terms of the availability of land, plant and machinery, raw-materials, technical know-how, etc. financial feasibility in terms of the availability of finance in required times; economic feasibility, in terms of the employment generation and development of backward areas and communities; and the management feasibility, in terms of the availability of the managerial personnel for the smooth implementation and running of the project. At this stage some of the projects identified in the first stage may be dropped if they do not meet the test of feasibility.

The feasible projects are then appraised in terms of their economic viability. This step is carried out through first projecting cash flows from each project during the project life and then comparing them through the use of measures of investment worth discussed in the previous section. The projects, which lead the list of the viable projects and which are within the capital constraint of the investor, if any, are then selected.

The selected project(s) is (are) then implemented in terms of arranging of finance, purchasing of land, plant and machinery, etc., construction of buildings, hiring of labour and other staff, etc. While implementing the project and after it is commissioned for commercial production, the investor must monitor the project on a regular basis. Under this activity, the investor would see to it that the project is commissioned as soon as possible, particularly within the stipulated time unless unwarranted by unforeseen and unavoidable circumstances. During the running period of the project, there is ample need for close monitoring. This is with regard to the quality and quantity of production, development of the market for the product, repayments of loans, recoveries of dues from the clients, payments of reasonable dividends, maintenance of good industrial relations and customers' goodwill, etc. In this regard, **project over-runs** in terms of either or both of capital cost and time are of utmost important. Needless to say, such over-runs often lead a viable project into troubles.

The last step in capital budgeting, called project evaluation is concerned with the post completion audit of the project. In this step, the investor evaluates the validity of his prior decision. This he does by re-computing the measures of investment worth, this time on the basis of actual cash flows rather than expected cash flows of the project appraisal stage. He then examines the actual worth of his chosen project vis-a-vis its expected worth and, to the extent possible, in relation to the projects he had rejected. This he does not because he can undo the project but simply to know what errors, if

any, he committed in that investment decision. This fact finding exercise helps him to learn, so that he does not commit such blunders in future. This step is, thus, like the post-mortem on a human body, which doctors often like to perform to understand the cause of the death and thereby to explore new lessons. Needless to say, over-runs in cost and/or time in projects, rendering the project non-viable or less profitable than expected could be due to the investors' inadequate appraisal and/or unexpected events. The former could be due to his/her incompetence and/or careless, and the latter could have been caused by unforeseen changes in inputs prices or efficiencies, foreign exchange rate, inflation, recession, industrial disturbances, weather, etc.

It would now be obvious that a proper investment analysis requires the knowledge of various disciplines; technical (engineering), economics, finance, business forecasting, etc. The text is basically an economics one and thus we could deal with that part of analysis which can be handled through the tools of economics only. However, since there is a good deal of overlap in economics and finance, some readers would see the use of finance concepts and techniques as well. We have accordingly concentrated on project selection stage, though other steps have also been examined briefly.

## Private and Social Benefit-Cost Analyses

The private and social benefit-cost (B-C) analyses are also known as financial and economic analyses, respectively. Under the former, the economic viability of a project is evaluated from the point of view of the individual investor while under the latter the same is examined from the point of view of the society as a whole. The two analyses differ on three important counts:

(a) Objective
(b) Inclusion/exclusion
(c) Valuation

While an individual investor aims at the maximum possible return on investment (ROI), the society normally pursues multiple goals. These include ROI, economic growth, employment generation, reduction in economic inequalities over regions and peoples from different categories, self-reliance, etc. Thus, it is possible that a project might be viable from the individual point of view but not from the society's point of view, and vice versa.

The second difference between financial and economic analyses is with regard to the inclusion and exclusion of benefit and costs. Investment projects involve some **transfer payments** in the form of taxes and subsidies, which are costs and benefits as far as the individual investor is concerned but not from the point of view of the society. Similarly, there are externalities (vide Chapter 10) associated with the projects, which are irrelevant for the private B-C analysis but are relevant for the social B-C analysis. These include benefits, such as recreation and picnic spots availability if the project happens to be of building of a dam, developing of a swimming pool or rose garden, etc., and costs, such as air pollution, traffic congestion, and accident hazards if the project leads to setting up of units in the hearts of a town, producing inflammable items (e.g. Union Carbide in Bhopal). While these benefits and costs are difficult to quantify, they must be given due consideration in the social B-C analysis.

The last significant difference between the private and social B-C analyses is with regard to the valuation of various benefits and costs associated with the project. The valuation must be performed on the basis of the price received and paid by the investor concerned. Thus, the individual investor receives and pays the market price for its outputs and inputs, respectively, and accordingly, the valuation is

carried out on the basis of market prices for the private B-C analysis. However, this is not the case for the society. For the society, benefits and costs must be based on, what is called as, the shadow prices, which measure valuation for the society, of the corresponding items produced/used by the project. In particular, the shadow price of unskilled labour could be very different than the corresponding market wage rate; and if the foreign exchange is scarce, the market rate of exchange could be a poor proxy for the true cost of foreign exchange to the society. There are two approaches for valuation for the social B-C analysis. The one is recommended by Little and Mirrlees (1974) and the other by the UNIDO Guidelines (1972).

Under the former approach, the benefits and costs are valued at the world price (exports f.o.b. and imports c.i.f) and under the latter they are valued at the shadow price (resource cost). The World Bank approach (1975) synthesizes the two approaches under which the tradable items are valued at the corresponding world prices and non-tradable items at the corresponding shadow prices. In both these approaches, foreign exchange earnings and uses are valued at the shadow exchange rate (SER) and not at the official exchange rate (OER), and the labour cost at the shadow wage rate (SWR) and not at the market wage rate (MWR). There are systematic methods to compute these shadow rates but that is beyond the scope of this book. Suffice it to say here that the Planning Commission would normally indicate both the shadow exchange rate and shadow wage rate from time to time. In a developing economy, in general, we observe the relationships like, SER > OER and SWR< MWR. However, with emphasis on globalization, the differences between these are narrowing over time.

One more factor which must be taken into account here is with regard to the discount rate. In the private B-C analysis, the weighted average cost of capital for the project would serve the purpose of the discount rate. However, in the case of the social B-C analysis, the relevant rate would be the cost to the society of the funds employed in the project. This perhaps would be a weighted average of the rate at which the country obtains funds from international market and the domestic cost of funds.

Thus, under the private B-C analysis, private benefits and costs are taken into account and they are valued at the respective market prices. These are then evaluated through the various investment techniques, using the weighted cost of capital as the discount/hurdle rate, to arrive at an investment decision. In contrast, under the social B-C analysis, benefits and cost to the society are considered, they are valued at the world or shadow prices and then evaluated on the basis of various measures of investment worth. Further, in the social B-C analysis, suitable adjustments are made in the benefits and costs to take care of non-profit objectives, such as, creation of employment and reduction of socio-economic disparities among regions and communities. Thus, for example, if a project is likely to generate employment of unskilled manpower or/and development of a backward region/community, or/and production of export or import substitute item its benefits are compounded to take care of these benefits. In contrast, if a project is likely to produce goods for conspicuous consumption, without producing any other desirable benefits, its benefits would be discounted to incorporate its evil effect on the society. To conclude, the difference in private and social B-C analyses lies in the estimates of benefits and cost, and the discount/hurdle rate. So far as the techniques of investment are concerned, they remain the same under the two analyses.

The methods of project appraisal discussed above implicitly assume no uncertainty and constant prices. Neither of these assumptions is true in the case of investment projects. In what follows, we shall therefore, discuss the significance of their presence and the methods of handling investment decisions in the world of uncertainty and inflation/deflation.

## Adjustments for Handling Risk and Uncertainty

Investments involve cost and benefits over a period of time. At the time investment decisions have be to be made, the costs and benefits associated with alternative investment opportunities are the expected costs and returns, and not the actual costs and returns, in different future periods. Since the future is never known with certainty, uncertainty is inherent in every investment project.* Recall that one of the reasons for money to have time value is uncertainty with regard to the honouring of the promise. Before going into it, it will be instructive to understand the kind and seriousness of uncertainty in investment projects.

Uncertainty in investments is with respect to variability of costs and returns, i.e., cash outflows and cash inflows during the life of the projects. Any factor which may influence either outflows or inflows is the cause or source of uncertainty. Thus, prices of plant and machinery, raw-materials and product, rates of wages and salaries, technology, competition, consumer preferences and market, industrial relations, fire or theft in the factory, government policy with respect to taxes, subsidies, price controls, foreign exchange rate, etc. are the risk factors in capital expenditures. Some of these factors like fire and theft, and government policy are beyond the control of investors (insurance can, of course, be purchased against fire and theft), while the others like various prices, industrial relations, consumer preferences, technology, are at least subject to some influence by investors.

Although nothing can be said for sure in the dynamic world in which we live but generally speaking uncertainty is more with respect to inflows than outflows. This is because the bulk of cash outflows take place in the near future (between investment decision and commissioning of the project) while cash inflows are spread more or less uniformly over the project life (post-completion period), and generally speaking, more distant is the future, more difficult it is to predict it. Further, like the return on investment varies from project to project, different projects are associated with different degrees of risk. Although it is difficult to make rules in this respect, generally speaking, the magnitude of risk increases as one moves from left to right in Chart 11.1 above, which delineates various kinds of investments. Thus, it is the least in replacement investments of the like for like and the most in strategic investment, involving expenditures on research and development. Since risk, without quid pro quo is undesirable, if a replacement investment promises to earn as high a return on investment (ROI) as an expansionary investment, the investor would prefer the former to the latter. However, if an expansionary investment is associated with a higher ROI than as a replacement investment, the choice becomes difficult to make and we need new tools to help us in such a situation.

Investment techniques under uncertainty are, in fact, investment techniques only, for all investments are characterised by some degree of uncertainty. Further, the investment techniques discussed above are the valid ones even in the presence of uncertainty. All that one needs to do is to either slightly modify the above measures of investment worth or to compute more than one value for each of those measures, one for each set of expected cash flows. Since the modification is the same in each technique, in what follows, the NPV method alone is used to illustrate the difference. As is usual with all decision techniques, there are several ways to incorporate risk and uncertainty in investment analysis. A discussion on important ones follows:

*Statisticians make a distinction between risk and uncertainty. By risk they mean a situation where all the possible events are known with their respective probabilities; while under uncertainty, either the exhaustive list of the possible events or their probability distribution or both are not known. Since investors are neither totally ignorant nor do they know the exact probabilities of such events, the terms risk and uncertainty are used interchangeably in this field. The text is no exception to this tradition.

**Informal Method** In this method, the NPV based on single-valued estimates of annual returns (using the firm's cost of capital as the discount rate) is calculated for each alternative project. The decision-maker would recognize that some projects are riskier than others. If the net present values on two mutually exclusive projects are reasonably close to each other, the less risky one is chosen. The extent by which the NPV of the riskier project must exceed that of the less risky project before the riskier project will be selected is not specified—the decision rules are strictly internal to the decision-maker. Thus, if the return on investment from a replacement investment and an expansion investment are the same, then the former is better than the latter.

**Pay-Back Period Method** The pay-back period method considers cash flows up to the pay-back period only. Thus, it assigns a probability of unity (i.e., assumes them as 100 per cent certain) to cash flows within the pay-back period and a probability of zero (i.e. assumes them as zero net returns) to those beyond the said period. Since the degree of uncertainty normally increases with the length of the period, the pay-back period method incorporates uncertainty in a crude way. In fact, this method would have no place in the measures of investment worth had there been no uncertainty in investments.

**Risk-Adjusted Discount Rate** For willingness to bear risk, economic theorists have for many years assumed that the businessmen require a premium over and above an alternative which was risk-free (they will accept uncertainty of the investment's NPV over its certain value (= capital investment) only if there is a positive reward on this trade-off). Accordingly, the more uncertain are the future costs/returns, the greater is the risk; and the greater would be the required premium. From this basic reasoning, various theorists have proposed that risk premium be incorporated into the capital budgeting calculations through the discount rate. The composite discount rate, then, is asked to do double role, viz. to allow for time preference and risk preference. Accordingly, it equals the sum of (a) the risk-free rate and (b) a rate reflecting the investor's attitude towards risk. For example, if the risk-free rate was 10%, and the compensation for bearing the risk of the investment was 5%, then the discount rate of 15% must be used. This is known as the Risk-Adjusted Discount Rate or the Market Rate of Discount. Riskier the investment, the higher is the risk adjusted discount rate. Obviously, the compensation for project risk would vary from project to project and also from investors to investors. Recall from the previous chapter that investors could be risk-neutral, risk-averse or risk seekers/lovers. While risk neutrals simply rely on expected returns/costs, risk-averse people charge extra for risk and risk-seekers attach a premium on risk. Most investors are risk averse entrepreneurs and accordingly, the risk adjusted discount rate is usually kept at a higher rate than the counterpart unadjusted discount rate. Accordingly, the method handles risk through increase in the denominator of the present value equation. It may be noted that the method under discussion raises two problems. One, it is difficult to decide by how much the discount rate should be jacked-up to take care of the risk in a particular investment. Two, the conservative investors may choose a rather high risk-adjusted discount rate and may reject the otherwise profitable ventures. In contrast, the liberals may choose a lower risk-adjusted discount rate and accordingly may undertake unprofitable projects. This explains the current scenario, where a given project is accepted by some and rejected by some others, and these "some" and "others" include not just the investors but also the banks and financial institutions.

**Conservative Forecasts** Another common method for dealing with the uncertain future is through reducing the forecasted net returns to some conservative levels. For example, if the decision-

maker, or his expert, states that his "best estimate" is a net return of Rs. 500 thousands next year, the decision-maker applies some intuitive correction factor and says he will work with Rs. 450 thousands "just to be safe". It is thus a version of the certainty equivalence method. In this method, the numerator of the present value equation is reduced and thereby the NPV.

The method suffers from the two limitations, similar to the ones discussed above under risk adjusted discount rate method. They are, (a) the procedure for reducing forecasts of net cash inflows is ambiguous and it is likely to vary from investment to investment, and (b) use of this method adversely affects investment prospects and thus tends to increase the chances of rejecting otherwise good investments.

**Sensitivity Analysis (Decision Tree)** In this method, the investor examines the sensitivity of the measures of investment worth with respect to all the possible values of all the critical variables which significantly affect the capital cost and/or the net cash inflows expected from an investment. Only the critical variables are considered, for otherwise computations become too much and the method loses its practicality. What is a critical variable? A critical variable in an investment project has two features, viz. (a) its exact value at a future date is not known, and (b) a small change in its value produces significant change in the worth of an investment. While most critical factors are general, some are project specific. In what follows, we shall explain the sensitivity method, together with the critical factors, through an example from sugar industry.

The critical factors for sugar industry include sugar price ($P_S$), sugarcane price ($P_C$), and the number of sugar crushing days in a given year ($D$). Sugar often has a dual market, where a part of the sugar has to be sold to the government at a levy price for distribution to the public through fair price shops and the other part is sold at a free market price. The free market price varies over time, and even levy price and levy amount are not constant over time. Obviously, the weighted average price of sugar is, thus, a variable over time. There is a minimum support price for sugarcane which is fixed year after year, and the levy quantity could vary over years as well. Accordingly, the composite or a weighted average price of sugarcane which the sugar manufacturers pay in future is uncertain. Since the supply of sugarcane is seasonal and its production depends on the monsoon, the availability of sugarcane to sugar factories is uncertain and hence the number of crushing days in a particular year is a critical factor.

Net cash flows from sugar mill obviously depend on the values of $P_S$, $P_C$, and $D$. Corresponding to each set of values of these variables (assuming all the other determinants of cash flows either take a given value or that their influence is insignificant), there will be one value of the measure of investment worth, be it pay-back period, NPV or IRR, etc. The hypothetical results in terms of NPV for each set of values of all the three critical factors are presented in Fig. (decision tree) 11.1. In the example, the capital cost of the project is assumed to be Rs. 500 millions, project life 15 years and the discount rate 12%. The critical variables are assumed to take more than one value. In particular, calculations are provided for $P_S$ = Rs. 5000, Rs. 4000 and Rs. 3000 per quintal; $P_C$ = Rs. 150 and Rs. 200 per quintal; and $D$ = 125 and 100. Thus, if the project is accepted, and $P_S$ = Rs. 5000/quintal, $P_C$ = Rs. 150/quintal, $D$ = 125 days, the NPV from the project will be Rs. 200 millions. For all other sets of values of these variables, the corresponding NPVs can be read from Fig. 11.1. Thus, depending on what values the critical factors take, the NPV could be as high as (Plus) Rs. 200 millions and as low as (minus) Rs. 100 millions; considering the values of the critical factors assumed in the figure as exhaustive. The figure indicates the sensitivity of the project worth (NPV) to the values of the critical factors.

For taking the decision as to whether to go for a sugar mill or not, the investor would have to compare zero NPV in relation to various NPVs ranging from plus Rs. 200 millions to minus Rs. 100 millions. All the 12 NPVs in Fig. 11.1 are possible outcomes of the project. Depending upon the investor's attitude towards risk and the ability to take risk (as measured by his income and wealth), the investor would decide as to whether to go in for a sugar factory or not.

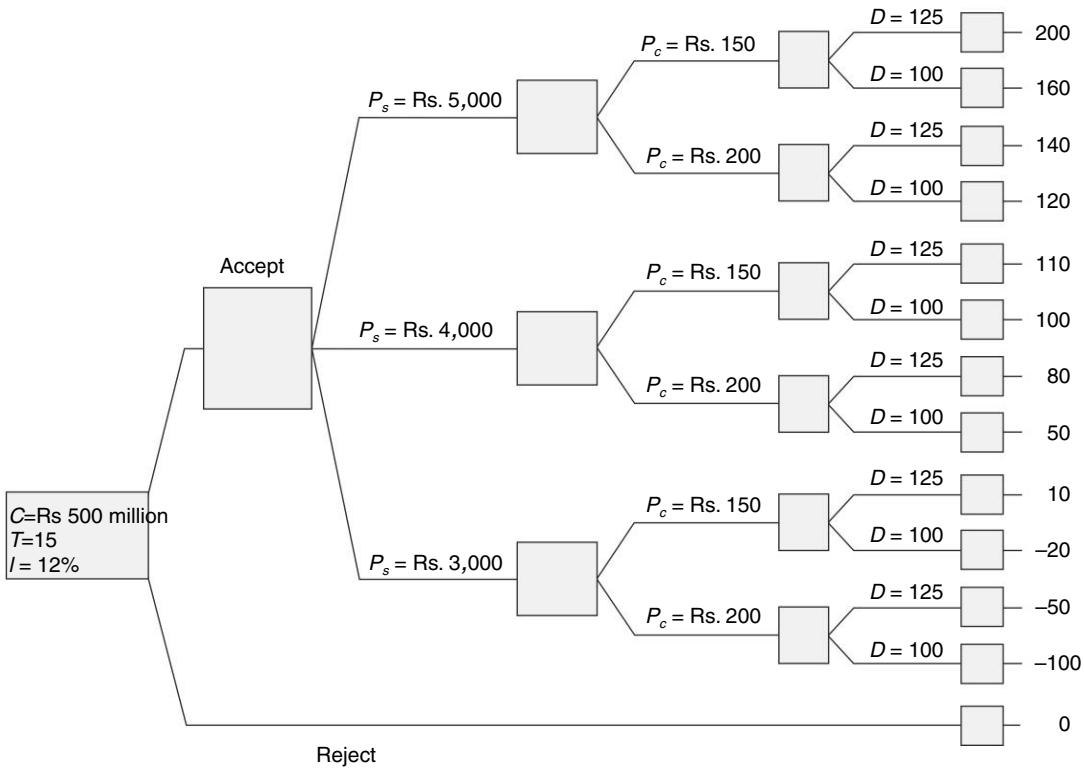



**Fig. 11.1** *Decision Tree for Sugar Factory*

The sensitivity analysis method is quite appropriate for investment decisions. This is because here the investor knows all possible outcomes of his decisions in advance and so he gets mentally and financially prepared to face both the best and the worst outcomes. Thus, it helps the investor in planning for future use of his surpluses and in meeting financial crisis, should it arise. Also, it avoids surprises and possible heart attacks!

This method is also known as the **decision tree method**. For, as would be seen in Figure 11.1, under this method a decision is literally mapped out in the form of the branches of a tree, and the various paths that the decision-maker and project can take, and the alternative outcomes that the project could generate are diagrammed. Along the various paths of the tree are inserted the possible present decisions that the firm may take and the chance events that the firm may face. On the top of each branch of the tree are the fruits (NPVs), some of which are very sweet (high positive NPVs) and some are very sour (high negative NPVs).

The sensitivity approach surely suffers from some limitations. For example, it ignores the chance (probability) associated with the different returns, it does not generally consider all the key variables,

and the critical variables under it are generally assumed to change on ad-hoc basis, e.g. 5 per cent, 10 per cent, etc. For taking care of some of these limitations, the **simulation method** is recommended, under which all possible permutations and combinations of all the key variables are considered and the results are obtained through computer, for manually it becomes impossible to handle.

**Expected Monetary Value Method** In the decision tree method, we compute cash flows and then NPV, etc. separately for each alternative set of the expected values of the critical variables. On the basis of the alternative outcomes, the investor takes the decision. Under the expected monetary value method, the investor assigns subjective probabilities of occurrence to each alternative outcome (NPV) and then combines them by calculating the expected value of the final outcome. For example, if the probabilities of the various possible NPVs in the above Chart were.

| *NPV* | *Probability* |
|---|---|
| 200 | .05 |
| 160 | .10 |
| 140 | .15 |
| 120 | .20 |
| 110 | .15 |
| 100 | .10 |
| 80 | .08 |
| 50 | .06 |
| 10 | .05 |
| –20 | .03 |
| –50 | .02 |
| –100 | .01 |
| | 1.00 |

The expected monetary value of NPV (EMV) would be given by

EMV = 200 (.05) + 160 (.10) + 140 (.15) + 120 (.20) + 110 (.15) + 100 (.10) + 80 (.08) + 50 (.06) + 10 (.05) – 20 (.03) – 50 (.02) – 100 (.01) = 104.8 = 105

If the EMV is positive, the project is acceptable; otherwise not. Higher the EMV, the better is the project. Since our hypothetical project gives a positive expected monetary value (=Rs. 105 million), it is worth accepting. The method is, of course, subject to some limitations. For example, it is based on average and thus suffers from the limitations of ignoring extreme values. The common way to explain this limitation is to cite an event, viz. a 6 feet, non-swimmer person getting drowned while crossing a river, whose average water depth was 3 feet but going up to 10 feet. To take care of such a possibility, one needs to consider the EMV (= Rs. 105 millions) along with some measure of risk that it could vary between –100 and + 200 millions of rupees. This is what the next method talks about.

**The Ideal Method** The ideal way of handling uncertainty is to use a combination of the last two methods, viz. sensitivity analysis and EMV method. Under the former, there are perhaps too many values of NPV, which may confuse the less sophisticated management. Under the latter, only the average

risk is taken into account. If one uses both these techniques, EMV would give a measure of central tendency of return while the range of NPVs would provide a measure of risk. Alternatively, one could compute the standard deviation (s) or the coefficient of variation (CV) to measure the risk. For the sugar factory example, these two latter measures are

$$\sigma = \sqrt{\begin{aligned}&.05(200-105)^2 + .10(160-105)^2 + .15(140-105)^2 \\ &+.20(120-105)^2 + .15(110-105)^2 + .10(100-105)^2 + 08(80-105)^2 \\ &+.06(50-105)^2.05(10-105)^2 + .03(-20-105)^2 \\ &+.02(-50-105)^2 + .01(-100-105)^2\end{aligned}}$$

$$= 55.15$$

and, $$CV = \frac{\sigma}{EMV} = \frac{55.15}{105} = 0.53$$

Thus, the sugar industry project, whose capital cost is Rs. 500 million, is associated with an average return (EMV) of Rs. 105 millions and the risk of the return varying in the range of + Rs. 200 millions to Rs. l00 millions, or a standard deviation of Rs. 55.15 millions or the coefficient of variation of 0.53. Depending upon the willingness and ability of the investors to bear the risk, the investor could decide whether to accept or reject the project. For ranking various projects, return and risk on all alternative projects can be computed and if the trade-off between return and risk is defined, the investor can easily rank them.

This method has the advantage of using all available information in taking investment decisions, which are crucial, particularly for investment decisions which are practically irreversible and make or break the investor's future.

In conclusion, it may be stated that none of the methods discussed in the note for handling uncertainty is useless. Of course, the ideal method, as termed here, is the best. However, there are situations where decisions have to be made in haste and alternative investment opportunities being considered are many. Under such circumstances, the use of ideal method may be impractical. In that case, methods in the order in which they are discussed in this note may be applied and some projects can be rejected in favour of others on the basis of even less sophisticated methods. The thorough analysis as required for the ideal method may be attempted for a couple of the most promising projects only.

## Adjustments for Handling Inflation/Deflation

In capital budgeting calculations, it is customary to assume constant prices for all inputs and outputs. If this assumption is not valid, the decisions based on these calculations could very well be erroneous. This is because, incorporation of inflation/deflation in the calculations of the measures of investment worth could lead to fundamental changes in the accept-reject and ranking decisions. This is for several reasons:

(a) Depreciation charges are based on original rather than replacement costs and accordingly they remain invariant even in the presence of inflation/deflation.

(b) Interest rate may be fixed at the time loan is sanctioned and, if so, interest cost does not change in the face of inflation/deflation.

(c) The cost items other than depreciation and interest costs, and revenue items (sales of Outputs) are influenced by changing prices but usually at the varying rates. This is because, during inflation and deflation, all prices do not change in the same direction and definitely not in the same proportion.

(d) Taxes are based on nominal rather than real income. As income grows (falls) with inflation (deflation), an increasing (decreasing) portion is taxed, with (go ahead of) inflation. This is called the 'bracket creep' effect of inflation/deflation.

(e) The discount rate in the NPV method and the hurdle rate in the IRR method are influenced by inflation/deflation. This is because they serve to account for the time value of money, which is subject to inflation/deflation.

Thus, it can be concluded that since the effect of inflation/deflation on various cash flows and the discount/hurdle rate is non uniform, assumption of constant prices in the face of significant inflation/deflation could lead to misleading investment decisions. Further, in view of the factors (a) and (b) above, even if all prices are expected to move in the same direction and same proportion, the assumption of constant prices in computing investment measures is unwarranted. The question then is, how to incorporate changing prices in the various measures of investment worth? This is simple and the procedure is explained below.

Each item of cash outflows and inflows except depreciation and interest cost must be adjusted for the expected inflation/deflation rate in that item. Thus, the cost of plant and machinery, if it is to be incurred in a future period must be adjusted by the expected inflation/deflation rate in the price of plant and machinery, the cost of raw-materials by the expected inflation/deflation rate in the price of raw materials, the wage and salary bill by the expected hike/fall in the wage-salary rate, the cost of electricity by the expected change in electricity tariffs, the sales proceeds by the expected price change in the output of the project and so on. The tax liabilities, over project life must accordingly be computed on the basis of expected profit after due considerations of the expected price changes in various cost and return items. Similarly, the discount rate to be used now must consist of both the charge for the earning power of money (measured by the weighted average cost of capital) as well as for the hedge against inflation. The expected inflation/deflation rate in the general price index (wholesale price index or consumer price index) could be used as the proxy for the inflation rate for adjusting the discount rate for changing prices. Once the various adjustments in cash flows and the discount rate are carried out to take account of the changing prices, the various measures of investment worth could be computed on the basis of the adjusted data. The results would then serve the purpose of guiding the investor with regard to "which project to accept" or "what to produce". This is how inflation and deflation can be handled in investment analysis.

## 11.5 CONCLUDING REMARKS

The subject of investment analysis is both very significant and difficult. The techniques available for this purpose are quite useful and versatile, in the sense that they are capable of handling uncertainty, inflation/deflation as well as the capital constraints, which are facts of life. In the foregoing sections, we have presented the ways of handling problem created by social constraints, if any. However, each problem was handled in isolation, though in the real life all of them could occur simultaneously. In view of this it may be worthwhile to spell out the procedure of investment analysis once again here briefly. For this purpose, we would use the decision tree method, together with the pay-back period, NPV and IRR methods, which we consider as the appropriate method for evaluating the worth of alternative projects. The analyst could proceed as follows:

(a) Identify alternative investment opportunities in relation to the size of investments he has in mind and the kind of expertise and interest he possesses.

(b) Examine the feasibilities of all the projects identified in step (a) in terms of the availability of land, plant and machinery, raw-materials, technical know-how, finance, foreign exchange requirements, etc.

(c) Project cash flows from each of the feasible projects during their life-times. In doing so, due consideration must be paid to the expected inflation/deflation and all the likely values for all the critical variables in each project. Thus, for each project, there could be alternative series of cash flows, depending on the various values of its critical variables. Depending upon the investor and the interest in financial/or economic analysis, cash flows for the individual investor and/or the society could be projected.

(d) Compare cash flows of alternative projects with the aid of the measures of investment worth, such as the pay-back period, NPV and IRR. There would be as many values for each of these measures for a project as the number of combinations of its critical variables. While comparing alternative projects, both the expected return (in the form of pay-back period, NPV and IRR) and expected risk (range/standard deviation/coefficient of variation in the pay-back period, NPV and IRR) must be taken in to account.

(e) Select one or more projects on the basis of the expected return from and risk involved in them, investors' willingness and ability to bear risk, and the funds available for investment.

(f) Review the investment decision after the project has run through for a few years and the extent to which the expected results have come true. Examine the difference and reasons, and learn from the mistakes, if any, so as to avoid them in future.

This is all that a chapter on investment analysis in a text in managerial economics can offer. For more, students are advised to go through the references listed below and some case studies on the subject, and to experiment real life investment decisions.

## REFERENCES

1. Bierman, H. and S. Smidt (1975): *The Capital Budgeting Decision,* 4th edn., Macmillan, New York.
2. Gittinger, J.P. (1984): *Economic Analysis of Agricultural Projects,* 2nd edition, Johns Hopkins University Press, Baltimore.
3. Gupta, G.S. (1977): "Techniques for Investment Decisions," *Dynamic Management,* 1 (March), 75–97.
4. Little I.M.D. and I.A. Mirrlees (1974): *Project Appraisal and Planning for Developing Countries,* Basic Books, New York.
5. Squire, Lyn and Herman G. Van der Tak (1975): *Economic Analysis of Projects,* Johns Hopkins University Press, Baltimore.
6. United Nations Industrial Development Organization (1972): *Guidelines for Project Evaluation,* United Nations, New York.
7. Van Horne, J.C. (1983): *Financial Management and Policy,* 6th edition, Prentice-Hall, New Delhi.

## CASELETS

1. The Murari Sugar Cooperatives (MSC) is considering an investment of Rs. 354 million for the production of white crystal sugar by double sulphitation from sugarcane. The depreciable capital is expected to be Rs. 300 millions. The promoters contemplate to raise the share capital in the amount of Rs. 130 million, the rest of the fund would come from term loans from financial institutions at the expected interest rate of 12 percent. On the assumption of constant prices, the expected cash-flows over the first five years during the operating period are as follows.

| *Year* | $\pi$ | *D* | *I* |
|---|---|---|---|
| 1 | (42.30) | 27.75 | 18.75 |
| 2 | (5.91) | 25.00 | 18.75 |
| 3 | 4.05 | 22.71 | 18.75 |
| 4 | 13.18 | 20.61 | 16.20 |
| 5 | 15.15 | 18.65 | 14.90 |

where, $\pi$ = Profit (loss) before tax but after all operating costs, depreciation and interest on borrowed fixed capital (net profit before tax), expressed in Rs. millions

*D* = Depreciation, in Rs. millions

*I* = Interest cost on borrowed fixed capital, in Rs. millions

MSC is subject to income tax at the rate of 30%, loss deductible. The project is not eligible for capital subsidy. The opportunity cost of own funds (share capital) is estimated at 15%. The project is expected to run without any difficulty for at least 10 years.

If you were the economist at MSC, would you advice it to undertake the project? If necessary, make reasonable assumptions about the missing information and state the same explicitly.

2. Goyal Brothers is multi-department partnership firm. It has been in the manufacturing and trading business for over 50 years. During this period, it has implemented a large number of projects, some of which have been wound up after the completion of their economic life, while others are still running. The company has done reasonably well in the past but currently there is problem threatening the partnership. In the past, projects were undertaken without a serious financial or economic appraisal. However, in a recent board meeting, it was resolved to thoroughly scrutinize all the alternative project proposals from all the departments on the basis of various measures of economic worth. The meeting also decided to limit all the new investments to a total of Rs. 500 million.

   The various project proposed by all the departments have the following projections for their investment costs and net cash flows (Rs. millions):

| *Project* | *Investment cost* | *Net cash flows at the end of the year* | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| *A* | 200 | –20 | 40 | 60 | 90 | 90 | 90 | 150 |
| *B* | 200 | –10 | 50 | 80 | 100 | 100 | 160 | — |
| *C* | 200 | 10 | 60 | 100 | 100 | 180 | — | — |
| *D* | 100 | –5 | 25 | 40 | 50 | 50 | 50 | 90 |
| *E* | 100 | 0 | 40 | 50 | 60 | 60 | 100 | — |
| *F* | 100 | 10 | 50 | 60 | 60 | 80 | — | — |
| *G* | 100 | 20 | 50 | 50 | 80 | — | — | — |

The net cash flows represent sales minus all costs, except depreciation and interest on fixed capital borrowings and the terminal year flows are inclusive of the corresponding salvage value. A blank (—) in net cash flows means that the project is expected to be concluded before that period.

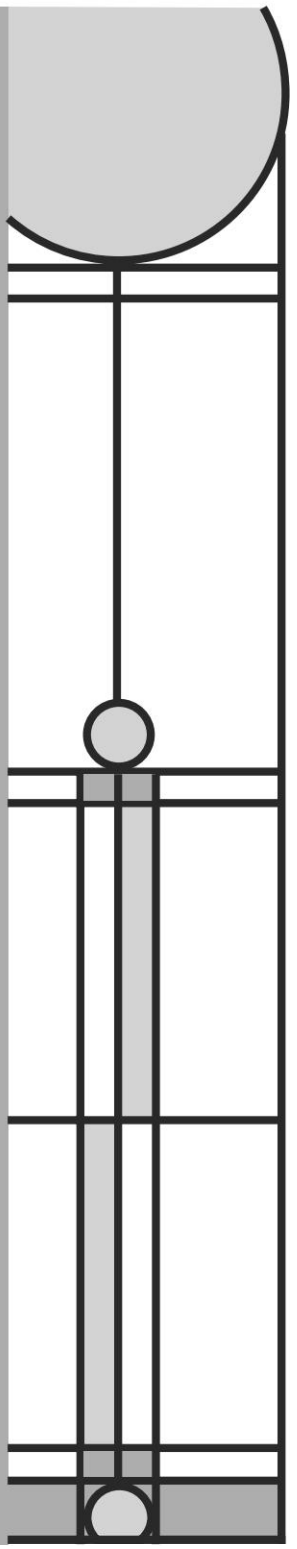
The company's current financial position and operating results are reflected in the magnitudes of the following important variables:

| | |
|---|---|
| Net worth | : Rs. 400 million |
| Equity debt ratio | : 1 : 1.5 |
| Net fixed assets | : Rs. 700 million |
| Net current assets | : Rs. 300 million |
| Profit after taxes (last year) | : Rs. 85 million |
| Dividend (last year) | : 20% |

The current opportunity cost of funds to the company is 15%. The company is in no special hurry to recover its investment costs.

If you are appointed as the investment consultant to this company, what would be your advice? Make sure that you compute all the useful measures of investment worth and comment on your findings, particularly if different measures give incompatible results. Make appropriate assumptions about the missing data and state the same explicitly.

The McGraw-Hill Companies



# ECONOMIC OPTIMIZATION

Economics is both a positive as well as a normative social science. As in the former, it helps in **explaining** why the various economic units behave the way they do. As the latter, it guides the decision-maker regarding **what it ought to do** under its given objectives and constraints. For example, **positive economics** would offer rationalization as to why a particular household buys a particular consumption basket, or why a producer produces a given quantity of its output or why the minimum wage regulation cause unemployment. In contrast, normative economics would help in determining what consumption basket the household ought to select, given its objective (maximization of utility or satisfaction) and the constraints (income, custom, religion, ethics, values, etc.); what production level the entrepreneur must choose, given its objective (profit maximization, values, etc.) and the constraints (technology, raw-material availability, input prices, etc.); and what should the minimum wage be, given the socio-political philosophy, etc. The latter role of economics is referred to as the one dealing with optimization. This involves, analysis of data/facts, as in positive economics, and, in addition, values cherished by the decision-makers, like ethics, religion, customs, political position, etc. The two roles are, of course, inter-related. Though normative role goes beyond economics, since we are concentrating only on economic fundamentals, we call it economic optimization.

## A.1 MEANING AND TYPES OF OPTIMIZATION PROBLEMS

Optimization deals with the determination of extreme values which could be maximum or minimum for the goal/objective variable. The goal variable could be just one (unique) or more (multiple). For example, a private firm might pursue profit-maximization as its **single goal**. If so, the optimization technique must determine the values of the variables, which are under the firm's control, called **choice variables**, so that they ensure the achievement of maximum possible profit to the firm. Alternatively, a public sector firm might aim at minimizing its average cost of production as the sole objective. In that case, the role of optimization techniques would be to find out the values of the variables under the firm's control that are

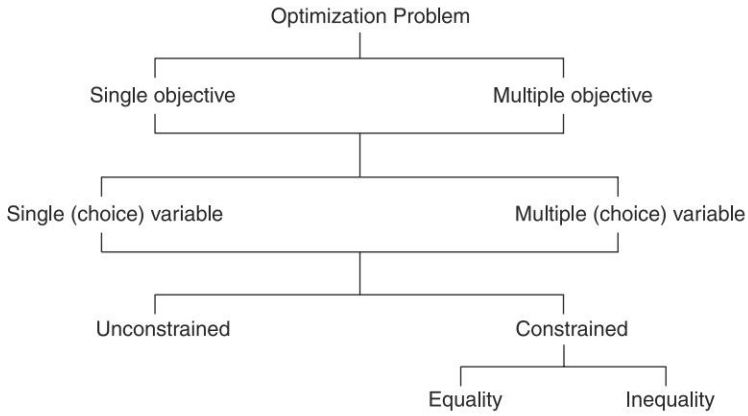
consistent with the minimum possible value for its average cost. In contrast, a government undertaking might have **twin goals**, namely, maximization of profit and maximization of employment of unskilled labour. Further, these two goals may or may not be compatible i.e., the achievement of one may lead or harm the attainment of the other. If the two or more goals are conflicting, one has to resort to the multiple variable optimization techniques which are outside the scope of this text. It will be sufficient to point out here, that the problem could be handled by choosing one goal variable as the objective and the remaining goal variables having some assigned specific (target) values to act as the constraints. If necessary, iterations could be carried out by changing the goal variables and/or by redefining the targets of the residual goal variables.

The value of the objective variable must be controllable by the decision-maker if it has any role in optimization. Thus, the profit level of a profit-maximizing firm might depend on the quantity of its output, price of its product, its production cost budget, its advertising budgets, etc. If none of these profit-determining variables are under the control of the firm, it has no role in profit-maximization. However, in the real world, the decision-maker enjoys flexibility with respect to at least some profit determining variables and hence it plays a significant role in making a given amount of profit (loss). This leads us to another classification of optimization problem. This is regarding whether, the profit (objective variable) determining variable is just one or more. If it is the former, we have **the single (choice) variable optimization problem** and if the latter, there is the multiple variable optimization problem. Thus, if profit depends only on the level of production, the profit-maximizing problem is a single variable optimization problem. In contrast, if the profit depends on the level of output as well as the advertisement budget (or in addition on some other variables as well), there is a **multiple variable optimization problem.**

The optimization problem facing a decision-making unit is further classified into **unconstrained and constrained problems**. In the former, the decision-maker optimizes subject to no constraints, internal or external. In the latter, it has one or more constraints (also called side conditions), imposed either, by itself (internal) or by outside agencies (external) such as government and or market conditions. An example of unconstrained optimization would be one where the firm aims at the maximum possible profit subject to no constraints of any kind. In contrast, under constrained optimization, the firm aims at maximum possible profit subject to one or more constraints such as a fixed production cost budget, a fixed quantity of output to be produced, availability of scarce raw-materials in fixed quantity, employment of minimum number of unskilled labour, etc.

The constrained optimization problems are further classified into **equality and inequality constrained problems**. For example, a profit-maximizing firm might be required to produce a specific quantity of output of all of its multiple products or if it is a single product firm, it might be faced with a fixed production cost budget or a fixed quantity of a particular scarce raw-material. Under these conditions, the optimizing firm must strictly adhere to the given value of the constrained variable. An example of optimization subject to inequality constraints would be the one where the firm is seeking maximum possible profits but there is a fixed quantity of skilled labour available in the market; the firm can employ all the skilled labour, or any quantity less than it but no more.

The various types of optimization problems can, thus, be represented as follows:




## A.2 PREREQUISITES FOR SOLVING OPTIMIZATION PROBLEMS

Before we proceed with the examples and solution of optimization issues, it is pertinent to emphasize that there are two essentials for solving any optimization problems. They are:

(a) There must be an objective function, indicating that the objective variable depends on at least one variable which is controllable by the decision-maker.

(b) The objective function must be expressed in a form suitable for analysis, called analytical form.

It is easy to see that if the objective variable has either a fixed value or that its value depends on some variables, all of which are parameters to the decision maker, there is nothing to optimize. For example, the government could assign a fixed quota (no more, no less) for importing a certain kind of Japanese car to a certain dealer and fix the selling price of those cars in India. If cars have a good market, the dealer would import but it has no optimization problem. There is no variable which affects its profit and on which it has any control. The import price is fixed by the Japanese firm and import duties, if any, are decided by the government of India. There is no choice and so there is no problem. However, if the government leaves the dealer free in setting its own selling price for the car in India and the Indian market is imperfect, *ceteris paribus*, there is an optimization problem. The profit the dealer would make depends on the price he charges. Since he has a role in setting the car price (choice), he can work out his optimum strategy.

After discussing the condition for the existence of the optimization problem, we proceed to the essential condition for the derivation of the solution for the problem. Mere dependence of profit (objective variable) on one or more of the controllable factors does not guarantee the derivation of the solution. For example, if the profit depends on price but the relationship between the two is not known in any convenient form, there is no technique which can help determine that price which gives the maximum possible profit. The convenient or **analytical form of a function** would be any one of the following:

(a) Table, (b) Graph, (c) quation

To explain these, consider a bi-variate function, which expresses profit as a function of output alone:

$$\pi = f(Q) \tag{A.1}$$

(where, $\pi$ = profit, $Q$ = output)

Function A.1 is silent about the nature of the relationship between profit and output. The relationship between these two variables could be (a) monotonic positive, (b) monotonic negative, or non-monotonic

of the either kind, like (c) positive to negative, or (d) negative to positive. In the first case, there is no solution to the problem of profit-maximization unless there is a constraint on the level of output. This is because, in the case of a positive monotonic relationship between profit and output, the more the firm produces the more is its profit and so there is no optimum output level. In the second case, where profit is a monotonic negative function of output, the less the firm produces the more is its profit, thus the solution is the zero level of output. Needless to say, the last is an altogether unrealistic example but is offered to illustrate the concept. Thus, it is clear that monotonic function poses no unconstrained (free) optimization problem. Since a linear function is necessarily a monotonic function, the objective function must be non-linear to warrant unconstrained optimization problem.

Most economic relationships are non-monotonic and pose issues for optimization. The usual relationship between profit and output would be of the following type, expressed in tabular form:

**Table A.1** *Hypothetical Profit-Output Relationship*

| *Output (Q)* | *Profit (II)* | *Output (Q)* | *Profit (II)* |
|---|---|---|---|
| 0 | –60 | 10 | 190 |
| 1 | 12.5 | 11 | 187.5 |
| 2 | 30 | 12 | 180 |
| 3 | 67.5 | 13 | 167.5 |
| 4 | 100 | 14 | 150 |
| 5 | 127.5 | 15 | 127.5 |
| 6 | 150 | 16 | 100 |
| 7 | 167.5 | 17 | 67.5 |
| 8 | 180 | 18 | 30 |
| 9 | 187.5 | 19 | – 12.5 |

The same relationship could be available in graphical form as presented in Fig. A.1.

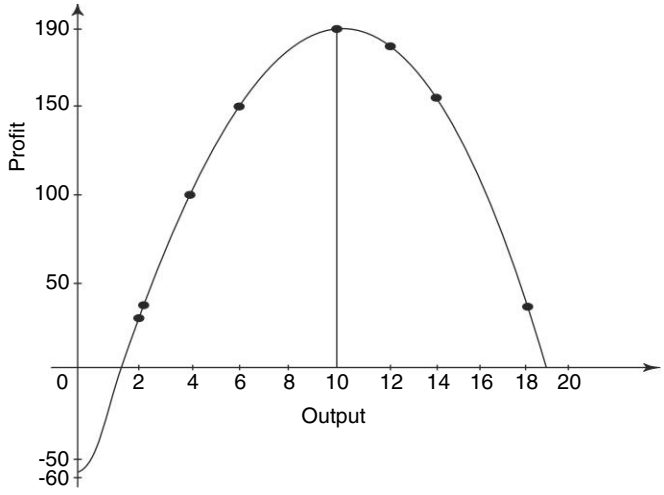



**Fig. A.1** *Profit function*

Yet another analytical form for the function would be the equation:

$$\pi = -60 + 50Q - 2.5Q^2 \quad \text{(A.2)}$$

It can be verified that Table A.1, Fig. A.1 and Eq. A.2 denote an identical relationship between profit and output. If the relationship is known in any one of these three forms, the optimization problem could be easily handled. To illustrate the technique, called the **Marginal Analysis**, the optimum solutions for different kinds of problems through all the three forms of the function are discussed in the following sections.

## A.3 OPTIMIZATION THROUGH TABLES AND GRAPHS

From the tabular form of the profit functions, it is clear that profit is at maximum when $Q = 10$, assuming that there are no fractions of a unit of output. Alternatively, one could obtain average incremental profit (AIP), which is like marginal profit, for various output levels. If this is done, the profit maximizing output would be the one at which the AIP is zero, or if at no output level the AIP = 0, then the one at which it is the least and after which it turns negative. The AIP is defined as change in profit divided by change in $Q$. In Table A.1, since output increases by one unit each time, AIP = $\Delta\pi$, where $\Delta$ stands for change. The AIP table is given below:

| *Output* | *AIP* | *Output* | *AIP* |
|---|---|---|---|
| 1 | 72.5 | 10 | 2.5 |
| 2 | 17.5 | 11 | –2.5 |
| 3 | 37.5 | 12 | –7.5 |
| 4 | 32.5 | 13 | –12.5 |
| 5 | 27.5 | 14 | –17.5 |
| 6 | 22.5 | 15 | –22.5 |
| 7 | 17.5 | 16 | –27.5 |
| 8 | 12.5 | 17 | –32.5 |
| 9 | 7.5 | 18 | –37.5 |
| | | 19 | –42.5 |

By this method also, at $Q = 10$, profit is maximum. In the above table, profit was directly given for various output levels. In some cases, tables are available in the forms of total revenue (*TR*) and total costs (*TC*) for various levels of output. If so, the profit column could be obtained simply as the difference between *TR* and *TC* for corresponding outputs levels. Once that is done, the above technique could be followed to obtain the profit-maximizing level of output. By the same methodology, the optimization problem corresponding to minimization could be handled. Instead of looking for the value of the controlled variable (output) that maximizes the objective variable (profit), one could now look for the value of the controlled variable (say output) that minimizes the objective variable (say average cost). If the objective function contains more than one choice variable, the tabulation method is still possible but is tedious to apply. Needless to say, the degree of inconvenience increases as the number of controllable variables increase and at an increasing rate.

The marginal or incremental method is available even if the function is given in the graphical form. In Fig. A.1, profit is shown graphically against output and if the optimization problem is to determine the

output level which maximizes profit, then all that one has to look for is the "top of the hill". This top is at $Q = 10$ and hence the profit-maximizing output equals 10. The problem, however, need not be so simple in all even single-variable optimization cases. For example, one may be given a total cost-output graph, and from that required to determine, say, the output level at which the average cost is the least. In that case, one would need to derive the average cost curve from the total cost curve before the optimization problem could be answered. This could be attempted graphically as illustrated in Fig. A.2. In Fig. A.2, *TC* denotes the total cost curve. It should be noted that, by definition, average cost at any point along the *TC* curve is given by the slope of a straight line from the origin to that point. Thus, in Fig. A.2, *AC* at point *A* is given by angle $\alpha_1$, which equals $AQ_1/OQ_1$, which in turn, equals *aq*, for $oq = 1$. Similarly, *AC* at point *B* is given by angle $\alpha_2$, which equals $BQ_2\ OQ_2$, which, in turn, equals $bq_1$, for $oq_1 = 1$; and so on. Thus, given the *AC* at various output levels, the *AC* curve could be drawn as in Fig. A.2. The solution of the optimization problem then is obvious: the "bottom of the hill" is the answer, that is, the output at which the *AC* curve has the least value ($Q = Q_1$), is the *AC* minimizing output level.

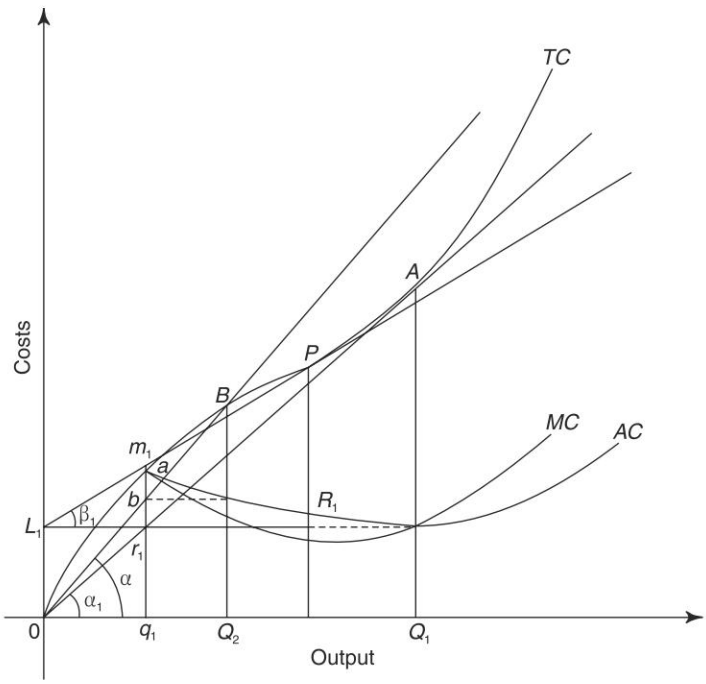



**Fig. A.2** *Derivation of AC and MC curves from TC curve*

At this stage, it will be useful to explain the derivation of the marginal curve from the total curve as well. The marginal cost at any point along the *TC* curve is given by the slope of a line drawn tangent to the *TC* curve at that point. Thus, in Fig. A.2, *MC* at point *P* on the *TC* curve is given by angle $\beta_1$,which equals $m_1r_1/L_1\ r_1$, which, in turn, equals, $m_1\ r_1$ for $L_1\ r_1 = 1$. In the same way, *MC* at various output levels could be determined and through them the entire *MC* curve.

It would be interesting to note that there is a definite relationship among the total, average and marginal curves. This is as follows:

(a) When the total reaches its top, the marginal touches the zero line.

(b) When the total attains the point of inflexion (where its curvature changes, at point $P$ in Fig. A.2), the marginal assumes its least (minimum) value.

(c) When the average takes its minimum (maximum) value, marginal equals average.

(d) When the marginal is less than the average, the average is falling.

(e) When the marginal is more than the average, the average is rising.

As with the tabulation form, the graphical form is amenable for handling the optimization problem when the objective function has more than one argument. However, it is equally cumbersome to deal with the multi-variate functions.

## A.4 OPTIMIZATION THROUGH EQUATIONS

The equation form of the function is the most appropriate one for solving the problem of optimization, be it a single variable or multi-variable function and an unconstrained or constrained optimization problem. The marginal principle holds good, so long as either it is an unconstrained optimization issue or if it is a constrained one, the constrained is of the equality type. However, one needs to take recourse to the calculus of differentiation to handle these problems and besides the marginal condition, there is (are) side condition(s) to distinguish between maxima and minima. Under the situation when the constraint is of an inequality type, the programming techniques have to be used to work out the optimum solution. In the following sections, examples are provided to explain the optimization procedure. As before, simple illustrations would precede the difficult ones.

### Unconstrained Single Variable Optimization Problems

The simplest optimization problem consists of the one where the objective function is a bi-variate one and the decision-maker has no constraint. The bi-variate function, by definition, has just one argument. To illustrate, let us once again consider the problem of profit-maximization of the previous section. The bi-variate profit function, in the equation form, is given in Eq. A.2. The choice (independent) variable is output ($Q$). The problem is to find out that level of output which maximizes profit. For this there are two conditions, namely, necessary and sufficient conditions.

The **necessary condition** for optimization states that the first derivative of the objective function with respect to the independent variable must equal zero. Stated mathematically, it is

$$\frac{d\pi}{dQ} = 0 \tag{A.3}$$

The necessary condition is concerned about the extreme values only. However, the extreme value could be either maximum or minimum, and hence the necessary condition is unable to distinguish between these two extremes. To take care of this, there is a **sufficient condition**, which for maximization, requires that the second derivative of the objective function with respect to its independent variable must be negative. Stated mathematically, it is

$$\frac{d^2\pi}{dQ^2} < 0 \tag{A.4}$$

If the optimization problem is one of minimization, then the necessary condition would be the same as in Eq. A.3, but the sufficient condition would be an inequality, which would be the opposite of inequality A.3 above. Specifically, it would be

$$\frac{d^2\pi}{dQ^2} > 0 \qquad \text{(A.5)}$$

Thus, meeting of conditions A.3 and A.4 would ensure maximization of $\pi$ with respect to $Q$ while that of conditions A.3 and A.5 the minimization of $\pi$ with respect to $Q$. The necessary and sufficient conditions are also known as **the first order and second order conditions**, respectively, as they involve the first and second derivatives.

Applying these conditions to our simple optimization problem, the first derivative of the objective function A.2 is*

$$\frac{d\pi}{dQ} = 50 - 5Q$$

Superimposing the necessary condition (vide Eq. A.3) on this, we have

$$50 - 5Q = 0$$

or, $$Q = 10$$

The second derivative of the objective function A.2 yields

$$\frac{d^2\pi}{dQ^2} = -5$$

Since the second derivative is negative, $Q = 10$ satisfies the sufficient condition for profit maximization. Thus, the firm must produce 10 units of its output to earn the maximum possible profit. Substitution of $Q = 10$ in the profit function A.2 gives, $\pi = 190$, which tallies exactly with the result obtained above through the tabular and graphical approaches.

## Unconstrained Multiple Variable Optimization Problems

The multiple variable optimization problem is characterized by two or more independent variables. The simplest example of such an optimization problem would, thus, have two independent variables. A possible example in this category would be the one where a profit-maximizing firm is engaged in the production of two goods and its profit depends on the outputs of those two goods only. Let the objective function of such a firm be

$$\pi = -125 + 100X_1 + 50X_2 - 5X_1^2 - 10X_2^2 + 10X_1X_2 \qquad \text{(A.6)}$$

where, $\pi$ = Profit

$X_1$ = Output of product 1

$X_2$ = Output of product 2

The firm aims at minimum possible profit and faces no constraints. The optimization procedure would be similar to that of a single variable case but it would involve partial derivatives.

The necessary condition for optimization would be that the total derivative of $\pi$ with respect to both $X$ and $Y$ be zero, i.e. $d\pi = 0$, and the sufficient condition for profit minimization would be that the

* For derivatives, please refer to any standard textbook on differential calculus or mathematical economics.

second total derivative be negative, i.e. $d^2\pi < 0$. For profit minimization, the necessary condition would be the same but the sufficient condition would entail that the second total derivative be positive, i.e. $d^2\pi > 0$. It can be mathematically proved that these conditions are equivalent to the following corresponding conditions*.

(I) Necessary $\partial\pi / \partial X_1 = 0$ (A.7)

Conditions for

Optimization $\partial\pi/\partial X_2 = 0$ (A.8)

(Maximum or Minimum)

(II) Sufficient $\partial^2\pi / \partial X_1^2 < 0$ (A.9)

Conditions for maximization $\partial^2\pi / \partial X_2^2 < 0$ (A.10)

$$\left(\frac{\partial^2\pi}{\partial X_1^2}\right)\left(\frac{\partial^2\pi}{\partial X_2^2}\right) > \left(\frac{\partial^2\pi}{\partial X_1 \partial X_2}\right)^2 \quad \text{(A.11)}$$

(III) Sufficient $\dfrac{\partial^2\pi}{\partial X_1^2} > 0$ (A.12)

Conditions for minimization $\dfrac{\partial^2\pi}{\partial X_2^2} > 0$ (A.13)

$$\left(\frac{\partial^2\pi}{\partial X_1^2}\right)\left(\frac{\partial^2\pi}{\partial X_2^2}\right) > \left(\frac{\partial^2\pi}{\partial X_1 \partial X_2}\right)^2 \quad \text{(A.14)}$$

It must be noted that the conditions A.11 and A.14 are identical.

Applying the above conditions to the optimization problem of our two product firms, we first take the first partial derivatives of the objective function A.6 and set each of them equal to zero:

$$\frac{\partial\pi}{\partial X_1} = 100 - 10X_1 + 10X_2 = 0$$

or, $X_1 - X_2 = 10$

$$\frac{\partial\pi}{\partial X_2} = 50 - 20X_2 + 10X_1 = 0$$

or, $X_1 - 2X_2 = -5$

Solving these two equations, we get

$$X_1 = 25 \text{ and } X_2 = 15$$

To verify the second order conditions, we take the second partial derivatives of the objective function A.6:

$\partial^2\pi / \partial X_1^2 = -10$

$\partial^2\pi / \partial X_2^2 = -20$

$\partial^2\pi / \partial X_1 \partial X_2 = 10 = \partial^2\pi / \partial X_2 \partial X_1$

These results meet with the three second order (sufficient) conditions of profit-maximization, viz.

* For the proofs as well as for the understanding of partial derivatives readers could look up Chiang. Alpha C. (1984): *Fundamental Methods of Mathematical Economics* or any other good text on mathematical economics.

A.9, A.10 and A.11. Hence, the solution $X_1 = 25$ and $X_2 = 15$ is the optimum solution for the firm in question. Substitution of these values into the objective function yields $\pi = 1500$, which denotes the maximum possible profit that the firm could make.

The presence of more than two independent variables in the objective function would simply complicate the solution but the procedure would remain the same. In particular, the following points are worth noting:

(a) Number of necessary conditions would equal the number of independent variables in the objective function.

(b) Each necessary condition would mean setting the partial derivative of the objective function with respect to one independent variable equal to zero.

(c) Sufficient conditions for maximization would require each second order partial derivative to be negative, while those for minimization would require each one of them to be positive; and

(d) Corresponding to condition A.11 (or A.14) above, there would be additional sufficient conditions for optimization. These would involve what is popularly known as, the Hessian determinants to alternate in sign, starting with negative in the case of maximization and to be all positive in the case of minimization.*

A detailed discussion of more than two independent variables would involve higher mathematics, which is beyond the scope of this text. The interested students are advised to refer to the list of references provided at the end of this Appendix.

## Constrained Multiple Variable Optimization Problems

In the above examples, the decision-maker was not faced by any constraints. In other words, it choice variables were independent of one another in the sense that the decision made regarding one variable did not impinge upon the choices of the remaining variables. However, in the real world, most firms face constraints and so are not free to choose any values of all the choice variables independently. These constraints might come from the availability of scarce raw materials, availability of funds for investment, limit on capacity stipulated by the license issuing authority or dictated by the market, availability of skilled manpower, etc. Under such situations, the decision-maker faces, what is called the constrained optimization problem. It must honour the constraints and within those strive to attain its stated objective(s). The constraints which a firm faces are of two kinds, viz., equality constraints and inequality constraints. Accordingly, optimization procedures pertaining to both kinds of constraints are discussed in what follows:

## Equality Constraints Optimization Problems

Under the equality constraints, two or more of the independent variables are bound by one or more equality relationships among them. Thus, to pursue the above example of the two product profit-maximizing firm further, suppose the firm is required to produce exactly 10 units of both its products. Then its optimization problem becomes

* The Hessian determinant is composed of all the second order partial derivatives of the objective function. Its order is exactly equal to the number of independent variables in the objective function. For details, see Chiang (1984), op. cit.

Maximize $\pi = -125 + 100X_1 + 50X_2 - 5X_1^2 - 10X_2^2 + 10X_1X_2$

Subject to $X_1 + X_2 = 10$ (A.15)

Here the firm's choice for $X_1$ and $X_2$ is constrained by their sum to be exactly equal to 10. Thus, its degrees of freedom is reduced from 2 to 1, i.e., earlier under unconstrained issue, it was free to choose any values for both $X_1$ and $X_2$, but now it can choose the value for either $X_1$ or $X_2$, value of the remaining variables ($X_2$ or $X_1$) would be given by the constraint $X_1 + X_2 = 10$.

There are two ways to solve this problem:

(a) Substitution Method

(b) Lagrangian Multiplier Method

Under the substitution method, which is also known as the **elimination of variables method**, the constraint is substituted in the objective function and then the procedure of unconstrained optimization is applied to the so derived objective function. By this, the number of choice variables in the objective function is reduced by the number of (equality) constraints:

Applying the substitution method to the above problem, the constraint can be solved to

$$X_2 = 10 - X_1$$

Substituting this value of $X_2$ in the objective function, we get

$$\pi = -125 + 100X_1 + 50\,(10 - X_1) - 5X_1^2 - 10(10 - X_1)^2 + 10X_1(10 - X_1)$$

$$= -125 + 100X_1 + 500 - 50X_1 - 5X_1^2 - 1000 - 10X_1^2 + 200X_1 + 100X_1 - 10X_1^2$$

or, $$\pi = -625 + 350X_1 - 25X_1^2$$

The constraint optimization problem now reduces to the free maximization of the above function with respect to $X_1$. Setting its first derivative to zero, we have

$$d\pi/\, dX_1 = 350 - 50X_1 = 0$$

or, $$X_1 = 7$$

Substitution of this value in the constraint ($X_1 + X_2 = 10$) yields $X_2 = 3$. The sufficient condition is met, for $\dfrac{d^2\pi}{dx_1^2} = -50 < 0$

The corresponding profit is obtained by the substitution of these values of $X_1$ and $X_2$ in the original objective function which comes out to be 600. If one compares this result with the result of the previous section, one would realize that the constraint has reduced the profit of the enterprise from 1500 to 600. This would always be the case so long as the constraint is a binding one. However, if the constraint is ineffective, there would be no change in the solution. Thus, if the constraint were $X_1 + X_2 = 40$, there would be no change, for the total output under unconstrained (free) profit-maximization was also exactly 40.

The substitution method is thus, quite simple. However, it is not always applicable. In cases, of more than one constraints or when the constraint is a complicated one, the substitution method is quite complicated. In such cases, the **Lagrangian multiplier method** is the ideal one to use.

The essence of the Lagrangian multiplier method is to convert a constrained optimization problem into a free optimization problem. This is done through constructing what is called, the Lagrangian expression, which has two parts: the first consists of the objective function and the second part of

the constraint multiplied by the parameter, called the Lagrangian multiplier (λ). The above mentioned constrained optimization problem is solved through this method as shown below. The Lagrangian expression would be

$$L^* = -125 + 100X_1 + 50X_2 - 5X_1^2 - 10X_2^2 + 10X_1X_2 + \lambda\,(10 - X_1 - X_2) \qquad \text{(A.16)}$$

It has been proved that optimization (maximization) of profit subject to the equity constraint is the same as the optimization of the Lagrangian expression subject to no constraint. Note that the new expression has an extra variable, λ. Thus, the method adds a variable to the optimization problem.

The necessary conditions for optimum solution would thus be,

$$\frac{\partial L^*}{\partial X_1} = 0 \frac{\partial L^*}{\partial X_2} = 0 \text{ and } \frac{\partial L^*}{\partial \lambda} = 0$$

Performing the first partial derivative on the expression A.16 and imposing the necessary conditions for optimization, we have

$$\frac{\partial L^*}{\partial X_1} = 100 - 10X_1 + 10X_2 - \lambda = 0$$

or, $$\lambda = -10X_1 + 10X_2 + 100$$

$$\frac{\partial L^*}{\partial X_2} = 50 - 20X_2 + 10X_1 - \lambda = 0$$

or, $$\lambda = 10X_1 - 20X_2 + 50$$

$$\frac{\partial L^*}{\partial \lambda} = 10 - X_1 - X_2 = 0$$

or, $$X_1 + X_2 = 10$$

Solving the first two equations, we have

$$-10X_1 + 10X_2 + 100 = 10X_1 - 20X_2 + 50$$

or, $$-20X_1 + 30X_2 = -50$$

or, $$2X_1 - 3X_2 = 5$$

The solution of equations

$$2X_1 - 3X_2 = 5$$

$$X_1 + X_2 = 10$$

Would yield

$$X_1 = 7 \text{ and } X_2 = 3$$

Which tallies exactly with the solution obtained earlier through the substitution method. There are, of course, sufficient conditions to verify the maximization, which are met but their proof is ignored here for the sake of simplicity.

The Lagrangian multiplier technique has an additional advantage over the substitution method. This comes through the value of the Lagrangian multiplier, λ. Substitution of the optimum values of $X_1$ and $X_2$

into either of the first two equations derived from the necessary conditions for optimization ($\lambda = -10X_1 + 10X_2 - 100$; $\lambda = 10X_1 - 20X_2 + 50$) would yield $\lambda = 60$. The $\lambda$ value is of special significance here. It denotes the marginal effect on the objective variable of an infinitesimally small change in the constraint variable. Thus, in this case, if $X_1 + X_2 = 10$ changes to $X_1 + X_2 = 11$, the maximum possible profit would increase by 60, i.e. from 600 to 660. Likewise, if the constraint changes to $X_1 + X_2 = 9$, the said profit would decline by 60, i.e. from 600 to 540. This information is very useful and hence the Lagrangian method is superior to the substitution method on this count as well.

It must be mentioned at this stage that the Lagrangian method is available even if the number of constraints are more than one. There would be need for as many Lagrangian multipliers as the number of constraints, one for each constraint. The various multipliers would, thus, indicate the marginal effects on the objective variable of a small change in the corresponding constraint.

## Inequality Constraints Optimization Problems (Linear Programming)

Under the inequality constraints, two or more of the choice variables are bound by one by or more of the inequality conditions among themselves. Such inequality conditions might exist due to some capacity constraints of scarce resources, such as investments, skilled manpower, scarce raw-materials, scarce machines, etc. To illustrate optimization, under such a situation, let us consider an example.

Suppose a firm is engaged in the production of two goods, $x$ and $y$. The production requires use of two scarce resources, viz. a particular machine and a particular type of engineer's time. The firm has a fixed quantity of these two scarce resources but is free to use them fully, partly or not at all. The fixed quantities are 120 hours of machines' time and 80 hours of engineers' time. The technology of the two products is such that a unit of product $x$ requires 2 hours of machine's time and 4 hours of engineer's time, while a unit of product $y$ requires 5 hours of machine's time and 2 hours of engineer's time. In such a resource and technology position, the firm has, thus, two constraints:

Machines' constraint: $2x + 5y \leq 120$ (A.17)

Engineers' constraint: $4x + 2y \leq 80$ (A.18)

In addition, there are non-negative constraints for the two products, because the firm cannot produce a negative quantity of either product. Thus,

$$x \geq 0, y \geq 0 \quad \text{(A.19)}$$

In addition to these, there must be the objective function of the firm. For the sake of simplicity, let the objective function be linear and as follows:

$$\text{Maximize } \pi = 3x + 4y \quad \text{(A.20)}$$

Function A.20 implies that the firm makes a profit of Rs. 3 per unit of product $x$ and Rs. 4 per unit of product $y$. The optimization problem is to maximize function A.20 subject to constraints A.17, A.18 and A.19. Since both the objective function and the constraints are linear, the problem is referred to as a linear programming problem. The linear programming problem can be solved through either geometry or algebra. To explain the geometry method first, plot the constraints on a graph. To do so, treat the inequality as equality, choose any two values for $x$ and find out the corresponding values for $y$ and identify the two points on the graph. Since the constraints are linear, the two points for each constraint would enable us to draw both the constraints. The procedure is elaborated below.

For constraint A.17: If $x = 10$, $y = 20$

If $x = 30$, $y = 12$

For constraint A.18: If $x = 10$, $y = 20$

If $x = 15$, $y = 10$

Constraints in A.17 and A.18 are marked by lines *AB* and *BD*, respectively, and constraints in A.19 by the *x*-axis and *y*-axis boundaries in Fig. A.3.

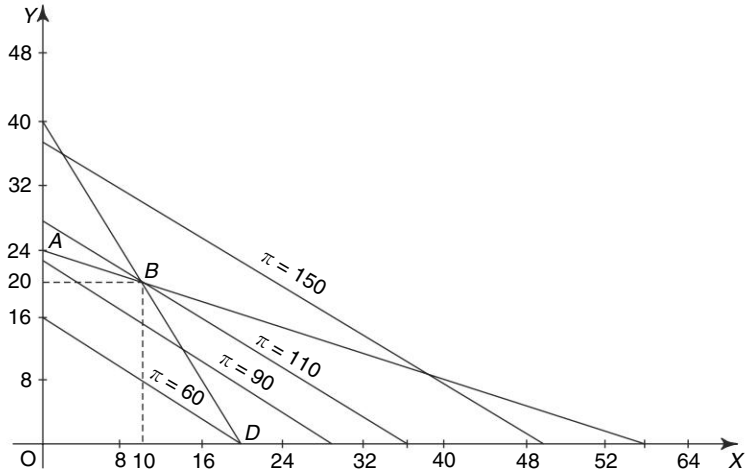



**Fig. A.3** *Solution of linear programming problem*

The feasible solution, the one which satisfies all the constraints, is given by the shaded region *OABD* in Fig. A.3. The optimum solution has to be somewhere in this region. In order to determine the optimum solution, one needs to draw the objective function for various arbitrary chosen values of the objective variable ($\pi$). In this case these would be the iso-profit lines. The same are also included in Fig. A.3 for $\pi = 60$, 90, 110 and 150. To explain the drawing of these profit lines, $\pi = 60$ line, for example, is obtained through the following function:

$$3x + 4y = 60$$

Thus, if $x = 10$, $y = 7.5$ and if $x = 12$, $y = 6$. Since the profit function is also linear, two points on it enable us to draw the full line. The firm would endeavour to attain as high a profit line as possible subject to its constraints, denoted by the feasible region. Thus, the point at which the iso-profit line is tangent to the feasible region denotes the optimum point. Thus, point *B* in Fig. A.3 marks the optimum solution. Its coordinates are $x = 10$, $y = 20$, which represents the solution to the problem in question. The maximum possible profit, which is known directly from the particular iso-profit line, stands at Rs. 110. Incidentally, note that under linear programming, the optimum solution would always lie at one of the corner points of the feasible region.

The graphical method is easy but inconvenient to use in the presence of many constraints. For this reason, the algebraic method is the popular one. The latter method uses the slack variables to convert inequalities into equalities. Thus, the constraints A.17 and A.18 are rewritten as

$$2x + 5y + s_1 = 120 \qquad \text{(A.21)}$$

$$4x + 2y + s_2 = 80 \qquad \text{(A.22)}$$

where, $s_1, s_2 \geq 0$

This gives us two additional variables ($s_1$ and $s_2$) but no additional equations. However, it has been proved that as many (two) variables would have zero values at the optimum point. Thus, of the 4 variables, two would have zero solutions. But it is not known which ones would have zero and which ones non-zero answers. Therefore, one has to try all possible combinations and then see which combination yields the optimum value for the objective variable. The possible combinations and the resulting values for the remaining variables, obtained through solving Eqs. A.21 and A.22, would be as follows:

| | *Possible Combinations* | *Resulting Solution to Residual Variables* |
|---|---|---|
| (a) | $x = 0, \quad y = 0$ | $s_1 = 120, \quad s_2 = 80$ |
| (b) | $x = 0, \quad s_1 = 0$ | $y = 24, \quad s_2 = 32$ |
| (c) | $x = 0, \quad s_2 = 0$ | $y = 40, \quad s_1 = -80$ |
| (d) | $y = 0, \quad s_1 = 0$ | $x = 60, \quad s_2 = -160$ |
| (e) | $y = 0, \quad s_2 = 0$ | $x = 20, \quad s_1 = 80$ |
| (f) | $s_1 = 0, \quad s_2 = 0$ | $x = 10, \quad y = 20$ |

Of the six possible solutions, solutions (c) and (d) are not feasible, for they lead to negative values for the slack variables, which violate the constraints. All the remaining solutions are feasible ones.

To determine the optimum among the four feasible ones, one has to evaluate each solution with regard to the value of the objective variable. This is done by substituting the values for $x$ and $y$ in the objective function A.20. Thus, if

(a) $x = 0, \quad y = 0, \quad \pi = 0$

(b) $x = 0, \quad y = 24, \quad \pi = 96$

(c) $x = 20, \quad y = 0, \quad \pi = 60$

(d) $x = 10, \quad y = 20, \quad \pi = 110$

Of the several values, solution (d) gives the maximum profit, hence $x = 10$ and $y = 20$ marks the optimum solution. It could now be verified that at the optimum values, the constraints are met. Further, it could also be checked as to whether the constraints are exactly met or there is excess capacity (slack) in any of the scarce resources. This would help in planning the alternative uses, if possible, for the unutilized part of the scarce resources. Checking on this, one finds that at $x_1 = 10$ and $y = 20$, the two constraints of engineers' time and machines' time are exactly met and thus, there is no slack in either of the scarce resources. Consequently, there is no scope for their alternative uses. Incidentally note that the algebraic method just describes is easily computerized while the geometric method discussed above has to be solved manually. For this reason also, the algebraic method has become the popular one today.

This concludes the discussion on linear programming but not on optimization techniques. As pointed out in the just concluded example, we took linear objective function as well as linear constraints. If either of these or both of them are non-linear, which is often the case in many real-life situations, what is the procedure of optimization? This calls for quadratic programming techniques, which is beyond the scope of this text.

## A.5 CONCLUSIONS

This appendix has dealt with various kinds of optimization issues and has provided procedures of handling the same. Since the text is intended to omit topics requiring higher mathematics, the advanced method have not been detailed. Further, just one issue of each type alone has been raised and answered. In particular, examples of maximization alone have been emphasized. As explained in the text, the procedure for minimization is the same except that the sufficient conditions have opposite signs. In the framework of linear programming, objective function would be minimized, constraints would have inequalities in the form of less than or equal to zero and the feasible region would be the area above the constraints in the case of minimization instead of the reverse in all these cases under maximization. It is believed that a thorough grasp of this material would enable the readers to understand all the optimization problems presented in this book.

## REFERENCES

1. Allen, R.G.D. (1962): *Mathematical Analysis for Economists*, London, Macmillan
2. Baumo1, W.J. (1982): *Economic Theory and Operations Analysis*, 4th edition, New Delhi, Prentice-Hall.
3. Chiang, Alpha C. (1984): *Fundamental Methods of Mathematical Economics*, 3rd edition, New Delhi, Tata McGraw-Hill Book Company.
4. Gupta, G.S. (1983–84): "Production Function and Optimum Input Mix in Fish Farming in India" *Vishleshan*, IX, 4 and X, 1 (Dec–March), 181–91.
5. Pappas, J.L., E.F. Brigham and Mark Hirschey, (1983): *Managerial Economics*, 4th edition, Chicago, Dryden Press.

## CASELETS

1. Goyal Enterprises manufactures an electronic good. The enterprise is a relatively new one and so is its product in the market. Currently, its sole objective is to maximize the sales of its product. With the help of a consultant, the enterprise has discovered that its sales depend solely on the money it would put on advertising the products. There are two media of advertising, namely television and newspapers, and the estimated sales function by the consultant is

$$S = 150 + 150A - 7.5A^2 + 24B - 3B^2$$

where, $S$ = sales (in number of units)

$A$ = advertising through television (in Rs. thousands)

$B$ = advertising through newspapers (in Rs. thousands)

(a) If you were the marketing manager of this enterprise, how much money would you like to spend on each of the two kinds of advertisements?

(b) Suppose the enterprise has a fixed constraint on the total advertisement budget in the amount of Rs. 10,500 (i.e. $A + B = 10.5$), what would be your suggestion regarding allocation of this budget between two advertising media?

(c) Suppose the sales of a unit of the product fetches a profit of Rs. 75, would you advise the enterprise to expand its total advertising budget?

(*Hint:* Use the Lagrangian multiplier value to answer the last question).

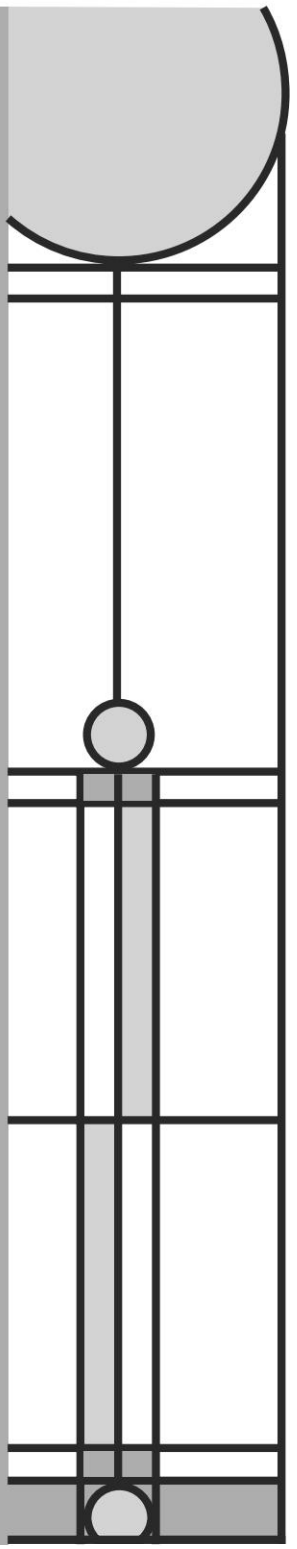
2. Kamal Electronics Ltd. (KEL) has just come out with a new two-in-one (radio-cum-cassette player) for installation in cars. It wishes to advertise the product widely among the current and prospective car owners. There are two advertising media that the company has decided to use for the purpose, namely television ($T$) and a popular magazine ($M$). The cost per advertisement in television is Rs. 500 and that in the magazine is Rs. 400. KEL has also decided that its advertisement must reach a certain minimum number of audiences in all important categories. The categories and the minimum requirements for each are as follows:

| *Audience Category* | *Minimum Target* |
|---|---|
| (a) Existing car owners | 6,00,000 |
| (b) Income tax payers | 1,80,000 |
| (c) Graduates | 2,60,000 |

The market survey has revealed that a television advertisement reaches 20,000 car owners, 15,000 income tax payers and 10,000 graduates, while a magazine advertisement reaches 30,000 car owners, 5,000 income tax payers and 10,000 graduates.

If you were the economist of KEL, what would be your recommendation regarding the number of advertisements, the company must sponsor in each of the two media? Assume that the sole objective of the company is to minimize the total cost of advertisement under the above constraints.

The McGraw-Hill Companies



APPENDIX

B

# TABLES FOR INVESTMENT ANALYSIS

The significance of the time value of money and alternative methods for handling the same were discussed in Chapter 11. Useful tables for handling problems involving time value of money are provided in this appendix. Six different types of tables have been included. Table B.1 gives the compound factors, Table B.2 the discount factors, Table. B.3 the future worth of annuity factors, Table B.4 the present worth of annuity factors, Table B.5 the sinking fund factors, and Table B.6 the capital recovery factors (also called the annuity for an investment for a given period) for various rates of interest and for various time periods. These have been derived through formulas formulated through the definitions of these concepts. In the following section, the derivations of these formulas and examples of the usefulness of these tables have been provided. Incidentally, note that all the formulas and tables assume that interest is reckoned once in a period.

## B.1 COMPOUND FACTORS (CF)

These give the value of a rupee after a specified period of time at various rates of interest. Thus, the formula is

$$CF = (1 + i)^T \tag{B.1}$$

where, $i$ = interest rate per period

$T$ = number of time periods

Thus, the value of a rupee at $i = 10\%$ per year, after, say 5 years, would equal

Rs. $(1 + .1)^5$ = Rs. 1.6105 (Table B.1)

If the amount were Rs. 5,000, its future sum (5 years hence) would equal

Rs. 5,000 $\times$ 1.6105 = Rs. 8,052.5

The table could be used even for monthly periods, if the interest rate is given on monthly basis. Thus, for example, if $i = 1\%$ per month, then the value of Re 1 after, say, 20 months, would equal

Rs. $(1 + .01)^{20}$ = 1.2202

And if the amount were Rs. 5,000, after 20 months it would equal

Rs. 5,000 × 1.2202 = Rs. 6,101.0

## B.2 DISCOUNT FACTORS (DF)

These factors help in determining the present value of a future sum, given the rate of interest and the future period. Its formula is the reciprocal of formula (B.1) given above:

$$DF = \frac{1}{(1+i)^T} \tag{B.2}$$

Thus, the present value of a rupee to be received after, say, 5 years, at interest rate equal to 10% per year, equals

$$\text{Rs. } \frac{1}{(1+0.1)^5} = \text{Rs. } 0.6209$$

If the amount were Rs. 5,000, its present worth would equal

Rs. 5,000 × 0.6209 = Rs. 3,104.5

As in case of the previous formula, the formula as well as its corresponding table (Table B.2) could be used for monthly basis if the interest rate were stipulated on monthly basis.

## B.3 FUTURE WORTH OF AN ANNUITY FACTOR (FWAF)

These factors give the future value of a uniform sum received each period for a certain period of time at various rates of interest. Thus, if one rupee is received at the end of each year for $T$ years, its total sum after $T$ years at an interest rate of $i$ per year would equal

$$S_1 = (1+i)^{T-1} + (1+i)^{T-2} + \ldots + (1+i) + 1$$

This is an expression in the geometrical progression. To reduce it into a manageable formula, let us re-write it and perform some mathematical operations:

$$S_1 = 1 + (1+i) + (1+i)^2 + \ldots + (1+i)^{T-2} + (1+i)^{T-1}$$

and,
$$S_1 (1+i) = (1+i) + (1+i)^2 + \ldots (1+i)^{T-2} + (1+i)^{T-1} + (1+i)^T$$

Subtracting the second equation from the first, we have

$$S_1 [1 - (1+i)] = 1 - (1+i)^T$$

or,
$$S_1 = \frac{(1+i)^T - 1}{i} \tag{B.3}$$

For a given rate of interest and the value of $T$, the values for $S_1$ can be obtained through this formula, which, in turn, denotes the future value factor for an annuity of rupee one. The various values for different values of $i$ and $T$ are given in Table B.3.

To illustrate the use of this table, one can consider the example of recurring deposit in a bank, suppose, an individual opens a recurring deposit account in bank where he puts amount of Rs. 40,000 per year in the bank for a period of 15 years and the bank pays an interest rate of 10% per annum on this amount. The amount on maturity would then equal to

Rs. 40,000 × 15 years FWAF at $i$ = 10%

= Rs. 40,000 × 31.772 (vide Table B.3)

= Rs. 1, 270,880

The FWAF is useful in all cases where the future value of a uniform annuity is to be calculated. This factor is, in fact, a cumulative summation of CF, as discussed above.

## B.4 PRESENT WORTH OF AN ANNUITY FACTOR (PWAF)

The PWAF measures the discounted (present) value of a uniform amount to be received every period over a certain period of time, given the rate of interest. Thus, the present value of one rupee expected to be received every year for $T$ number of years at an interest rate of $i$ per year would be given by

$$S_2 = \frac{1}{1+i} + \frac{1}{(1+i)^2} + \ldots + \frac{1}{(1+i)^T}$$

This expression is also in the geometrical progression and hence its sum could be found out by the same procedure as above.

Writing an expression for, $S_2\left(\frac{1}{1+i}\right)$, we have

$$S_2\left(\frac{1}{1+i}\right) = \frac{1}{(1+i)^2} + \frac{1}{(1+i)^3} + \ldots + \frac{1}{(1+i)^{T+1}}$$

Subtracting this equation from the equation of $S_2$, we have

$$S_2\left[1 - \frac{1}{1+i}\right] = \frac{1}{1+i} - \frac{1}{(1+i)^{T+1}}$$

or,
$$S_2 = \frac{1+i}{i}\left[\frac{1}{1+i} - \frac{1}{(1+i)^{T+1}}\right]$$

or,
$$S_2 = \frac{1 - \frac{1}{(1+i)^T}}{i} \qquad \text{(B.4)}$$

For a given value of $i$ and $T$, the value of $S_2$, could easily be deduced the formula (B.4). Table B.4 is obtained using this formula.

To illustrate its use, consider an option between a lump sum amount on retirement and a yearly pension in a certain amount. Suppose, on superannuation an employee is offered the following choice. Take a lump sum of Rs. 2 million or a yearly pension of Rs. 150,000 for life. Suppose the employee decides to compare this option on the assumption of 20 years more of his life time and rate of interest applicable in this case is 10% per annum. In that case, the employee could compute the present value of his life time pension and then compare that amount with the down payment of Rs. 2 million.

The present value of the pension, using the PWAF would be given by

Rs. 150,000 × 20 years PWAF at $i$ = 10%

= Rs. 150,000 × 8.5135 (vide Table B.4)

= Rs. 1,277,025

Since this amount is less than the down payment of the lump sum, the employee would opt for the lump sum amount. The PWAF is useful in all other computations involving the calculations of the present value of an annuity. This factor is simply a cumulative sum of DF.

## B.5 SINKING FUND FACTOR (SFF)

The SFF is simply the reciprocal of FWAF. Thus, it is defined as

$$\text{SFF} = \frac{i}{(1+i)^T - 1} \tag{B.5}$$

It denotes the uniform amount that a person must save every year for a certain number of years if he desires to accumulate a given sum at the end of that period, given the rate of interest. Thus, if a father desires to send his son for higher studies abroad after, say 10 years, and the cost of foreign education 10 years hence is estimated at Rs. 1.5 million and the rate of interest is 10% per year, then he must save annually for the next 10 years a uniform amount equal to

Rs. 1,500,000 × 10 years SFF at $i$ = 10%
= Rs. 1,500,000 × 0.0627 (vide Table B.5)
= Rs. 94,050

## B.6 CAPITAL RECOVERY FACTOR (CRF)

The CRF is simply the inverse of PWAF. Its formula would, thus, be

$$\text{CRF} = \frac{i}{1 - \frac{1}{(1+i)^T}} \tag{B.6}$$

It measures the uniform amount an investment must earn each period over a given period of time, given the rate of interest, to recover its original cost (Break-even). Thus, if an individual constructs a house costing Rs. 5 millions and decides to rent it out on such a uniform rent that he recovers the full cost of the house through rent in, say, 20 years, interest rate being 10%, then the yearly rent he must charge equals

Rs. 5,000,000 × 20 years CRF at $i$ = 10%
= Rs. 5,000,000 × 0.11745 (vide Table B.6)
= Rs. 587, 250

The CRF is highly useful in investment analysis.

Tables B.1 through B.6 present the values of all these six factors for various rates of interest and for different values of the time period. Use of these tables would facilitate calculations needed for investment analysis. Intelligent students could use these tables for the computations of the present values and also the internal rates of return for evaluating the economic worth of various investment opportunities.

**Table B.1** *Tables for Investment Analysis*

$CF = (I + i)^T$

(Interest Rate in Percentages)

| *Period* | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* | *10* | *11* | *12* | *13* | *14* | *15* | *18* | *20* | *25* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ..1 | 1.0100 | 1.0200 | 1.0300 | 1.0400 | 1.0500 | 1.0600 | 1.0700 | 1.0800 | 1.0900 | 1.1000 | 1.1100 | 1.1200 | 1.1300 | 1.1400 | 1.1500 | 1.1800 | !.2000 | 1.2500 |
| ..2 | 1.0201 | 1.0404 | 1.0609 | 1.0816 | 1.1025 | 1.1236 | 1.1449 | 1.1664 | 1.1881 | 1.2100 | 1.2321 | 1.2544 | 1.2769 | 1.2996 | 1.3225 | 1.3924 | 1.4400 | 1.5625 |
| ..3 | 1.0303 | 1.0612 | 1.0927 | 1.1249 | 1.1576 | 1.1910 | 1.2250 | 1.2597 | 1.2950 | 1.3310 | 1.3676 | 1.4049 | 1.4429 | 1.4815 | 1.5209 | 1.6430 | 1.7280 | 1.9531 |
| ..4 | 1.0406 | 1.0824 | 1.1255 | 1.1699 | 1.2155 | 1.2625 | 1.3108 | 1.3605 | 14116 | 1.4641 | 1.5181 | 1.5735 | 1.6305 | 1.6890 | 1.7490 | 1.9388 | 2.0736 | 2.4414 |
| ..5 | 1.0510 | 1.1041 | 1.1593 | 1.2167 | 1.2763 | 1.3382 | 1.4026 | 1.4693 | 1.5386 | 1.6105 | 1.6851 | 1.7623 | 1.8424 | 1.9254 | 2.0114 | 2.2878 | 2.4883 | 3.0518 |
| ..6 | 1.0615 | 1.1262 | 1.1941 | 1.2653 | 1.3401 | 1.4185 | 1.5007 | 1.5869 | 1.6771 | 1.7716 | 1.8704 | 1.9738 | 2.0820 | 2.1950 | 2.3131 | 2.6996 | 2.9860 | 3.8147 |
| ..7 | 1.0721 | 1.1487 | 1.2299 | 1.3159 | 1.4071 | 1.5036 | 1.6058 | 1.7138 | 1.8280 | 1.9487 | 2.0762 | 2.2107 | 2.3526 | 2.5023 | 2.6600 | 3.1855 | 3.5832 | 4.7684 |
| ..8 | 1.0829 | 1.1717 | 1.2688 | 1.3886 | 1.4 775 | 1.5938 | 1.7182 | 1.8509 | 1.9926 | 2.1436 | 2.3045 | 2.4760 | 2.6584 | 2.8526 | 3.0590 | 3.7589 | 4.2998 | 5.9605 |
| ..9 | 1.0973 | 1.1951 | 1.3048 | 1.4233 | 1.5513 | 1.6895 | 1.8385 | 1.9990 | 2.1719 | 2.3579 | 2.5580 | 2.7731 | 3.0040 | 3.2519 | 3.5179 | 4.4355 | 5.1598 | 7.4506 |
| 10 | 1.1046 | 1.2190 | 1.3439 | 1.4802 | 1.6289 | 1.7908 | 1.9672 | 2.1589 | 2.3674 | 2.5937 | 2.8394 | 3.1058 | 3.3946 | 3.7072 | 4.0456 | 5.2338 | 6.1917 | 9.3132 |
| 11 | 1.1157 | 1.2434 | 1.3642 | 1.5395 | 1.7103 | 1.8983 | 2.1049 | 2.3316 | 2.5804 | 2.8531 | 3.1518 | 3.4785 | 3.8359 | 4.2262 | 4.6524 | 6.1759 | 7.4301 | 11.6415 |
| 12 | 1.1268 | 1.2682 | 1.4258 | 1.6010 | 1.7959 | 2.0122 | 2.2522 | 2.5182 | 2.8127 | 3.1384 | 3.4985 | 3.8960 | 4.3345 | 4.8179 | 5.3503 | 7.2876 | 8.9161 | 14.5519 |
| 13 | 1.1381 | 1.2936 | 1.4685 | 1.6851 | 1.8856 | 2.1329 | 2.4098 | 2.7196 | 3.0658 | 3.4523 | 3.8833 | 4.3635 | 4.8980 | 5.4924 | 6.1528 | 8.5994 | 10.6993 | 18.1899 |
| 14 | 1.1495 | 1.3195 | 1.5126 | 1.7317 | 1.9799 | 2.2609 | 2.5785 | 2.9372 | 3.3417 | 3.7975 | 4.3104 | 4.8871 | 5.5348 | 6.2613 | 7.0757 | 10.1472 | 12.8392 | 22.7374 |
| 15 | 1.1610 | 1.3459 | 1.5580 | 1.8009 | 2.0789 | 2.3968 | 2.7590 | 3.1722 | 3.6425 | 4.1772 | 4.7846 | 5.4736 | 6.2543 | 7.1379 | 8.1371 | 11.9737 | 15.4070 | 28.4217 |
| 16 | 1.1726 | 1.3728 | 1.6047 | 1.8730 | 2.1829 | 2.5404 | 2.9522 | 3.4259 | 3.9703 | 4.5950 | 5.3109 | 6.1304 | 7.0673 | 8.1372 | 9.3576 | 14.1290 | 18.4884 | 35.5271 |
| 17 | 1.1843 | 1.4002 | 1.6528 | 1.9479 | 2.2920 | 2.6928 | 3.1588 | 3.7000 | 4.3276 | 5.0545 | 5.8951 | 6 . 8660 | 7.9861 | 9.2765 | 10.7613 | 16.6722 | 22.1861 | 44.4089 |
| 18 | 1.1961 | 1.4282 | 1.7024 | 2.0258 | 2.4068 | 2.8543 | 3.3799 | 3.9960 | 4.7171 | 5.5599 | 6.5436 | 7.6900 | 9.0243 | 10.5752 | 12.3755 | 19.6733 | 26.6233 | 55.5112 |
| 19 | 1.2081 | 1.4568 | 1.7535 | 2.1068 | 2.5270 | 3.0256 | 3.6165 | 4.3157 | 5.1417 | 6.1159 | 7.2633 | 8.6128 | 10.1974 | 120557 | 14.2318 | 23.2144 | 31.9460 | 69.3889 |
| 20 | 1.2202 | 1.4859 | 1.8061 | 2.1911 | 2.6533 | 3.2071 | 3.8697 | 4.6610 | 5.6044 | 6.7275 | 8.0623 | 9.6443 | 11.5231 | 13.7435 | 16.3665 | 27.3930 | 38.3376 | 86.7362 |
| 21 | 1.2324 | 1.5157 | 1.8603 | 2.2788 | 2.7860 | 3.3996 | 4.1406 | 5.0338 | 6.1088 | 7.4002 | 8.9492 | 10.8038 | 13.0211 | 15.6676 | 18.8215 | 32.3238 | 46.0051 | 108.4202 |
| 22 | 1.2447 | 1.5460 | 1.9161 | 2.3699 | 2.9253 | 3.6035 | 4.4304 | 5.4365 | 6.6586 | 8.1403 | 9.9336 | 12.1003 | 14.7138 | 17.8610 | 21.6447 | 38.1421 | 55.2061 | 135.5253 |
| 23 | 1.2572 | 1.5769 | 1.9736 | 2.4647 | 3.0715 | 3.8197 | 4.7405 | 5.8715 | 7.2579 | 8.9543 | 11.0263 | 13.5523 | 16.6266 | 20.3616 | 24.8915 | 45.0076 | 66.2474 | 169.4066 |
| 24 | 1.2697 | 1.6084 | 2.0328 | 2.5633 | 3.2251 | 4.0489 | 5.0724 | 6.3412 | 7.9111 | 9.8497 | 12.2392 | 15.1786 | 18.7881 | 23.2122 | 28.6252 | 53.1090 | 79.4968 | 211.7582 |
| 25 | 1.2824 | 1.6406 | 2.0938 | 2.6658 | 3.3864 | 4.2919 | 5.4274 | 6.8485 | 8.6231 | 10.8347 | 13.5855 | 17.0001 | 21.2305 | 26.4619 | 32.9190 | 62.6686 | 95.3962 | 264.6978 |
| 30 | 1.3478 | 1.8114 | 2.4273 | 3.2434 | 4.3219 | 5.7435 | 7.6123 | 10.0627 | 13.2677 | 17.4494 | 22.8923 | 29.9599 | 39.1159 | 50.9502 | 66.2118 | 14.3.3706 | 237.3763 | 807.7936 |
| 35 | 1.4166 | 1.9999 | 2.8139 | 3.9461 | 5.5160 | 7.6861 | 10.6766 | 14.7853 | 20.4140 | 28.1024 | 38.5749 | 52.7996 | 72.0685 | 98.1002 | 133.1755 | 327.9973 | 590.6682 | 2465.1903 |
| 40 | 1.4889 | 2.2080 | 3.2620 | 4.8010 | 7.0400 | 10.2857 | 14.9745 | 21.7245 | 31.4094 | 45.2593 | 65.0009 | 93.0510 | 132.7816 | 188.8835 | 267.8635 | 750.3783 | 1469.7716 | 7523.1638 |
| 45 | 1.5648 | 2.4379 | 3.7816 | 5.8412 | 8.9850 | 13.7646 | 21.0025 | 31.924 | 48.3273 | 72.8905 | 109.5302 | 163.9876 | 244.6414 | 363.6791 | 538.7693 | 1716.6839 | 3657.2620 | 22958.8740 |
| 50 | 1.6446 | 2.6916 | 4.3839 | 7.1067 | 11.4674 | 18.4202 | 29.4570 | 46.9016 | 74.3575 | 117.3909 | 184.5648 | 289.0022 | 450.7359 | 700.2330 | 1080.6574 | 3927.3569 | 9100.4382 | 70064.9232 |





The McGraw-Hill Companies

**Table B.2** *Tables for Investment Analysis*

$$DF = \frac{1}{(1+i)^T}$$

(Interest Rate in Percentages)

| *Period* | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* | *10* | *11* | *12* | *13* | *14* | *15* | *18* | *20* | *25* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.9901 | 0.9804 | 0.9709 | 0.9615 | 0.9524 | 0.9434 | 0.9346 | 0.9259 | 0.9174 | 0.9091 | 0.9009 | 0.8929 | 0.8850 | 0.8772 | 0.8696 | 0.8475 | 0.8333 | 0.8000 |
| 2 | 0.9803 | 0.9612 | 0.9426 | 0.9246 | 0.9070 | 0.8900 | 0.8734 | 0.8573 | 0.8417 | 0.8264 | 0.8116 | 0.7972 | 0.7831 | 0.7695 | 0.7561 | 0.7182 | 0.6944 | 0.6400 |
| 3 | 0.9706 | 0.9423 | 0.9151 | 0.8890 | 0.8638 | 0.8396 | 0.8163 | 0.7938 | 0.7722 | 0.7513 | 0.7312 | 0.7118 | 0.6931 | 0.6750 | 0.6575 | 0.6086 | 0.5787 | 0.5120 |
| 4 | 0.9610 | 0.9238 | 0 8885 | 0.8548 | 0.8227 | 0.7921 | 0.7629 | 0.7350 | 0.7084 | 0.6830 | 0.6587 | 0.6355 | 0.6133 | 0.5921 | 0.5718 | 0.5158 | 0.4823 | 0.4096 |
| 5 | 0.9515 | 0.9057 | 0.8626 | 0.8219 | 0.7835 | 0.7473 | 0.7130 | 0.6806 | 0.6499 | 0.6209 | 0.5935 | 0.5674 | 0.5428 | 0.5194 | 0.4972 | 0.4371 | 0.4019 | 0.3277 |
| 6 | 0.9420 | 0.8880 | 0.8375 | 0.7903 | 0.7462 | 0.7050 | 0.6663 | 0.6302 | 0.5963 | 0.5645 | 0.5646 | 0.5066 | 0.4803 | 0.4556 | 0.4323 | 0.3704 | 0.3349 | 0.2621 |
| 7 | 0.9327 | 0.8706 | 0.8131 | 0.7599 | 0.7107 | 0.6651 | 0.6227 | 0.5835 | 0.5470 | 0.5132 | 0.4817 | 0.4523 | 0.4251 | 0.3996 | 0.3759 | 0.3139 | 0.2791 | 0.2097 |
| 8 | 0.9235 | 0.8535 | 0.7894 | 0.7307 | 0.6768 | 0.6274 | 0.5820 | 0.5403 | 0.5019 | 0.4665 | 0.4339 | 0.4039 | 0.3762 | 0.3506 | 0.3269 | 0.2660 | 0.2326 | 0.1678 |
| 9 | 0.9143 | 0.8368 | 0.7664 | 0.7026 | 0.6446 | 0.5919 | 0.5439 | 0.5002 | 0.4604 | 0.4241 | 0.3909 | 0.3606 | 0.3329 | 0.3075 | 0.2843 | 0.2255 | 0.1938 | 0.1342 |
| 10 | 0.9053 | 0.8203 | 0.7441 | 0.6756 | 0.6139 | 0.5584 | 0.5083 | 0.4632 | 0.4224 | 0.3855 | 0.3522 | 0.3220 | 0.2946 | 0.2697 | 0.2472 | 0.1911 | 0.1615 | 0.1074 |
| 11 | 0.8963 | 0.8043 | 0.7224 | 0.6496 | 0.5847 | 0.5268 | 0.4751 | 0.4289 | 0.3875 | 0.3505 | 0.3173 | 0.2875 | 0.2607 | 0.2366 | 0.2149 | 0.1619 | 0.1346 | 0.0859 |
| 12 | 0.8874 | 0.7885 | 0.7014 | 0.6246 | 0.5568 | 0.4970 | 0.4440 | 0.3971 | 0.3555 | 0.3186 | 0.2858 | 0.2567 | 0.2307 | 0.2076 | 0.1869 | 0.1372 | 0.1122 | 0.0687 |
| 13 | 0.8787 | 0.7730 | 0.6810 | 0.6006 | 0.5303 | 0.4688 | 0.4150 | 0.3677 | 0.3262 | 0.2897 | 0.2575 | 0.2242 | 0.2042 | 0.1821 | 0.1625 | 0.1163 | 0.0935 | 0.0550 |
| 14 | 0.8700 | 0.7579 | 0.6611 | 0.5775 | 0.5051 | 0.4423 | 0.3878 | 0.3405 | 0.2992 | 0.2633 | 0.2320 | 0.2046 | 0.1807 | 0.1597 | 0.1413 | 0.0985 | 0.0779 | 0.0440 |
| 15 | 0.8613 | 0.7430 | 0.6419 | 0.5553 | 0.4810 | 0.4173 | 0.3624 | 0.3152 | 0.2745 | 0.2394 | 0.2090 | 0.1827 | 0.1599 | 0.1401 | 0.1229 | 0.0835 | 0.0649 | 0.0352 |
| 16 | 0.8528 | 0.7284 | 0.6232 | 0.5339 | 0.4581 | 0.3936 | 0.3387 | 0.2919 | 0.2519 | 0.2176 | 0.1883 | 0.1631 | 0.1415 | 0.1229 | 0.1069 | 0.0708 | 0.0541 | 0.0281 |
| 17 | 0.8444 | 0.7142 | 0.6050 | 0.5134 | 0.4363 | 0.3714 | 0.3166 | 0.2703 | 0.2311 | 0.1978 | 0.1696 | 0.1456 | 0.1252 | 0.1078 | 0.0929 | 0.0600 | 0.0451 | 0.0225 |
| 18 | 0.8360 | 0.7002 | 0.5874 | 0.4936 | 0.4155 | 0.3503 | 0.2959 | 0.2502 | 0.2120 | 0.1799 | 0.1528 | 0.1300 | 0.1108 | 0.0946 | 0.0808 | 0.0508 | 0.0376 | 0.0180 |
| 19 | 0.8277 | 0.6864 | 0.5703 | 0.4746 | 0.3957 | 0.3305 | 0.2765 | 0.2317 | 0.1945 | 0.1635 | 0.1377 | 0.1161 | 0.0981 | 0.0829 | 0.0703 | 0.0431 | 0.0313 | 0.0144 |
| 20 | 0.8195 | 0.6730 | 0.5537 | 0.4564 | 0.3739 | 0.3118 | 0.2584 | 0.2145 | 0.1784 | 0.1486 | 0.1240 | 0.1037 | 0.0868 | 0.0728 | 0.0611 | 0.0365 | 0.0261 | 0.0115 |
| 21 | 0.8114 | 0.6598 | 0.5375 | 0.4388 | 0.3589 | 0.2942 | 0.2415 | 0.1987 | 0.1637 | 0.1351 | 0.1117 | 0.0926 | 0.0768 | 0.0638 | 0.0531 | 0.0309 | 0.0217 | 0.0092 |
| 22 | 0.8034 | 0.6468 | 0.5219 | 0.4220 | 0.3418 | 0.2775 | 0.2257 | 0.1839 | 0.1502 | 0.1228 | 0.1007 | 0.0826 | 0.0680 | 0.0560 | 0.0462 | 0.0262 | 0.0181 | 0.0074 |
| 23 | 0.7954 | 0.6342 | 0.5067 | 0.4057 | 0,;3256 | 0.2618 | 0.2109 | 0.1703 | 0.1378 | 0.1117 | 0.0907 | 0.0738 | 0.0601 | 0.0491 | 0.0402 | 0.0222 | 0.0151 | 0.0059 |
| 24 | 0.7876 | 0.6217 | 0.4919 | 0.3901 | 0.3101 | 0.2470 | 0.1971 | 0.1577 | 0.1264 | 0.1015 | 0.0817 | 0.0659 | 0.0532 | 0.0431 | 0.0349 | 0.0188 | 0.0126 | 0.0047 |
| 25 | 0.7798 | 0.6095 | 0.4776 | 0.3751 | 0.2953 | 0.2330 | 0.1842 | 0.1460 | 0.1160 | 0.0923 | 0.0736 | 0.0588 | 0.0471 | 0.0378 | 0.0304 | 0.0160 | 0.0105 | 0.0038 |
| 30 | 0.7419 | 0.5521 | 0.4120 | 0.3083 | 0.2314 | 0.1741 | 0.1314 | 0.0994 | 0.0754 | 0.0573 | 0.0473 | 0.0334 | 0.0256 | 0.0196 | 0.0151 | 0.0070 | 0.0042 | 0.0012 |
| 35 | 0.7059 | 0.5000 | 0.3554 | 0.2534 | 0.1813 | 0.1301 | 0.0937 | 0.0676 | 0.0490 | 0.0356 | 0.0259 | 0.0189 | 0.0139 | 0.0102 | 0.0075 | 0.0030 | 0.0017 | 0.0004 |
| 40 | 0.6717 | 0.4529 | 0 . 3066 | 0.2083 | 0.1420 | 0.0972 | 0.0668 | 0.0460 | 0.0318 | 0.0221 | 0.0154 | 0.0107 | 0.0075 | 0.0053 | 0.0037 | 0.0013 | 0.0007 | 0.0001 |
| 45 | 0.6391 | 0.4102 | 0.2644 | 0.1712 | 0.1113 | 0.0727 | 0.0476 | 0.0313 | 0.0207 | 0.0137 | 0.0091 | 0.0061 | 0.0041 | 0.0027 | 0.0019 | 0 . 0006 | 0.0003 | 0.0000 |
| 50 | 0.6080 | 0.3715 | 0.2281 | 0.1407 | 0.0872 | 0.0543 | 0.0339 | 0.0213 | 0.0134 | 0.0085 | 0.0054 | 0.0035 | 0.0022 | 0.0014 | 0.0009 | 0.0003 | 0.0001 | 0.0000 |

**Table B.3** *Tables for Investment Analysis*

$$FWAF = \frac{(1+i)^T - 1}{i}$$

(Interest Rate in Percentages)

| *Period* | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* | *10* | *11* | *12* | *13* | *14* | *15* | *18* | *20* | *25* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 2 | 2.0100 | 2.0200 | 2.0300 | 2.0400 | 2.0500 | 2.0600 | 2.0700 | 2.0800 | 2.0900 | 2.1000 | 2.1100 | 2.1200 | 2.1300 | 2.1400 | 2.1500 | 2.1800 | 2.2000 | 2.2500 |
| 3 | 3.0301 | 3.0604 | 3.0909 | 3.1216 | 3.1525 | 3.1836 | 3.2149 | 3.2464 | 3.2781 | 3.3100 | 3.3421 | 3.3744 | 3.4069 | 3.4396 | 3.4725 | 3.5724 | 3.6400 | 3.8125 |
| 4 | 4.0604 | 4.1216 | 4.1836 | 4.2465 | 4.3101 | 4.3746 | 4.4399 | 4.5061 | 4.5731 | 4.6410 | 4.7097 | 4.7793 | 4.8498 | 4.9211 | 4.9934 | 5.2154 | 5.3680 | 5.7656 |
| 5 | 5.1010 | 5.2040 | 5.3091 | 5.4163 | 5.5256 | 5.6371 | 5.7507 | 5.8666 | 5.9847 | 6.1051 | 6.2278 | 6.3528 | 6.4803 | 6.6101 | 6.7424 | 7.1542 | 7.4416 | 8.2070 |
| 6 | 6.1520 | 6.3081 | 6.4684 | 6.6330 | 6.8019 | 6.9753 | 7.1533 | 7.3359 | 7.5233 | 7.7156 | 7.9129 | 8.1152 | 8.3227 | 8.5355 | 8.7537 | 9.4420 | 9.9299 | 11.2588 |
| 7 | 7.2135 | 7.4343 | 7.6625 | 7.8983 | 8.1420 | 8.3938 | 8.6540 | 8.9228 | 9.2004 | 9.4872 | 9.7833 | 10.0890 | 10.4047 | 10.7305 | 11.0668 | 12.1415 | 12.9159 | 15.0735 |
| 8 | 8.2857 | 8.5830 | 8.8923 | 9.2142 | 9.5491 | 9.8975 | 10.2598 | 10.6366 | 11.0285 | 11.4359 | 11.8594 | 12.2997 | 12.7573 | 13.2328 | 13.7268 | 15.3270 | 16.4991 | 19.8419 |
| 9 | 9.3685 | 9.7546 | 10.1591 | 10.5828 | 11.0266 | 11.4913 | 11.9780 | 12.4876 | 13.0210 | 13.5795 | 14.1640 | 14.7757 | 15.4157 | 16.0853 | 16.7858 | 19.0859 | 20.7989 | 25.8023 |
| 10 | 10.4622 | 10.9497 | 11.4639 | 12.0061 | 12.5779 | 13.1808 | 13.8164 | 14.4866 | 15.1929 | ‘15.9374 | 16.7220 | 17.5487 | 18.4197 | 19.3373 | 20.3037 | 23.5213 | 25.9587 | 33.2529 |
| 11 | 11.5668 | 12.1687 | 12.8078 | 13.4864 | 1 4.2068 | 14.9716 | 15.7836 | 16.6455 | 17.5603 | 18.5312 | 19.5614 | 20.6546 | 21.8143 | 23.0445 | 24.3493 | 28.7551 | 32.1504 | 42.5661 |
| 12 | 12.6825 | 13.4121 | 14.1920 | 15.0258 | 15.9171 | 16.8699 | 17.8885 | 18.9771 | 20.1407 | 21.3843 | 22.7132 | 24.1331 | 25.6502 | 27.2702 | 29.0017 | 34.9311 | 39.5805 | 54.2077 |
| 13 | 13.8093 | 14.6803 | 15.6178 | 16.6268 | 17.7130 | 18.8821 | 20.1406 | 21.4953 | 22.9534 | 24.5227 | 26.2116 | 28.0291 | 29.9847 | 32.0887 | 34.3519 | 42.2187 | 48.4966 | 68.7596 |
| 14 | 14.9474 | 15.9739 | 17.0863 | 18.2919 | 19.5986 | 21.0151 | 22.5505 | 24.2149 | 26.0192 | 27.9750 | 30.0949 | 32.3926 | 34.8827 | 37.5811 | 40.5047 | 50.8180 | 59.1959 | 86.9495 |
| 15 | 16.0969 | 17.2934 | 18.5989 | 20.0236 | 21.5786 | 23.2760 | 25.1290 | 27.1521 | 29.3609 | 31.7725 | 34.4054 | 37.2797 | 40.4175 | 43.8424 | 47.5804 | 60.9653 | 72.0351 | 109.6868 |
| 16 | 17.2579 | 18.6393 | 20.1569 | 21.8245 | 23.6575 | 25.6725 | 27.8881 | 30.3243 | 33.0034 | 35.9497 | 39.1899 | 42.7533 | 46.6717 | 50.9804 | 55.7175 | 72.9390 | 87.4421 | 138.1085 |
| 17 | 18.4304 | 20.0121 | 21.7616 | 23.6975 | 25.8404 | 28.2129 | 30.8402 | 33.7502 | 36.9737 | 40.5447 | 44.5008 | 48.8837 | 53.7391 | 59.1176 | 65.0751 | 87.0680 | 105.9306 | 173.6357 |
| 18 | 19.6147 | 21.4123 | 23.4144 | 25.8454 | 28.1324 | 30.9057 | 33.9990 | 37.4502 | 41.3013 | 45.5992 | 50.3959 | 55.7497 | 61.7251 | 68.3941 | 75.8384 | 103.7403 | 128.1167 | 218.0446 |
| 19 | 20.8109 | 22.8406 | 25.1169 | 27.6712 | 30.5390 | 33.7600 | 37.3790 | 41.4463 | 46.0185 | 51.1591 | 56.9395 | 63.4397 | 70.7494 | 78.9692 | 88.2118 | 123.4135 | 154.7400 | 273.5558 |
| 20 | 22.0190 | 24.2974 | 26.8704 | 29.7781 | 33.0660 | 36.7856 | 40.9955 | 45.7620 | 51.1601 | 57.2750 | 64.2028 | 72.0524 | 80.9468 | 91.0249 | 102.4436 | 146.6280 | 186.6880 | 342.9447 |
| 21 | 23.2392 | 25.7833 | 28.6765 | 31.9692 | 35.7193 | 39.9927 | 44.8652 | 50.4229 | 56.7645 | 64.0025 | 72.2651 | 81.6987 | 92.4699 | 104.7684 | 118.8101 | 174.0210 | 225.0265 | 429.6809 |
| 22 | 24.4716 | 27.2990 | 30.5368 | 34.2480 | 38.5052 | 43.3923 | 49.0057 | 55.4588 | 62.8733 | 71.4027 | 81.2143 | 92.5026 | 105.4910 | 120.4360 | 137.6316 | 206.3448 | 271.0307 | 538.1011 |
| 23 | 25.7163 | 28.8450 | 32.4529 | 36.6179 | 41.4305 | 46.9958 | 53.4361 | 60.8933 | 69.5319 | 79.5430 | 91.1479 | 104.6029 | 120.2048 | 138.2970 | 159.2764 | 244.4868 | 326.2369 | 673.6264 |
| 24 | 26.9735 | 30.4219 | 34.4265 | 39.0826 | 44.5020 | 50.8156 | 58.1767 | 66.7648 | 76.7898 | 88.4973 | 102.1742 | 118.1552 | 136.8315 | 158.6586 | 184.1678 | 289.4945 | 392.4842 | 843.0329 |
| 25 | 28.2432 | 32.0303 | 36.4593 | 41.6159 | 47.7271 | 58.8645 | 63.2490 | 73.1059 | 87.7009 | 98.3471 | 114.4133 | 133.3339 | 155.6196 | 181.8708 | 212.7930 | 342.6035 | 471.9811 | 1054.7912 |
| 30 | 34.7849 | 40.5681 | 47.5754 | 56.0849 | 66.4388 | 79.0582 | 94.4608 | 113.2832 | 136.3075 | 164.4940 | 199.0209 | 241.3327 | 293.1992 | 356.7868 | 434.7451 | 790.9480 | 1181.8816 | 3227.1743 |
| 35 | 41.6603 | 49.9945 | 60.4621 | 73.6522 | 90.3203 | 111.4348 | 138.2369 | 172.3168 | 215.7108 | 271.0244 | 341.5896 | 431.6635 | 546.6808 | 693.5727 | 881.1702 | 1816.6516 | 2948.3411 | 9856.7613 |
| 40 | 48.8864 | 60.4020 | 75.4013 | 95.0255 | 120.7998 | 154.7620 | 199. 6351 | 259.0565 | 337.8824 | 442.5926 | 581.8261 | 767.0914 | 1013.7042 | 1342.0251 | 1779.0903 | 4163.2130 | 7343.8578 | |
| 45 | 56.4811 | 71.8927 | 92.7199 | 121.0294 | 159.7002 | 212.7435 | 285.7493 | 386.5056 | 525.8587 | 718.9048 | 986.6386 | 1358.2300 | 1874.1646 | 2590.5648 | 3585.1285 | 9531.5771 | | |
| 50 | 64.4632 | 84.5794 | 112.7969 | 152.6671 | 209.3480 | 290.3359 | 406.5289 | 573.7702 | 815.0836 | | | | | | | | | |



The McGraw-Hill Companies

**Table B.4** *Tables for Investment Analysis*

$$PWAF = \frac{1 - \frac{1}{(1+i)^T}}{i}$$

(Interest Rate in Percentages)

| *Period* | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* | *10* | *11* | *12* | *13* | *14* | *15* | *18* | *20* | *25* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.9901 | 0.9804 | 0.9709 | 0.9615 | 0.9524 | 0.9434 | 0.9346 | 0.9259 | 0.9174 | 0.9091 | 0.9009 | 0.8929 | 0.8850 | 0.8772 | 0.8696 | 0.8475 | 0.8333 | 0.8000 |
| 2 | 1.9704 | 1.9416 | 1.9135 | 1.8861 | 1.8594 | 1.8334 | 1.8080 | 1.7833 | 1.7591 | 1.7355 | 1.7125 | 1.6901 | 1.6681 | 1.6467 | 1.6257 | 1.5656 | 1.5278 | 1.4400 |
| 3 | 2.9410 | 2.8839 | 2.8286 | 2.7751 | 2.7232 | 2.673O | 2.6243 | 2.5771 | 2.5313 | 2.4869 | 2.4437 | 2.4018 | 2.3612 | 2.3216 | 2.2832 | 2.1743 | 2.1065 | 1.9520 |
| 4 | 3.9020 | 3.8077 | 3.7171 | 3.6299 | 3.5460 | 3.4651 | 3.3872 | 3.3121 | 3.2397 | 3.1699 | 3.1024 | 3.0373 | 2.9745 | 2.9137 | 2.8550 | 2.6901 | 2.5887 | 2.3616 |
| 5 | 4.8534 | 4.7135 | 4.5797 | 4.4518 | 4.3295 | 4.2124 | 4.1002 | 3.9927 | 3.8897 | 3.7908 | 3.6959 | 3.6048 | 3.5172 | 3.4331 | 3.3522 | 3.1272 | 2.9906 | 2.6893 |
| 6 | 5.7955 | 5.6014 | 5.4172 | 5.2421 | 5.0757 | 4.9173 | 4.7665 | 4.6229 | 4.4859 | 4.3553 | 4.2305 | 4.1114 | 3.9975 | 3.8887 | 3.7845 | 3.4976 | 3.3255 | 2.9514 |
| 7 | 6.7282 | 6.4720 | 6.2303 | 6.0021 | 5.7864 | 5.5824 | 5.3893 | 5.2064 | 5.0330 | 4.8684 | 4.7122 | 4.5638 | 4.4226 | 4.2883 | 4.1604 | 3.8115 | 3.6046 | 3.1611 |
| 8 | 1.6517 | 7.3255 | 7.0197 | 6.7327 | 6.4632 | 6.2098 | 5.9713 | 5.7466 | 5.5348 | 5.3349 | 5.1461 | 4.9676 | 4.7988 | 4.6389 | 4.4873 | 4.0776 | 3.8372 | 3.3289 |
| 9 | 8.5660 | 8.1622 | 7.7861 | 7.4353 | 7.4078 | 6.8017 | 6.5152 | 6.2469 | 5.9952 | 5.7590 | 5.5370 | 5.3282 | 5.1317 | 4.9464 | 4.1716 | 4.3030 | 4.0310 | 3.4631 |
| 10 | 9.4713 | 8.9826 | 8.5302 | 8.1109 | 7.7217 | 7.3601 | 7.0236 | 6.7101 | 6.4177 | 6.1446 | 5.8892 | 5.6502 | 5.4262 | 5.2161 | 5.0188 | 4.4941 | 4.1925 | 3.5705 |
| 11 | 10.3676 | 9.7868 | 9.2526 | 8.7605 | 8.3064 | 7.8869 | 7.4987 | 7.1390 | 6.8052 | 6.4951 | 6.2065 | 5.9377 | 5.6869 | 5.4527 | 5.2337 | 4.6560 | 4.3271 | 3.6564 |
| 12 | 11.2551 | 10.5753 | 9.9540 | 9.3851 | 8.8633 | 8.3838 | 7.9427 | 7.5361 | 7.1607 | 6.8137 | 6.4924 | 6.1944 | 5.9176 | 5.6603 | 5.4206 | 4.7932 | 4.4392 | 3.1251 |
| 13 | 12.1337 | 11.3484 | 10.6350 | 9.9856 | 9.3936 | 8.8527 | 8.3577 | 7.9038 | 7.4869 | 7.1034 | 6.7499 | 6.4235 | 6.1218 | 5.8424 | 55831 | 4 . 0095 | 4.5327 | 3.7801 |
| 14 | 13.0037 | 12.1062 | 11.2961 | 10.5631 | 9.8986 | 9. 2950 | 8.7455 | 8.2442 | 7.7862 | 7.3667 | 6.9819 | 6.6282 | 6.3025 | 6.0021 | 5.7245 | 5.0081 | 4.6106 | 3.8241 |
| 15 | 13.8651 | 12.8493 | 11.9379 | 11.1184 | 10.3797 | 9.7122 | 9.1079 | 8.5595 | 8.0607 | 7.6061 | 7.1909 | 6.8109 | 6.4624 | 6.1422 | 5.8474 | 5.0916 | 4.6755 | 3.8593 |
| 16 | 14.7179 | 13.5777 | 12.5611 | 11.6523 | 10.8378 | 10.1059 | 9.4466 | 8.8514 | 8.3126 | 7.8237 | 7.3792 | 6.9740 | 6.6039 | 6.2651 | 5.9542 | 5.1624 | 4.7296 | 3.8874 |
| 17 | 15.5623 | 14.2919 | 13.1661 | 12.1657' | 11.2741 | 10.4773 | 9.7632 | 9.1216 | 8.5436 | 8.0216 | 7.5488 | 7.1196 | 6.7291 | 6.3729 | 6.0472 | 5.2223 | 4.7746 | 3.9099 |
| 18 | 16.3983 | 14.9920 | 13.7535 | 12.6593 | 11.6896 | 10.8276 | 10.0591 | 9.3719 | 8.7556 | 8.2014 | 7.7016 | 7.2497 | 6.8399 | 6.4674 | 6.1280 | 5.2732 | 4.8122 | 3.9279 |
| 19 | 17.2260 | 15.6785 | 14.3238 | 13.1339 | 12.0853 | 11.1581 | 10.3356 | 9.6036 | 8.9501 | 8.3649 | 7.8393 | 7.3658 | 6.9380 | 6.5504 | 6.1982 | 5.3162 | 4.8435 | 3.9424 |
| 20 | 18.0456 | 16.3514 | 14.8775 | 13.5903 | 12.4622 | 11.4699 | 10.5940 | 9.8181 | 9.1285 | 8.5136 | 7.9633 | 7.4694 | 7.0248 | 6.6231 | 6.2593 | 5.3527 | 4.8696 | 3.9539 |
| 21 | 18.8570 | 17.0112 | 15.4150 | 14.0292 | 12.8212 | 11.7641 | 10.8355 | 10.0168 | 9.2922 | 8.6487 | 8.0751 | 7.5620 | 7.1016 | 6.6870 | 6.3125 | 5.3837 | 4.8913 | 3.9631 |
| 22 | 15.6604 | 17.6580 | 15.9369 | 14.4511 | 13.1630 | 12.0416 | 11.0612 | 10.2007 | 9.4424 | 8.7715 | 8.1757 | 7.6446 | 7.1695 | 6.7429 | 6.3587 | 5.4099 | 4.9094 | 3.9705 |
| 23 | 20.4558 | 18.2922 | 16.4436 | 14.8568 | 13.4886 | 12.3034 | 11.2722 | 10.3711 | 9.5802 | 8.8832 | 8.2664 | 7.7184 | 7.2297 | 6.7921 | 6.3988 | 5.4321 | 4.9245 | 3.9764 |
| 24 | 21.2434 | 18.9139 | 16.9355 | 15.2470 | 13.7986 | 12.5504 | 11.4693 | 10.5288 | 9.7066 | 8.9847 | 8.3481 | 7.7843 | 7.2829 | 6.8351 | 6.4338 | 5.4509 | 4.9371 | 3.9811 |
| 25 | 22.0232 | 19.5235 | 17.4131 | 15.6221 | 14.0939 | 12.7834 | 11.6536 | 10.6748 | 9.8226 | 9.0770 | 8.4211 | 7.8431 | 7.3300 | 6.8729 | 6.4641 | 5.4669 | 4.9476 | 3.9849 |
| 30 | 25.8077 | 22.3965 | 19.6004 | 17.2920 | 15.3725 | 13.7648 | 12.4090 | 11.2578 | 10.2737 | 9.4269 | 8.6938 | 8.0552 | 7.4957 | 7.0027 | 6.5660 | 5.5168 | 4.9789 | 3.9950 |
| 35 | 29.4086 | 24.9986 | 21.4872 | 18.6646 | 16.3742 | 14.4982 | 12.9477 | 11.6546 | 10.5668 | 9.6442 | 8.8552 | 8.1755 | 7.5856 | 7.0700 | 6.6166 | 5.5386 | 4.9915 | 3.9984 |
| 40 | 32.8347 | 27.3555 | 23.1148 | 19.7928 | 17.1591 | 15.0463 | 13.3317 | 11.9246 | 10.7574 | 9.7791 | 8.9511 | 8.2438 | 7.6344 | 7.1050 | 6.6418 | 5.5482 | 4.9966 | 3.9995 |
| 45 | 36.0945 | 29.4902 | 24.5187 | 20.7200 | 17.7741 | 15.4558 | 13.6055 | 12.1084 | 10.8812 | 9.8628 | 9.0079 | 8.2825 | 7.6609 | 7.1232 | 6.6543 | 5.5523 | 4.9966 | 3.9998 |
| 50 | 39.1961 | 31.4236 | 25.7296 | 21.4822 | 18.2559 | 15.7619 | 13.8007 | 12.2335 | 10.9617 | 9.9148 | 9.0417 | 8.3045 | 7.6752 | 7.1327 | 6.6605 | 5.5541 | 4.9995 | 3.9999 |

**Table B.5** *Tables for Investment Analysis*

$$SFF = \frac{i}{(1+i)^T - 1}$$

(Interest Rate in Percentages)

| *Period* | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* | *10* | *11* | *12* | *13* | *14* | *15* | *18* | *20* | *25* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 2 | 0.4975 | 0.4950 | 0.4926 | 0.4902 | 0.4878 | 0.4854 | 4.4831 | 0.4808 | 0.4785 | 0.4762 | 0.4739 | 0.4717 | 0.4695 | 0.4673 | 0.4651 | 0.4587 | 0.4545 | 0.4444 |
| 3 | 0.3300 | 0.3268 | 0.3235 | 0.3203 | 0.3172 | 0.3141 | 0.3111 | 0.3080 | 0.3051 | 0.3021 | 0.2992 | 0.2963 | 0.2935 | 0.2907 | 0.2880 | 0.2799 | 0.2747 | 0.2623 |
| 4 | 0.2463 | 0.2426 | 0.2390 | 0.2355 | 0.2320 | 0.2286 | 0.2252 | 0.2219 | 0.2187 | 0.2155 | 0.2123 | 0.2092 | 0.2062 | 0.2032 | 0.2003 | 0.1917 | 0.1863 | 0.1734 |
| 5 | 0.1960 | 0.1922 | 0.1884 | 0.1846 | 0.1810 | 0.1774 | 0.1739 | 0.1705 | 0.1671 | 0.1638 | 0.1606 | 0.1574 | 0.1543 | 0.1513 | 0.1483 | 0.1398 | 0.1344 | 0.1218 |
| 6 | 0.1625 | 0.1585 | 0.1546 | 0.1508 | 0.1470 | 0.1434 | 0.1398 | 0.1363 | 0.1329 | 0.1296 | 0.1264 | 0.1232 | 0.1202 | 0.1172 | 0.1142 | 0.1059 | 0.1007 | 0.0888 |
| 7 | 0.1386 | 0.1345 | 0.1305 | 0.1266 | 0.1228 | 0.1191 | 0.1156 | 0.1121 | 0.1087 | 0.1054 | 0.1022 | 0.0991 | 0.0961 | 0.0932 | 0.0904 | 0.0824 | 0.0774 | 0.0663 |
| 8 | 0.1207 | 0.1165 | 0.1125 | 0.1085 | 0.1047 | 0.1010 | 0.0975 | 0.0940 | 0.0907 | 0.0874 | 0.0843 | 0.0813 | 0.0784 | 0.0756 | 0.0729 | 0.0652 | 0.0606 | 0.0504 |
| 9 | 0.1067 | 0.1025 | 0.0984 | 0.0945 | 0.0907 | 0.0870 | 0.0837 | 0.0801 | 0.0768 | 0.0736 | 0.0706 | 0.0677 | 0.0647 | 0.0622 | 0.0596 | 0.0524 | ‘0.0481 | 0.0388 |
| 10 | 0.0956 | 0.0913 | 0.0872 | 0.0833 | 0.0795 | 0.0759 | 0.0724 | 0.0690 | 0.0658 | 0.0627 | 0.0598 | 0.0570 | 0.0543 | 0.0517 | 0.0493 | 0.0425 | 0 ,0385 | 0.0301 |
| 11 | 0.0865 | 0.0822 | 0.0781 | 0.0741 | 0.0704 | 0.0668 | 0.0634 | 0.0601 | 0.0569 | 0.0540 | 0.0511 | 0.0484 | 0.0458 | 0.0434 | 0.0411 | 0.0348 | 0.0311 | 0.0235 |
| 12 | 0.0788 | 0.0746 | 0.0705 | 0.0666 | 0.0628 | 0.0593 | 0.0559 | 0.0527 | 0.0497 | 0.0468 | 0.0440 | 0.0414 | 0.0390 | 0.0367 | 0.0345 | 0.0286 | 0.0253 | 0.0184 |
| 13 | 0.0724 | 0.0681 | 0.0640 | 0.0601 | 0.0565 | 0.0530 | 0.0497 | 0.0465 | 0.0436 | 0.0408 | 0.0382 | 0.0357 | 0.0334 | 0.0312 | 0.0291 | 0.0237 | 0.0206 | 0.0145 |
| 14 | 0.0669 | 0.0626 | 0.0585 | 0.0547 | 0.0510 | 0.0476 | 0.0443 | 0.0413 | 0.0384 | 0.0357 | 0.0332 | 0.0309 | 0.0287 | 0.0266 | 0.0247 | 0.0197 | 0.0169 | 0.0115 |
| 15 | 0.0621 | 0.0578 | 0.0538 | 0.0499 | 0.0463 | 0.0430 | 0.0398 | 0.0368 | 0.0341 | 0.0315 | 0.0291 | 0.0268 | 0.0247 | 0.0228 | 0.0210 | 0.0164 | 0.0139 | 0.0091 |
| 16 | 0.0579 | 0.0537 | 0.0496 | 0.0458 | 0.0423 | 0.0390 | 0.0359 | 0.0330 | 0.0303 | 0.0278 | 0.0255 | 0.0234 | 0.0214 | 0.0196 | 0.0179 | 0.0137 | 0.0114 | 0.0072 |
| 17 | 0.0543 | 0.0500 | 0.0460 | 0.0422 | 0.0387 | 0.0354 | 0.0324 | 0.0296 | 0.0270 | 0.0247 | 0.0225 | 0.0205 | 0.0186 | 0.0169 | 0.0154 | 0.0115 | 0.0094 | 0.0058 |
| 18 | 0.0510 | 0.0467 | 0.0427 | 0.0390 | 0.0355 | 0.0324 | 0.0294 | 0.0267 | 0.0242 | 0.0219 | 0.0198 | 0.0179 | 0.0162 | 0.0146 | 0.0132 | 0.0096 | 0.0078 | 0.0046 |
| 19 | 0.0481 | 0.0438 | 0.0398 | 0.0361 | 0.0327 | 0.0296 | 0.0268 | 0.0241 | 0.0217 | 0.0195 | 0.0176 | 0.0158 | 0.0141 | 0.0127 | 0.0113 | 0.0081 | 0.0065 | 0.0037 |
| 20 | 0.0454 | 0.0412 | 0.0372 | 0.0336 | 0.0302’ | 0.02:2 | 0.0244 | 0.0219 | 0.0195 | 0.0175 | 0.0156 | 0.0139 | 0.0124 | 0.0110 | 0.0098 | 0.0068 | 0.0054 | 0.0029 |
| 21 | 0.0430 | 0.0388 | 0.0349 | 0.0313 | 0.0280 | 0.0250 | 0.0223 | 0.0198 | 0.0176 | 0.0156 | 0.0138 | 0.0122 | 0.0108 | 0.0095 | 0.0084 | 0.0057 | 0,0044 | 0.0023 |
| 22 | 0.0409 | 0.0366 | 0.0327 | 0.0292 | 0.0260 | 0.0230 | 0.0204 | 0.0180 | 0.0159 | 0.0140 | 0.0123 | 0.0108 | 0.0095 | 0.0083 | 0.0073 | 0.0048 | 0.0037 | 0.0019 |
| 23 | 0.0389 | 0.0347 | 0.0308 | 0.0273 | 0.0241 | 0.0213 | 0.0187 | 0.0164 | 0.0144 | 0.0126 | 0.0110 | 0.0096 | 0.0083 | 0.0072 | 0.0063 | 0.0041 | 0.0031 | 0.0015 |
| 24 | 0.0371 | 0.0329 | 0.0290 | 0.0256 | 0.0225 | 0.0197 | 0.0172 | 0.0150 | 0.0130 | 0.0113 | 0.0098 | 0.0085 | 0.0073 | 0.0063 | 0.0054 | 0.0035 | 0.0025 | 0.0012 |
| 25 | 0.0354 | 0.0312 | 0.0274 | 0.0240, | 0.0210 | 0.0182 | 0.0158 | 0.0137 | 0.0118 | 0.0102 | 0.0087 | 0.0075 | 0.0064 | 0.0055 | 0.0047 | 0.0029 | 0.0021 | 0.0009 |
| 30 | 0.0287 | 0.0246 | 0.0210 | 0.0178 | 0.0151 | 0.0126 | 0.0106 | 0.0088 | 0.0073 | 0.0061 | 0.0050 | 0.0041 | 0.0034 | 0.0028 | 0.0023 | 0.0013 | 0.0008 | 0.0003 |
| 35 | 0.0240 | 0.0200 | 0.0165 | 0.0136 | 0.0111 | 0.0090 | 0.0072 | 0.0058 | 0.0046 | 0.0037 | 0.0029 | 0.0023 | 0.0018 | 0.0014 | 0.0011 | 0.0006 | 0.0003 | 0.0001 |
| 40 | 0.0205 | 0.0166 | 0.0133 | 0.0105 | 0.0083 | 0.0065 | 0.0050 | 0.0039 | 0.0030 | 0.0023 | 0.0017 | 0.0013 | 0.0010 | 0.0007 | 0.0006 | 0.0002 | 0.0001 | 0.0000 |
| 45 | 0.0177 | 0.0139 | 0.010.8 | 0.0083 | 0.0063 | 0.0047 | 0.0035 | 0.0026 | 0.0019 | 0.0014 | 0.0010 | 0.0007 | 0.0005 | 0.0004 | 0.0003 | 0.0001 | 0.0001 | 0.0000 |
| 50 | 0.0155 | 0.0118 | 0.0089 | 0.0066 | 0.0048 | 0.0034 | 0.0025 | 0.0017 | 0.0012 | 0.0009 | 0.0006 | 0 0004 | 0.0003 | 0.0002 | 0.0001 | 0.0000 | 0.0000 | 0 0000 |



The McGraw-Hill Companies

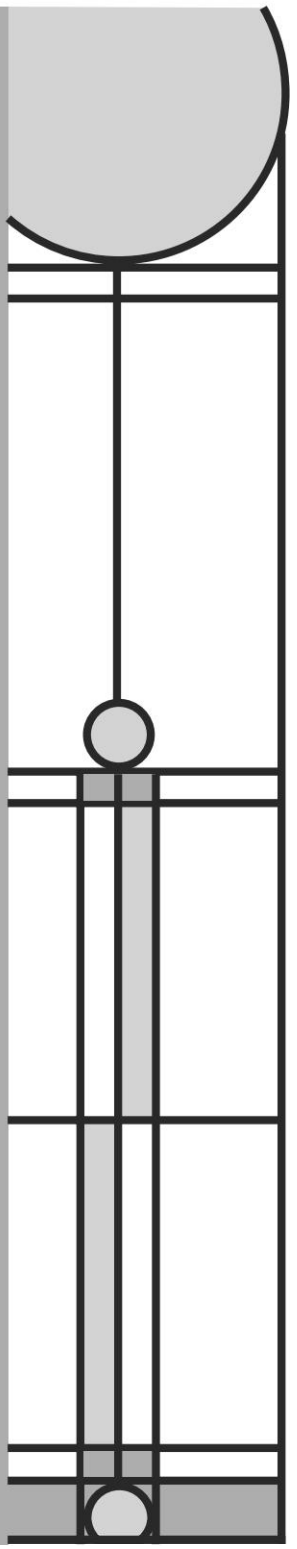
**Table B.6** *Tables for Investment Analysis*

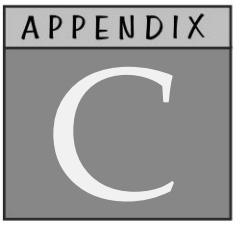
$$CRF = \frac{i}{1 - \frac{1}{(1+t)^T}}$$

(Interest Rate in Percentages)

| *Period* | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9* | *10* | *11* | *12* | *13* | *14* | *15* | *18* | *20* | *25* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.0100 | 1.0200 | 1.0300 | 1.0400 | 1.0500 | 1.0600 | 1.0700 | 1.0800 | 1.0900 | 1.1000 | 1.1100 | 1.1200 | 1.1300 | 1.1400 | 1.1500 | 1.1800 | 1.2000 | 1.2500 |
| 2 | 0.5075 | 0.5150 | 0.5226 | 0.5302 | 0.5378 | 0.5454 | 0.5531 | 0.5608 | 0.5685 | 0.5762 | 0.5839 | 0.5917 | 0.5995 | 0.6073 | 0.6151 | 0.6387 | 0.6545 | 0.6944 |
| 3 | 0.3400 | 0.3468 | 0.3535 | 0.3603 | 0.3672 | 0.3741 | 0.3811 | 0.3880 | 0.3951 | 0.4021 | 0.4092 | 0.4163 | 0.4235 | 0.4307 | 0.4380 | 0.4599 | 0.4747 | 0.5123 |
| 4 | 0.2563 | 0.2626 | 0.2690 | 0.2755 | 0.2820 | 0.2886 | 0.2952 | 0.3019 | 0.3087 | 0.3155 | 0.3223 | 0.3292 | 0.3362 | 0.3432 | 0.3503 | 0.3717 | 0.3863 | 0.4234 |
| 5 | 0.2060 | 0.2122 | 0.2184 | 0.2246 | 0.2310 | 0.2374 | 0.2439 | 0.2505 | 0.2571 | 0.2638 | 0.2706 | 0.2774 | 0.2843 | 0.2913 | 0.2983 | 0.3198 | 0.3344 | 0.3718 |
| 6 | 0.1725 | 0.1785 | 0.1846 | 0.1908 | 0.1970 | 0.2034 | 0.2098 | 0.2163 | 0.2229 | 0.2296 | 0.2364 | 0.2432 | 0.2502 | 0.2572 | 0.2642 | 0.285'9 | 0.3007 | 0.3388 |
| 7 | 0.1486 | 0.1545 | 0.1605 | 0.1666 | 0.1728 | 0.1791 | 0.1856 | 0.1921 | 0.1987 | 0.2054 | 0.2122 | 0.2191 | 0.2261 | 0.2332 | 0.2404 | 0.2624 | 0.2774 | 0.3163 |
| 8 | 0.1307 | 0.1365 | 0.1425 | 0.1485 | 0.1547 | 0.1610 | 0.1675 | 0.1740 | 0.1807 | 0.1874 | 0.1943 | 0.2013 | 0.2084 | 0.2156 | 0.2229 | 0.2452 | 0.2606 | 0.3004 |
| 9 | 0.1167 | 0.1225 | 0.1284 | 0.1345 | 0.1407 | 0.1470 | 0.1535 | 0.1601 | 0.1668 | 0.1736 | 0.1806 | 0.1877 | 0.1949 | 0.2022 | 0.2096 | 0.2324 | 0.2481 | 0.2888 |
| 10 | 0.1056 | 0.1113 | 0.1172 | 0.1233 | 0.1295 | 0.1359 | 0.1424 | 0.1490 | 0.1558 | 0.1627 | 0.1698 | 0.1770 | 0.1843 | 0.1917 | 0.1993 | 0.2225 | 0.2385 | 0.2801 |
| 11 | 0.0965 | 0.1022 | 0.1081 | 0.1141 | 0.1204 | 0.1268 | 0.1334 | 0.1401 | 0.1469 | 0.1540 | 0.1611 | 0.1684 | 0.1758 | 0.1934 | 0.1911 | 0.2148 | 0.2311 | 0.2735 |
| 12 | 0.0888 | 0.0946 | 0.1005 | 0.1066 | 0.1128 | 0.1193 | 0.1259 | 0.1327 | 0.1397 | 0.1468 | 0.1540 | 0.1614 | 0.1690 | 0.1767 | 0.1845 | 0.2086 | 0.2253 | 0.2684 |
| 13 | 0.0824 | 0.0881 | 0.0940 | 0.1001 | 0.1065 | 0.1130 | 0.1197 | 0.1265 | 0.1336 | 0.1408 | 0.1482 | 0.1557 | 0.1634 | 0.1712 | 0.1791 | 0.2037 | 0.2206 | 0.2645 |
| 14 | 0.0769 | 0.0826 | 0.0885 | 0.0947 | 0.1010 | 0.1076 | 0.1143 | 0.1213 | 0.1284 | 0.1357 | 0.1432 | 0.1509 | 0.1587 | 0.1666 | 0.1747 | 0.1997 | 0.2169 | 0.2615 |
| 15 | 0.0721 | 0.0778 | 0.0838 | 0.0899 | 0.0963 | 0.1030 | 0.1098 | 0.1168 | 0.1241 | 0.1315 | 0.1391 | 0.1468 | 0.1547 | 0.1628 | 0.1710 | 0.1964 | 0.2139 | 0.2591 |
| 16 | 0.0679 | 0.0737 | 0.0796 | 0.0858 | 0.0923 | 0.0990 | 0.1059 | 0.1130 | 0.1203 | 0.1278 | 0.1355 | 0.1434 | 0.1514 | 0.1596 | 0.1679 | 0.1937 | 0.2114 | 0.2572 |
| 17 | 0.0643 | 0.0700 | 0.0760 | 0.0822 | 0.0887 | 0.0954 | 0.1024 | 0.1096 | 0.1170 | 0.1247 | 0.1325 | 0.1405 | 0.1486 | 0.1569 | 0.1654 | 0.1915 | 0.2094 | 0.2558 |
| 18 | 0.0610 | 0.0667 | 0.0727 | 0.0790 | 0.0855 | 0.0924 | 0.0994 | 0.1067 | 0.1142 | 0.1219 | 0.1298 | 0.1379 | 0.1462 | 0.1546 | 0.1632 | 0.1896 | 0.2078 | 0.2546 |
| 19 | 0.0581 | 0.0638 | 0.0698 | 0.0761 | 0.0827 | 0.0896 | 0.0968 | 0.1041 | 0.1117 | 0.1195 | 0.1276 | 0.1358 | 0.1441 | 0.1527 | 0.1613 | 0.1881 | 0.2065 | 0.2537 |
| 20 | 0.0554 | 0.0612 | 0.0672 | 0.0736 | 0.0802 | 0.0872 | 0.0944 | 0.1019 | 0.1095 | 0.1175 | 0.1256 | 0.1339 | 0.1424 | 0.1510 | 0.1598 | 0.1868 | 0.2054 | 0.2529 |
| 21 | 0.0530 | 0.0588 | 0.0649 | 0.0713 | 0.0780 | 0.0850 | 0.0923 | 0.0998 | 0.1076 | 0.1156 | 0.1238 | 0.1322 | 0.1408 | 0.1495 | 0.1584 | 0.1857 | 0.2044 | 0.2523 |
| 22 | 0.0509 | 0.0566 | 0.0627 | 0.0692 | 0.0760 | 0.0830 | 0.0904 | 0.0980 | 0.1059 | 0.1140 | 0.1223 | 0.1308 | 0.1395 | 0.1483 | 0.1573 | 0.1848 | 0.2037 | 0.2519 |
| 23 | 0.0489 | 0.0547 | 0.0608 | 0.0673 | 0.0741 | 0.0813 | 0.0887 | 0.0964 | 0.1044 | 0.1126 | 0.1210 | 0.1296 | 0.1383 | 0.1472 | 0.1563 | 0.1841 | 0.2031 | 0.2515 |
| 24 | 0.0471 | 0.0529 | 0.0590 | 0.0656 | 0.0725 | 0.0797 | 0.0872 | 0.0950 | 0.1030 | 0.1113 | 0.1198 | 0.1285 | 0.1373 | 0.1463 | 0.1554 | 0.1835 | 0.2025 | 0.2512 |
| 25 | 0.0454 | 0.0512 | 0.0574 | 0.0640 | 0.0710 | 0.0782 | 0.0858 | 0.0937 | 0.1018 | 0.1102 | 0.1187 | 0.1275 | 0.1364 | 0.1455 | 0.1547 | 0.1829 | 0.2021 | 0.2509 |
| 30 | 0.0387 | 0.0446 | 0.0510 | 0.0578 | 0.0651 | 0.0726 | 0.0806 | 0.0888 | 0.0973 | 0.1061 | 0.1150 | 0.1241 | 0.1334 | 0.1428 | 0.1523 | 0.1813 | 0.2008 | 0.2503 |
| 35 | 0.0340 | 0.0400 | 0.0465 | 0.0536 | 0.0611 | 0.0690 | 0.0772 | 0.0858 | 0.0946 | 0.1037 | 0.1129 | 0.1223 | 0.1318 | 0.1414 | 0.1511 | 0.1806 | 0.2003 | 0.2501 |
| 40 | 0.0305 | 0.0366 | 0.0433 | 0.0505 | 0.0583 | 0.0665 | 0.0750 | 0.0839 | 0.0930 | 0.1023 | 0.1117 | 0.1213 | 0.1310 | 0.1407 | 0.1506 | 01802 | 0.2001 | 0.2500 |
| 45 | 0.0277 | 0.0339 | 0.0408 | 0.0483 | 0.0563 | 0.0647 | 0.0735 | 0.0826 | 0.0919 | 0.1014 | 0.1110 | 0.1207 | 0.1305 | .0.1404 | 0.1503 | 0.1801 | 0.2001 | 0.2500 |
| 50 | 0.0255 | 0.0318 | 0.0389 | 0.0466 | 0.0548 | 0.0634 | 0.0725 | 0.0817 | 0.0912 | 0.1009 | 0.1106 | 0.1204 | 0.1303 | 0.1402 | 0.1501 | 0.1800 | 0.2000 | 0.2500 |

The McGraw-Hill Companies






# MULTIPLE CHOICE QUESTIONS*

## CHAPTER 1

1. Economics
   (a) is the science of choice under scarcity.
   (b) is concerned with the wealth of nations, individuals and groups.
   (c) is the study of humankind in the ordinary business of life.
   (d) deals with cause-effect relationships.
   (e) All of the above.
2. The distinction between microeconomics and macroeconomics is due to
   (a) Ragnar Frisch
   (b) Alfred Marshall
   (c) Lionel Robbins
   (d) Adam Smith
3. Microeconomics does not deal with the
   (a) price and output of cars.
   (b) behaviour of individual buyers and sellers of cars.
   (c) excise duty on cars.
   (d) inflation in India.
4. Managerial economics does not use the concepts and techniques from
   (a) microeconomics.
   (b) statistics.
   (c) mathematics.
   (d) None of the above.

* Students must choose the BEST answer from the alternatives available in each question.

5. Economic profit is given by
   (a) total revenue minus total explicit cost.
   (b) total revenue minus total accounting cost.
   (c) total revenue minus economic cost.
   (d) total revenue minus total implicit cost.
6. Opportunity cost of a thing is defined as the
   (a) sacrifice one has to make to get that thing.
   (b) loss of the next best alternative while selecting that thing.
   (c) price paid for getting that thing.
   (d) All of the above.
7. Value of a firm varies negatively with
   (a) profits that the firm makes during its life-time.
   (b) discount rate.
   (c) growth rate in the firm's profit.
   (d) sales that the firm makes during its life-time.
8. Many firms aim at maximum sales subject to some target profit level because of
   (a) agency problem.
   (b) economies of mass production.
   (c) dichotomy between the owners and managers.
   (d) Both (a) and (c) above.
9. Economics does not aid firms in decisions like
   (a) What and how much to produce?
   (b) How to produce?
   (c) What should be its goal?
   (d) For whom to produce?
10. Mixed economic system prevails because
   (a) some goods are public goods.
   (b) under certain situations, Government rules could promote efficiency.
   (c) free market often leads to inequalities.
   (d) free market often promotes efficiency.
   (e) none of the above.
11. Which of the following statements is incorrect?
   (a) Economic principles assume that decision-makers act rationally.
   (b) Economics is concerned with benefits and costs.
   (c) All firms seek maximum possible profit.
   (d) Government is an inevitable part of any economy.

## CHAPTER 2

12. Demand is defined as the
   (a) schedule giving the quantities of a good that a consumer will buy at various prices.
   (b) schedule giving the maximum prices that a buyer will pay for various quantities of a good.
   (c) quantity of a good that a person buys at a given price.
   (d) All of the above.

13. Demand for car is a negative function of
    (a) national income
    (b) price of petrol/diesel
    (c) price of two-wheelers
    (d) expected growth rate of GDP
    (e) population.
14. Demand increases when
    (a) more is demanded at the same price.
    (b) same is demanded at a higher price.
    (c) more is purchased at a lower price.
    (d) Both (a) and (b) above.
    (e) All of the above.
15. Demand for MBA education is increasing over time not because
    (a) number of business schools have been increasing.
    (b) tuition fees for MBA education in real terms has been falling.
    (c) return on MBA education has a positive trend.
    (d) per capita income has been rising.
16. Market (Industry) demand curve is obtained through summation of individual consumers' demand curves
    (a) at each of various given prices.
    (b) at each of various given quantities.
    (c) horizontally i.e., across quantities.
    (d) Both (a) and (b) above.
    (e) Both (a) and (c) above.
17. Point elasticity is more appropriate than arc elasticity when the
    (a) change in the cause variable is small.
    (b) change in the cause variable is large.
    (c) change in the effect variable is small.
    (d) change in the effect variable is large.
18. The cross elasticity is useful to measure the
    (a) inter-dependence between the related goods.
    (b) degree of competition between the competing firms.
    (c) effectiveness of promotional activities of the firm.
    (d) Both (a) and (b) above.
19. The total revenue (TR) test of the price elasticity of demand (E) indicates that
    (a) if TR and price move in the same direction, demand is price elastic.
    (b) if TR and price move in the opposite direction, absolute value of $E < 1$.
    (c) if TR and price move in the opposite direction, absolute value of $E > 1$.
    (d) the product is an item of necessity.
20. If income elasticity of demand for a product is greater than unity, the product is
    (a) luxury
    (b) necessary

(c) giffen good
(d) inferior

21. State which of the following statements is correct?
(a) Demand curve represents the marginal utility curve when the latter is measured in rupees.
(b) Demand curve is necessarily linear.
(c) Price elasticity of demand is constant on a linear demand curve.
(d) Price elasticity of demand is greater than one, in absolute value, when marginal revenue is zero.

## CHAPTER 3

22. Which of the following is not a feature of indifference curves?
(a) Indifference curves slope downward from left to right.
(b) Slope of an indifference curve decreases as one move along it from left to right.
(c) Higher the indifference curve, the more is the satisfaction.
(d) Indifference curves are subjective.

23. An indifference curve is the
(a) locus of various combinations of two goods which yield the same satisfaction to the consumer.
(b) locus of those combinations of two inputs which yield the same output.
(c) locus of those combinations of two goods which costs the same amount of money.
(d) locus of those combinations of two goods which can be produced with the same resources.

24. When the price of the good on the X-axis falls, ceteris paribus,
(a) budget line shifts to the right.
(b) indifference curves shift to left.
(c) budget line rotates clock wise.
(d) budget line rotates anticlock wise.

25. Slope of the consumer's budget constraint is
(a) given by the ratio of prices of the two goods.
(b) a constant.
(c) given by the ratio of the marginal utilities of the two goods.
(d) positive.

26. The decision rule for the consumer under the indifference curve approach is that
(a) marginal utility per rupee must be equal across various goods.
(b) indifference curve must be tangent at the budget constraint.
(c) marginal utility must equal price for each good.
(d) Both (a) and (b) above.

27. Suppose Mohan's income of Rs. 5,000/week and the market prices of goods X and Y stand at Rs.10 and Rs. 5, respectively. Under the situation, the marginal rate of substitution of good X (horizontal axis) for good Y (vertical axis) at consumer's equilibrium point equals
(a) 2
(b) 0.5
(c) –0.5
(d) –2

28. Which of the following statements is incorrect?
    (a) The coordinates of the price consumption curve (PCC) describes the demand schedule/ curve for the product whose price is changing.
    (b) The coordinates of the income consumption curve (ICC) describes the Engel schedule/ curve for the product whose price is changing.
    (c) Indifference curves are convex to the origin because of the law of diminishing marginal rate of substitution.
    (d) Budget line describes the consumer's affordable bundle of goods.

## CHAPTER 4

29. An unspecified demand function is not adequate to determine the
    (a) direction of change in demand due to economic growth.
    (b) potential factors which affect the demand.
    (c) magnitude of the loss in sales due to inflation.
    (d) Both (a) and (b) above.
30. Equations are preferred to tables because
    (a) tables are finite.
    (b) tables are continuous.
    (c) tables are discrete.
    (d) Both (a) and (c) above.
31. Time Series data do not refer to the observations
    (a) on a given population over time arranged chronologically.
    (b) on GDP in 2009 across different countries.
    (c) on India's GDP during 1970 through 2009.
    (d) on Spain's sovereign debt during the decade 2000-09.
32. The demand for chicken in India function was estimated using the state-wise data for a year as follows:

    $D_C = 0.41 + 0.012\,Y - 0.013\,P_C + 0.004\,P_F$

    And average values of per capita consumption of chicken ($D_C$), per capita income (Y), price of chicken ($P_C$) and price of fish ($P_F$) were 0.08, 9.28, 44.23, and 26.02, respectively; where demand for chicken is measured as per capita consumption in kilograms, Y is measured in thousands of rupees and both $P_C$ and $P_F$ in price per kilogram. In this function, which of the following is not correct?
    (a) Law of demand holds good for chicken.
    (b) Fish is a substitute for chicken.
    (c) Fish is an inferior good.
    (d) Both (a) and (b).
33. For the function in Question number 32, which of the following statements is not true?
    (a) Average demand for fish equals 0.08 kilograms.
    (b) Cross elasticity equals 1.3.
    (c) Chicken demand equation is given by $D_C = 0.62 - 0.013\,P_C$.
    (d) None of the above.

34. Trend method of forecasting is appropriate when the variable under forecast
    (a) is stationary.
    (b) is seasonal.
    (c) is highly correlated with India's population.
    (d) is cyclical.
35. Single smoothing methods (simple as well as exponential) of forecasting are subject to a downward bias when the time series
    (a) contains a significant positive trend.
    (b) contains a significant negative trend.
    (c) is stationary.
    (d) is cyclical.
36. Regression method of forecasting could yield inaccurate forecasts due to
    (a) errors in the forecasts of explanatory variables.
    (b) inaccuracies in the estimates of parameters.
    (c) stochastic nature of the regression equation.
    (d) All of the above.
37. Regression method of forecasting is the most popular because
    (a) it is prescriptive.
    (b) it is descriptive/analytical.
    (c) requires the least data.
    (d) it always gives the best forecasts.
38. The accuracy of ex anti (genuine) forecasts could be best evaluated
    (a) through MAPE (mean absolute percentage error).
    (b) through $R^2$ value.
    (c) through a graph between the actual and forecasts values of the variable during the sample period.
    (d) None of the above.

## CHAPTER 5

39. Production function describes a relationship between
    (a) total revenue and factors of production.
    (b) total cost and inputs.
    (c) monetary value of output and physical quantities of factors of production.
    (d) physical quantity of output and physical quantities of factors of production.
40. Microeconomics distinction between short- and long-run is described as
    (a) less than one year is short run and more than one year is long run.
    (b) prices are rigid in short-run and flexible in the long-run.
    (c) short-run is the period when at least one of the factors of production (FOP) is fixed and at least one of the FOP is a variable, while in long-run all FOP are variable.
    (d) Both (a) and (b) above
41. The law of diminishing marginal returns states that
    (a) marginal physical product of labour declines as additional units of labour is employed with a fixed quantity of capital input.

(b) total physical product increases at a diminishing rate as more and more of the variable input is employed with a given units of the fixed input.
(c) average physical product of labour curve is above the marginal physical product curve of labour.
(d) All of the above.

42. Production function cannot be linear not because
(a) it is inconsistent with the law of diminishing marginal returns.
(b) it implies that production is possible even without any employment of labour.
(c) factors of production are perfect substitutes to each other.
(d) marginal physical product of labour depends positively on the quantity of capital employed.

43. If marginal physical product of labour ($MPP_L$) is increasing.
(a) Average physical product of labour ($APP_L$) > $MPP_L$.
(b) $APP_L < MPP_L$.
(c) $APP_L = MPP_L$.
(d) Total physical product of labour is at its peak.

44. Decision rule for "how to produce?" under perfect competition is not given by
(a) the ratio of marginal physical product to price must be equal for each factors of production.
(b) marginal rate of (technical) substitution of labour for capital must equal the ratio of the price of capital to the price of labour.
(c) marginal rate of (technical) substitution of labour for capital must equal the ratio of the price of labour to the price of capital.
(d) marginal revenue (value) product of labour must equal the nominal wage rate.

45. Total physical product of labour curve
(a) is a straight line passing through the origin.
(b) is non-linear such that it is initially convex and then concave to the labour axis.
(c) is non-linear such that it is initially concave and then convex to the labour axis.
(d) is the locus of those combinations of labour and capital which yield the same output.

46. Increasing returns to scale is the situation in which
(a) the expansion path is a straight line passing through the origin.
(b) output increases at a rate faster than the rate at which each of the various inputs increase.
(c) labour elasticity of output is greater than unity.
(d) the law of diminishing marginal rate of technical substitution holds.

47. The Cobb-Douglas production is inconsistent with the
(a) law of diminishing marginal returns.
(b) diminishing returns to scale.
(c) convexity of iso-quants.
(d) None of these.

48. An iso-quant represents
(a) a short-run production function.
(b) a given level of output that can be produced from alternative combinations of two inputs.
(c) a given budget that can purchase the alternative combinations of two inputs.
(d) a given level of utility that a consumer gets from alternative combinations of two goods.

49. Expansion path is the
    (a) locus of those combinations of two inputs which represent optimum input mix for various levels of output.
    (b) locus of those combinations of two inputs which give the same level of output.
    (c) locus of those combinations of two inputs which could be hired at the same total cost.
    (d) locus of those combinations of two goods which represent the optimum consumption basket under various levels of consumer's income.
50. A firm is operating where the $MPP_L = 20$ and $MPP_K = 10$, and the wage rate and capital rental rate are Rs. 100 and Rs. 80, respectively. Under the situation, the firm
    (a) is minimising total cost for a given output.
    (b) is maximising output for a given total cost.
    (c) is maximising profit.
    (d) should hire more labour and less capital.
51. Suppose a firm is facing the production function $Q = 25\ K^{0.7}L^{0.5}$, and wage and capital rental rates equal Rs. 100 and Rs. 80, respectively. Under the situation, the firm
    (a) is enjoying the increasing returns to scale.
    (b) is operating under a capacity (to produce) constraint.
    (c) is not subject to the law of diminishing marginal returns.
    (d) is not subject to the law of diminishing marginal rate of technical substitution.

## CHAPTER 6

52. Cost of production represents the
    (a) payments the firm makes to all the hired factors of production.
    (b) all the sacrifices that the firm make to produce its output.
    (c) firm's sunk cost plus the payments it makes to all the hired factors of production.
    (d) decrease in the book value of its all assets net of all liabilities.
53. Sunk cost is not the same as
    (a) cost that is lost for ever after it has been paid.
    (b) expenditure committed in the past and is now irreversible.
    (c) cost that is irrelevant in decision-making.
    (d) fixed cost.
54. Total cost of production varies inversely with
    (a) prices of factors of production.
    (b) size of output.
    (c) technical progress.
    (d) industrial unrest.
55. A linear total cost function implies
    (a) an U-shaped average cost curve.
    (b) a falling marginal cost curve.
    (c) economies of scale.
    (d) economies of scope.

56. Average total cost curve is U-shaped because of
    (a) economies of scope.
    (b) monotonically falling average fixed cost.
    (c) the law of so-called diminishing (variable) marginal returns.
    (d) Both (b) and (c) above.
57. The duality principle in production and cost analysis suggests that
    (a) if production function and input prices are known, cost function can be derived.
    (b) if input prices are fixed and labour is the only variable input, then average variable cost is inversely related with average physical productivity of labour.
    (c) if input prices are fixed, then increasing returns to scale implies economies of scale.
    (d) All of the above.
58. To reap economies of scope, the firm must
    (a) enhance its capacity in the same product line.
    (b) integrate vertically.
    (c) broaden its product line in the related products.
    (d) earn more experience.
59. To verify the validity of the U-shape of both the average cost and marginal cost curves, the total cost function must be hypothesized as a
    (a) linear function.
    (b) quadratic function.
    (c) double log (log linear) function.
    (d) cubic function.
60. Supply function of cars does not shift with a change in the
    (a) price of cars.
    (b) price of two-wheelers.
    (c) excise duty on cars.
    (d) wage rate.
61. Which of the following statements is incorrect?
    (a) Under perfect competition, short-run supply curve is that part of marginal cost curve which lies above the average variable cost curve.
    (b) Non-perfectly competitive industries have no supply curve.
    (c) Supply curve is a straight line passing through the origin.
    (d) Supply curve denotes the minimum prices that the firm/industry would charge for supplying the various quantities of its product.
62. Suppose a firm's total cost (TC) function is estimated as $TC = 25 - 2\,Q + 4\,Q^2$. Under the situation, which of the following statements in incorrect"
    (a) Firm's total fixed cost = 25.
    (b) Firm's marginal cost at output ($Q$) equals 2 is 14.
    (c) Firm's average variable cost function = $25/Q - 2 + 4\,Q$.
    (d) Firm's marginal cost function is linear.

## CHAPTER 7

63. Price of a product is not affected by the
    (a) demand for the product.
    (b) marginal cost of its production.
    (c) government regulations on the product.
    (d) None of the above.
64. Profit maximisation does not require
    (a) marginal revenue = marginal cost
    (b) marginal cost curve to intersect marginal revenue curve from below
    (c) price is not less than the shut-down price
    (d) average cost to be minimum possible
65. Intra-industry competition (market structure) does not depend on the
    (a) number of firms in the industry
    (b) degree of product differentiation in the industry
    (c) firms' objectives
    (d) entry/exit barriers in the industry
66. Which of the following is not a necessary condition of the oligopoly market?
    (a) significant entry/exist barriers.
    (b) product differentiation.
    (c) presence of at least one "big" firm.
    (d) None of the above.
67. Natural monopoly exists in which of the following industries?
    (a) Food grains.
    (b) Airports.
    (c) Aircrafts.
    (d) Business degrees (MBA, DBA, etc.)
68. Barriers to entry do not come from
    (a) sunk cost
    (b) legal matters
    (c) economic factors
    (d) deliberate obstacles from incumbents
69. Firms seek vertical integration in order to enjoy benefits of
    (a) economies of scale.
    (b) economies of scope.
    (c) economies of transactions cost.
    (d) experience curve.
70. The four industries have the following features. Which of them has the highest concentration?
    (a) Four-firm concentration ratio = 80 percent
    (b) Herfindahl–Hirschman index = 2,000
    (c) Industry has just four firms
    (d) Industry has no entry/exit barrier

71. Break-even analysis is useful to determine the
    (a) break-even output for a given price.
    (b) break-even price for a given output.
    (c) target profit output for a given price.
    (d) All of the above.
72. Government regulations on product prices exist because of its role as the
    (a) provider of public goods.
    (b) moderator of equity.
    (c) moderator of efficiency.
    (d) Both (a) and (b) above.
    (e) All of the above.

## CHAPTER 8

73. Under perfect competition, which of the following statements is incorrect?
    (a) Firm is a price-taker
    (b) Firm is an output-taker
    (c) Price is determined by invisible hands
    (d) P = marginal cost
74. Effective price ceilings do not cause
    (a) Dead weight loss
    (b) Shortage and black marketing
    (c) Glut/surplus
    (d) Fall in product price
75. The burden of the excise duty on cigarettes fall more on the smokers than on the manufacturers of cigarettes because the
    (a) price elasticity of demand for cigarettes is lower than the price elasticity of supply for cigarettes.
    (b) price elasticity of demand for cigarettes is higher than the price elasticity of supply for cigarettes.
    (c) cigarettes are injurious for health.
    (d) smoking is banned in public places.
76. The monopolist is not free to decide the
    (a) price or output of its product.
    (b) price and output of its product.
    (c) its research and development (R & D) budget.
    (d) whether to advertise or not to advertise its product.
77. Which of the following statements is incorrect under pure monopoly?
    (a) The firm enjoys market power
    (b) $P > MC$
    (c) The firm never incurs a loss
    (d) The firm always operates on the elastic part of the demand curve
78. The virtues of monopoly vis-à-vis perfect competition include
    (a) monopoly output is more than competitive output.

(b) monopoly leads to socially optimum outcome.
(c) monopolist always operates at the least average cost.
(d) monopolist is motivated towards innovation (R & D).

79. Natural monopoly is a special case of monopoly because under it the
(a) average cost curve slopes downward for all relevant output range.
(b) marginal cost curve slopes downward for all relevant output range.
(c) demand curve slopes downward monotonically.
(d) firm always makes economic profit.

80. The salient feature of monopolistic competition is
(a) falling demand curve.
(b) product differentiation.
(c) U-shaped average cost curve.
(d) excess capacity in the long-run.

81. The principal difference between pure monopoly and monopolistic competition is in terms of the
(a) degree of the market power.
(b) size of the price elasticity of demand.
(c) degree of the slope of the demand curve.
(d) All of the above.

82. Monopolistic competition is a popular form of market structure not because
(a) it is associated with excess capacity in the long-run.
(b) consumers like varieties.
(c) while firms necessarily break-even in the long-run , they have opportunities to make economic profits in the short-run.
(d) Firms find that their promotional activities are usually rewarding.

83. Market is contestable when
(a) sunk cost in the industry is insignificant.
(b) there are no legal barriers to entry/exit in the industry.
(c) the industry is not subject to natural monopoly.
(d) All of the above.

84. A multi-plant firm sets the price and output of its product such that
(a) marginal costs at all the plants are equal and they are equal to marginal revenue .
(b) average costs at all plants are equal and they are equal to marginal revenue.
(c) marginal costs at all plants are equal.
(d) average costs at all plants are equal.

85. Which of the following statements is incorrect?
(a) Perfect competition results in socially optimum outcomes.
(b) To induce a monopolist to produce perfectly competitive outcomes, governments must impose a price ceiling such that the firm is forbidden to charge a price higher than its marginal cost.
(c) To insure socially optimum outcomes under natural monopoly, governments would have to provide subsidy to the natural monopolist.
(d) None of the above.

## CHAPTER 9

86. The unique feature of oligopoly markets is
    (a) differentiated product.
    (b) entry/exit deterrents.
    (c) interdependence of firms.
    (d) Both (a) and (c) above.
87. Oligopolists face kinky demand curve because
    (a) firms behave asymmetrically.
    (b) firms form cartel.
    (c) firms enter into a tactic collusion.
    (d) oligopoly is the most complicated type of market structure.
88. Optimum price in oligopoly markets is often indeterminate because
    (a) firms' fortunes are inter-dependent.
    (b) cartels are illegal.
    (c) firms cheat under tacit collusion.
    (d) All of the above.
89. The Cournot model is based on the assumption that
    (a) there are no significant barriers to entry/exit.
    (b) other firms match price decreases but not price increases.
    (c) rival firms keep their outputs constants.
    (d) rival firms keep their prices constants.
90. The Stackelberg-Cournot model of oligopoly
    (a) is not a kind of the leader-follower model.
    (b) assumes that the follower firms take the leader's output as a constant.
    (c) assumes that the follower firms take the leader's price as a constant.
    (d) assumes that firms behave asymmetrically.
91. Which of the following oligopoly model is likely to yield the maximum profit for the industry?
    (a) Perfect collusion model.
    (b) Tacit collusion model.
    (c) Bertrand model.
    (d) Non-cooperative game theory model.
92. Which of the following oligopoly model is the most desirable model for consumers?
    (a) Cournot model.
    (b) Non-competitive game theory model.
    (c) Bertrand model.
    (d) Cartel model.
93. Hurdles in cartel formations arise from
    (a) competition policy.
    (b) product differentiation.
    (c) scopes for cheating.
    (d) All of the above.

94. The residual market model of oligopoly assumes that the
    (a) product is differentiated.
    (b) leader maximises its profit with respect to its output.
    (c) followers set their outputs such that their joint profit is maximum.
    (d) each follower earns the maximum possible profit.
95. The three ingredients of a game do not include the
    (a) players.
    (b) full list of strategies for each player.
    (c) Pay-off matrix for all possible mix of strategies of all players.
    (d) existence of Nash equilibrium.
96. Game theory assumes that
    (a) each player acts rationally.
    (b) all players have full information.
    (c) the game contains risk but not uncertainty.
    (d) All of the above.
97. A dominant strategy is the one which
    (a) guarantees the best pay-off irrespective of the rivals' actions.
    (b) guarantees the best pay-off under the worst situation.
    (c) is selected by the dominant firm.
    (d) guarantees the best pay-off to each player.
98. Nash equilibrium is the one which
    (a) guarantees the best pay-off to each player.
    (b) maximises the joint pay-off of all the players.
    (c) is a perfect equilibrium.
    (d) guarantees the best pay-off to each player, given the strategic choices of the rivals.
99. Which of the following is a limitation of Nash equilibrium?
    (a) It is unstable.
    (b) A game may not have an unique equilibrium.
    (c) A game may have a perfect equilibrium but not a Nash equilibrium.
    (d) None of the above.
100. A threat is credible if
    (a) it is in the best interest of the threatening player.
    (b) it leads to Nash equilibrium.
    (c) it tallies with the maximin strategy.
    (d) Both (a) and (b)
101. Game theory is of little use because
    (a) pay-offs are hard to quantify
    (b) players may not act rationally
    (c) players may cheat
    (d) All of the above

## CHAPTER 10

102. The standard price-output determination model implicitly assume that the
    (a) law of one price holds.
    (b) buyers and sellers have full information.
    (c) there are no externalities.
    (d) All of the above.
103. A firm has market power if it
    (a) operates in an industry experiencing decreasing average cost.
    (b) faces a downward sloping demand curve.
    (c) operates in a perfectly competitive market.
    (d) has a strong Board of Directors.
104. Price discrimination across market segments is viable not because
    (a) market is distinctly separable into some segments.
    (b) no resale is permitted.
    (c) firms seek maximum possible profit.
    (d) different market segments have different values for the price elasticity of demand.
105. Under inter-temporal price discrimination, firms charge
    (a) different prices for a good from different customers at different times.
    (b) different prices for a good from different customers at different locations.
    (c) different prices for similar goods distinguished by cost differences.
    (d) All of the above.
106. Peak-load pricing policy is generally suitable when
    (a) firm faces capacity constraint.
    (b) demand is seasonal.
    (c) there are significant economies of scale.
    (d) Both (a) and (b) above.
107. Business schools offer tuition waivers and even stipends to bright students. This is an example of
    (a) price discrimination.
    (b) random pricing.
    (c) benevolence.
    (d) price penetration.
108. Suppose the price elasticity of demand for movies by senior citizens was –3 and that for general public was –2. If the movie ticket costs Rs. 120 for the general public, what should it cost to the senior citizens?
    (a) Rs. 80
    (b) Rs. 90
    (c) Rs. 160
    (d) None of the above
109. Which of the following pricing strategies would not enable the firm with market power to extract the entire consumer surplus?
    (a) First degree price discrimination.

(b) Commodity bundling.
(c) Peak-load pricing.
(d) Two-part pricing.

110. Which of the following pricing strategies does not usually enhance the profits of firms enjoying market power?
(a) Block pricing.
(b) Marginal cost pricing.
(c) Cross subsidising.
(d) Tying.

111. Which of the following statements is correct?
(a) The more price elastic the demand, the lower is the profit-maximising mark-up.
(b) The more price elastic the demand, the higher is the profit-maximising mark-up.
(c) The higher is the marginal cost, the lower is the profit maximising price.
(d) Price skimming is a strategy to get an entry in the market.

112. Health insurances are purchased more by unhealthy people than the healthy people because
(a) insurance companies know more about the people's health conditions than the people themselves.
(b) people know more about their health conditions than the insurance companies.
(c) healthy people are risk-seekers.
(d) insurance companies offer discounts to unhealthy people.

113. Some people drive recklessly and at times do not bother to lock their vehicles because their vehicles are insured against accidents and thefts. This is an example of
(a) adverse selection.
(b) risk-averse.
(c) asymmetric information.
(d) moral hazard.

114. Negative externalities do not lead to
(a) lower price and higher output.
(b) higher price and lower output.
(c) private cost lower than the social cost.
(d) economic inefficiency.

115. Which of the following statements is incorrect?
(a) Tobin tax on international transactions is a tax to cope with negative externalities arising from those transactions.
(b) Coase theorem was advanced to minimise dead weight loss under laissez faire.
(c) People invest in portfolios than in a single security because they are risk averse.
(d) A firm is risk-averse if it takes price-output decisions on the basis of expected demand and cost functions.

## CHAPTER 11

116. Capital expenditure decisions are significant because
(a) investment projects are almost irreversible.
(b) investments are expected to be self-financing.

(c) capital is scarce in relation to investment options.
(d) All of the above.

117. Techniques for project appraisal/evaluation are several because
(a) no technique is useless, and no technique is perfect.
(b) different techniques give different and yet very relevant results.
(c) most techniques require essentially the same data.
(d) All of the above.

118. In project appraisals, financial institutions, who appraise projects for the total investment cost, use the yearly net cash flows which are
(a) profits before depreciation, interest on fixed capital borrowings and loan repayments but after interest on working capital loans and taxes, besides including the salvage value at the terminal year.
(b) profits before depreciation but after interest on fixed capital and working capital loans, loan repayments and taxes, besides including the salvage value at the terminal year.
(c) profit after depreciation, interest on fixed capital and working capital loans, loan repayments and taxes, besides including the salvage value at the terminal year.
(d) None of the above.

119. Internal rate of return method is popular because
(a) it gives the rate of profit which could be easily compared across alternative investment opportunities.
(b) it has no use of the discount rate.
(c) its requirements for data are relatively less.
(d) None of the above.

120. Social benefit-cost analysis differ from the private benefit-cost analysis not because
(a) the goal of a private enterprise usually differ from social goals.
(b) cash flows to a private enterprise differ from those to the society.
(c) what is a good investment technique for a private enterprise may not be a good investment technique for the society.
(d) All of the above.

## MISCELLANEOUS

121. Firms exist because of
(a) increasing returns to scale.
(b) economies of scale.
(c) advantages of specialisation.
(d) All of the above.

122. Markets exist not because of
(a) specialisation.
(b) government is inefficient.
(c) people are self-sufficient.
(d) Both (a) and (b) above.

123. Government exist not because
   (a) market always produce inefficient results.
   (b) imperfect competition prevails in most markets.
   (c) many products cause externalities.
   (d) property rights promote efficiency.

124. Which of the following statements is incorrect?
   (a) Many industries have large firms because they have significant economies of scale and/or economies of scope.
   (b) Many industries have small firms because their promoters lack technical know-how and/or capital.
   (c) Many firms are vertically integrated because they save on transactions costs.
   (d) Firms go conglomerates to minimise risk and uncertainty in business.

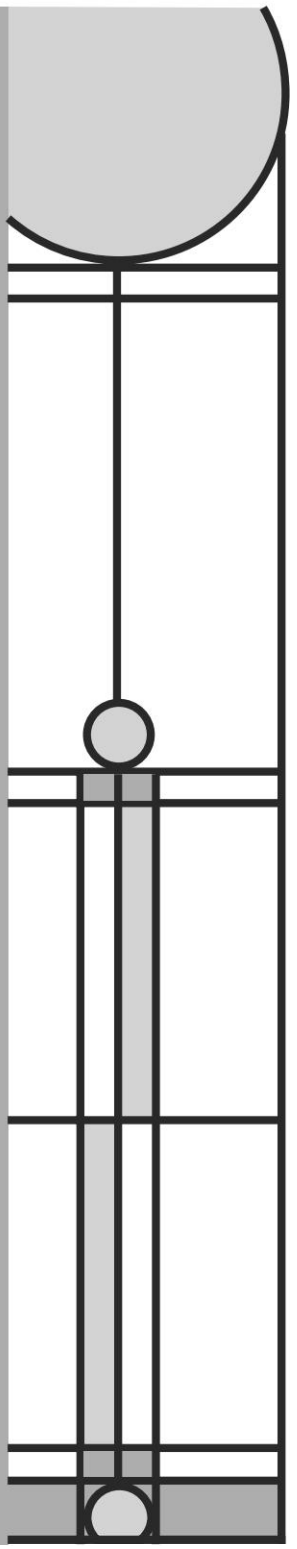
125. Which of the following statements is incorrect?
   (a) Monopoly exists in some industries because it is the best form of market structure to promote innovation.
   (b) Monopolistic competition is a popular form of market structure because it promotes competition among firms and brings variety to consumers.
   (c) Oligopoly is a popular form of market structure because it enables firms to exploit economies of scale.
   (d) Perfect competition prevails in just a few markets because firms enjoy market power.

## ANSWERS TO MULTIPLE CHOICE QUESTIONS

The **best answer** for each question is given in the following matrix. These are to be read through rows and columns. For example, the answer to question number 11 is available in row number 1 and column number one, which is "c", and that for question number 115 is given in row number 11 and column number 5, which is "d".

| # | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | - | e | a | d | d | c | d | b | d | c |
| 1 | e | c | b | b | d | a | e | a | d | c |
| 2 | a | a | b | a | d | a | d | d | b | c |
| 3 | d | b | c | d | c | a | d | b | d | D |
| 4 | c | d | c | b | c | b | b | d | b | a |
| 5 | d | a | b | d | c | c | d | d | c | D |
| 6 | a | c | c | d | d | c | b | b | a | c |
| 7 | c | d | e | b | c | a | b | c | d | a |
| 8 | b | d | a | d | a | d | c | a | d | c |
| 9 | b | a | c | d | b | d | d | a | d | B |
| 10 | a | d | d | b | c | a | d | a | b | c |
| 11 | b | a | b | d | b | d | d | d | a | D |
| 12 | c | d | c | a | b | d | - | - | - | - |

The McGraw-Hill Companies



# FILL IN BLANKS QUESTIONS

## CHAPTER 1

1. Economics is the science of _____.
2. While Micro economics is concerned with _____, Macro economics deals with _____.
3. Economic theory assumes that decision-makers are _____.
4. The trade-off that a consumer faces is between _____ and _____, while the trade off that a worker faces is between _____ and _____.
5. The four primary factors of production are land, _____, _____, and _____.
6. Economics aids firms in decisions like, (a) What to produce?, (b) _____, and (c) _____.
7. All decision-makers are supposed to honour two kinds of rules, viz. (a) market rules and (b) _____.

## CHAPTER 2

8. In economics, demand means _____.
9. The factors that would shift the demand curve for car to the right include (a) _____, (b) _____ and (c) _____.
10. Demand curve for a consumer good slopes down primarily due to the law of _____.
11. The law of diminishing marginal utility states that _____.
12. The Engle curve describes a relationship between _____and _____, *ceteris paribus*.

13. Consider a hypothetical relationship among certain demand related variables:

| *Price of X* | *Income* | *Price of Y* | *Price of Z* | *Advertisement budget* | *Demand for X* |
|---|---|---|---|---|---|
| 10 | 100 | 5 | 12 | 40 | 4 |
| 8 | 100 | 5 | 12 | 40 | 5 |
| 10 | 120 | 5 | 12 | 40 | 6 |
| 10 | 120 | 4 | 12 | 40 | 5 |

Compute the following three demand elasticities:

(a) Price _____

(b) Income _____

(c) Cross _____

14. The three factors that affect the magnitude of the price elasticity of demand are (a) _____, (b) _____ and (c) _____.

15. A consumer good is luxury if the _____ elasticity of demand for it is _____ than unity.

16. A consumer good is a Giffen good if the demand curve for it is _____ sloping.

## CHAPTER 3

17. An indifference curve is the locus of those combinations of _____that yield the same _____ to the concerned _____.

18. Slope of the consumers' budget line is given by the ratio of _____.

19. Consumers' optimum point is given by the tangency between _____ and _____.

20. The three properties of indifference curves are, (a) _____, (b) _____ and (c) _____.

21. If Mohan considers oranges and apples as perfect substitutes, his indifference curve between the two goods would be _____.

## CHAPTER 4

22. The three alternative forms in which a function could be specified are (a) _____, (b) _____, and (c) _____.

23. Firms need a quantified demand function because _____.

24. Under the least-squares method of estimation, estimates for parameters of an equation are obtained by minimizing the sum of _____ with respect to each of the parameters in the equation.

25. Trend method of forecasting suffers from downward bias when the time series is subject to a _____ trend and upward bias when the time series is subject to a _____ trend.

26. Exponential smoothing method of forecasting is superior to simple smoothing method of forecasting because _____.
27. Forecasts through Regression method could be wrong due to reasons like (a) _____ and (b) _____.
28. Degree of the accuracy of ex-post forecasts could be evaluated by the measure like _____.

## CHAPTER 5

29. The decision rule with regard to "how to produce?" for a firm in the long-run is given by _____.
30. The marginal physical product of labour is _____ at the point of inflexion and it is _____ at the point of saturation on the total physical product curve for labour.
31. Production function cannot be linear because it violates the law of _____ and it is based on an unrealistic assumption that _____.
32. In a Cobb-Douglas production function, labour elasticity of output is _____ and the return to scale's measure is given by _____.
33. The expansion path denotes the locus of those combinations of labour and capital _____.

## CHAPTER 6

34. Fixed cost is _____ cost in the short-run and zero in the _____.
35. Normal profit is the opportunity cost of _____.
36. Social cost differs from the private cost due to two reasons, viz. (a) _____ and (b) _____.
37. Economies of scale means _____.
38. Economies of scope means _____.
39. Learning curve describes the relationship between _____ and _____.
40. Transaction cost is the cost of _____.
41. By the duality theorem of production and cost, the relationship between average variable cost and average physical productivity of labour is given by _____.
42. Two of the factors that would shift a firm's supply curve to the left are (a) _____ and (b) _____.
43. The two requirements to be able to obtain a firm's supply function are, (a) _____ and (b) _____.
44. Two of the factors that source/cause economies of scale are, (a) _____ and (b) _____.
45. From the supply side, the firm's optimum output is given by _____.

## CHAPTER 7

46. The two of the barriers to exit from an industry are (a) _____ and (b) _____.
47. Many firms have become conglomerates because _____.
48. The necessary and sufficient conditions for profit maximisation ignore the possibility of minimizing loss through the option of _____.
49. Profit maximisation generally results in _____ output and _____ price than sales maximisation.
50. White goods' industries in India are characterised as oligopolies because (a) _____ and (b) _____.
51. The two limitations of the Herfindahl–Hirschman index as a measure of industry concentration are (a) _____ and (b) _____.
52. Governments resort to price ceilings to safeguard the interests of _____ (consumers or firms) and to price floors to help _____ (consumers or firms).

## CHAPTER 8

53. In the perfectly competitive market structure, the long-run equilibrium is at a point where price = MC = MR = AR = _____ and there is full utilisation of _____.
54. Incidence of sales tax is shared by buyers and sellers as per the ratio of _____.
55. Competition policy aims to regulate all monopolies because otherwise they would produce _____ and charge _____ price.
56. The primary difference between monopolistic competition and perfect competition is _____.
57. Firms enjoy market power in monopoly markets and _____ markets.
58. The degree of monopoly power can be measured by _____ index, whose formula is given by _____.
59. Patents are not owned under perfect competition and _____.
60. Many dairy farms raise cows and buffalos, and produce milk as well as cheese because of economies of _____.
61. Most firms produce multiple products in order to take advantages of _____ and _____.
62. Transaction cost is zero under _____ competition.
63. There is no market supply curve under _____ and _____.

## CHAPTER 9

64. The basic difference between oligopoly and monopoly is with regard to _____.
65. There are several models for price-output determination in oligopoly markets because _____.

66. The Sweezy model of oligopoly explains _____ and not _____.
67. Cartel under oligopoly is hard to form because (a) _____ and (b) _____.
68. Stackelberg–Bertrand model assumes firm compete on _____.
69. Under the kinky demand curve model, the demand curve is _____ at higher prices than at lower prices.
70. A strategy is dominant if it guarantees the best pay-off irrespective of the _____.
71. A strategy is secure if it guarantees the best pay-off under the _____.
72. An equilibrium is Nash equilibrium if it guarantees the best pay-off to each player, given the strategic choices of the _____.
73. A threat is credible if _____.
74. Two limitations of the Nash equilibrium are (a) _____ and (b) _____.
75. Under the leader-follower model, the oligopolists move _____ (simultaneously or sequentially).
76. Dynamic games differ from static games because in the former _____ while in the latter _____.
77. Economists do not find game theory very attractive because (a) _____ and (b) _____.

## CHAPTER 10

78. A firm has market power if _____.
79. An example of first degree price discrimination is _____.
80. The basic difference between price discrimination across market segments and inter-temporal price discrimination is _____.
81. The special reason for a firm to adopt peak-load pricing strategy is _____.
82. Transfer pricing is required only if the firm is _____ integrated.
83. An example where the two-part pricing is practiced is _____, where the two parts are _____ and _____.
84. Block pricing allows firms to enhance its profits through increasing its _____.
85. Tying is a special case of commodity bundling because _____.
86. Price matching policy helps a firm enhance its profits through at least two ways, viz. (a) _____ and (b) _____.
87. Randomised pricing is sometime practiced by a firm to _____ which generally brings more profits to firms than otherwise.
88. Limit pricing strategy is followed to _____ and it is set below _____.
89. Predatory pricing is resorted to _____ and it is set below _____.

90. Externalities cause market failures (economic inefficiencies) because they cause _____ cost or benefit to deviate from _____ cost or benefit and their impact could be minimised through _____.
91. Asymmetric information results in market failures (economic inefficiencies) through several factors including (a) _____ and (b) _____.
92. A person is risk-averse if _____.
93. While a risk-neutral firm would decide the price of its product on the basis of the expected values of the concerned parameters/functions only, a risk- averse firm would additionally consider _____.
94. Market fails to attain economic efficiency (i.e. cause dead weight loss) for various reasons including (a) _____, (b) _____, and (c) _____.

## CHAPTER 11

95. Money has time value because (a) _____, (b) _____, and (c) _____.
96. Payback period is quite popular in project appraisal because _____.
97. The only difference between the net present value method and internal rate of return method is _____.
98. Benefit – cost ratio technique is superior to the net present value method when _____.
99. Pay-back method is useful only under uncertainty because _____.
100. The best practical method for project appraisal under uncertainty is _____.

# INDEX


**A**

Accounting cost 8, 400
Accounting profit 9
Accounting rate of return 341, 350
Adverse selection 314–15
Agency problem 313
Akerlef, George, A 313
Annuity 387
Anti–trust law 260
Arbitrage 40–41, 49, 291
Arc elasticity 40–41, 46, 49, 401
Arc marginal cost 152, 179–80
Arc marginal revenue 47
ARIMA method 99
Asymmetric information 185–87, 193, 213, 254, 313–17, 414–16, 424
Autonomous demand 27
Auto–regressive method 99
Average annual rate of return 341–42, 350

**B**

Backward–bending labour supply curve 137
Bandwagon effect 29
Bargaining power 228, 295, 313
Barometric method of forecasting 100
Barriers to entry 207–8, 237–38, 408, 410
Basic principles 15–20
Behavioural economics 104
Benefit–cost analysis–private Vs. social 356
Bertrand, Joseph 261
Bertrand model 255, 260–63
Benefit–cost ratio 343
Bilateral monopoly 228
Block pricing 187, 294, 306, 309, 327
Book value 352
Break–even analysis 151, 194–200, 409
  linear 195
  non–linear 196
Break–even rate 344
Budget constraints 12, 55, 61–66
Budget line 62–67, 69, 402–3, 420
Business profit 9

**C**

Capital budgeting 332, 339, 354, 359, 363, 365
Capital cost 339, 341
Capital rental 120
Cardinal utility approach 61
Cartels 240–41, 255–60
Causal methods of forecasting 101
Ceiling prices 185, 189–90, 193, 221–24, 233–36
Census 73, 86
Certainty equivalence 360
Choice 4–6
Coase Ronald 323
Coase theorem 323–24, 414
Cobb–Douglas production function 119, 421
Coefficient of determination 76
Coefficient of variation 326, 363, 365
Collusion
  perfect 256, 259
  tacit 263, 267, 276–77
Common cost 151–52, 241
Competition
  monopolistic 208
  non–price 208
  pure 202, 203, 215, 216, 242, 253, 282
Competition policy 260
Competitive markets 42, 135, 186, 203, 213–27, 231

Complementary goods 25, 27, 35, 39
Completely elastic demand 298
Completely inelastic demand 42, 225
Compounding principle 16, 335–37
Compound value 337, 345, 351
  factor 337
Concentration ratio 204–7, 209, 408
Conservative forecasts 359
Constant returns to scale 129–30, 155, 161
Constrained optimization 123–25, 137, 299, 370, 378–83
Consumer behaviour 55, 67
Consumer choice 55, 63–67
Consumer intentions survey method 87
Consumer preferences 55, 57, 358
Consumers' reserve price 294
Consumer satisfaction maximisation 5, 63–67
Consumer surplus 222–227, 290, 306–9, 413
Consumers' goods 25
Contestable markets 237–39
Contribution 195
Cooperative games 270, 274, 277
Copy rights 207
Corner solutions 59, 65
Corporate takeovers 208
Cost
  average 152
  economic 147–148
  common 151
  explicit 8, 148
  fixed 151, 153
  historical 149
  implicit 8, 148
  incremental 17, 150
  long–run 152, 153, 155, 165
  marginal 17, 18
  opportunity 4, 15
  private 152, 318
  relevant 149
  replacement 149–50
  short–run 152–53, 162
  social 152
  sunk 150–51, 238
  total 17, 50, 153
  variable 151
Cost analysis 19, 147–79, 407, 415
Cost–benefit analysis 354
Cost curves
  long–run 157, 165
  short–run 162–66
Cost estimation methods
  engineering method 170
  statistical method 170–172
  survivorship method 170
Cost functions 153, 158
Cost–output relationships 155–156, 161
Cost based pricing methods
  average cost pricing 278
  cost plus pricing 278–79
  full–cost pricing 278
  marginal–cost pricing 278–79
Cost of capital 120, 353
Cournot model 255, 260–61, 268–69, 273, 285, 411
Cournot–Nash equilibrium 272
Cross–price eleasticity of demand 41, 45
Cross–section data 75
Cubic cost function 168
Customers reserve prices 309–10
Cut–off rate 342

**D**

Dead–weight loss 222–27, 232–33
Decision process 14
Decision tree 276, 286, 360–62
Decreasing returns to scale 129–30, 155–56, 161, 204
Delphi method of forecasting 89–90
Demand
  cross–elasticity of 41, 45
  definition of 23–24
  estimation of 71–80
  factors determining 30–39
  income elasticity of 41, 44
  price elasticity of 41–44
  elasticity of 41
  type of 25–30
Demand curve 24, 28–29, 34, 38–39, 42–43, 48, 266, 282
Demand estimation methods
  consumer interviews 72–73

market experiments 73
regression method 74–80
Demand forecasting 30, 71, 84–107
Demand function 31–39, 74–84, 101, 107–11
Depreciation 120, 149, 150, 338, 353–54, 363–64, 415
Derived demand 27–28
Descarte's rule 344
Differentiated products 258, 264, 281
Diminishing marginal returns 119, 133, 135, 143–44, 167, 176, 404–6
Discount factor 388
Discount rate 16, 338, 341–44, 351–53, 415
Discounted cash flow rate of return 344
Discounted cash flow techniques 6, 346
Discounting principle 16, 337–38
Diseconomies of scale 156, 158, 160–61
Discriminating monopoly 228, 298–301
Dominant firm model 263–65
Dominant strategy 271–76
Duality principle/theorem 162, 166, 421
Dumping 295, 298
Duopoly 258, 261–62, 267–68
Durable goods 25–26
Dynamic games 276

**E**

Econometric methods of forecasting 104
Econometric models 86
Economic cost 147–48, 400
Economic efficiency 249, 304, 424
Economic optimization 14, 19, 369–85
Economic profit 9, 148, 231, 244, 400
Economic rent 316
Economics, definition of 3–4
Economies of scale 9, 158–61, 167, 207, 238, 284
Economies of scope 168–69, 207
Economies of transaction cost 169–70, 207
Elasticity
arc 40, 41
point 40, 41
Elasticity of demand
cross 41, 45
income 41, 44
price 41–44
promotional 41, 45
Elasticity of cost with respect to output 158
Elasticity of output with respect to inputs 129
Elasticity of factor substitution 138–39
End use method 87–88
Engel curve 32, 66, 111
Entry barriers 207–8, 237–38, 408, 410
Entry fee 308–9
Equi–marginal principle 18–19
Equilibrium 125, 215–26
Estimation of cost function 170–72
Ex–ante forecasts 84
Ex–post forecasts 84, 109
Excess demand 186, 223–24
Excess supply 221, 223
Expansion path 128–29, 137–38, 155, 405–6, 421
Expected monetary value method 362–63
Expert opinion method of forecasting demand 88
Explicit cost 8, 148–49, 400
Externalities 5, 152, 289, 317–23, 327, 338, 356, 413, 424
Extrapolation 90, 103

**F**

Factors of production 7, 113–15, 147, 154, 176, 215, 229
Factor productivity 129, 131, 154
Feasible solution 382
Firms' constraints 13–14
Firm's demand 38, 84
Firms' objectives 8–12
Firms' value maximization 10–11
First degree price discrimination 293–94, 413, 423
First mover advantage 267, 269, 286
Fixed costs 151, 166, 175, 191, 194, 200, 230, 278
Fixed inputs 115–16, 119, 162–63
Forecasting
accuracy of 104
active forecasts 84
definition of 84
demand forecasting 84–104
ex–antie 84
ex–post 84
methods of 85–104
paradox of 85

passive forecasts 84
Free entry and exit 203, 213
Functions
cost 153–55
demand 36–38
production 113–20
supply 176–80

**G**

Game theory 255, 270–77, 412
Giffen good 33–34, 67, 111, 402, 420
Government, role in product pricing 188–192
Graphical method of forecasting 90
Gross domestic product 108, 109

**H**

Homogeneous product 213, 294
Horizontal integration 168
Human capital 7
Hurdle rate 342, 345, 349, 351, 353

**I**

Implicit cost 8, 148, 172, 400
Import tariffs and quotas 14
Imputed cost 8, 209
Income consumption curve 66
Income effect 33–34, 67
Income elasticity of demand 41, 44, 46, 81, 111, 401
Increasing returns to scale 129–30, 155–57, 161, 204, 406–7, 415
Incremental cost 150
Incremental principle 17–18
Indifference curves and maps 56–60, 64–67
Individual demand 29, 214, 266, 306
Industry demand 25, 29, 38, 84, 107, 214, 216, 228, 247–48, 264–66, 318–22
Inelastic demand 42, 231, 281
Inferior goods 25–26, 32–33, 44, 66
Infinitely elastic demand 214
Inflation, adjustment in investment analysis 363–64
Inputs 113
Inputs elasticity of output 129–30
Input–output relationships 114, 170
Internal rate of return 344–45, 348, 415, 424
International trade 4
Inter–temporal price discrimination 292, 302–04
Investment analysis 331–65
Investment decision 333–34
classification of 332
appraisal of 354–64
Iso–cost lines 121–26, 128
Iso–quants 117–18, 128, 405

**K**

Kinky demand curve 282–83

**L**

Labour demand 135–36
Labour supply 133, 135–37
Lagrangian multiplier technique 63, 123–24, 137, 299, 379–81, 385
Law of demand 32–34, 51, 56, 81
Law of Diminishing Marginal Returns 119, 133, 143–44, 167, 176, 404–6
Law of supply 176
Law of variable marginal returns 119, 163
Leadership, price 263
Leading indicator method of forecasting 86, 100
Learning curve 153, 167, 170, 421
Least–cost input combination 120–25
Least–squares method of estimation 76–77, 91, 100, 420
Lemon theory 313
Lerner Index 232, 247, 291
Linear programming 14, 381–3
Long–run cost analysis 19, 156–61, 165–66
Long–run cost curves 157, 165
Long–run production analysis 120–31
Long–run supply curve 175, 180
Long–run survival 8, 11–12

**M**

Macroeconomics 4, 6, 399
Management utility maximization 8, 12
Managerial economics, definition of 6
Marginal cost 17–18, 152, 162, 194, 218, 264, 296,

327, 406–10, 414
Marginal efficiency of capital 344
Marginal physical product 131, 142, 404–5, 421
Marginal principle 17–18
Marginal revenue 47–50
Marginal revenue product 18, 127–28, 134–36
Marginal rate of substitution 59–60, 65, 402
Marginal rate of technical substitution 122, 124, 405–6
Marginal utility 18, 56, 222, 294, 322, 402, 419
Market concentration 213
Market demand 28–29
Market experiments method 73, 90
Market failure 5, 188, 314, 320, 323, 424
Market power 125, 186, 214, 247, 289–91, 311–12, 409–10, 413–14, 422–23
Market price 24, 61, 189–91, 213–14, 356, 360, 402
Market structure 186, 201–4, 213, 229, 253, 289, 308
Market survey 86–90, 385
Mark up pricing 279
Maximin strategy 272, 274, 412
Maximum–output input combination 125, 129, 140
Maximum–profit input combination 126–28
Measures of investment worth 338–64
Microeconomics 4, 6
Minimum efficient scale 160
Minimum–maximorum rule 272, 274
Money, time value of 16, 349, 387
Monopolies and restrictive trade policies 260
Monopolistic competition 135, 202, 213, 242–47, 410, 416, 422
Monopoly
  discriminating 228, 291–301
  pure 202, 214, 232
  simple 228–36
Monopsony 204
Moral hazard 315, 414
Multi–plant firms 239–40
Multiple prices 184
Multiple regression analysis 77, 107

**N**

Nash equilibrium 272, 274–76, 286, 412, 422
Nash, John 272, 285
Natural monopoly 207, 228, 236–37
Necessary goods 25, 44
Negative externalities 318–21, 323, 327, 414
Net present value 342–43, 359, 424
Net terminal value 345–46
Non–cooperative games 270, 281, 285
Non–price competition 245
Non–profit objectives 12–14
Normal goods 26
Normal profit 149, 215, 412

**O**

Objective function 371, 376
Oligopoly 202, 253–284, 411–12, 422
Oligopsony 204
Opportunity cost 4, 15–16, 149, 334, 400, 421
Opportunity cost principle 15–16
Optimal input combinations 120–27
Optimum output 133, 153, 160, 172–73, 236–37, 246, 307, 322, 421
Optimum scale 173
Optimization 14, 63, 76, 122, 175, 194, 239, 277, 296, 300, 318, 369–84
Ordinal utility analysis 61
Out of pocket cost 149
Output effect 123
Output elasticity of cost 158

**P**

Pass through rule 280
Patents 207, 422
Pay–back period 340–41, 346, 348–50, 359,
Pay–off matrix 273, 412
Peak–load pricing 187, 292, 302, 304–06, 414, 423
Perfect competition 136, 140, 186, 213–27, 233, 249, 300, 307–9, 318, 407, 409, 416, 422
Perfectly competitive markets 135, 186, 213–27, 318
Perfect complements 27, 60
Perfect price discrimination 293, 294
Perfect substitutes 60, 116, 118–19, 243
Perishable goods 26
Personalised pricing 294

Pigovian tax/subsidy 324
Point elasticity 40, 41, 401
Pollution 14, 152, 317, 323, 338, 356
Post–completion audit 354
Predatory pricing 187, 238, 423
Price
  definition of 184
  determination of 19, 213–328
  multiple 184
Price ceiling 185, 188–90, 223–24, 233–35, 409, 422
Price competition 262–63
Price consumption curve 67
Price controls 188–90, 220–24, 233–35
Price discrimination 187, 235, 241, 291–301
Price effect 33, 67, 122–23
Price elasticity 41–44
Price floor 185, 189, 221–23
Price leadership 263–70
Price maker 228
Price–output determination under–
  alternative objectives 200–201
  discriminating monopoly 291–301
  monopolistic competition 242–47
  multiple products 241–42
  oligopoly 253–85
  perfect competition 213–27
  pure/simple monopoly 228–37
Price elasticity of demand 41–44
Price matching 295
Price penetration 281, 302
Price regulation 188–92
Price rigidity 282–84
Price skimming 281, 302
Price supports programme 223
Price taker 214, 228, 264, 289
Pricing practices 278–81
Prisoners' dilemma 272–74, 278–81
Private cost 152, 318, 323, 354
Producers' goods 25, 37, 88
Producers' surplus 222–33
Production function 113–19
  Cobb–Douglas 119, 130
  long–run 116
  short–run 115–16
Profit
  business 9
  economic 9
Profit, theories of
  frictional theory 9
  innovation theory 9
  managerial efficiency theory 9–10
  monopoly theory 9
  risk–bearing theory 9
Profit maximization conditions193–94
Project appraisal 354–55
Project cost 339
Project evaluation 354–55
Project life 351–52
Promotional elasticity of demand 45–46
Public goods 5, 12, 188, 400, 409
Pure competition 202–3, 215–16, 248, 253, 282
Pure monopoly 232, 248, 282, 409–10

**Q**

Quadratic cost function 172
Quantity discounts 159, 294
Quasi rent 216

**R**

Randomized pricing 303
Reaction functions 254, 261
Regression method of forecasting 101–4
Relationship between average and marginal 132–33, 164
Research and development (R & D) 148
Returns to scale 120, 128–30, 155, 161, 204
Revealed preference theory 61
Ridge lines 129
Risk adjusted discount rate 359
Risk analysis 324–26, 358–63
Risk averse 314, 325, 359, 414
Risk lover 325–26
Risk neutral 314, 325, 359
Risk seeker 325–26, 359
Rothschild index 279

**S**

Saddle point 271–72, 274, 286

Sales–maximization 8, 11, 198
Salvage value 351–54, 367, 415
Sample 73, 75–76, 86–87, 102, 104, 171, 404
Satisfying theory 12
Scale economies 158
Scarcity 3–4, 399
Scope economies 153–54, 168
Second degree price discrimination 293–94, 306
Sensitivity analysis 360–63
Separable costs 151
Sequential games 270, 274
Short–run production function 115–16, 118–19, 131–38, 405, 407
Short–run supply curve 175
Shut–down price 174, 194, 408
Simulation method 73, 362
Simultaneous equation method 104
Size maximization 8, 11
Smith, Adam 3, 158, 167, 183
Smoothing methods 93–99, 404
Snob effect 29
Social costs 152, 191, 356
Social welfare 318
Stackelberg disequilibrium 269
Standard deviation 325–26, 363, 365
Static games 270
Sticky–price theory 282–83
Strategic pricing 187
Strategic decisions 218
Subsidies 5, 152, 185, 218, 227, 324, 338, 356
Substitute goods 27, 35, 43
Substitution effect 33, 43, 67, 123
Sunk costs 150–51, 194, 238
Superior goods 25–26, 32–33, 44, 66
Supply
  curve 174–75, 214–15, 229
  law 176
Survey methods of forecasting 86–90

**T**

Tacit collusion 255, 263–70, 411
Target profit pricing 199–200
Tariffs 14, 184, 191
Taxes and price
  ad valorem sales 235–36
  incidence of 225–27
  lump–sum 235
  profit 235
  real estate 235
  specific sales 225–27, 235
Tax shield 354
Technology 114, 207, 369, 381
Time perspective principle 19
Time series data 74–75, 107, 171, 403
Time value of money 334–38
Tobin tax 324
Trend method 90–93, 101, 107, 109–10, 404, 420
Transfer payments 356
Transfer pricing 310–12
Two–part pricing 187, 291, 308–9, 327, 414, 423

**U**

Uncertainty 4–5, 16, 213, 270–71, 324–26, 334–35, 358–63, 412, 416, 424
Uncertainty analysis 326, 358–63
Unemployment 4, 114, 190, 316–17
Utility maximisation 8, 12

**V**

Value maximization 10–11
Variable costs 151, 153, 195–200
Vertical integration 169, 241, 311, 408

**W**

Written down value 352, 354

**Y**

Yield 57, 344