

MICROECONOMICS

Satya R. Chakravarty

Microeconomics

Microeconomics

Satya R. Chakravarty



ALLIED PUBLISHERS PVT. LTD.

*New Delhi • Mumbai • Kolkata • Lucknow • Chennai
Nagpur • Bangalore • Hyderabad • Ahmedabad*

ALLIED PUBLISHERS PVT. LIMITED

Regd. Off. : 15 J.N. Heredia Marg, Ballard Estate, Mumbai-400001, Ph.: 022-22626476

E-mail: mumbai.books@alliedpublishers.com

12 Prem Nagar, Ashok Marg, Opp. Indira Bhawan, Lucknow-226001, Ph.: 0522-2614253

E-mail: appltdlko@sify.com

Prarthna Flats (2nd Floor), Navrangpura, Ahmedabad-380009, Ph.: 079-26465916

E-mail: ahmbd.books@alliedpublishers.com

3-2-844/6 & 7 Kachiguda Station Road, Hyderabad-500027, Ph.: 040-24619079

E-mail: hyd.books@alliedpublishers.com

5th Main Road, Gandhinagar, Bangalore-560009, Ph.: 080-22262081

E-mail: bngl.books@alliedpublishers.com

1/13-14 Asaf Ali Road, New Delhi-110002, Ph.: 011-23239001

E-mail: delhi.books@alliedpublishers.com

17 Chittaranjan Avenue, Kolkata-700072, Ph.: 033-22129618

E-mail: cal.books@alliedpublishers.com

81 Hill Road, Ramnagar, Nagpur-440010, Ph.: 0712-2521122

E-mail: ngp.books@alliedpublishers.com

751 Anna Salai, Chennai-600002, Ph.: 044-28523938

E-mail: chennai.books@alliedpublishers.com

Website: www.alliedpublishers.com

© 2002, ALLIED PUBLISHERS PVT. LIMITED

Third Reprint : June, 2008

No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical including photocopying, recording or by any information storage and retrieval system, without prior written permission from the copyright owners.

ISBN 81-7764-348-7

Price : Rs. 275.00

Published by Sunil Sachdev and printed by Ravi Sachdev at Allied Publishers Pvt. Ltd., (Printing Division), A-104 Mayapuri Phase II, New Delhi - 110064

To my mother

Preface

This book has grown out of my microeconomics teaching at the University of British Columbia, Vancouver in Canada, and at the Indian Statistical Institute, Calcutta. There are certain innovative features of the book that are worth noting. First, methodology of the book is analytical and geometric. However, mathematics has been kept out of the main body of the text. Therefore, many of the chapters are supplemented by mathematical appendices. Second, each chapter contains at least one applied example directly related to the theoretical issues discussed in the text. Integration of theory and practice helps students to grasp the theoretical issues first and then see their relevance in practical circumstances. Third, each chapter contains different types of problems. The multiple choice and prove-disprove questions are designed to test students' understanding of the subject. For a multiple choice question, each option deserves thought, and selection of the correct answer will often not be easy. A prove-disprove question involves a chain of logical reasoning that is either true or false. Each chapter also contains a set of quantitative problems and questions. They can all be solved with the economics discussed in the text. They are intended to enhance students' understanding of applied as well as theoretical issues. A workbook containing solutions to all the exercises will be available after sometime. The workbook will also contain many additional exercises and their solutions. Finally, the book covers not only traditional aspects of microeconomics but also summarizes some recent developments like game theory, externalities, public goods, information and law. Another important point is that a few topics of importance that

are usually not available in recent texts are also covered. Examples are linear models, theory of distribution and international trade.

Given the wide coverage of the book and treatment of topics, it will be useful to both undergraduate and graduate students of economics and business. Particularly, the exercises, although assembled primarily for undergraduates would also benefit many graduate students.

My greatest debt is to the late Professor Prabhat Sarbadhikari, a great teacher of quantitative economics. Professor Sarbadhikari taught me microeconomics when I was a B.Stat (Hons) student at the Indian Statistical Institute, Calcutta during 1972-76. It is a pleasure for me to cherish his lectures, some of the finest lectures I have ever heard. I am grateful to Professor Hirendra Nath Roy who went through several chapters and offered many suggestions. Much of the credit for the publication of the book belongs to the students who persuaded me to write it. My gratitude goes to Gita Ghosh, who speedily typed a tedious manuscript with extreme care and unbelievable patience and to Allied Publishers for undertaking publication of the book. Thanks are also due to Diganta Mukherjee for typing the first two chapters of the book and arranging the bibliography alphabetically. Most of the figures in the typed draft were drawn by my wife Sumita. The remaining figures were drawn by Amita Majumder. I wish to thank them for their cooperation. Finally, I sincerely thank my son Ananyo and my wife Sumita for enduring the boring period of preparation of the manuscript. Without their patience and support this book would never have been completed.

Nobel Laureates in Economics

Beginning in 1969, a Nobel prize has been awarded each year for outstanding contributions to economics. The list of Nobel prize winners in economics is as follows:

1969	Ragner Frisch	Norway
1969	Jan Tinbergen	Holland
1970	Paul A. Samuelson	USA
1971	Simon Kuznets	USA
1972	Kenneth J. Arrow	USA
	Sir John R. Hicks	England
1973	Wassily Leontief	USA
1974	Gunner Myrdahl	Sweden
	Friedrich von Hayeck	Austria
1975	Tjalling C. Koopmans	USA-Holland
	Leonid Kantorovich	USSR
1976	Milton Friedman	USA
1977	James E. Meade	England
	Bertil Ohlin	Sweden
1978	Herbert A. Simon	USA
1979	Sir Arthur Lewis	England
	Theodore W. Schultz	USA

x Contents

1980	Lawrence R. Klien	USA
1981	James Tobin	USA
1982	George J. Stigler	USA
1983	Gerard Debreu	USA
1984	Sir Richard Stone	England
1985	Franco Modigliani	USA
1986	James Buchanan	USA
1987	Roberst M. Solow	USA
1988	Mauruce Allais	France
1989	Trygve Haalvelmo	Norway
1990	Harry M. Markowitz	USA
	Merton H. Miller	USA
	William F. Sharpe	USA
1991	Ronald H. Coase	USA
1992	Gary S. Becker	USA
1993	Douglass C. North	USA
	Robert William Fogel	USA
1994	John C. Harsanyi	USA
	John F. Nash	USA
	Reinherd Selten	Germany
1995	Robert E. Lucas Jr.	USA
1996	William Vickery	USA
	James A. Mirlees	England
1997	Robert C. Merton	USA
	Myron S. Scholes	USA
1998	Amartya Sen	India
1999	Robert A. Mundell	Canada
2000	James J. Heckman	USA
	Daniel L. Mc Fadden	USA
2001	George A. Akerlof	USA
	A. Michael Spence	USA
	Joseph E. Stiglitz	USA

Contents

<i>Preface</i>	<i>vii</i>
<i>Nobel Laureates in Economics</i>	<i>ix</i>
1. Introduction	1
1.1 Basic Problems of an Economy	1
1.2 The Construction of Economic Theory	2
1.3 Verification of Economic Theory	4
1.4 Microeconomic Theory	5
1.5 Equilibrium Analysis	5
2. Utility Theory	8
2.1 Cardinal Utility	8
2.2 Marginal Utility	10
2.3 Derivation of a Demand Curve	15
2.4 Limitations of the Cardinal Approach	16
2.5 Recent Developments: A Brief Introduction	16
2.6 The Consumer's Preferences	17
2.7 Properties for a Consumer's Preferences	18
2.8 Indifference Map	20
2.9 Marginal Rate of Substitution	22
2.10 Some Particular Cases	25
2.11 Ordinal Utility	31

2.12	Construction of a Utility Function	33
2.13	Examples of Utility Functions	34
2.14	Non-existence of a Utility Function	37
2.15	Marginal Utility	38
2.16	Marginal Rate of Substitution and Marginal Utility	39
	Example	42
	Applied Example	43
	Appendix	44
	Exercises	47
3.	Utility Maximization and Consumer Demand	52
3.1	The Budget Constraint	53
3.2	Shifts in the Budget Line	55
3.3	Budget Sets for Taxed, Subsidized and Rationed Goods	56
3.4	Utility Maximizing Choice	58
3.5	Corner Solutions	61
3.6	Consumer Demand Function	62
3.7	Homogeneity of Degree Zero of Consumer Demand Function	63
3.8	Some Demand Functions	63
3.9	Indirect Utility Function	67
3.10	Expenditure Function	68
	Example	70
	Applied Example	71
	Appendix	73
	Exercises	76
4.	Further Topics on Individual Demand	80
4.1	Changes in Income	81
4.1.1	Normal and Inferior Goods	82
4.1.2	Income Consumption Curve and the Engel Curve	83
4.1.3	Some Examples	84
4.2	Changes in a Good's Own Price	87
4.2.1	The Substitution Effect and the Income Effect: Both Goods Normal	88
4.2.2	The Substitution Effect and the Income Effect: One Good Normal and One Good Inferior	91
4.3	Examples of Substitution and Income Effects	93
4.4	The Law of Demand	95

4.5	Deriving a Consumer's Demand Curve and Price Offer Curve	95
4.6	Some Examples of Demand and Price Offer Curves	98
4.7	Compensated Demand Curves	100
4.8	Gross Substitutes and Complements	102
4.9	Net Substitutes and Net Complements	104
4.10	Another Substitution Effect	105
4.11	Theory of Revealed Preference	107
4.11.1	Idea of Revealed Preference	107
4.11.2	Weak Axiom of Revealed Preference	109
4.11.3	Weak Axiom of Revealed Preference and Negativity of Hicks Substitution Effect	111
4.11.4	The Strong Axiom of Revealed Preference	112
4.12	Price Index Numbers	112
4.13	Quantity Index Numbers	114
4.14	Consumer's Surplus	115
4.15	Compensating and Equivalent Variations	118
	Example	119
	Applied Example	123
	Appendix	124
	Exercises	129
5.	Market Demand	138
5.1	Market Demand Curves	138
5.2	Elasticity	141
5.3	Elasticity and Revenue	144
5.4	Constant Elasticity Demand Curves	146
5.5	Determinants of Elasticity	147
5.6	Cross-Elasticity of Demand	148
5.7	Income Elasticity of Demand	148
5.8	Empirical Estimation of Market Demand	149
	Example	154
	Applied Example	154
	Appendix	155
	Exercises	156
6.	Intertemporal Issues	164
6.1	Rate of Interest	164
6.2	Intertemporal Budget Constraint	166

6.3	Time Preferences	169
6.4	Utility Maximization and Choice between Two Periods	170
6.5	Comparative Statics: Effects of Change in Rate of Interest on Savings	171
6.6	Saving and Rate of Interest	174
6.7	Asset Markets	175
6.8	Discounted Present Value Criterion	175
6.9	Annuities and Perpetuities	177
6.10	Special Case of Bond	178
6.11	Applications of Discounted Present Values	178
6.12	Discounted Present Value and Rentals of Assets	181
6.13	Market for New Capital Goods	182
6.14	Human Capital	184
6.14	Demand for Durable Goods	185
	Example	188
	Applied Example	190
	Appendix	192
	Exercises	192
7.	Uncertainty	197
7.1	Uncertainty and Probability	198
7.2	Measurable Utility	199
7.3	Expected Utility Maximization	201
7.4	Risk Aversion	203
7.5	Risk Aversion and Insurance	206
7.6	More on Probability Distributions	208
7.7	Stochastic Dominance and Decisions Under Uncertainty	210
7.8	Mean Variance Analysis	211
	Example	215
	Applied Example	216
	Appendix	216
	Exercises	218
8.	Technology	223
8.1	Profit Maximization and Firm	223
8.2	Production Function	225
8.3	Production Process	226
8.4	The Short-Run and Long-Run	227

8.5	Variations in One Input	228
8.6	The Law of Variable Proportion	232
8.7	Isoquant Map	232
8.8	Input Substitution and Marginal Rate of Technical Substitution	234
8.9	Input Substitution: Two Polar Cases	236
8.10	Returns to Scale	238
8.11	Elasticity of Substitution	240
8.12	Alternative Models of a Firm: A Brief Discussion	242
	Example	244
	Applied Example	244
	Appendix	245
	Exercises	247
9.	Theory of Costs	253
9.1	Opportunity Cost	254
9.2	Short Run and Long Run Cost Functions	254
9.3	Short Run Analysis	256
9.3.1	Short Run Cost Minimization Problem	256
9.3.2	Revealed Cost Minimization	258
9.3.3	Dual to the Cost Minimization Problem	259
9.3.4	The Firm's Expansion Path	260
9.3.5	Average and Marginal Cost Functions	261
9.3.6	Cost Curves	262
9.4	Long Run Analysis	266
9.4.1	Returns to Scale and Cost Curves	266
9.4.2	Long Run and Short Run Cost	270
9.4.3	Long Run Marginal Cost	271
9.5	Elasticity of Substitution Reconsidered	272
	Example	273
	Applied Example	274
	Appendix	276
	Exercises	280
10.	Perfect Competition	286
10.1	Assumptions	286
10.2	Short Run Analysis	288
10.2.1	Revenue, Average Revenue and Marginal Revenue	288

10.2.2	Firm's Short Run Profit Maximization	290
10.2.3	Loss Minimization	293
10.2.4	Profit Maximization, Output Supply and Input Demand	294
10.2.5	Revealed Profitability	294
10.2.6	Homogeneity of Degree Zero of Firm's Output Supply Function	295
10.2.7	Short Run Output Supply Curve	296
10.2.8	Producer's Surplus	298
10.2.9	Short Run Market Supply Curve	299
10.2.10	Elasticity of Short Run Supply	300
10.2.11	Short Run Equilibrium	301
10.2.12	Stability of Equilibrium	304
10.2.13	Comparative Statics	308
10.2.14	Tax Incidence Analysis	309
10.3	Long Run Analysis	311
10.3.1	Long Run Supply Curve of a Firm	311
10.3.2	Short Run and Long Run Elasticities of Supply	313
10.3.3	Long Run Equilibrium	314
10.3.4	The Adjustment Process	317
10.3.5	Long Run Market Supply with External Economies and Diseconomies	318
10.4	The Theory of Contestability	320
	Example	321
	Applied Example	322
	Appendix	323
	Exercises	330
11.	Monopoly	343
11.1	Costs and Revenues Under Monopoly	344
11.2	Linear Demand Curve and Monopoly Revenues	347
11.3	Monopoly Profit Maximization	348
11.4	Monopoly Supply of Output	351
11.5	Monopoly Pricing and Measure of Market Power	352
11.6	Comparative Statics Analysis	352
11.7	Price Discrimination	353
11.7.1	First-Degree Price Discrimination	354
11.7.2	Second-Degree Price Discrimination	355
11.7.3	Third-Degree Price Discrimination	356

11.8	Inefficiency of Monopoly	359
11.9	Monopoly Deadweight Loss	360
11.10	Why Monopolies Exist	362
11.11	Natural Monopoly	363
11.12	Regulation of Monopoly	365
11.12.1	Marginal Cost Pricing and Natural Monopoly	365
11.12.2	Two-Tier Pricing System	366
11.12.3	Rate of Return Regulation	367
11.13	Benefits of Monopoly	368
11.14	Monopolistic Competition	368
11.15	Hotelling's Model of Location and Product Differentiation	373
	Example	374
	Applied Example	375
	Appendix	375
	Exercises	377
12.	Oligopoly	389
12.1	Pricing in Homogeneous Good Oligopoly	390
12.1.1	Quantity Leadership Model	391
12.1.2	Price Leadership Model	393
12.1.3	Cartel Model	395
12.1.4	Cournot's Model	398
12.1.5	Stability of Cournot Equilibrium	400
12.1.6	Market Power in Cournot Model	400
12.1.7	Bertrand's Model	402
12.2	Comparison of Solutions	404
12.3	Kinked Demand Curve Model	405
12.4	Industrial Organization Theory	407
12.4.1	Market Structure	407
12.4.2	Firm Conduct	410
12.4.3	Industry Performance	410
	Example	410
	Applied Example	413
	Appendix	414
	Exercises	417
13.	Game Theory	425
13.1	Basic Elements of a Game	426

13.2	Concepts of Equilibrium	427
13.3	Representation of Games	428
13.4	Dominant and Nash Equilibria	429
13.5	Matching Pennies Game	431
13.6	Mixed Strategies	431
13.7	Battle of the Sexes	433
13.8	The Prisoner's Dilemma	434
13.9	Games with Cournot and Bertrand Equilibria	435
13.10	Repeated Games and Enforcing Cooperation	437
13.11	Sequential Games	440
13.12	Sunk Costs, First Mover Advantage and Entry Deterrence	443
13.13	Limit Pricing and Predatory Pricing	445
	Appendix	447
	Exercises	448
14.	Perfectly Competitive Factor Markets	452
14.1	Economic Rent	452
14.2	Economic and Opportunity Costs	454
14.3	Land Rent	455
14.4	Ricardian Theory of Rent	456
14.5	Marginal Productivity Theory of Factor Demand	458
	14.5.1 One-Input Case	458
	14.5.2 Two-Input Case	460
14.6	Market Demand Curve	463
14.7	Market Supply of a Variable Factor of Production	464
14.8	Determination of Factor Prices	467
	Example	468
	Applied Example	468
	Appendix	469
	Exercises	470
15.	Imperfectly Competitive Factor Markets	475
15.1	Monopoly in the Output Market	476
15.2	Monopsony	477
15.3	Discrimination in Hiring by Monopsonist	480
15.4	Monopoly in Supply of Inputs	481
	15.4.1 Monopoly Power and Labour Unions	482
	15.4.2 Bilateral Monopoly	483

15.5 Upstream and Downstream Monopolies	483
Example	485
Applied Example	486
Appendix	487
Exercises	488
16. Exchange	492
16.1 The Edgeworth Box	493
16.2 Pareto Efficient Allocations	494
16.3 Gains from Exchange	496
16.4 Core of an Exchange Economy	497
16.5 Market Trade and Competitive Equilibrium	499
16.6 Existence of Competitive Equilibrium	502
16.7 First Fundamental Theorem of Welfare Economics	505
16.8 Offer Curves and Efficiency in Exchange	506
16.9 Second Fundamental Theorem of Welfare Economics	508
Example	511
Applied Example	512
Appendix	512
Exercises	515
17. Production	519
17.1 Production Possibilities	519
17.2 Efficiency in Production	524
17.3 Competitive Equilibrium	526
17.4 First Fundamental Theorem of Welfare Economics: Production	528
17.5 Second Fundamental Theorem of Welfare Economics: Production	528
17.6 Joint Production Possibilities and Comparative Advantage	528
17.7 Combining Production and Exchange	529
17.8 The Two Fundamental Theorems in Combined Model	531
Example	532
Applied Example	533
Appendix	534
Exercises	536

18. Welfare Economics	540
18.1 Preliminaries	541
18.2 Pareto Efficiency Principle Reconsidered	541
18.3 Compensation Criteria	543
18.4 Non-envy and Fairness	545
18.5 Social Preference and Arrow's Impossibility Theorem	547
18.6 Comparability Requirements	550
18.7 Social Welfare Functions	551
18.8 Maximization of Social Welfare	553
18.9 Samuelson Improvement Criterion	557
Example	558
Applied Example	558
Appendix	560
Exercises	561
19. Linear Models	565
19.1 Input-Output Analysis	567
19.1.1 Numerical Illustration	567
19.1.2 Analytical Formulation	569
19.1.3 Solving an Input-Output System	571
19.1.4 Hawkins-Simon Conditions	572
19.1.5 Linear Programming Interpretation	574
19.1.6 Price and Income Determination	576
19.1.7 Substitution Theorem	577
19.2 Linear Programming	579
19.2.1 A Linear Programming Problem and Its Graphical Representation	579
19.2.2 Slack Variables, Feasible Solutions and Basic Solutions	583
19.2.3 Simplex Method of Solving Linear Programming Problems	584
19.2.4 Simplex Method Applied	585
19.2.5 Duality and the Pricing of Inputs	590
Example	592
Applied Example	593
Appendix	
Exercises	594

20. Theory of Distribution	599
20.1 On The Marginal Productivity Theory	600
20.2 Marginal Productivity Analysis and The Determinants of Input Shares	601
20.3 The Bliss Model	603
20.4 The Ricardian Model	605
20.5 The Kaldor Model	608
Example	610
Applied Example	610
Appendix	611
Exercises	612
21. International Trade	615
21.1 Pre-Ricardian Economic Views on Basis for Trade	616
21.2 Ricardo and The Principle of Comparative Advantage	617
21.3 Analytical Formulation of the Ricardian Theory	620
21.4 Elimination of Ricardo's Simplifications	623
21.5 The Heckscher-Ohlin Theory	627
21.6 The Factor Price Equalization Theorem	631
21.7 The Leontief Paradox and Factor Intensity Reversal	632
21.8 Barriers to Trade	633
21.9 Import Tariff and the Stolper-Samuelson Theorem	634
Example	635
Applied Example	637
Appendix	638
Exercises	646
22. Market Failures and the Theory of Second Best	651
22.1 General Discussion	651
Appendix	654
Exercises	655
23. Externalities	656
23.1 Production Externalities	657
23.2 Resolution of Production Externalities: Traditional Ways	659
23.2.1 Merger and Internalization	660
23.2.2 Pigouvian Taxation	662
23.3 Property Rights and the Coase Theory	663

23.4	Consumption Externalities	667
23.5	The Problem of Commons	673
	Applied Example	674
	Exercises	675
24.	Public Goods	678
24.1	Efficient Supply of Discrete Public Good	680
24.2	Private Financing of Discrete Public Good and the Free Rider Problem	683
24.3	Efficient Supply of a Continuous Public Good	684
24.4	Private Financing of Continuous Public Good and the Free Rider Problem	687
24.5	Lindahl's Tax Scheme	690
24.6	Voting	692
	Example	698
	Applied Example	698
	Appendix	699
	Exercises	701
25.	Economics of Information	705
	Limited Information about Quality	706
	The Market for Lemons	707
	Adverse Selection	708
	Moral Hazard	709
	Solving the Problem: Cases of Equal Information	711
	Signalling	712
	The Principal-Agent Problem	716
	Applied Example 25.1 Village Panchayats and Information	719
	Value of Information	720
	Moral Hazard	721
	Principle-Agent Problem	723
	Exercises	724
	Multiple Choice	724
	Quantitative Problems and Questions	725
26.	Law and Economics	728
	Crime and Punishment	729
	Liability Law and Its Economic Considerations	734

Antitrust Violations and Multiple Damages	738
Contract Law	740
Applied Example 26.1 Probability of Conviction and Severity of Punishment	743
Minimization of Social Loss from Crime	744
Minimization of Social Loss from Accident	745
Exercises	746
Multiple Choice	746
Quantitative Problems and Questions	747
27. Mathematics	749
Functions	749
Properties of Functions	750
Continuity and Derivatives	754
Rules for Finding Derivatives	755
Partial Derivatives and Total Differential	756
Homogeneous Functions	758
Optimization	759
Closed and Open Sets	760
<i>References</i>	761
<i>Index</i>	767

1

Introduction

BASIC PROBLEMS OF AN ECONOMY

A fundamental problem that an economy faces is that the wants of its members exceed its capacities for satisfying those wants. Consider, for example, one of the personal economic problems of a consumer: how to use his limited income, his limited financial resources. With the possible exception of the very rich, none of us can afford to consume everything we would like to have. Each of us can think of a virtually unlimited number of products we want: food, shelter, clothing, medicines, a private car, a personal computer, and so on. Quite often the satisfaction of a want through the consumption of a product creates new wants. For instance, purchase a tape recorder, and soon you wish to buy audio cassettes, headphones, storage cabinets, and so on.

In order to satisfy unlimited human demands, societies or nations require the use of economic resources, the inputs that are used for creating a good or providing a service. Every society's stock of productive resources is limited or scarce—resources do not exist in sufficient amounts to satisfy all human demands. 'Scarcity' is a relative term, relative to demand. For example, good eggs may be scarce in a normal market, while bad eggs may not, because there does not exist any demand for bad eggs in the same market.

At any given point of time even the richest economies of the world possess a certain amount of resources to be used in producing goods and services. This scarcity imposes a variety of constraints both on the choices available to a society and on the opportunities open to its members. Thus, an economy's ability to produce goods and services is limited, just as an individual's ability to satisfy his personal wants is limited.

The inability to satisfy all human wants forces us to make choices about how we can best use our limited resources. This is what economics is about: wise allocation of the scarce resources among competing needs. Therefore, economics can be defined as the study of how to use our scarce resources to satisfy our alternative wants as fully as possible. By examining the activities of consumers, producers, resource suppliers, governments and voters, and by focusing on interactions among these economic agents, economists try to understand how resources are allocated.

THE CONSTRUCTION OF ECONOMIC THEORY

To make a systematic analysis of economic activity, the interaction among different economic agents involved in production, consumption, and exchange of goods and services, we need to make use of economic theory. Economic theory, like the theory of any other science, consists of a set of principles or causal relationships among the important variables that reflect economic activity.

Any theory relies on a set of conditions that are taken as given. These are referred to as postulates, upon which the theory is constructed. For instance, the theory of consumer behavior in economics relies on consumer rationality, which demands that the general desire of a consumer is to secure as much satisfaction as he can in spending his income. Thus, the first step in the construction of a theory is the specification and definition of its postulates.

The second step is the observation of the characteristics of the activity which we want to theorize. As we make continuous and repeated observations, it will become clear that some characteristics are irrelevant and can be discarded, while others are obviously significant.

The next step is to specify hypotheses which are tentative statements of cause and effect relationships among variables. For instance, repeated observations may indicate that the reduction in the consumption of a good and the decrease in the prices of its substitutes occur simultaneously. We,

therefore, reach the tentative conclusion that the decreased prices of the substitutes of a good cause its consumption to decline.

The final step is the establishment of a set of principles. Principles are hypotheses that have undergone and survived thorough testing. Once hypotheses have been formulated, they need to be tested to determine the extent to which they are valid. The tools of statistics are of particular value in this respect. The hypotheses that may be found to hold most of the time in most of the circumstances in which they are relevant, are taken as principles. Economic theory is a set of principles that are the best available statements of causal relationships. As argued above, since generally statements concerning principles are made through model formulation, theory is also referred to as model.

The variety of economic models in current use is very large. Specific assumptions vary depending on the problem being addressed. Despite this variety, practically all economic models incorporate three common elements: (i) the *ceteris paribus* (other things remaining the same) assumption; (ii) the supposition that economic decision makers seek to maximize something; and (iii) a distinction between positive and normative issues.

Let us now try to explain these three features. A model of the market for rice, for example, may seek to explain rice prices with a small number of factors, such as wages of farm workers, rainfall and consumer incomes. A model specification will allow us to understand here how rice price gets affected by specific forces. Although any researcher will recognize that many additional forces (for instance, changes in the price of fertilizers, shifts in consumer attitudes about eating rice) affect the price of rice, these other forces are held constant in the formulation of the model. This is what is meant by the *ceteris paribus* assumption. It is important to note that we are not assuming that other factors do not affect rice prices: instead these factors are assumed to be unchanged during the period of study. Examples of the second feature which says that economic decision makers optimize something are: consumers maximize their own well-being (utility); firms maximize profits and minimize costs; and government regulators attempt to maximize public welfare. Optimization models occupy a prominent position in modern economic theory. Such models seem to be fairly good at explaining reality because of their apparent empirical validity. The final feature attempts to differentiate between positive and normative questions. Positive economics is supposed to be completely objective. It is the study of the economic relationships that exist. On the other hand, normative economics is the study

of what *ought* to be. For example, an economist engaged in positive economics might investigate how income is distributed in India. But when an economist advocates that the income distribution in India should be more equal, he has implicitly moved into normative analysis.

VERIFICATION OF ECONOMIC THEORY

It becomes necessary to verify the theory. We must be able to differentiate bad models from the good ones. Two general methods can be employed for verification of economic theory: (i) the direct method, which attempts to establish the appropriateness of the assumptions on which the theory (model) is constructed, and (ii) the indirect method, which seeks to establish validity by demonstrating that a theory (model) correctly predicts real world situations.

The most common procedure adopted for verifying the validity of an economic model is to send a questionnaire to concerned economic agents and to decide on the basis of the answers received whether the model formulation is correct or not. For instance, suppose that we wish to verify whether a firm maximizes profit. The model assumes that profits are the only relevant goal of a firm. In the questionnaire, it is asked what goals the firms pursue. The results of such studies may vary. Businessmen often say that they only try to do the best they can. In fact, in general, the direct approach often turns out to be inconclusive.

Some economists object to the idea that a theory can be tested by enquiring into the appropriateness of its assumptions. They, therefore, argue that we should adopt indirect empirical tests to verify a theory. They conclude that the only way to determine the validity of a theory is to see whether it is capable of explaining and predicting real world situations. According to their view the ultimate test of a theory in economics comes when it is verified with data from the economy itself. According to the indirect approach, a test of a profit-maximizing hypothesis model would be given by trying to predict the behavior of the firms in reality by assuming that the firms behave as if they are profit maximizers. If these predictions are reasonably consistent with reality, we may accept the profit maximizing hypothesis. If the predictions are not in accord with reality, then we must investigate whether it is because the theory is invalid, or because the assumptions of *ceteris paribus* have been violated.

Theories employ abstract deductive reasoning, whereby conclusions are drawn from the set of initial assumptions. Purely empirical studies are inductive in nature. The two approaches are complementary because theories show the way for empirical studies, and empirical studies provide tests of the assumptions and conclusions of theories.

MICROECONOMIC THEORY

The main object of the theory in which we are interested is the analysis of the market activities of individual economic units such as consumers, resource owners and firms. It is called microeconomics because it respects the individuality of each good and of an economic agent. It deals with the flow of goods and services from firms to consumers, the compositions of the flow, and the process for determining the relative prices of components of the flow. It is also concerned with the flow of resources from resource owners to firms, with their evaluation and with their allocation among different uses. In contrast, macroeconomics treats the economic system as a whole. The value of the overall flow of goods and the value of the overall flow are the focus of attention in macroeconomics.

As a theoretical example of a microeconomic model, we now discuss the familiar supply-demand model of price determination.

EQUILIBRIUM ANALYSIS

In his *Principles of Economics* published in 1890, English economist Alfred Marshall (1842–1924) argued that demand and supply jointly operate to determine price. One cannot say that either supply or demand alone is sufficient to determine the price. This is analogous to claiming that one cannot tell which blade of a pair of scissors does the cutting.

In Fig. 1.1, the quantity of a good purchased per period is shown on the horizontal axis, and its price appears on the vertical axis. The curve $D(p)$ shows the quantity of the good demanded by some economic agents per period at each possible price p . This curve is the consequence of some underlying maximization, or at least can be modelled as such. $D(p)$ depends on many other variables which for the time being we assume to be fixed (*ceteris paribus* assumption). The negative slope of the curve reflects the principle that as price increases people are willing to buy fewer and fewer units of the good. Similarly, another group of economic agents determine by some maximization

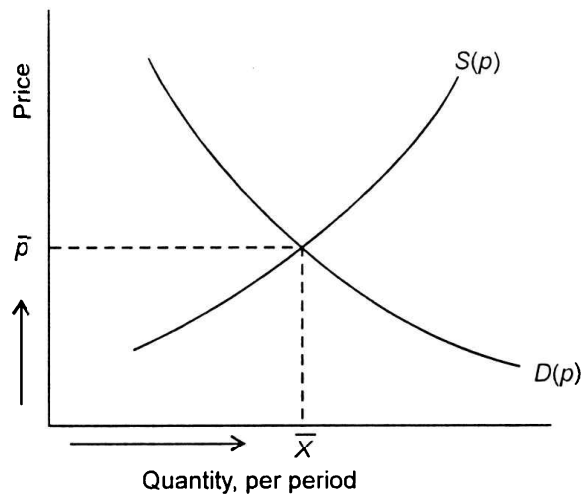


Fig. 1.1 Supply Demand Analysis

principle their desired supply of the good under consideration. We denote this supply curve by $S(p)$ under the *ceteris paribus* assumption. This curve is in fact the aggregate supply curve of firms in a perfectly competitive industry. A firm is an organization that engages in production. The term industry, which is often used interchangeably with a market, refers to a group of firms which produce goods that are identical or close substitutes. A perfectly competitive industry, loosely speaking, is an industry in which none of the participants has any control over price. We will see later that the supply curve $S(p)$ shows how the production costs of the competitive industry rises as output production increases by one more unit.

At any arbitrary price, it is unlikely that the quantity of the good demanded will be equal to the quantity supplied. In such a case the market is said to be in disequilibrium. In a situation of disequilibrium, market forces may operate on demand and supply to establish compatibility between them. (This issue is discussed in Chapter 10.) The point of compatibility between demand and supply is an equilibrium point—the demand and supply curves intersect at this point. At equilibrium, both buyers and sellers are content with the quantity being traded, and the price at which it is traded. In Fig. 1.1, economic equilibrium occurs at the price $\bar{p}$ and the quantity $\bar{X}$. Thus, price and quantity are simultaneously determined by the joint operation of supply and demand.

The supply-demand framework described above looks at only one market at a time. That is, we have considered the market behavior in a partial equilibrium framework. For some questions, this narrow viewpoint may give

valuable insights. However, for broader questions we must have a model of the entire economy that considers interrelationships among different markets and various economic agents. Analysis of equilibrium in a model of this type is called general equilibrium analysis.

Given that an equilibrium exists, we may be interested in examining how it responds to changes in economic conditions. Since we are interested in comparing the two equilibrium positions, not in studying the dynamics of movement from one position to another, this kind of analysis is called comparative statics analysis. For instance, imposition of a per unit tax of t in a market would raise the price of the product from p to $p + t$. If p is to be an equilibrium price, we must have $D(p + t) = S(p)$. Obviously, in this case we determine p as a function of t . Consequently, it might be of interest to investigate how p changes as t changes.

Another interesting issue that can be investigated about an example of equilibrium is to see as to what extent it succeeds in promoting the objectives of the economic agents. For instance, we say that a Pareto efficient situation exists, when no other situation can increase the well-being of one agent without decreasing the well-being of someone else. If a situation is not Pareto efficient, one or more agents can be made better off without making anyone else worse off. (The term Pareto efficiency is named after the early-twentieth century Italian economist Vilfredo Pareto who originated the idea.) In this book we will ask, if it is reasonable to ask, how well various economic institutions perform with respect to Pareto efficiency. Since equilibria are consequences of market institutions, we will try to investigate how well they fulfil conditions for Pareto efficiency. In fact, this book is primarily concerned with welfare in the Pareto efficiency sense, that is, it will have much to say about the extent to which the institutions under investigation perform in fulfilling the objectives (needs and wants) of the economic agents.

2

Utility Theory

Individuals in a society demand different kinds of goods and services. Presumably, they derive some welfare from those items that they demand. The theory of consumer behavior studies how choices are made by a consumer from among available alternatives. That is, it deals with a consumer's decision making process. It is designed to explain the demands for goods and services made by a consumer from different perspectives. In this chapter and the next, we analyze different aspects of consumer behavior.

We begin the chapter with a discussion of the derivation of a consumer's demand in the cardinal utility framework. The modern approach to consumer theory is taken up after that.

CARDINAL UTILITY

Cardinal measurement of an item refers to measurable quantities of it. Examples are abundant: temperature can be measured in degrees, distance can be measured in kilometers, and so on. The most important characteristic of cardinal measurement is that we can talk about differences in the measurement scale. Such differences will have the same meaning regardless of where or to what the scale is applied. A difference of 12 degrees in tempera-

ture is the same in Delhi as it is in Calcutta. Similarly, the gap between 100° Centigrade and 90° Centigrade is recorded as twice that between 85° Celsius and 80° Celsius, no matter whether we express these temperatures in Celsius or Fahrenheit (in which they correspond to 212° , 194° , 185° and 176° respectively).

The notion of cardinal utility in economics goes back at least as far as Jeremy Bentham (1748–1832). Bentham regarded utility of an action as a magnitude that is, in principle, measurable. Contributions along this line also came from Leon Walras, Stanley Jevons, Alfred Marshall and others. In the cardinal framework, the total utility obtained from a commodity refers to the amount of satisfaction that a consumer gets from consuming it. Up to a certain point, the more units of an item an individual consumes per unit of time, the greater will be the total utility or satisfaction received from it. At some level of consumption, total utility reaches a maximum. It does not yield higher satisfaction even though the consumer possesses more of it. This state is called the saturation point for the commodity. Thus, up to the saturation point, utility is an increasing function of consumption levels. It may be possible that more units of the good beyond the saturation point forced upon the consumer can cause total utility to decrease simply because of storage problems. That is, up to the saturation point the commodity is a 'good' to the consumer, beyond that point it is a 'bad'. However, for our purpose the possibility of reduction in total utility beyond the point of saturation is not of much importance.

A hypothetical utility curve is shown in panel (a) of Fig. 2.1

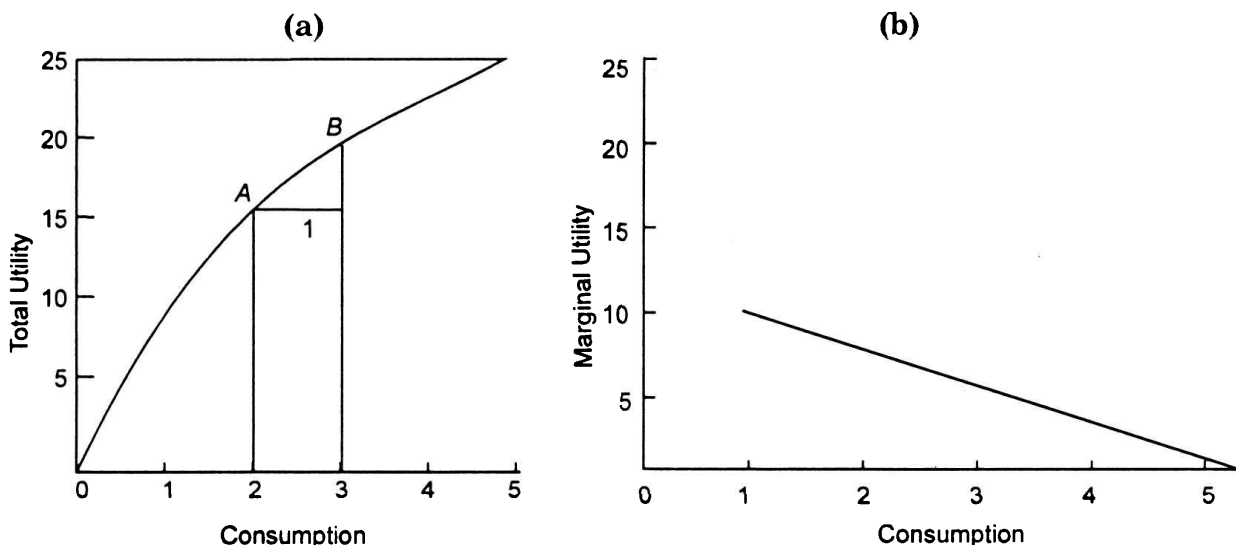


Fig. 2.1 Total Utility and Marginal Utility

As stated earlier, we assume that utility can be quantified, measured and that different quantities of utility can be added to arrive at a meaningful total. The point of saturation is reached at a consumption level of five units of the commodity. Up to five units of consumption, utility increases as consumption increases—beyond that point utility decreases.

The hypothetical utility curve in panel (a) of Fig. 2.1 is based on the following table which shows total utility values for different units of commodity consumed.

Table 2.1 Total Utility

<i>Units of Commodity Consumed</i>	<i>Total Utility</i>
0	0
1	9
2	16
3	21
4	24
5	25

MARGINAL UTILITY

The marginal utility associated with consuming a unit of a commodity is the change in utility resulting from one unit increase in consumption of the commodity. From Table 2.1 we conclude that the marginal utility of the first unit consumed is nine. To understand this, we note from the table that the total utility when the consumer does not consume anything is zero, and the total utility when the consumer buys the first unit is nine. Now, since marginal utility is the change in total utility associated with consumption of one more unit, and since the first unit adds nine to the total utility, we can say that the marginal utility of the first unit is nine. Similarly, the marginal utilities of the second, third, fourth and the fifth units are seven, five, three and one respectively. By plotting these marginal utility figures against the **corresponding units of consumption**, we get the marginal utility curve, which is shown in panel (b) of Fig. 2.1.

The above calculation reveals one important feature of the marginal utility. The marginal utility of a commodity decreases as the consumer acquires larger quantities of it. That is, the marginal utility of the first unit is larger than that of the second unit, and the second unit has a higher marginal utility

than the third unit, and so on. This behavior of marginal utility is known as the Law of Diminishing Marginal Utility. We state this law as:

LAW 2.1 *A consumer experiences diminishing marginal utility from consumption of additional units of a commodity.*

Equivalently, we can say that:

Each additional unit consumed of a commodity provides less additional satisfaction (utility) to the consumer.

Technically, the Law of Diminishing Marginal Utility means that utility is a strictly concave function of consumption levels.

The reasoning behind this law is quite clear. A consumer who has only a few units of a commodity will tend to use them to fulfil his most pressing wants. As more units of the commodity are available, the purposes for which he wants to use these units are of less and less importance to him, and hence yield less and less increase in total utility. For instance, if a person gets some minimal amount of water everyday, he will use it for drinking. If some more amount of water is available, he will use it for drinking, cooking and washing. When a greater amount of water becomes available, he will use it for watering the lawn also. After some stage, water becomes a bad rather than a good. This shows how marginal utility decreases as additional units become available.

It may be pointed out that diminishing marginal utility need not be the case for all levels of consumption. For instance, a family with several children may not be very happy with a home having only one bed-room. Two bed-rooms—one for the parents and one for the children—may yield more than twice the satisfaction of one. But the successive increases in total utility yielded by three, four, etc. will certainly be successively smaller. Thus, through some range of consumption levels, marginal utility may rise as consumption increases. However, if a saturation point for a commodity exists for a given individual, then as his consumption level approaches that point, marginal utility must be decreasing even though it may be increasing at lower levels of consumption.

Before we investigate how a cardinal utility function can be employed to derive a consumer's demand for various commodities, let us consider total and marginal utilities technically. Let x denote the number of units of a commodity consumed. We write $U(x)$ for the total utility associated with the level of consumption x , and $MU(x)$ for marginal utility. We assume that the commodity is continuously divisible (such as water or oil). Let us denote the

change in total utility when consumption increases from x to $x + \Delta x$ by $\Delta U(x)$. That is, $\Delta U(x) = U(x + \Delta x) - U(x)$. Then $\Delta U(x)/\Delta x$ is the rate of change in utility associated with the change Δx in the quantity of the commodity consumed. If Δx is marginal (a very small change), then the rate

$$MU(x) = \frac{\Delta U(x)}{\Delta x} \quad (2.1)$$

is called the marginal utility at the consumption level x . Given that Δx is infinitesimal, $MU(x)$ in Eq. (2.1) is measured by the slope of the total utility function at any given level of consumption x .

An example of (total) utility function can be

$$U(x) = 10x - x^2, \quad 0 \leq x \leq 5. \quad (2.2)$$

This function is increasing and strictly concave up to $x = 5$. The consumption level $x = 5$ is a point of saturation; beyond this point the function is decreasing. The marginal utility function associated with $U(x)$ in Eq. (2.2) is

$$MU(x) = 10 - 2x, \quad 0 \leq x \leq 5. \quad (2.3)$$

This function is uniformly decreasing in x .

The law of diminishing marginal utility can be used to explain the way consumers allocate their expenditure among goods and services. Let us imagine a consumer who wishes to spend his income m on the consumption of two goods, 1 and 2. Let x_1 and x_2 denote the number of units of goods 1 and 2 that the consumer purchases. Total utility from the consumption of two goods can be viewed as the sum $U_1(x_1) + U_2(x_2)$, where $U_i(x_i)$ is the utility of x_i , $i = 1, 2$. This additivity condition implies that the utilities are independent. Thus, the marginal utility of each commodity is independent of the consumption of the other commodity. (The additivity assumption, which has been made in much of the Marginalist literature, makes the analysis simple. It is not strictly necessary for the cardinal utility theory.)

Suppose that p_1 and p_2 are given prices of the commodities 1 and 2. Then the consumer's expenditure on the consumption of two goods is $p_1x_1 + p_2x_2$. Since he cannot spend more than m on his purchases of the two goods, it must be the case that $p_1x_1 + p_2x_2 \leq m$. This is known as the consumer's budget constraint; the equation $p_1x_1 + p_2x_2 = m$ is called the budget line. It contains all those commodity bundles that exhaust the income m . (See Chapter 3 for further discussions on the budget constraint.) We assume that the consumer

wishes to maximize total utility, that is, he wishes to reach the highest level of satisfaction subject to the budget constraint. Given that the consumer's utility function plays the only role in determining his consumption bundle, we can say that the consumer is the best judge of his own interest. Further, by maximizing utility subject to the budget constraint, the consumer behaves rationally.

We now formally state the consumer's utility maximizing choice condition.

THEOREM 2.1 *The law of diminishing marginal utility implies that the consumer will allocate his income m on the purchases of the two goods 1 and 2 to satisfy the following condition.*

$$\frac{MU(x_1)}{MU(x_2)} = \frac{p_1}{p_2} \quad (2.4)$$

where $MU(x_i)$ is the marginal utility of good i and p_i is its given price. $i = 1, 2$.

Proof. Suppose the two goods are purchased in amounts such that the ratio of marginal utilities is smaller than the price ratio. That is, $\frac{MU(x_1)}{MU(x_2)} < \frac{p_1}{p_2}$.

Since marginal utility is diminishing, we can say that the consumer has purchased more x_1 and less x_2 than if there were equality between the ratio of marginal utilities and the price ratio.

Suppose now the consumer decreases his consumption of good 1 by Δx_1 . It should be clear from Eq. (2.1) that this change decreases his utility by $\Delta x_1 MU(x_1)$. Now the consumer's budget line is $p_1 x_1 + p_2 x_2 = m$. (As stated, we are assuming that the consumer spends the entire income m on the consumption of the two goods.) Since p_1 , p_2 and m are fixed, by reducing x_1 by Δx_1 , the consumer can increase his consumption of x_2 by some amount, say, Δx_2 . From the budget constraint we have $p_1 \Delta x_1 + p_2 \Delta x_2 = 0$, from which it follows that

$$\Delta x_2 = -p_1 \Delta x_1 / p_2.$$

The increase in utility due to Δx_2 increase in consumption of x_2 is $\Delta x_2 MU(x_2)$. Therefore, the net change in utility ΔU is

$$\begin{aligned} \Delta U &= \Delta x_1 MU(x_1) + \Delta x_2 MU(x_2) \\ &= \Delta x_1 MU(x_1) - \frac{p_1}{p_2} \Delta x_1 MU(x_2) \end{aligned}$$

$$= \Delta x_1 MU(x_2) \left(\frac{MU(x_1)}{MU(x_2)} - \frac{p_1}{p_2} \right).$$

But by assumption $\frac{MU(x_1)}{MU(x_2)} < \frac{p_1}{p_2}$. Since $\Delta x_1 < 0$ and $MU(x_2) > 0$, $\Delta U > 0$. Therefore, without spending any additional amount of money, the consumer is able to increase utility by moving his purchase away from good 1 and toward good 2, until the ratio of marginal utilities becomes equal to the price ratio.

If we assume that $\frac{MU(x_1)}{MU(x_2)} > \frac{p_1}{p_2}$, then the law of diminishing marginal utility implies that the consumer has bought less of x_1 and more of x_2 than the amounts that would make the ratio of marginal utilities equal to the price ratio. We can apply arguments similar to those employed above to show that we can increase total utility by moving consumption away from good 2 and toward good 1, until the marginal utility ratio equals the price ratio. Therefore, it follows that if the consumer wishes to achieve the highest level of utility subject to a given budget constraint, then he will adjust his expenditure so that the ratio of marginal utilities is equal to the price ratio. This completes the proof of the theorem.

The ratio p_1/p_2 gives the number of units of good 2 that exchanges for one unit of good 1. Thus, if $p_1/p_2 = 3$, then three units of good 2 would exchange for one unit of good 1. Note that we can rewrite the equality $\frac{MU(x_1)}{MU(x_2)} = \frac{p_1}{p_2}$ as $\frac{MU(x_1)}{p_1} = \frac{MU(x_2)}{p_2}$. This condition means that the utility derived from spending an additional unit of money must be the same for both the commodities.

An interesting observation concerning cardinal utility is that if $U(x)$ is the utility of x , then a positive linear transformation of $U(x)$, such as $V(x) = a + bU(x)$, where $b > 0$ and a are constants, can also be used as a utility function. (Positive linear transformations are also called affine transformations. We discuss them in greater detail in the mathematics chapter.) The marginal utility associated with $V(x)$ becomes $bMU(x)$. However, the saturation point does not change under affine transformations. Given that all utility values are changed by the same affine transformation, the ratio of marginal utilities remains unaltered. This means that the consumer's utility maximizing condition Eq. (2.4) remains invariant with respect to affine transformations of utilities.

DERIVATION OF A DEMAND CURVE

We can employ the law of diminishing marginal utility and the consumer's utility maximizing condition to derive his demand curve for a particular commodity. A consumer's demand curve for a commodity graphically shows the relationship between the price of the commodity and the quantity purchased of that commodity under the assumption that the prices of the other commodities, the consumer's income and the tastes of the consumer remain constant. Suppose we wish to find the demand curve for good 1. To do this, we begin from the utility maximizing condition. Let $(\bar{x}_1, \bar{x}_2)$ be the units of commodities 1 and 2 that solve $\frac{MU(x_1)}{MU(x_2)} = \frac{p_1}{p_2}$ and $p_1x_1 + p_2x_2 = m$. Then $(\bar{x}_1, p_1)$ is a point on the consumer's demand curve.

Next, suppose p_1 increases to p'_1 . Thus, the ratio $\frac{MU(\bar{x}_1)}{MU(\bar{x}_2)}$ is smaller than p'_1/p_2 . The law of diminishing marginal utility demands that he must decrease $\bar{x}_1$ to, say $\bar{x}_1 + \Delta\bar{x}_1$, and increase $\bar{x}_2$ to say $\bar{x}_2 + \Delta\bar{x}_2$, to achieve the equality $\frac{MU(\bar{x}_1 + \Delta\bar{x}_1)}{MU(\bar{x}_2 + \Delta\bar{x}_2)} = \frac{p'_1}{p_2}$. Clearly $\Delta\bar{x}_1 < 0$ and $\Delta\bar{x}_2 > 0$. Assuming that $p'_1(\bar{x}_1 + \Delta\bar{x}_1) + p_2(\bar{x}_2 + \Delta\bar{x}_2) = m$, we can say that $(\bar{x}_1 + \Delta\bar{x}_1, p'_1)$ is another point on the demand curve for good 1. Since $\bar{x}_1 + \Delta\bar{x}_1 < \bar{x}_1$, but $p_1 < p'_1$, it follows that the quantity demanded falls as the price increases. The process can be repeated any number of times. By joining the resulting points, we get the consumer's demand curve for commodity 1.



Fig. 2.2 Price-Quantity Relation

Sometimes prices are plotted on the vertical axis and the quantities demanded are shown on the horizontal axis. The resulting curve is called the inverse demand curve. The inverse demand curve is the demand curve that views price as a function of quantity. For each level of demand for good 1, this curve shows the price of good 1 that would have to be in order for the consumer to choose that level of the good. Therefore, like the direct demand curve, the inverse demand curve also shows relationship between price and quantity, but from a different perspective.

LIMITATIONS OF THE CARDINAL APPROACH

According to the cardinal approach, utility is a concept as real as the commodities themselves. But there are certain basic weaknesses of the cardinal approach. We cannot conceive of a way to measure satisfaction derived from consumption of various commodities objectively. The suggestion made by Walras to use subjective units (utils) for the measurement of utility is also not a satisfactory solution. There is no way to determine how one person's utilities can be compared to another's. The axiom of diminishing marginal utility, which can be regarded as a psychological law, is also based on the questionable measurability assumption. A consumer cannot tell the marginal utility of the last unit of a commodity consumed.

RECENT DEVELOPMENTS: A BRIEF INTRODUCTION

It is possible to draw all the conclusions of the cardinal utility approach to consumer behavior without assuming the law of diminishing marginal utility. Developments along this line took place in two separate but related directions. One of these developments focuses on the structure of a consumer's preferences and the way it affects his choices. The second development concentrates directly on the choices and lays stress on testable propositions about them. These two approaches will be referred to as the modern utility theory and the revealed preference theory respectively.

Modern utility theory antedates the revealed preference theory. Although the modern utility theory dates from the 1880s, it was not integrated into the main body of economic thought until the 1930s. The idea was introduced by Francis Y. Edgeworth in 1881. Edgeworth was followed by Vilfredo Pareto, Irving Fisher and Eugen Slutsky in the first decades of this century. Then in the 1930s Nicholas Georgescu-Roegen and later John R. Hicks and R.G.D. Allen made significant contributions to this approach. Axiomatic foundations of modern utility theory were developed among others, by Kenneth J. Arrow,

Gerard Debreu and Tjalling C. Koopmans. Revealed preference theory was proposed independently by Abraham Wald and Paul A. Samuelson. Important contributions to this literature also came from Hendrik Houthakker. (See the Bibliography for a few references along this line.)

In the remainder of this chapter we will discuss the modern utility theory of consumer behavior. The revealed preference approach to consumer theory will be taken up in Chapter 4.

THE CONSUMER'S PREFERENCES

We wish to study an individual consumer's behavior by examining his preferences. We reject at the outset the Marginalists' notion of utility and the way in which utilities are added up to obtain total utility. The modern utility theory begins by assuming that the consumer possesses preferences over market baskets of goods, called commodity bundles. A commodity bundle or a consumption bundle is a listing of the amounts of each of the commodities consumed by a consumer. By concentrating on commodity bundles, the modern approach assumes that the consumer chooses all goods simultaneously, rather than selecting one commodity after another as in the cardinal approach. Its focus on consumer choice amongst commodity bundles underlies the importance of substitution amongst commodities.

To discuss different properties of a consumer's preferences over commodity bundles, let us assume that there are only two goods, say 1 and 2, available to the consumer. Let x_1 denote the amount of the first good and x_2 be the amount of the other. Therefore, $(x_1, x_2) = x$, say, is a commodity bundle.

We will assume that given any two bundles $x = (x_1, x_2)$ and $y = (y_1, y_2)$, the consumer can rank them according to their desirability. That is, the consumer can decide whether one bundle is better than the other, or he finds them equally satisfactory. If the consumer thinks that x is better than y , then symbolically we write this as $x \succ y$. In this case x is said to be strictly preferred to y . That is, given that the two bundles x and y are available to the consumer, he wishes to choose bundle x instead of y . Therefore, the idea of preference is based on the consumer's behavior. If the consumer is indifferent between two bundles of goods x and y , we express this symbolically as $x \sim y$. Indifference means that the consumer finds x and y equally satisfactory according to his preferences.

If the consumer prefers, or is indifferent, between the two bundles x and y , we say that he prefers x to y and write $x \succeq y$. If $x \succeq y$ and $y \succeq x$, then

$x \sim y$. That is, if x is at least as good as y and y is at least as good as x , then the consumer is indifferent between the two bundles. Similarly, if the consumer thinks that x is at least as good as y and y is not at least as good as x , then x must be strictly preferred to y . Symbolically, if $x \succeq y$ and it is not the case that $y \succeq x$, then $x \succ y$.

PROPERTIES FOR A CONSUMER'S PREFERENCES

A consumer's preference relations are assumed to possess certain properties. They are necessary in order to set up a formal analysis of consumer behavior.

Completeness

Completeness means that any two bundles can be compared. That is, given any two bundles x and y , the individual can always specify one of the following:

- i) $x \succeq y$,
- ii) $y \succeq x$,
- iii) both (i) and (ii).

Reflexivity

Any bundle is at least as good as itself. That is, for any commodity bundle x , $x \succeq x$.

Transitivity

For any three bundles $x = (x_1, x_2)$, $y = (y_1, y_2)$ and $z = (z_1, z_2)$, if $x \succeq y$ and $y \succeq z$, then $x \succeq z$. That is, if the consumer thinks that x is at least as good as y and y is at least as good as z , then he thinks x is at least as good as z .

Monotonicity

Let $x = (x_1, x_2)$ be a bundle of goods and $y = (y_1, y_2)$ be another bundle of goods, where $x_i \geq y_i$ for $i = 1, 2$, with strict inequality for at least one i . Then $x \succ y$. That is, between two bundles x and y , if x has at least as much as of both goods and more of at least one good, then x is strictly preferred to y .

Strict Convexity

For any two distinct bundles (x_1, x_2) and (y_1, y_2) , if $(x_1, x_2) \sim (y_1, y_2)$, then

$$[tx_1 + (1 - t)y_1, tx_2 + (1 - t)y_2] \succ (x_1, x_2)$$

for any t such that $0 < t < 1$. That is, the weighted average of two distinct bundles (x_1, x_2) and (y_1, y_2) about which the consumer is indifferent, must be strictly preferred to (x_1, x_2) and hence to (y_1, y_2) because (x_1, x_2) and (y_1, y_2) are indifferent. Note that since $0 < t < 1$, the weighted average cannot be (x_1, x_2) or (y_1, y_2) .

By completeness we mean that the consumer is able to make a choice between any two bundles. Therefore, the consumer is not paralyzed by indecision. He can always make up his mind about the desirability of any two alternatives. The second postulate is self-explanatory. Any bundle is always at least as good as an identical bundle. Clearly, transitivity is a consistency requirement. If preferences are not transitive, there may exist a set of bundles for which there is no best choice. Therefore, if we ask a consumer about his best choice it becomes problematic for him to reply if his preferences are not transitive. Monotonicity means that the consumer strictly prefers more of any good or service to less of it, that is, the consumer is not satiated with any one. Satiation with a good is not impossible. For instance, it may occur temporarily with food. However, economics would not be a very interesting subject in a situation where everyone was satiated in the consumption of every good. Convexity of a preference relation means that the consumer prefers averages to extremes. For any t lying between 0 and 1, $tx_1 + (1 - t)y_1$ is a weighted average of the quantities of the first commodity in the two bundles $x = (x_1, x_2)$ and $y = (y_1, y_2)$, where the weights are given by t and $(1 - t)$ respectively. Similarly, $tx_2 + (1 - t)y_2$ is a weighted average of the quantities of the second commodity. Strict convexity demands that the weighted average bundle $[tx_1 + (1 - t)y_1, tx_2 + (1 - t)y_2]$ is strictly preferred to x . Note that $[tx_1 + (1 - t)y_1, tx_2 + (1 - t)y_2]$ lies on the straight line connecting the points (x_1, x_2) and (y_1, y_2) . Since by strict convexity of preference relation the weighted average bundle is strictly preferred to x , and since t is arbitrary, this line joining the two bundles x and y must belong to the set of bundles that are strictly preferred to x . Thus, we are demanding that the set of bundles strictly preferred to x be convex. A convex set has the property that if you take any two points belonging to the set, then the straight line connecting these two points lies entirely in the set.

In panel (a) of Fig. 2.3 we show a convex set. But the set shown in panel (b) of the figure is not convex. We see that the line joining the points v_1 and v_2 in panel (b) does not belong to the set.

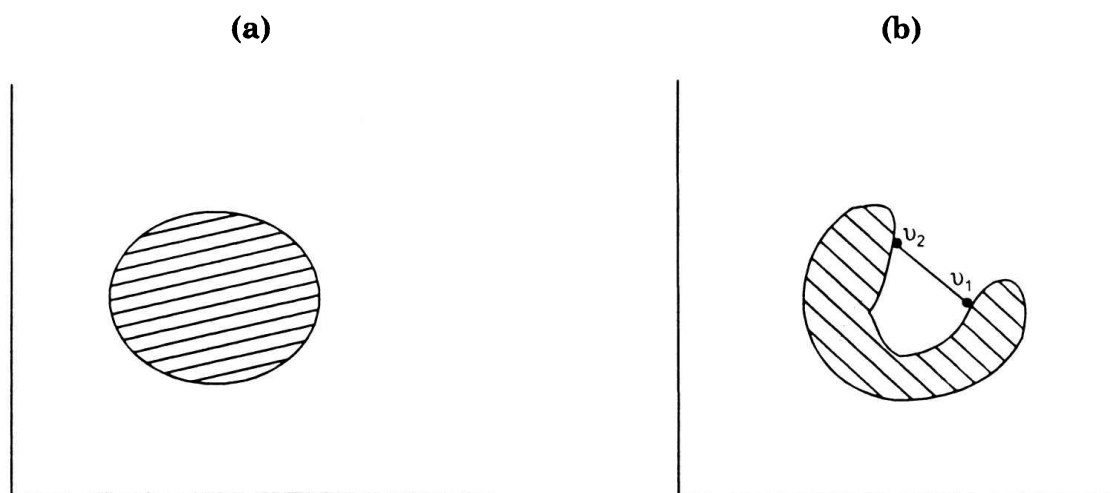


Fig. 2.3 Convex and Non-Convex Sets

But why do we want strict convexity of preference relation? Because the goods are consumed together. Let us consider a consumer's preferences for monthly consumption. Each month he prefers having some of each of the two goods rather than specializing in consuming either one for the whole month. That is, instead of buying extreme quantities of one of the two goods, the consumer prefers to choose some of each. This is strict convexity of preferences.

INDIFFERENCE MAP

The above postulates enable us to construct an individual consumer's indifference map. An indifference map is a graph of a family of indifference curves, each of which is a locus of commodity bundles that are indifferent to one another.

In Fig. 2.4, two axes represent a consumer's consumption of goods 1 and 2. Let us consider a particular commodity bundle $x = (x_1, x_2)$. The indifference curve through x consists of all bundles of goods that leave the consumer indifferent to the given bundle x .

Let us now try to look at the implications of the postulates for a consumer's preference relations about the shape of his indifference curves. If we start at $x = (x_1, x_2)$ and move up and to the right, then by the monotonicity property we will be moving to a strictly preferred position. Similarly, if we move down and to the left, we must move to a worse position. Therefore, movement along an indifferent position requires that we must move either left and up, or right and down. This means that the indifference curve is negatively sloped.

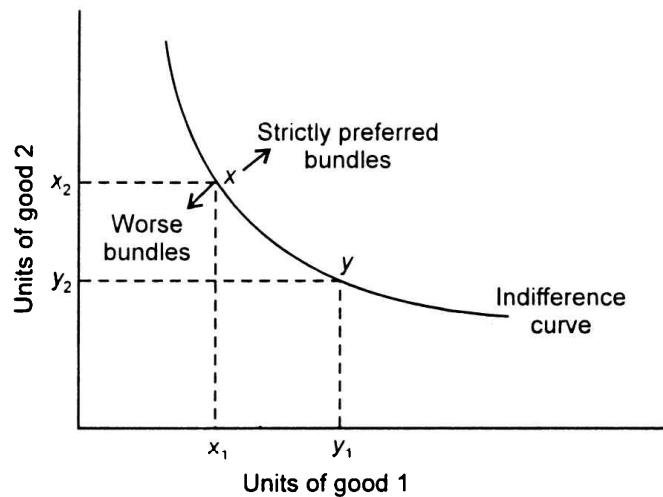


Fig. 2.4 Indifference Curve

By moving right and down, let us now choose another bundle $y = (y_1, y_2)$. Assume that the consumer is indifferent between x and y . Consider the straight line joining x and y . By strict convexity of preference relation, any point on the line joining x and y , excepting the extreme points x and y , must be strictly preferred to x (and y). Now, since y has been assumed to be indifferent to x , y must lie on the indifference curve passing through x . Therefore, for any bundle lying on the line joining x and y , barring the bundles x and y , to be strictly preferred to any bundle lying on the indifference curve passing through x and y , it must be the case that the indifference curve lies below the line connecting x and y . Since x and y are arbitrary, it follows that the indifference curve is strictly convex to the origin. We show this in Fig. 2.4. In view of our earlier discussion it is clear that this is the same as the condition that the indifference curve bounds a convex set.

Drawing an indifference curve as a continuous curve implies that the consumer regards the two commodities 1 and 2 as perfectly divisible. That is, the commodities are available in as fine a fractional unit as desired. Perfect divisibility along with completeness implies that there exists an indifference curve passing through every point representing some commodity bundle.

Next, using transitivity of a preference relation, we show that the indifference curves cannot cross. To demonstrate this, let us consider three bundles of goods x , y and z , such that x lies on one indifference curve, y lies on another indifference curve and z lies at the point of intersection of the two. This is shown in Fig. 2.5. Now, indifference curves represent distinct levels of preference. Since x and y lie on different indifference curves, it must be

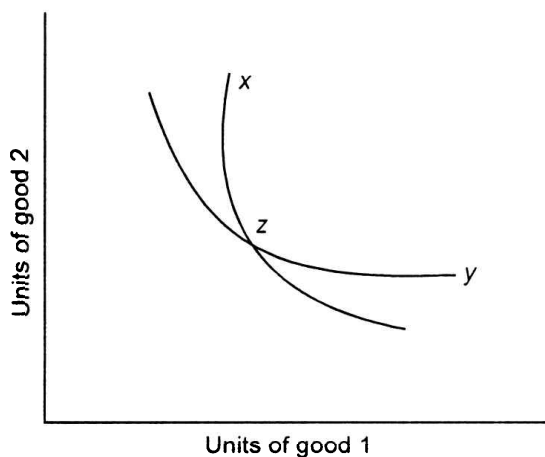


Fig. 2.5 Indifference Curves cannot Cross

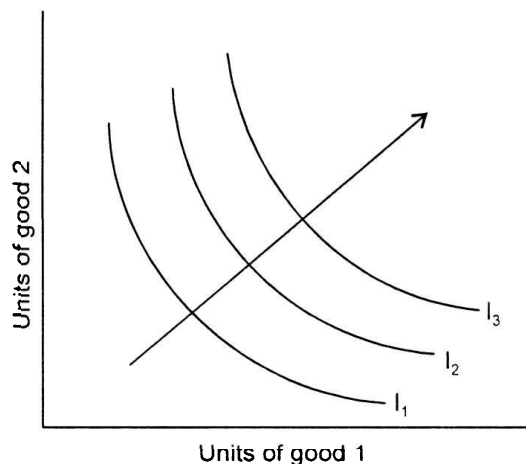


Fig. 2.6 Indifference Map

true that either $x \succ y$ or $y \succ x$. Assume that $x \succ y$. But given that z is the point of intersection of the two curves, we know that $x \sim z$ and $z \sim y$. Therefore, by transitivity $x \sim y$. This is a contradiction to the assumption that $x \succ y$. Hence indifference curves cannot cross.

Given that indifference curves are non-intersecting, we are now in a position to draw a consumer's indifference map. In Fig. 2.6 we draw several indifference curves. All commodity combinations on higher indifference curves—those farther from the origin—are strictly preferable to those lying on lower indifference curves. The arrow indicates the direction of preference. Thus, a bundle on the indifference curve I_2 is strictly preferred to any bundle on I_1 , and each bundle on I_3 is strictly preferred to any bundle on I_2 . The whole set of indifference curves constitutes the indifference map of the consumer. (For illustration, we have drawn only three indifference curves, but we can draw as many as we wish.)

MARGINAL RATE OF SUBSTITUTION

As drawn, the indifference curves are downward sloping and they become flatter as they move downward. This property reflects the substitutability of one good for another in the consumer's preference. We will show this now.

Suppose we take a small amount of good 1 (Δx_1) away from the consumer. We then give him an amount Δx_2 of good 2, so that he remains on the same indifference curve. That is, substitution of the amount Δx_2 for Δx_1 is sufficient to keep the consumer just as well off after this substitution, as he was bef-

ore. Thus, decreasing the amount of good 1 by Δx_1 is exactly offset by increasing the amount of good 2 by Δx_2 . We show this diagrammatically in Fig. 2.7.

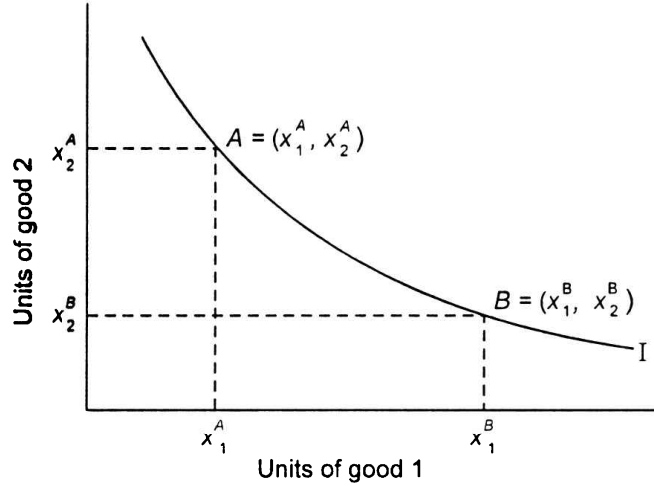


Fig. 2.7 Marginal Rate of Substitution

Both $A = (x_1^A, x_2^A)$ and $B = (x_1^B, x_2^B)$ lie on the indifference curve I, showing that the consumer is indifferent between consuming A or B. The change in good 2, $\Delta x_2 = (x_2^A - x_2^B) > 0$, accompanied by the change in good 1, $\Delta x_1 = (x_1^A - x_1^B) < 0$, leaves the consumer indifferent. We can regard the ratio $-\Delta x_2 / \Delta x_1$ as the rate at which the consumer is willing to substitute good 2 for good 1. That is, it indicates the trade-off between bundles of goods among which the consumer is indifferent. If the change Δx_1 is very small—a marginal change—then the ratio $-\Delta x_2 / \Delta x_1$ measures the marginal rate of substitution of good 2 for good 1.

Formally, the marginal rate of substitution of good 2 for good 1, MRS_{21} for short, along an indifference curve, is defined as the ratio $-\Delta x_2 / \Delta x_1$ i.e. the amount of good 2 that must be substituted for a loss of Δx_1 unit of good 1, if the loss Δx_1 of good 1 is marginal. That is,

$$MRS_{21} = -\Delta x_2 / \Delta x_1 \quad (2.5)$$

when Δx_1 is very small. Since Δx_1 is very small, the numerator Δx_2 of the ratio $\Delta x_2 / \Delta x_1$ is also small. This follows from the continuity of the indifference curve. The indifference curve has been drawn as a continuous curve of x_1 and x_2 . Therefore a very minor change Δx_1 in x_1 , will generate a very minor change Δx_2 in x_2 . It thus turns out that the ratio $\Delta x_2 / \Delta x_1$ is the slope of the indifference curve. The negative of the slope of the indifference curve, the marginal

rate of substitution, measures the rate at which the consumer is just on the margin of trading or not trading good 1 for good 2. At any rate of exchange other than the marginal rate of substitution, the consumer wants to trade one good for another. If the rate of exchange is equal to the marginal rate of substitution, he wants to stay put at the original consumption bundle. We can also say that MRS_{21} measures the consumer's marginal willingness to pay, because MRS_{21} says that the consumer is on the margin of being willing to pay some of good 1 in order to buy some more of good 2.

Let us now point out some important features of the marginal rate of substitution:

1. The marginal rate of substitution is positive.
2. We have noted that monotonic preferences imply that the indifference curves must be negatively sloped. Therefore, for monotonic preferences the marginal rate of substitution involves a reduction in the consumption of one good if the consumer wishes to get more of the other.
3. For strictly convex indifference curves, the marginal rate of substitution, the absolute value of the slope of the indifference curve decreases as x_1 increases. It takes less of good 2 to substitute for a given loss of good 1, more of good 1 that the consumer is consuming. Thus, the indifference curves exhibit a diminishing marginal rate of substitution. This is formalized as the Law of Diminishing Marginal Rate of Substitution:

LAW 2.2. As the number of units of a commodity consumed by a person increases, the marginal rate of substitution of any other commodity for the commodity under consideration decreases.

This asserts that the rate at which a consumer is willing to trade x_1 for x_2 decreases as x_1 increases. The intuitive reasoning behind this is quite clear. The more of one good and the less of another a consumer has, the more important units of the second become relative to units of the first. Therefore, the consumer is willing to give up more of the first good to get an extra unit of the second good.

We have demonstrated that monotonicity accompanied with strict convexity of a consumer's preference relation implies that the indifference curves are strictly convex to the origin, which in turn implies that the marginal rate of substitution is decreasing. In Appendix 2.1 we show that the converse is also true, that is, if the law of diminishing marginal rate of substitution holds, then the indifference curves become strictly convex to the origin.

Now, of two conditions (a) and (b), if (a) implies (b) and (b) implies (a), then in mathematical language we say that (a) and (b) are equivalent. Alternatively, we say that (a) holds *if and only if* (b) holds. (Here 'if' refers to the condition that [b] implies [a] and 'only if' refers to the requirement that [a] implies [b].) A third way of stating this equivalence is that (a) is *necessary and sufficient* for (b). (Condition [a] is sufficient for [b] means that [a] implies [b], while [a] is necessary for [b] means that [b] implies [a].)

We can formally state the interrelationship between strict convexity of the indifference curve and the marginal rate of substitution as:

THEOREM: *A consumer's indifference curves are strictly convex to the origin if, and only if, the law of diminishing marginal rate of substitution holds.*

SOME PARTICULAR CASES

In all the cases we discuss in this section, the indifference curves are not strictly convex to the origin. We begin by analyzing the case of perfect substitutes.

Perfect Substitutes

Two commodities are called perfect substitutes if the consumer is willing to trade one good for another at a constant rate. The simplest case of perfect substitution arises if the consumer substitutes one unit of one good, say good 2, for one unit of the other good, good 1. Examples of perfect substitutes are red pencils and yellow pencils, butter and margarine, sugar and saccharine, electric razors and safety razors, and many others.

Suppose two goods are perfect substitutes. Suppose also that one unit of one good can be traded for one unit of the other good. Consider a consumption bundle (5,5). Then any other consumption bundle that has a total of 10 units of the two goods will be indifferent to (5,5). For instance, the bundle (6,4) is indifferent to (5,5). This follows from the fact that if the consumer is at (5,5) and he increases his consumption of good 1 by one unit to six, he must give up one unit of good 2 to remain on the same indifference curve. Therefore, the quantity of the second good becomes four. Technically any consumption bundle (x_1, x_2) such that $x_1 + x_2 = 10$ will lie on the consumer's indifference curve through (5,5). Thus the indifference curves for this consumer are all parallel straight lines with a constant slope of -1 . Bundles with a larger total of the two goods are preferred to bundles with a smaller total. That is,

bundles for which $x_1 + x_2$ is higher is preferred to bundles with lower $x_1 + x_2$. We show this in Fig. 2.8. The marginal rate of substitution in this case is 1.

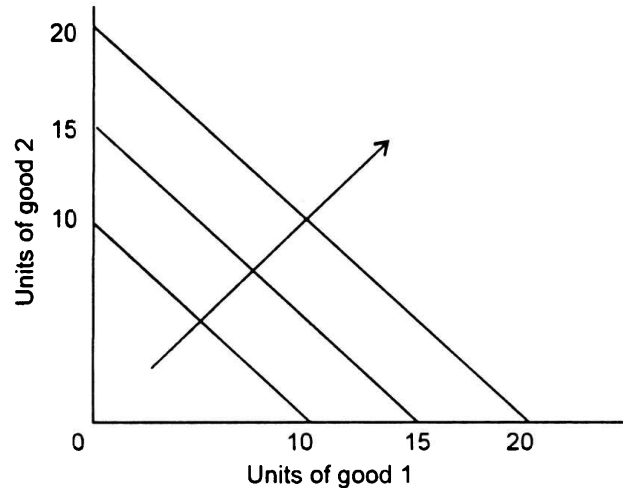


Fig. 2.8 Perfect Substitutes

The important thing about perfect substitution is that the indifference curves have a constant slope and hence the marginal rate of substitution is a constant. In general, the graph of an equation of the type $ax_1 + bx_2 = c$ where a, b, c are positive constants, is an indifference curve with constant slope $-a/b$. Therefore, the marginal rate of substitution of good 2 for good 1 is a/b . That is, the consumer is willing to give up b units of good 1 to get a units of good 2.

The indifference curves for perfect substitutes are convex to the origin, but not strictly so. That is, the preference relations satisfy convexity, but not *strict* convexity. A preference relation $\succeq$ is called convex if for any $(x_1, x_2) \sim (y_1, y_2)$ and for any t , such that $0 \leq t \leq 1$, $[tx_1 + (1-t)y_1, tx_2 + (1-t)y_2] \succeq (x_1, x_2)$. This means that any weighted average of two indifferent bundles (x_1, x_2) and (y_1, y_2) is at least as good as (x_1, x_2) , and hence (y_1, y_2) also.

Perfect Complements

Two commodities are said to be perfect complements if they are always consumed together in a fixed ratio. Suppose a consumer buys two goods, 1 and 2, in 1:1 ratio. Given that the consumer consumes one unit of good 1, then any amount of good 2 in excess of one unit is simply superfluous for the consumer, that is, he is indifferent between having one unit of good 2 and

any amount greater than that. Therefore, the consumer is indifferent between the bundles (1,1) and (1,2). Similarly, given that the consumer consumes one unit of good 2, any amount of good 1 in excess of one unit becomes needless. Thus, the consumer is indifferent between the bundles (1,1) and (2,1). These observations lead to the fact that the indifference curves are L-shaped, with the vertex of L occurring at the point where the number of units of the first good equals the number of units of the second good. This is shown in Fig. 2.9. An example where the two goods are bought in 1:1 ratio is, good 1 is a right shoe and good 2 is a left shoe.

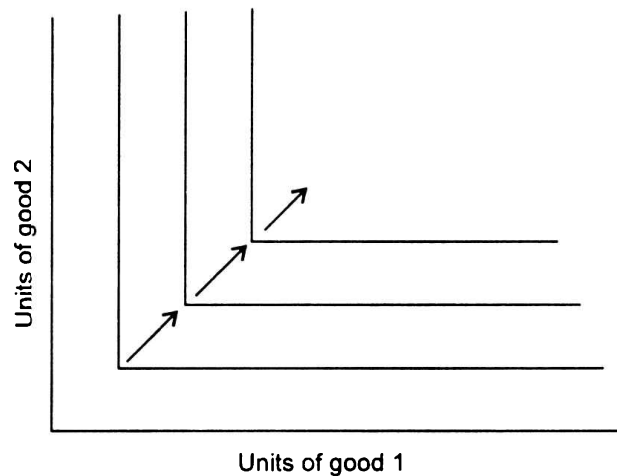


Fig. 2.9 Perfect Complements

The important thing about perfect complements is that the goods are bought in a fixed proportion, but the proportion is not necessarily one-to-one. If a consumer buys two goods in a fixed proportion, the indifference curves will be L-shaped. The corners of L will indicate the proportion in which the two goods are bought. Increasing the number of good 1 and the number of good 2 by the same percentage will move the consumer to a more preferred position. For instance, if a person uses two teaspoons of sugar in one cup of tea, then he will use four teaspoons of sugar in two cups of tea. In this case the consumer uses sugar and tea in an unequal ratio. When he increases tea from one cup to two cups, the amount of sugar increases from two teaspoons to four teaspoons. Thus, both sugar and tea increase by 100 percent and the consumer moves to a higher indifference curve.

In addition to the two examples considered above, sets of complementary goods include tennis rackets and tennis balls, bread and jelly, automobiles and gasoline, and many more.

Bads

A *bad* is a commodity which the consumer does not like. Suppose that a consumer likes to have an open balcony in his house. The balcony however becomes dusty and the consumer dislikes dust. That is, dust is a bad. The amount of dust increases with the size of the balcony. There is some trade-off between the size of the balcony and the amount of dust. That is, some size of the balcony would compensate the consumer for having to tolerate a given amount of dust. How do we draw the indifference curves here?

Consider a bundle (x_1, x_2) , where x_1 indicates the amount of dust and x_2 denotes the size of the balcony. Suppose we increase the size of the balcony. What do we have to do with the amount of dust to keep the consumer on the same indifference curve? Clearly, the consumer has to tolerate some extra amount of dust. Thus, the consumer's indifference curves will slope up and to the right. Obviously, the direction of increasing preference is toward the direction of increased balcony size and decreased dust amount. We show this explicitly in Fig. 2.10.

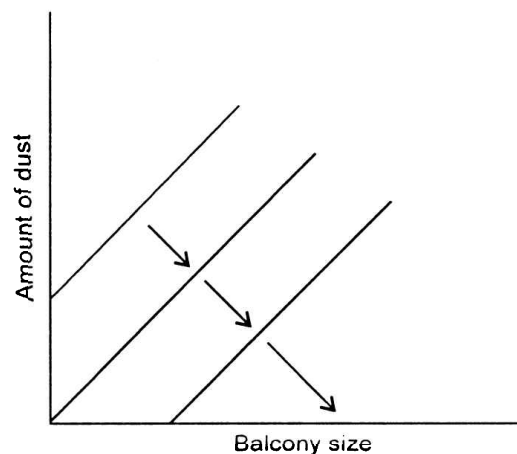


Fig. 2.10 Bads

Neutrals

A good is called a *neutral* if the consumer is completely indifferent about it. That is, he does not care about it one way or the other. Consider a miser's choice between money and 'all other goods'. The miser cares only about the amount of money and does not care at all about the quantities of all other goods. The more money he has the better he feels, but adding more of all other goods does not affect him at all. Thus, 'all other goods' is a neutral here.

If the horizontal axis represents the amount of money and the vertical axis denotes the quantities of all other goods, then his indifference curves will be vertical lines as shown in Fig. 2.11.

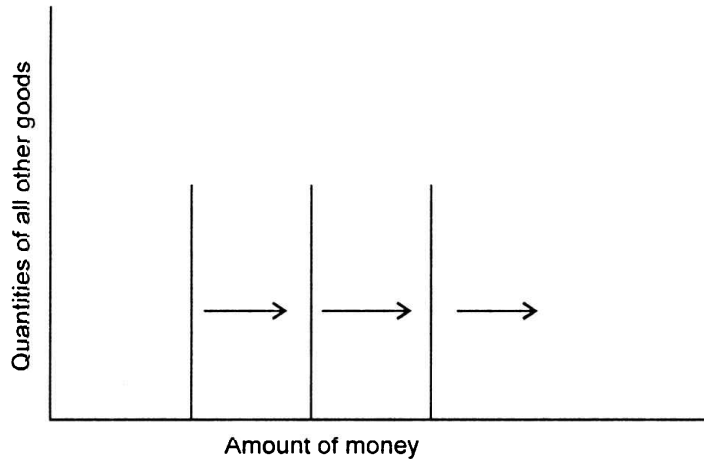


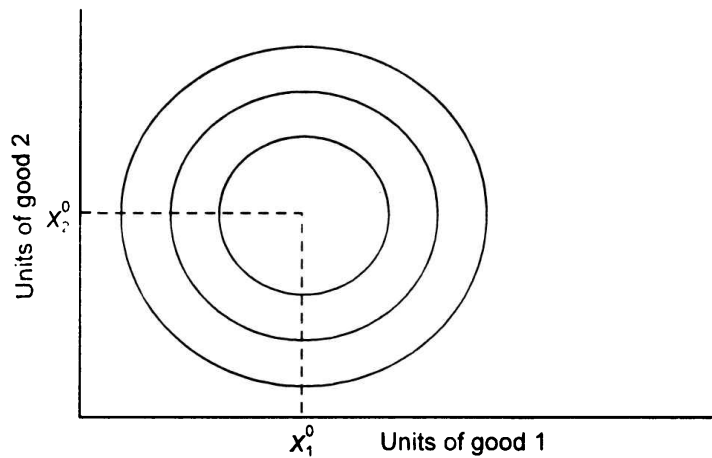
Fig. 2.11 Neutrals

Satiation

For many goods, a consumer may become satiated if he consumes enough of it. Suppose for a consumer there is some most preferred bundle of goods (x_1^0, x_2^0) , and the more are the deviations of amounts of consumption of the two goods from this bundle, the worse off he is. Then the point (x_1^0, x_2^0) is called a bliss point or a point of satiation. Clearly, the points farther away from the bliss point will lie on indifference curves that represent lower preference levels.

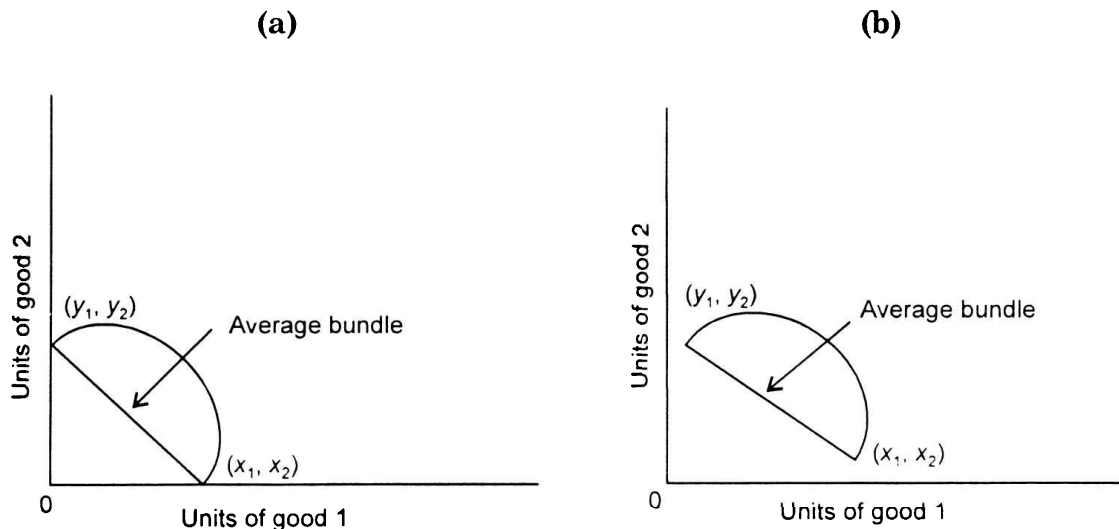
As we have indicated earlier, satiation may occur temporarily with food items. Consider two food items, rice and fish. A consumer may like to have some optimal amount of rice and fish every day. Any amount less than that would make him worse off, but any amount more than that also would make him worse off.

Thus, if he has too little of the two goods, he would like to have more of both. When he has too much of both goods, he will demand smaller amounts of them. Therefore, if he has too much of both goods, they are bads. Similarly, too much of one good also means that it is a bad. The indifference curves are negatively sloped if he has too little or too much of both goods, and positively sloped if he has too much of one good.

**Fig. 2.12 Satiation**

Strictly Concave Preference Relations

Consider a consumer who has preference for two goods, 1 and 2. The consumer likes both the goods but does not like to have them together. In case he needs to buy them together he will strictly prefer extreme points to any weighted average. That is, suppose he is indifferent between the two bundles $(3,2)$ and $(1,6)$. Then he will strictly prefer $(3,2)$ (and hence $(1,6)$) to $(2,4)$ or any other weighted average. Such preference relations are called strictly concave. Formally, for any two distinct bundles (x_1, x_2) and (y_1, y_2) if $(x_1, x_2) \sim (y_1, y_2)$, then $(x_1, x_2) \succ [tx_1 + (1-t)y_1, tx_2 + (1-t)y_2]$ for all $0 < t < 1$. We depict this kind of preferences in Fig. 2.13.

**Fig. 2.13 Strictly Concave Preference**

In this case the indifference curves become strictly concave to the origin. Similarly, for concave preference relations, the indifference curves are concave to the origin but need not be strictly so. Concavity of a preference relation $\succeq$ demands that for any (x_1, x_2) and (y_1, y_2) if $(x_1, x_2) \sim (y_1, y_2)$, then $(x_1, x_2) \succeq [tx_1 + (1 - t)y_1, tx_2 + (1 - t)y_2]$ for all $0 \leq t \leq 1$.

Discrete Goods

If commodities are only available in discrete units, the feasible consumption bundles consist of isolated points in the $x_1 - x_2$ plane. Suppose x_1 is the quantity demanded of a good and these quantities are available only in integral amounts. An example is a TV. Let x_2 denote the money to be spent on other goods. Then the bundles indifferent to a given bundle will be a set of discrete points. Here the points A, B and C are indifferent. Similarly, D, E and F are also indifferent. However, the consumer prefers D (hence E and F) to A, B or C . From the figure it follows that the set of bundles that are at least as good as a particular bundle will be a set of line segments.

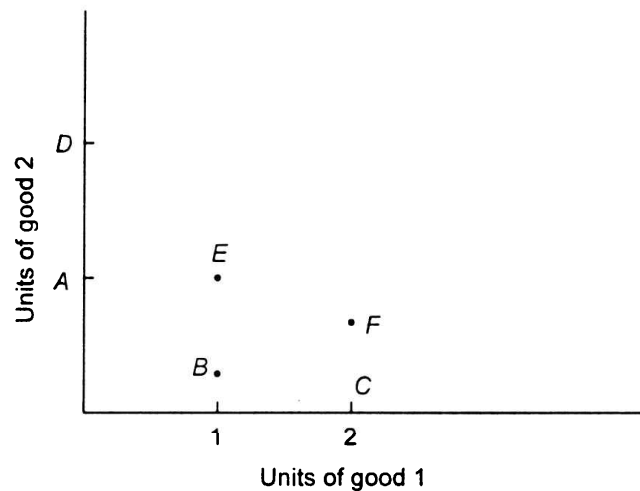


Fig. 2.14 Discrete Goods

ORDINAL UTILITY

We are now in a position to discuss the modern approach to utility. It has been argued that the indifference map can be used to indicate the order of a consumer's preference among alternative commodity bundles. This forms the basis of the modern approach to utility theory.

In the modern approach, utility is seen only as a way of describing preferences. To illustrate this explicitly, suppose we attach an identifying num-

ber to each indifference curve. The numbers attached to different indifference curves should be such that a higher curve in the preference ranking gets a higher number. That is, the ordering of numbers should correspond to the ordering of indifference curves in terms of preference. For instance, in Fig. 2.6 we may attach the number 2 to the indifference curve I_1 , 5 to the indifference curve I_2 and 7 to the indifference curve I_3 , and so on.

Now, let us associate with each point on I_1 the number 2, with each point on I_2 number 5, and with each point on I_3 the number 7. In modern utility theory, 2 is then called the utility of any commodity bundle lying on the curve I_1 , 5 is the utility of any bundle lying on the curve I_2 , and 7 is the utility of any bundle that lies on I_3 .

Thus, a utility function is a way of labelling indifference curves. Since all the bundles on an indifference curve must have the same utility, a utility function is a way of assigning numbers to indifference curves such that higher indifference curves get assigned larger numbers. This means that the only property of a utility assignment in which we are interested is how it orders different bundles of goods. The size of the utility difference between any two bundles of goods does not really matter. This kind of utility is referred to as *ordinal utility* because it is concerned with ordering of commodity bundles. This utility is nothing more than an assignment of numbers to bundles in a way that reflects a consumer's preferences.

Symbolically, any commodity bundle (x_1, x_2) is strictly preferred to another bundle (y_1, y_2) if and only if the utility $U(x_1, x_2)$ of (x_1, x_2) is larger than the utility $U(y_1, y_2)$ of (y_1, y_2) . More technically $(x_1, x_2) \succ (y_1, y_2)$ if and only if $U(x_1, x_2) > U(y_1, y_2)$.

If $U(x_1, x_2)$ is a way of assigning utility numbers to (x_1, x_2) , then any increasing transformation $f[U(x_1, x_2)]$ of $U(x_1, x_2)$ is also a way of assigning utility numbers to (x_1, x_2) . Increasing transformations are essentially increasing functions. Examples of such functions are $f(U) = aU + b$ (where $a > 0$ and b are constants), $f(U) = U^5$, $f(U) = e^U$, and so on. Increasing transformations require that if the value of U increases, then the value of $f(U)$ also increases.

We can establish rigorously the above claim that if a utility function U represents some preferences $\succ$, then any increasing transformation $f(U)$ of U also represents the same preferences. The demonstration runs as follows:

- a) U represents some preferences means that $(x_1, x_2) \succ (y_1, y_2)$ if and only if $U(x_1, x_2) > U(y_1, y_2)$.

- b) If f is an increasing transformation, then $U(x_1, x_2) > U(y_1, y_2)$ if and only if $f[U(x_1, x_2)] > f[U(y_1, y_2)]$.
- c) From (a) and (b) we have $(x_1, x_2) \succ (y_1, y_2)$ if and only if $f[U(x_1, x_2)] > f[U(y_1, y_2)]$. Hence $f(U)$ represents the same preferences as U .

To illustrate the problem numerically, let (x_1, x_2) and (y_1, y_2) be two bundles such that $U(y_1, y_2) = 3$ and $U(x_1, x_2) = 4$, so that $U(x_1, x_2) > U(y_1, y_2)$. Consequently, $(x_1, x_2) \succ (y_1, y_2)$. Now suppose $f(U) = U^3$. then $f[U(y_1, y_2)] = 27$ and $f[U(x_1, x_2)] = 64$. Note that $f[U(y_1, y_2)]$ and $f[U(x_1, x_2)]$ are just the renumbering of utility levels $U(y_1, y_2) = 3$ and $U(x_1, x_2) = 4$, ensuring the inequality $f[U(x_1, x_2)] > f[U(y_1, y_2)]$. Therefore, it follows that $(x_1, x_2) \succ (y_1, y_2)$ for the transformed utility values also. We can easily establish the reverse side as well.

It may be important to note that in many texts the term ‘monotonic transformation’ is used in place of ‘increasing transformations’. But decreasing transformations can also be regarded as monotonic transformations. Now, given that U represents some preference, only increasing transformations of U represent the same preference. Therefore, since we are interested only in increasing transformations of U , in order to avoid confusion we will use the term ‘increasing transformation’ rather than ‘monotonic transformation’.

CONSTRUCTION OF A UTILITY FUNCTION

We have assumed that a consumer’s preferences can be summarized by means of a utility function. But does there exist such a utility function? That is, is there a utility function that describes any reasonable preference relation?

It has been demonstrated that if the preference ordering is complete, reflexive, transitive, monotonic and continuous, then there exists a continuous, increasing utility function U which represents the preference. To understand the continuity of a preference ordering, suppose $(x_1, x_2) \succ (y_1, y_2)$ and $(y_1, y_2) \succ (z_1, z_2)$. Then continuity requires any curve connecting (x_1, x_2) and (z_1, z_2) to cross the indifference curve containing (y_1, y_2) . That is, there are no sudden jumps from being better than (y_1, y_2) to being worse than (y_1, y_2) . The most important consequence of continuity is that if (x_1, x_2) is strictly preferred to (v_1, v_2) and (y_1, y_2) is quite close to (x_1, x_2) , then (y_1, y_2) is also strictly preferred to (v_1, v_2) . Continuity of U demands that small changes in x_1 and x_2 will generate small changes in $U(x_1, x_2)$. Thus, minor observational errors in the quantities of two goods demanded by the consumer will not make $U(x_1, x_2)$ oversensitive. The increasingness of U means that if we in-

crease x_1 keeping x_2 fixed, then $U(x_1, x_2)$ increases. Similarly, if we fix x_1 and increase x_2 , then $U(x_1, x_2)$ increases. That is, if the consumer increases the consumption of one good, keeping the consumption of another good fixed, then his utility increases. Equivalently, we say that the consumer is not satiated with any commodity. Clearly, the increasingness of the utility function parallels the monotonicity of the preference ordering.

EXAMPLES OF UTILITY FUNCTIONS

We have seen how preferences can be represented by utility functions. If we are given a utility function U , it is easy to derive the indifference curves. We just plot all the points (x_1, x_2) such that $U(x_1, x_2) = k$, where k is a constant.

Example 1

A very commonly used utility function is the Cobb-Douglas utility function

$$U(x_1, x_2) = x_1^\alpha x_2^\beta \quad (2.6)$$

where α and β are positive constants that describe the preferences of the consumer. Any increasing transformation of the Cobb-Douglas utility function will represent the same preferences. Let us consider some transformations of this type.

First, let us consider the transformation $V_1(x_1, x_2) = \log U(x_1, x_2)$. Since the function $f(z) = \log z$, $z > 0$ is an increasing function of z , U and V_1 represent the same preferences.

Note that

$$V_1(x_1, x_2) = \alpha \log x_1 + \beta \log x_2 \quad (2.7)$$

This function V_1 is linear in logarithms of the quantities consumed.

An alternative of interest arises if we consider the transformation $V_2(x_1, x_2) = [U(x_1, x_2)]^{\frac{1}{\alpha+\beta}}$.

Since α and β are positive, V_2 is obtained from U through an increasing transformation.

$$V_2(x_1, x_2) = x_1^{\frac{\alpha}{\alpha+\beta}} x_2^{\frac{\beta}{\alpha+\beta}}$$

$$= x_1^c x_2^{1-c} \quad (2.8)$$

where $c = \frac{\alpha}{\alpha + \beta}$. Therefore we can always take an increasing transformation of the Cobb-Douglas utility function that make the exponents add up to one. The above examples show that the Cobb-Douglas utility function can be expressed in a variety of ways.

Let us now use a particular case of the utility function Eq. (2.8) to see how a utility function can be employed for constructing indifference curves. For this, put $c = \frac{1}{2}$ in Eq. (2.8). Then we have $V_2(x_1, x_2) = \sqrt{(x_1 x_2)}$. A typical indifference curve is a plot of all points (x_1, x_2) such that $\sqrt{(x_1 x_2)} = I$, where $I > 0$ is a constant. Solving for x_2 as a function of x_1 , we have

$$\sqrt{(x_2)} = \frac{I}{\sqrt{(x_1)}} \quad (2.9)$$

The function $\sqrt{x_1 x_2} = I$ is a rectangular hyperbola. We depict the function using the reformulation Eq. (2.9) for $I = 1, 2, 3, \dots$ in Fig. 2.15.

Now, let us consider the increasing transformation $V_3(x_1, x_2) = [V_2(x_1, x_2)]^2$ of the utility function $V_2(x_1, x_2) = \sqrt{(x_1 x_2)}$, that is $V_3(x_1, x_2) = x_1 x_2$. The labelling of indifference curves will then be 1, 4, 9, ... in place of 1, 2, 3, ... But $V_3(x_1, x_2)$ describes the same preferences as $V_2(x_1, x_2)$ since they order all the bundles in the same way. For instance, the set of bundles that has $V_2(x_1, x_2) = \sqrt{x_1 x_2} = 2$ is exactly the same set of bundles that has $V_3(x_1, x_2) = x_1 x_2 = 4$.

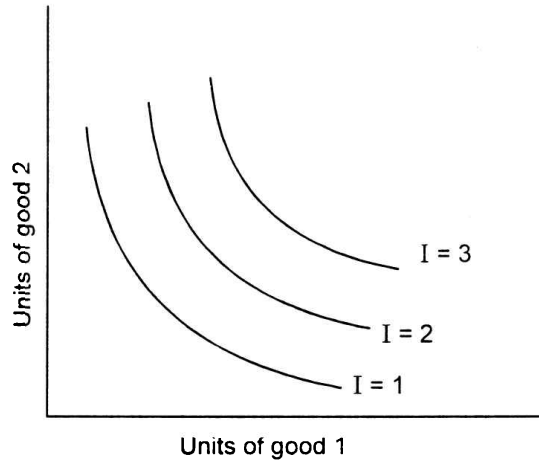


Fig. 2.15 Indifference Curves

Example 2: Perfect Substitutes

For perfect substitutes the indifference curves are straight lines. For such goods, the marginal rate of substitution is a constant. As stated earlier, if the consumer is willing to trade b units of good 1 for a units of good 2, then the marginal rate of substitution is a/b . In this case the consumer's preferences are represented by the utility function

$$U(x_1, x_2) = ax_1 + bx_2 \quad (2.10)$$

where a and b are positive constants. This is the general form of utility function when goods 1 and 2 are perfect substitutes.

Example 3: Perfect Complements

Suppose goods 1 and 2 are perfect complements and they are purchased in 1:1 ratio. In this case the bundles (1, 1), (2, 1) and (1, 2) are indifferent. Therefore, the number of complete bundles of goods (for instance, pairs of shoes) we have the minimum of the number of units of good 1 (right shoes) and the number of units of good 2 (left shoes). Thus, the utility function for perfect complements, when the goods are bought in 1:1 ratio, is $U(x_1, x_2) = \min \{x_1, x_2\}$.

Suppose the two goods are consumed in an unequal ratio. Then the utility function can be $U(x_1, x_2) = \min \{x_1/2, x_2\}$. For instance, let x_2 be the number of cups of tea and x_1 be the number of teaspoons of sugar available. The consumer uses two teaspoons of sugar in one cup of tea. If the number of cups of tea is greater than half the number of teaspoons of sugar available, then we cannot put two teaspoons of sugar in each cup. Therefore, the number of cups that can be correctly sweetened is $x_1/2$. Hence the utility function is of the form $U(x_1, x_2) = \min \{x_1/2, x_2\}$. If we multiply this utility function by 2, then the resulting utility function, which is given by $V(x_1, x_2) = 2U(x_1, x_2) = 2 \min \{x_1/2, x_2\} = \min \{x_1, 2x_2\}$, represents the same preferences.

In general, when goods 1 and 2 are perfect complements, the utility function is of the form

$$U(x_1, x_2) = \min \{ax_1, bx_2\} \quad (2.11)$$

where the positive numbers a and b denote the proportion in which the goods are consumed.

NON—EXISTENCE OF A UTILITY FUNCTION

Consider the lexicographic ordering $\succ l$ defined by

$$(x_1, x_2) \succ l (y_1, y_2) \text{ if either (a) } x_1 > y_1$$

$$\text{or} \quad (b) \ x_1 = y_1 \text{ and } x_2 > y_2. \quad (2.12)$$

That is, between any two commodity bundles $x = (x_1, x_2)$ and $y = (y_1, y_2)$, the consumer strictly prefers the one with more of good 1. If two bundles have the same quantity of good 1, then he strictly prefers the bundle containing more of good 2. This ordering is called lexicographic since it resembles that found in a dictionary. For a lexicographic ranking, the utility function does not exist because there are no distinct consumption bundles that are indifferent to one another.

Let us consider all bundles containing c amount of good 1. Each point on the line drawn upward from c is strictly preferred to any point below it (since points below it have the same quantity of good 1 but less of good 2). Therefore, large utility numbers must be assigned to higher points on this line. Similarly, a smaller utility number must be given to any point to the left of this line. (Refer Fig. 2.16.)

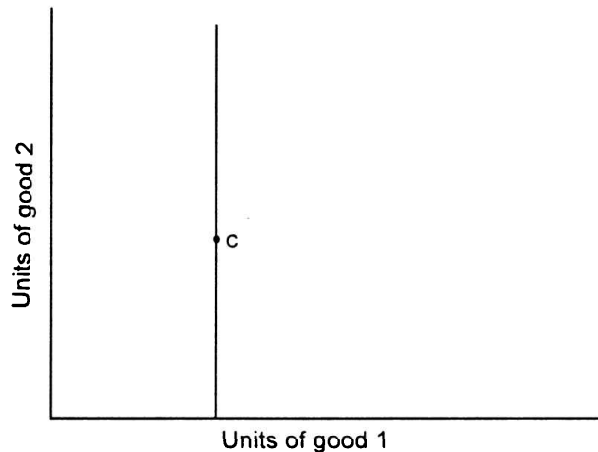


Fig. 2.16 Lexicographic Ordering

To demonstrate discontinuity of the lexicographic ranking, let us consider three bundles $x = (x_1, x_2)$, $y = (y_1, y_2)$ and $z = (z_1, z_2)$. In panel (a) of Fig. 2.17, we note that $x \succ y$ and $y \succ z$. Therefore, by transitivity $x \succ z$. As consumption moves from x to z , we move from more preferred to less preferred bundles in a continuous manner. In the process of movement we pass through a point indifferent to y . Next, in panel (b) of the figure, we have $x \succ l y$ and $y \succ l z$.

Therefore $x \succ l z$. In our movement from x to z we skip all the bundles above the point z on the line drawn up from c , including the bundle y . Thus, in the lexicographic case while passing from a more preferred bundle x to a less preferred bundle z , we are passing through neither the intermediate bundle y , nor any bundle which is indifferent to y . This therefore establishes discontinuity of the lexicographic ordering.

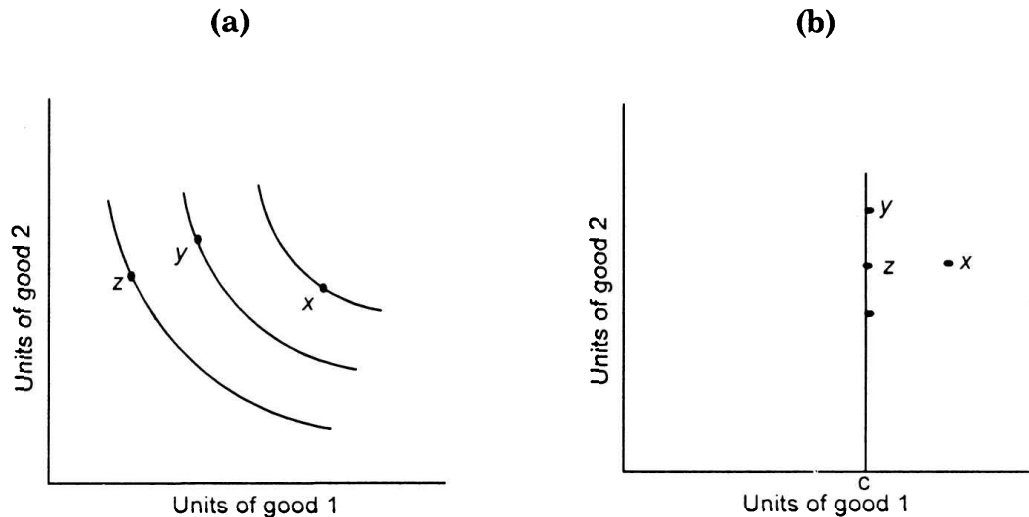


Fig. 2.17 Discontinuity of Lexicographic Ordering

The following is an example of a lexicographic ordering: suppose a person likes honesty, and among honest people he likes those with a sense of humor. Let us suppose that good 1 is the level of honesty, and good 2 is the sense of humor. Then a diagrammatic exposition shows that this ordering is a lexicographic ordering.

MARGINAL UTILITY

Consider a person who consumes two goods, 1 and 2. The rate of change in the person's utility associated with a small change in the amount of good 1, keeping the amount of good 2 fixed, is called his marginal utility with respect to good 1. More precisely, let (x_1, x_2) be a bundle of goods. Then the change in utility when x_1 changes by a small amount Δx_1 , keeping x_2 fixed, is given by $U(x_1 + \Delta x_1, x_2) - U(x_1, x_2) = \Delta U$. Therefore, the rate of change in utility when the quantity of good 1 changes by Δx_1 is

$$MU_1 = \frac{\Delta U}{\Delta x_1} = \frac{U(x_1 + \Delta x_1, x_2) - U(x_1, x_2)}{\Delta x_1} \quad (2.13)$$

Since Δx_1 is marginal or infinitesimal, MU_1 is called the marginal utility with respect to good 1. Clearly, MU_1 is the change in utility when consumption of good 1 changes by one unit, assuming that Δx_1 is very small.

From (2.13) it follows that the change in utility associated with a small change Δx_1 in the consumption of good 1 is

$$\Delta U = MU_1 \Delta x_1, \quad (2.14)$$

the product of the marginal utility of good 1 and the change in consumption.

Similarly, the marginal utility of good 2 is defined as

$$MU_2 = \frac{\Delta U}{\Delta x_2} = \frac{U(x_1, x_2 + \Delta x_2) - U(x_1, x_2)}{\Delta x_2} \quad (2.15)$$

In calculating the marginal utility of good 2, we keep the quantity of good 1 fixed. The change in utility resulting from a small change Δx_2 in the quantity of consumption of good 2 is

$$\Delta U = MU_2 \Delta x_2. \quad (2.16)$$

Clearly, the magnitude of marginal utility depends on the functional form of utility. Since a consumer's preference ordering is concerned with ranking of different bundles, we cannot calculate marginal utility from preference structure. For calculating marginal utility we need a specified form of utility function. (A calculus treatment of marginal utility is available in Appendix 2.1.)

MARGINAL RATE OF SUBSTITUTION AND MARGINAL UTILITY

We need to show how marginal utility becomes helpful in calculating the marginal rate of substitution. The marginal rate of substitution of good 2 for good 1, MRS_{21} , is the rate at which the consumer is willing to trade good 1 for good 2 along an indifference curve.

We recall that given a consumer's utility function U , an indifference curve can be obtained by plotting all points (x_1, x_2) such that $U(x_1, x_2) = k$, where $k > 0$ is the constant level of utility along the indifference curve. Now, consider a change in the consumption of the two goods $(\Delta x_1, \Delta x_2)$ that keeps utility constant. That is, the consumer moves from the original bundle (x_1, x_2) to the new bundle $(x_1 + \Delta x_1, x_2 + \Delta x_2)$ along the indifference curve. Then, since $k > 0$ is a constant, we must have

$$MU_1 \Delta x_1 + MU_2 \Delta x_2 = 0 \quad (2.17)$$

from which it follows that

$$MRS_{21} = -\frac{\Delta x_2}{\Delta x_1} = \frac{MU_1}{MU_2} \quad (2.18)$$

We have noted that the magnitude of the marginal utility function will depend on the form of the utility function. Thus, if we multiply the utility by some constant $b > 0$, marginal utility also gets multiplied by b . But the ratio of marginal utilities remains unaltered if the utility function is subjected to any increasing transformation. That is, MRS_{21} , the ratio of marginal utilities, is independent of any particular transformation of the utility function we wish to choose. Now, any increasing transformation just relabels the indifference curves in order preserving manner. Therefore, even though the marginal utilities get altered under increasing transformations, the marginal rate of substitution, the ratio of marginal utilities, remains independent of the particular way of representation of preferences. (See Appendix 2.1 for a rigorous demonstration of this.)

We have demonstrated earlier that if a consumer's preferences satisfy monotonocity and further, if they meet strict convexity, then the Law of Diminishing Marginal Rate of Substitution holds. Since in Eq. (2.18) we have expressed the marginal rate of substitution in terms of marginal utilities, we should be able to develop a condition for the marginal rate of substitution to decline involving the consumer's utility function.

In fact, increasingness and strict quasiconcavity of a consumer's utility function ensure that the Law of Diminishing Marginal Rate of Substitution holds. Exhibiting this interrelationship involves some degree of mathematical complication, and it will therefore be pursued in the Appendix. We formally state this result as:

Theorem: If the consumer's utility function is increasing and strictly quasiconcave then his indifference curves will be strictly convex to the origin, that is, the Law of Diminishing Marginal Rate of Substitution will hold.

An alternative simpler proof of the theorem, which is given as follows, will probably explain the relevance of the theorem in greater detail.

Alternative Proof: First, note that the increasingness of the utility function makes the indifference curve negatively sloped. To demonstrate this explicitly, we may employ arguments similar to those used for showing that monotonocity of a preference relation generates a negatively sloped indifference curve (see the section 'Indifference Map').

Next, a consumer's utility function is strictly quasiconcave if for any two distinct commodity bundles (x_1, x_2) and (y_1, y_2) , the weighted average $[tx_1 + (1 - t)y_1, tx_2 + (1 - t)y_2]$, where $0 < t < 1$ is arbitrary, generates higher utility than the minimum of utility values $U(x_1, x_2)$ and $U(y_1, y_2)$. That is,

$$U[tx_1 + (1 - t)y_1, tx_2 + (1 - t)y_2] > \min [U(x_1, x_2), U(y_1, y_2)].$$

Now, suppose (x_1, x_2) and (y_1, y_2) are on the same indifference curve. That is, $U(x_1, x_2) = U(y_1, y_2) = (\text{say}) c$. Then strict quasiconcavity means that $U[tx_1 + (1 - t)y_1, tx_2 + (1 - t)y_2] > c$. That is, the line joining (x_1, x_2) and (y_1, y_2) lies entirely above the indifference curve containing (x_1, x_2) and (y_1, y_2) . Thus, the indifference curves associated with the strictly quasiconcave utility function cannot contain any flat section.

Therefore increasingness and strict quasiconcavity of a utility function show that its indifference curves are strictly convex to the origin. This completes the proof of the theorem.

The above alternative proof shows that using the definitions of increasingness and strict quasiconcavity of a utility function, we generate negatively sloped strictly convex indifference curves. It also reveals that strict convexity of a consumer's preference relation is analogous to strict quasiconcavity of his utility function.

Now suppose the indifference curve has a flat shape. That is, it is of the form when the two goods are perfect substitutes (see Fig. 2.8). This indifference curve cannot be associated with a strictly quasiconcave utility function. In fact, the utility function corresponding to this indifference curve is quasiconcave, which demands that for any (x_1, x_2) , (y_1, y_2) and for any t , $0 \leq t \leq 1$.

$$U[tx_1 + (1 - t)y_1, tx_2 + (1 - t)y_2] \geq \min[U(x_1, x_2), U(y_1, y_2)].$$

That is, the utility function for perfect substitutes, which is of the form $U(x_1, x_2) = ax_1 + bx_2$, where a and b are positive constants, is quasiconcave, but not strictly so.

Finally, as we have stated in the mathematics chapter, there is an important characteristic of quasiconcave utility functions. If the utility function U is quasiconcave (strictly quasiconcave), then $f(U)$, where f is increasing, is also quasiconcave (strictly quasiconcave). That is, quasiconcavity/strict quasiconcavity remains preserved under increasing transformations.

Example 2.1

Let us consider the Cobb-Douglas utility function

$$U(x_1, x_2) = x_1^\alpha x_2^\beta$$

where α and β are positive constants. We want to determine the marginal utilities and marginal rate of substitution associated with this utility function.

The marginal utility of good 1 is

$$MU_1 = \frac{\delta U(x_1, x_2)}{\delta x_1} = \alpha x_1^{\alpha-1} x_2^\beta. \quad (2.19)$$

The marginal utility of good 2 is

$$MU_2 = \frac{\delta U(x_1, x_2)}{\delta x_2} = \beta x_1^\alpha x_2^{\beta-1}. \quad (2.20)$$

Let $I > 0$ be some constant. The indifference curve, along which the constant utility level is I , is the locus of all (x_1, x_2) , such that

$$x_1^\alpha x_2^\beta = I.$$

Differentiating totally, we have

$$\frac{\delta(x_1^\alpha x_2^\beta)}{\delta x_1} dx_1 + \frac{\delta(x_1^\alpha x_2^\beta)}{\delta x_2} dx_2 = dI.$$

Since I is a constant, the above equation becomes

$$\alpha x_1^{\alpha-1} x_2^\beta dx_1 + \beta x_1^\alpha x_2^{\beta-1} dx_2 = 0$$

from which it follows that

$$MRS_{21} = -\frac{dx_2}{dx_1} = \frac{\alpha x_1^{\alpha-1} x_2^\beta}{\beta x_1^\alpha x_2^{\beta-1}} = \frac{\alpha x_2}{\beta x_1}. \quad (2.21)$$

Taking the ratio between MU_1 in Eq. (2.19) and MU_2 in Eq. (2.20), we verify that $MRS_{21} = MU_1 / MU_2$.

Now, suppose the Cobb-Douglas utility function is subjected to a logarithmic transformation:

$$V(x_1, x_2) = \log [U(x_1, x_2)] = \alpha \log x_1 + \beta \log x_2.$$

In this case $MU_1 = \alpha/x_1$, $MU_2 = \beta/x_2$. Thus, these two marginal utilities are different from those given by Eqs (2.19) and (2.20). But the MRS_{21} for the logarithmic utility is $\alpha x_2 / \beta x_1$, which is the same as MRS_{21} in (2.21). Therefore, though marginal utilities change under increasing transformations of utility, the marginal rate of substitution remains invariant.

Applied Example 2.1 Cardinal Utility and Charitable Donation

Charitable activities can be regarded as a non-market source of satisfaction to many persons. Charity can be defined as an intentional trade-off of resources at negligible prices or even zero prices. The main objective of charity is to improve the position of the recipient(s).

It is well-known that charitable activity increases after disasters like earthquakes, floods, famines, war, epidemic, etc. There is enough evidence that following such sudden misfortunes non-affected persons tend to supply more food, medicines, shelter and other articles to the victims at below-market prices or zero prices than they are willing to supply before the misfortunes occurred. Why does charity increase after such events?

This happens because the donors obtain additional utility not only when their income increases, but also when the recipients become better off. In panel (a) of Fig. 2.18, the donor's total utility curve U_D^1 shows the utility he obtains from charitable donations to the recipient whose initial income is m_R^1 . The curve U_D^2 shows the donor's utility from giving to the recipient whose income is now at a higher level m_R^2 . Both increase at decreasing rates, reflecting that the donor's marginal utility from charitable giving is decreasing. The associated marginal utility curves are shown in panel (b). From panel (b) it emerges that the worse off the recipient is, the more contributions the donor will be willing to make before becoming satiated with gifts. The point of saturation for the donor is lower if the recipient's income is higher. Clearly this seems quite reasonable. We denote the points of saturation of the donor by g_1 and g_2 when the recipients' incomes are m_R^1 and m_R^2 respectively. The increment on U_D^1 per unit of increment of gift exceeds, on the average, the increments on U_D^2 per unit of increment of gift up to g_1 . Equivalently, we say that for any level of contribution below g_1 , the marginal utility per unit of contribution, shown in panel (b), is greater the worse off the recipient is, that is MU_R^1 lies above MU_R^2 .

The above discussions explain why charitable donations may increase following some disastrous events. Such donations increase the satisfaction of the donors because they provide assistance to more needy persons. There is even a possibility of friends or relatives of the contributors becoming recipients. In some countries, like India, charitable donations are encouraged by allowing the donors to enjoy some tax deductions.

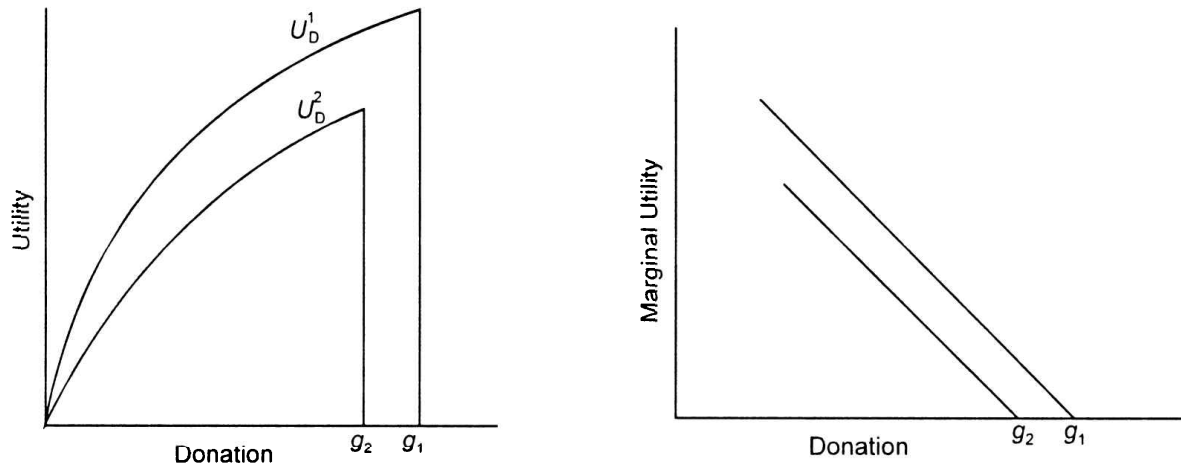


Fig. 2.18 Cardinal Utility and Donations

Appendix 2.1

MARGINAL UTILITY AND MARGINAL RATE OF SUBSTITUTION

In this appendix we first calculate the marginal utilities and the marginal rate of substitution using calculus. The marginal utility of good 1,

$$\begin{aligned} MU_1 &= \lim_{\Delta x_1 \rightarrow 0} \frac{U(x_1 + \Delta x_1, x_2) - U(x_1, x_2)}{\Delta x_1} \\ &= \frac{\delta U(x_1, x_2)}{\delta x_1}. \end{aligned}$$

Similarly the marginal utility of good 2, $MU_2 = \frac{\delta U(x_1, x_2)}{\delta x_2}$.

To determine the marginal rate of substitution, let

$$U(x_1, x_2) = I$$

where $I > 0$ is a constant. Differentiating totally, we get

$$\frac{\delta U(x_1, x_2)}{\delta x_1} dx_1 + \frac{\delta U(x_1, x_2)}{\delta x_2} dx_2 = 0.$$

Therefore, the marginal rate of substitution of good 2 for good 1

$$MRS_{21} = -\frac{dx_2}{dx_1} = \frac{\delta U(x_1, x_2) / \delta x_1}{\delta U(x_1, x_2) / \delta x_2} = \frac{MU_1}{MU_2}$$

STRICTLY CONVEX INDIFFERENCE CURVES AND MARGINAL RATE OF SUBSTITUTION

We are required to show that strict convexity of the indifference curve to the origin is implied by the Law of Diminishing Marginal Rate of Substitution. The indifference curve expresses x_2 as a function of x_1 for a given utility level. The slope of the indifference curve is dx_2/dx_1 . For strict convexity of the indifference curve to the origin, we need $d^2x_2/dx_1^2 > 0$. Now,

$$\begin{aligned} \frac{d^2x_2}{dx_1^2} &= \frac{d}{dx_1} \left(\frac{dx_2}{dx_1} \right) \\ &= \frac{d}{dx_1} (-MRS_{21}) \\ &= -\frac{d}{dx_1} (MRS_{21}). \end{aligned}$$

Therefore, $\frac{d^2x_2}{dx_1^2} > 0$ is implied by the condition $\frac{d}{dx_1} (MRS_{21}) < 0$. But

$\frac{d}{dx_1} (MRS_{21}) < 0$ means decreasingness of the marginal rate of substitution.

Hence, for the indifference curves to be strictly convex to the origin, we require the Law of Diminishing Marginal Rate of Substitution to hold.

INCREASING TRANSFORMATIONS OF UTILITY AND MARGINAL RATE OF SUBSTITUTION

We want to show that the marginal rate of substitution is independent of how utility is measured. Let $W(x_1, x_2) = f[U(x_1, x_2)]$ be an increasing transformation of $U(x_1, x_2)$, that is, $df/dU > 0$. Then for the revised utility function,

$$\begin{aligned} MRS_{21} &= \frac{\delta W(x_1, x_2) / \delta x_1}{\delta W(x_1, x_2) / \delta x_2} \\ &= \frac{df(U(x_1, x_2)) / dU \cdot \delta U(x_1, x_2) / \delta x_1}{df(U(x_1, x_2)) / dU \cdot \delta U(x_1, x_2) / \delta x_2} \\ &= \frac{\delta U(x_1, x_2) / \delta x_1}{\delta U(x_1, x_2) / \delta x_2} \end{aligned}$$

which is MRS_{21} for the original utility function.

STRICTLY QUASICONCAVE UTILITY FUNCTION AND MARGINAL RATE OF SUBSTITUTION

Next, we need to demonstrate that increasingness and strict quasiconcavity of a utility function guarantee that the marginal rate of substitution is decreasing. Now,

$$MRS_{21} = - \frac{dx_2}{dx_1} = \frac{MU_1}{MU_2}.$$

Let us now write U_1 for $MU_1 = \frac{\delta U(x_1, x_2)}{\delta x_1}$. Similarly let $U_2 = MU_2$. We also write

$$U_{11} \text{ for } \frac{\delta}{\delta x_1} \left(\frac{\delta U(x_1, x_2)}{\delta x_1} \right),$$

$$U_{22} \text{ for } \frac{\delta}{\delta x_2} \left(\frac{\delta U(x_1, x_2)}{\delta x_2} \right),$$

$$U_{12} \text{ for } \frac{\delta}{\delta x_2} \left(\frac{\delta U(x_1, x_2)}{\delta x_1} \right)$$

and U_{21} for $\frac{\delta}{\delta x_1} \left(\frac{\delta U(x_1, x_2)}{\delta x_2} \right)$.

The assumption of a diminishing MRS_{21} means that $d(MRS_{21})/dx_1 < 0$. But

$$\frac{d(MRS_{21})}{dx_1} = \left[U_2 \left(U_{11} + U_{12} \frac{dx_2}{dx_1} \right) - U_1 \left(U_{21} + U_{22} \frac{dx_2}{dx_1} \right) \right] / U_2^2.$$

Using the fact that $\frac{U_1}{U_2} = - \frac{dx_2}{dx_1}$, we have

$$\frac{d(MRS_{21})}{dx_1} = \frac{\left[U_2 \left(U_{11} - U_{12} \frac{U_1}{U_2} \right) - U_1 \left(U_{21} - U_{22} \frac{U_1}{U_2} \right) \right]}{U_2^2}.$$

Recognizing that $U_{12} = U_{21}$, we can rewrite the above as

$$\frac{d(MRS_{21})}{dx_1} = \frac{U_2 U_{11} - 2U_1 U_{12} + U_{22} (U_1)^2 / U_2}{(U_2)^2}$$

Multiplying numerator and denominator by U_2 ,

$$\frac{d(MRS_{21})}{dx_1} = \frac{(U_2)^2 U_{11} - 2U_1 U_2 U_{12} + U_{22} (U_1)^2}{(U_2)^3}$$

Given increasingness of the utility function, $U_2 > 0$. Then, MRS_{21} will diminish provided that

$$(U_2)^2 U_{11} - 2U_1 U_2 U_{12} + U_{22} (U_1)^2 < 0.$$

The above inequality holds if and only if the utility function is strictly quasiconcave (see Madden, 1986). Therefore, increasingness and strict quasiconcavity of a utility function are sufficient for a diminishing marginal rate of substitution.

EXERCISES

Multiple Choice

1. At the point of saturation for a commodity, the marginal utility is

- a) Positive.
 - b) Negative.
 - c) Zero.
 - d) Any one of the above.
2. If the marginal utility of good 1 is twice the marginal utility of good 2, then the Law of Diminishing Marginal Utility implies that the consumer will maximize his total utility only if
- a) The price of good 1 is twice the price of good 2.
 - b) The price of good 1 is greater than the price of good 2.
 - c) Price of good 1 is half the price of good 2.
 - d) None of the above.
3. Which of the following assumptions about a consumer's preferences ensures that it is possible to compare any two bundles?
- a) Transitivity.
 - b) Convexity.
 - c) Completeness.
 - d) Reflexivity.
4. Which of the following assumptions is not satisfied by the lexicographic ordering?
- a) Completeness.
 - b) Reflexivity.
 - c) Continuity.
 - d) Transitivity.
5. Suppose good 1 is measured along the horizontal axis and good 2 is measured along the vertical axis. Then a vertical indifference curve implies that the consumer
- a) Is saturated with good 1.
 - b) Is saturated with good 2.
 - c) Dislikes good 1.
 - d) Is saturated with both good 1 and good 2.

Quantitative Problems and Questions

1. Prove or disprove the following.

- a) For a two-commodity consumer, suppose good 1 is a neutral. Then the marginal rate of substitution of good 1 for good 2 is positive.
 - b) In a world with two goods, if the marginal utility of each is positive, then the indifference curve is strictly convex to the origin.
 - c) If the marginal rate of substitution of good 1 for good 2 is diminishing, then the marginal utility of each good is diminishing.
 - d) If a consumer's preferences satisfy monotonicity, then the indifference curves will be negatively sloped.
 - e) An indifference curve can never cross itself.
 - f) Suppose commodity 1 is a good and commodity 2 is a bad. Then the indifference curves satisfy the law of diminishing marginal rate of substitution.
 - g) Suppose a consumer's utility function is of the form $U(x_1, x_2) = \sqrt{(x_1 x_2)}$. Then the two goods are not perfect substitutes.
 - h) If a consumer's preferences satisfy monotonicity, then a diagonal line through the origin will intersect each indifference curve only once.
 - i) The indifference curve associated with a strictly quasiconcave utility function satisfy the law of diminishing marginal rate of substitution.
2. Plot the indifference maps for each of the following cases, indicating the direction of preference in each case.
- a) Two commodities purchased by a consumer are bads.
 - b) A consumer is saturated with some positive amount of commodity 1 but for commodity 2, which is a good, a saturation point does not exist.
 - c) Goods 1 and 2 are substitutes for a consumer. If the quantity of good 1 increases by 100% and that of good 2 decreases by 50%, he is just as well off.
 - d) Good 1 is labor hours and good 2 is a composite consumption good.
 - e) A person likes both bananas and apples, but he does not wish to have them together.
 - f) A person does not like jelly or butter alone, but he likes butter-and-jelly sandwiches.
 - g) Good 1 is good and good 2 is garbage.
 - h) I will spend all my time in city A or in city B, rather than dividing my time between the two.
 - i) I prefer a mixture of city A and city B life to being restricted to one or the other.
 - j) The marginal rate of substitution of good 2 for good 1 is increasing.

3. Demonstrate rigorously that the following conditions are equivalent.
 - a) Indifference curves are strictly convex to the origin.
 - b) The marginal rate of substitution decreases along an indifference curve.
4. Show why only strict convexity of a preference relation cannot guarantee that the indifference curves are strictly convex to the origin.
5. Consider the utility function $U(x_1, x_2) = \sqrt{(x_1 x_2)}$.
 - a) Draw indifference curves when $U = 2$ and $U = 3$.
 - b) Show that the law of diminishing marginal rate of substitution holds.
 - c) Calculate the marginal utilities when $x_1 = 4$ and $x_2 = 9$. Determine the marginal rate of substitution of x_2 for x_1 at these consumption levels. Suppose now x_1 increases from 4 to 6. How does the marginal rate of substitution change?
 - d) Consider the increasing transformation $V(x_1, x_2) = [U(x_1, x_2)]^2$. What is the marginal rate of substitution now when $x_1 = 4$ and $x_2 = 9$?
 - e) Is the function $W(x_1, x_2) = x_1 \sqrt{(x_2)}$ an increasing transformation of U ?
6. A consumer's utility function for goods 1 and 2 is given by $U(x_1, x_2) = (x_1 + x_2)x_2$. Assume the existence of two other goods, 3 and 4. Suppose each unit of good 3 provides 1 unit of good 2 and 1 unit of good 1, while 1 unit of good 4 provides 1 unit of good 1 and 0 units of good 2. Determine his utility function in terms of goods 3 and 4.
7. A consumer is choosing between bundles (x_1, x_2) and (y_1, y_2) . State two conditions which will ensure that $(x_1, x_2) \succ (y_1, y_2)$.
8. Explain why strictly concave preference means that the extremes are strictly preferred to averages.
9. Determine the point of saturation of the cardinal utility function $U(x) = 12x - x^2$. How does the point of saturation change under the transformation $V(x) = -10 + 2U(x)$? How does the marginal utility function change?
10. What is meant by the Law of Diminishing Marginal Rate of Substitution? What will be the shape of the indifference curve if the marginal rate of substitution is a constant?
11. Establish rigorously that any increasing function of a quasiconcave utility function is also quasiconcave.
12. Clearly explain why an increasing transformation of a utility function does not change the marginal rate of substitution.

13. Draw the indifference map associated with the utility function $U(x_1, x_2) = \log x_1 + x_2$. Show that the height of an indifference curve is a function of x_1 . Is the utility function $W(x_1, x_2) = x_1 e^{x_2}$ an increasing transformation of U ?
14. Discuss the limitations of the cardinal utility approach to consumer behavior.
15. Clearly state and explain the assumptions about a consumer's preferences.
16. Suppose a consumer can allocate his income among any number of goods. Does it make sense for the consumer to purchase that quantity of any one good that will maximize his utility for that good?

3

Utility Maximization and Consumer Demand

We have seen that a consumer's preferences or tastes are represented by his indifference map. What the consumer is *able* to choose has thus been left aside. His choices are determined by his preferences and the constraints imposed upon choices by his income and the prices of goods and services.

The choices are assumed to satisfy a basic consistency requirement. According to this requirement, the consumer chooses the best bundle he can afford. More precisely, given an assignment of utility numbers to bundles which the consumer can afford, he will choose the bundle that has the highest utility. That is, the consumer maximizes utility. This can be rephrased in terms of preference ordering as follows: given the set of commodity bundles which the consumer can afford, he will choose the most preferred bundle. Clearly, this basic consistency requirement is a kind of rationality assumption. Further, by assuming that the consumer's choices are determined by his tastes, we suppose that the consumer is the best judge of his own interest.

In this chapter we examine the implications of utility maximization in a very simple set-up. We will see how an individual allocates a fixed amount of income in purchasing goods so as to maximize utility from those goods.

THE BUDGET CONSTRAINT

The constraints on the individual's consumption activities are shown by budget lines or constraints, sometimes referred to as lines of attainable combinations. To show how the budget line is established, let us again limit the consumer to a two-commodity world.

We write $x = (x_1, x_2)$ for the consumer's commodity bundle, where x_i is the quantity of good i chosen by the consumer, $i = 1, 2$. The consumer treats the prices of the two goods (p_1, p_2) as constants, independent of the amounts of either good purchased. Let m denote the consumer's income.

The budget constraint of the consumer can be written as

$$p_1 x_1 + p_2 x_2 \leq m. \quad (3.1)$$

Here $p_1 x_1$ is the amount of expenditure on good 1, and $p_2 x_2$ is the amount of money spent on the consumption of good 2. Therefore, the budget constraint says that total expenditure (the sum of expenditures on goods 1 and 2) cannot exceed the consumer's income. The budget constraint, together with the constraint that the two commodities can be purchased only in non-negative amounts, generates the consumer's budget set. The budget set gives the consumer's affordable consumption bundles—the bundles that do not cost more than m .

The lined area in Fig. 3.1 is the consumer's budget set. The budget line is the set of consumption bundles that cost exactly m :

$$p_1 x_1 + p_2 x_2 = m. \quad (3.2)$$

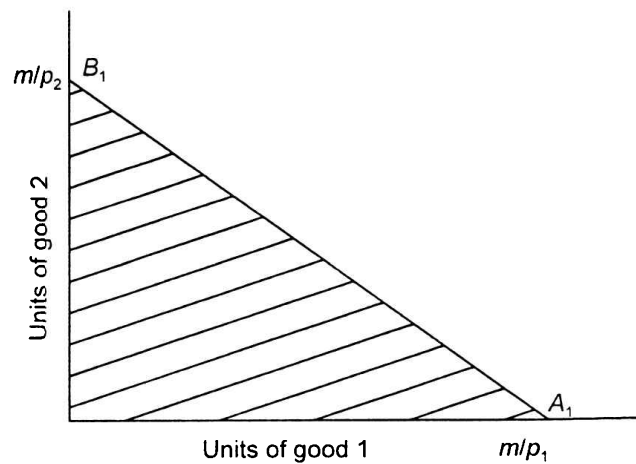


Fig. 3.1 The Budget Set

Any bundle inside the budget set will cost less than m . The budget line separates the feasible combinations—what the consumer can purchase—from those combinations beyond his financial reach.

The slope of the budget line will depend on the ratio p_1/p_2 . Given prices (p_1, p_2) and income m , if the consumer decides to spend all of his income on good 1, then m/p_1 is the total number of units that he could purchase. If he were to spend all of his income on good 2, then m/p_2 would show the number of units of good 2 he could purchase. The budget line joins the two extreme points m/p_1 (denoted by A_1) and m/p_2 (denoted by B_1).

We can rewrite the budget line equation as

$$x_2 = \frac{m}{p_2} - \frac{p_1}{p_2}x_1. \quad (3.3)$$

This is the equation of a straight line with a vertical intercept of m/p_2 and a slope of $-p_1/p_2$. The slope will indicate the rate at which the consumer would be willing to give up good 1 for good 2. To see this, let $\Delta x_1 > 0$ be the amount of good 1 that the consumer is ready to forego. We write $\Delta x_2 < 0$ for the change in consumption of good 2. Since (p_1, p_2) and m are constants, it follows from (3.3) that

$$\Delta x_2 = - \frac{p_1}{p_2} \Delta x_1 \quad (3.4)$$

from which we have

$$- \frac{\Delta x_2}{\Delta x_1} = \frac{p_1}{p_2}. \quad (3.5)$$

Thus, p_1/p_2 is the rate at which the consumer can trade good 1 for good 2, and still remain on the budget line. The ratio p_1/p_2 also measures the opportunity cost of consuming good 1. In order to consume more of good 2, the consumer has to give up some amount of good 1. Giving up the opportunity of buying good 1 is the actual cost of more of good 2 consumption, and this cost is given by p_1/p_2 .

Note that we can rearrange Eq. (3.2) as

$$x_1 + \frac{p_2}{p_1} x_2 = \frac{m}{p_1} \quad (3.6)$$

which can be written as

$$x_1 + p'_2 x_2 = m' \quad (3.7)$$

where $p'_2 = p_2/p_1$, $m' = m/p_1$. Eq. (3.7) is a normalization of Eq. (3.2) and in this version the price of good 1 has been normalized at unity. If we divide the budget line Eq. (3.2) by m , then income gets normalized at 1.

If we set one of the prices to 1, as has been done above, we refer to that price as the numeraire price. Clearly, the numeraire price is the price relative to which we measure the other price and income. In Eq. (3.7) the price of good 1 has been chosen as numeraire, but we can as well choose the other price as numeraire.

In some cases, one good may be a composite good that stands for everything else the consumer wishes to buy, excepting one of the goods. For instance, let x_1 be the quantity of rice purchased by the consumer. Then x_2 denotes the composite good that indicates everything other than rice. Sometimes x_2 may be the amount of money that the consumer wants to spend on the other goods. Then the budget constraint Eq. (3.1) will take the form

$$p_1x_1 + x_2 \leq m. \quad (3.8)$$

Thus, under composite good assumption the two-good budget constraint describes a very general framework (since if the consumer faces n goods, then the first good is rice and the remaining $[n-1]$ goods constitute the composite good).

SHIFTS IN THE BUDGET LINE

Changes in the consumer's income and changes in the prices of the goods will shift the individual's budget line. Referring to Fig. (3.2) suppose that the consumer's income is m , initially, and that the prices of goods 1 and 2 are still p_1 and p_2 respectively, so that his initial budget line is still B_1A_1 . If the price of good 1 increases to p'_1 , while his income and the price of good 2 remain constant, the budget line becomes a $B_1A'_1$. There is no change in the amount of good 2 that his income will purchase if it is all spent on good 2. However, the price of good 1 reduces OA_1 to OA'_1 , the amount of the good that he could purchase if his money were all spent on its consumption. The new budget line therefore joins B_1 and A'_1 . Thus, increasing p_1 makes the budget line steeper. (The ratio p'_1/p_2 is larger than p_1/p_2).

Suppose now that from the initial budget line B_1A_1 , the consumer's income decreases from m to m' while the prices of goods 1 and 2 remain constant. The budget line shifts to the left parallel to itself as B_2A_2 . The lower income forces him to purchase smaller amounts of good 1 if good 1 alone is

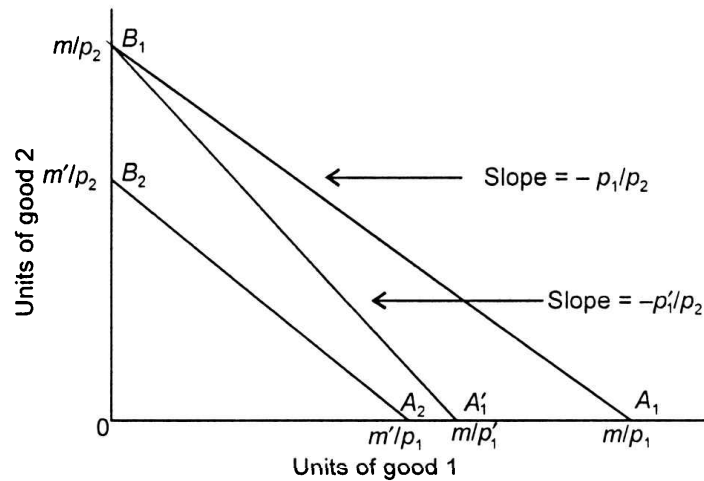


Fig. 3.2 Changes in the Budget Line

purchased, or smaller amounts of good 2 if good 2 alone is purchased, so that A_2 lies to the left of A_1 and B_2 lies to the left of B_1 . Since the prices of the two goods have not changed, both budget lines have a slope of $-p_1/p_2$, and are therefore parallel. Similarly, an increase in income will generate an outward parallel shift of the budget line.

Next, suppose p_1 and p_2 are increased equiproportionately. That is, both p_1 and p_2 are multiplied by some 'constant' $k > 1$. Then $p_1x_1 + p_2x_2 = m$ becomes $kp_1x_1 + kp_2x_2 = m$ which can be rewritten as $p_1x_1 + p_2x_2 = m/k$. This is the same as dividing the income by k . Therefore in this case also the budget line will shift to the left parallel to itself. It also follows that if we multiply both prices and income by the same constant k , then the budget line, hence the budget set, will not change. Thus, the budget line is homogeneous of degree zero in p_1 , p_2 and m . Given that the two intercepts of the budget line are m/p_1 and m/p_2 and its slope is $-p_1/p_2$, we can say that the budget set depends on the relative price p_1/p_2 and the purchasing power of income (m/p_1 , m/p_2).

Prices and income may change simultaneously in different directions also. For instance, suppose p_1 and p_2 increase and m decreases. Then since m/p_1 and m/p_2 change, the budget line will take a leftward or rightward shift depending on the magnitude of the change of the purchasing power of income. The budget line will become flatter or steeper as the ratio p_1/p_2 decreases or increases.

BUDGET SETS FOR TAXED, SUBSIDIZED AND RATIONED GOODS

Suppose the government imposes a quantity tax on some good. That is, the consumer is required to pay a certain amount to the government for each

unit of the good he purchases. A quantity tax on good 1 of Rs. t per unit would raise its price to $(p_1 + t)$. As demonstrated above, this implies that the budget line will be steeper, and hence the budget set shrinks. Another kind of tax is an ad valorem tax or percentage tax. For instance, if the sales tax is 4 percent, and if good 1 is priced at Rs. 2.00 per unit, then its selling price will be Rs. 2.08. Clearly, in this case also the budget set shrinks.

The converse of tax is subsidy. In the case of a quantity subsidy of Rs. s per unit of consumption of good 1, the price of the good would be $(p_1 - s)$. This would make the budget line flatter and the budget set larger. On the other hand, if good 1 is subjected to an ad valorem subsidy at the rate r , then the actual price of the good would be $p_1 (1 - r)$, which, given p_2 , would flatten the budget line.

We can also consider a lumpsum tax on general purchasing power. This means that the government collects Rs. T in revenue from the consumer and leaves the individual with $(m - T)$ to spend on the consumption of goods 1 and 2. Thus, a lumpsum tax will generate a leftward shift of the budget line. Similarly, with a lumpsum subsidy the budget line will shift outward.

Next, suppose that one good, say good 2, faces a rationing constraint. This means that the level of consumption of good 2 cannot be larger than some specified amount, say $\bar{x}_2$. The budget set for this case is shown in Fig. 3.3 (a). In panel (b), we combine rationing with taxes. Here the consumer can consume good 2 at the price p_2 up to $\bar{x}_2$, and after $\bar{x}_2$ he has to pay a quantity tax of Rs. t . In this case the consumer's budget line has a slope of $-p_1/p_2$ below $\bar{x}_2$, and a slope of $-p_1/(p_2 + t)$ above $\bar{x}_2$.

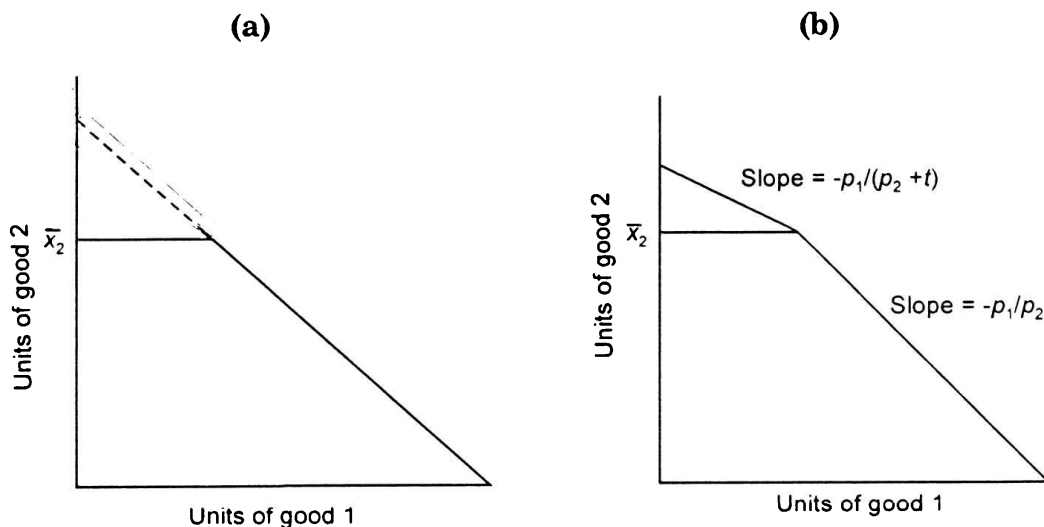


Fig. 3.3 Budget Set with Rationing

UTILITY MAXIMIZING CHOICE

The theory of consumer behavior is built on the premise that, given the budget constraint, the consumer seeks to maximize utility, that is, to remain on the highest possible indifference curve. To show the conditions under which the consumer can achieve this objective, his preference factors (the indifference map) and the restraining factors (the budget constraint) are brought together in Fig. 3.4. Any commodity bundle such as *A*, *B* or *C* on the budget line is available; so is any bundle such as *D* lying to the left of, or below it. Because of the budget constraint, consumption bundles, such as *E*, lying to the right of or above the budget line, are not available.

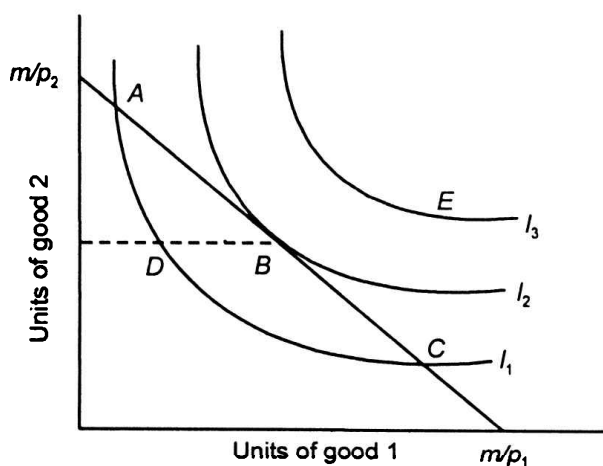


Fig. 3.4 Utility Maximizing Choice

The commodity bundle chosen must lie on the budget line. A consumer who consumes the bundle *D* violates the postulate that more of a good is preferred to less. (We are assuming that the consumer is spending his income m entirely on the consumption of the two goods, 1 and 2). By moving from *D* to *B*, the consumer obtains more of good 1 without sacrificing any amount of good 2 and consequently reaches a higher indifference curve. This sort of movement, or a movement that increases quantities of both the goods, is always possible for any consumption bundle below the budget line. That is, since extra units of goods provide extra utility, and there is no other use for income here, leaving any income unspent would fail to maximize utility.

Of the commodity bundles on the budget line, the consumer is expected to choose the one on the highest indifference curve touched by that line; this will be the commodity bundle *B*. (Commodity bundles *A* and *C* lie on a lower indifference curve). Bundle *B* is on the highest indifference curve that the consumer can reach with his income m : no higher level of utility is attainable.

Thus, the optimal or the utility maximizing consumption bundle is that at which the consumer's budget line is tangent to an indifference curve. In Fig. 3.4, this bundle is B.

Therefore at B, the slope of the budget line equals the slope of the indifference curve. The slope of an indifference curve at any point on it is the negative of the consumer's marginal rate of substitution, $-MRS_{21}$, at that point. The slope of a budget line at any point on it is $-p_1/p_2$. Therefore, as stated above, at the point of tangency—point B—the slopes of the two curves are the same, that is,

$$MRS_{21} = \frac{p_1}{p_2}. \quad (3.9)$$

But we have already noted that, along an indifference curve, the marginal rate of substitution between goods 1 and 2 is

$$MRS_{21} = -\frac{\Delta x_2}{\Delta x_1} = \frac{MU_1}{MU_2}. \quad (3.10)$$

Thus, the tangency condition between the budget line and the highest attainable indifference curve reduces to the condition

$$MRS_{21} = -\frac{\Delta x_2}{\Delta x_1} = \frac{MU_1}{MU_2} = \frac{p_1}{p_2}. \quad (3.11)$$

The condition of Eq. (3.11) means that the rate at which the consumer is willing to give up good 1 for good 2 (MRS_{21}) is equal to the rate at which he would be required by the market to give up good 1 to obtain good 2 (p_1/p_2). To see the intuitive reasoning behind this result, consider point C in Fig. 3.4. At this point the slope of the indifference curve I_1 is less than the slope of the budget line. Let MRS_{21} at C be 1. That is, the consumer is willing to trade one unit of good 1 for one unit of good 2 and remain equally well off. Assume that $p_1 = 2$ and $p_2 = 1$. The individual can give up one unit of good 1 and trade it in the market for two units of good 2. But only one extra unit of good 2 was needed to keep the person as happy as before the trade—the second unit of good 2 is a net addition to welfare. Therefore, the individual's money could not have been allocated rationally at C. A similar reasoning can be applied when $MRS_{21} > p_1/p_2$. The condition for maximum utility must be the equality of these two magnitudes.

We state the utility maximizing condition formally as follows.

Optimality Principle (Utility Maximization)

In order to maximize utility, given a fixed amount of income to spend, an individual will buy those quantities of goods which exhaust his total income and for which the rate of exchange between the goods (the marginal rate of substitution) is equal to the rate at which the goods can be traded one for the other at the market (the price ratio).

That is, for utility maximization all income should be spent and MRS_{21} should equal the ratio of prices p_1/p_2 . (A calculus approach to this is available in Appendix 3.1). We have developed this condition for a particular consumer only. Preferences will differ from consumer to consumer. That is why the amounts of goods 1 and 2 purchased will also differ across consumers. But at the margin, the trade-off between the goods in terms of personal preferences is the same for all the consumers (since prices are constant). The optimality or the utility maximizing condition is also referred to as the equilibrium condition of the consumer.

The tangency rule is only a necessary condition for utility maximization. That is, if the optimal choice involves consuming positive amounts of both goods, then the indifference curve will be necessarily tangent to the budget line. More precisely, for a commodity bundle involving positive amounts of both goods to be utility maximizing or optimal, we require the budget line to be a tangent to the indifference curve at the point of optimal choice. But, is the tangency condition sufficient for a bundle to be optimal? That is, if we find a bundle where the indifference curve is tangent to the budget line, can we conclude that the bundle is utility maximizing?

Let us look at Fig. 3.5. Here a point of tangency (C) is inferior to a point of non tangency (D). Of the three bundles A , B , C where the tangency condition is fulfilled, only A and B are optimal. This shows that in general, the tangency condition is a necessary condition for optimality, not a sufficient condition.

The failure of the tangency condition to produce an unambiguous maximum can be attributed to the peculiar shape of the indifference curves in Fig. 3.5. If the indifference curves are shaped as are those in Fig. 3.4, no such problem arises. That is, we are saying that if the indifference curves are strictly convex to the origin, any consumption bundle that satisfies the tangency condition must be an optimal or utility maximizing bundle. This is quite clear geometrically: since strictly convex indifference curves must curve away from the budget line, they cannot bend back to touch it again.

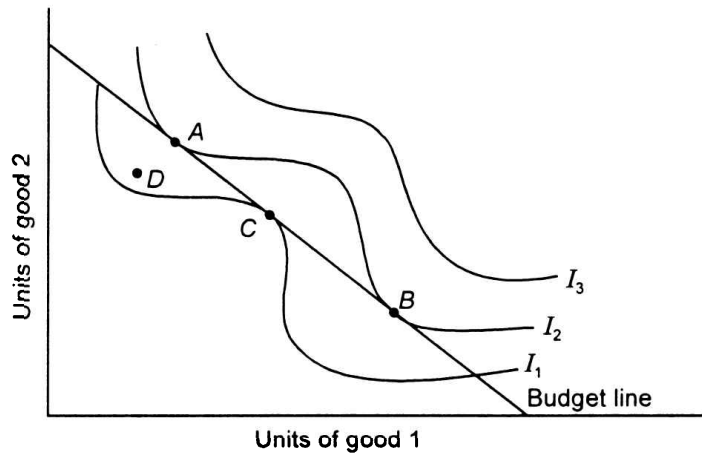


Fig. 3.5 The Tangency Condition is Necessary but not Sufficient

Another feature that emerges from Fig. 3.5 is that there may be more than one consumption bundle that satisfies the tangency condition. However, strict convexity again imposes a restriction. If the indifference curves are strictly convex to the origin, they cannot contain any flat section. Consequently, there will be only one point of tangency between the budget line and an indifference curve. In view of our discussions in Chapter 2 on strict convexity of indifference curves to the origin, we can now state the following:

THEOREM 3.1: *Any one of the following can be regarded as a sufficient condition for a consumer's optimal or utility maximizing bundle to be unique:*

- a) *The consumer's preferences satisfy monotonicity and they also satisfy strict convexity.*
- b) *The Law of Diminishing Marginal Rate of Substitution holds.*
- c) *The consumer's indifference curves are strictly convex to the origin.*
- d) *The consumer's utility function is increasing and strictly quasiconcave.*

CORNER SOLUTIONS

The utility maximization problem illustrated in Fig. 3.4 resulted in an interior maximum in which positive amounts of both commodities were purchased. In some cases a consumer's preferences may be such that he can obtain maximum utility by purchasing zero amount of some good. If a person does not like a good, there is no reason to allocate any income to its purchase. This situation is shown in Fig. 3.6. There utility is maximized at A, where $x_2 = \bar{x}_2$ and $x_1 = 0$ —any point on the budget line where positive amounts of good 1 are consumed generates a lower utility than does the point A.

Note that at A the budget line is not tangent to the indifference curve I_2 . In fact, at A the budget line is flatter than I_2 , showing that the rate at which good 1 can be traded for good 2 in the market is higher than the individual's trade-off rate (the marginal rate of substitution). Thus, at the prevailing market prices the consumer is more than willing to exchange good 1 for good 2. Since it is not possible to consume negative amounts of good 1, the physical limit for this process is the vertical axis, along which purchases of good 1 are zero. Therefore it is necessary to amend the tangency condition for utility maximization a bit to allow corner or boundary solutions of the type shown in Fig. 3.6. We discuss this issue in Appendix 3.1.

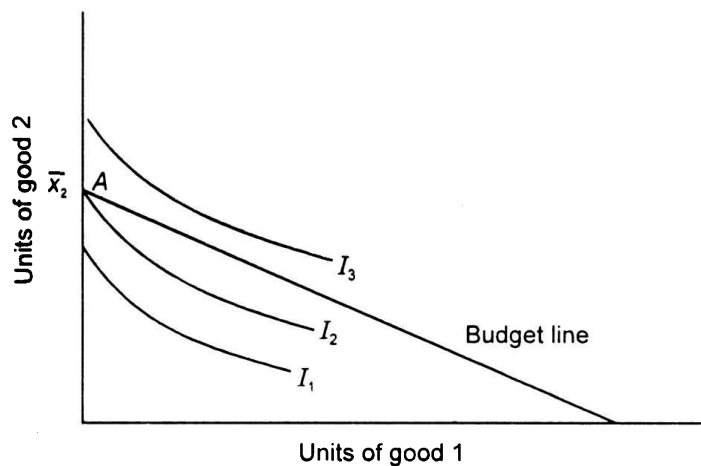


Fig. 3.6 Corner or Boundary Solution

CONSUMER DEMAND FUNCTION

The utility maximizing choice of goods 1 and 2 at some set of prices and income is referred to as the consumer's demanded bundle. When prices and income change, the consumer's optimal choice will change. The demand function for a commodity is the function that relates the quantities demanded of the good to the different values of prices and income. Thus, the demand function for good 1 is a schedule showing the quantities of the good demanded for each combination of values of p_1 , p_2 , and m .

We write the demand function for good 1 as $x_1 = f_1(p_1, p_2, m)$. Similarly, the demand function for good 2 is $x_2 = f_2(p_1, p_2, m)$. Clearly, different preferences lead to different kinds of demand functions. The *ceteris paribus* assumption that underlies the demand functions for goods 1 and 2 is that all things other than p_1 , p_2 , and m are being held constant in specifying f_1 and f_2 , including the consumers' tastes.

HOMOGENEITY OF DEGREE ZERO OF CONSUMER DEMAND FUNCTION

Suppose p_1, p_2 , and m are all doubled. Then both relative prices p_1/p_2 and the purchasing power of income m/p_1 and m/p_2 remain unchanged. But then the budget line and hence the budget sets are left unchanged as well. (This is verified directly from the fact that the slope and the intercepts of the budget line remain the same.)

In our discussion of consumer preferences, we implicitly assumed that preferences are not influenced by the prices of the goods and services consumed. This assumption is referred to as the absence of money illusion. Since the consumer's preferences are characterized by absence of money illusion, the indifference map remains unaffected under any changes in prices and income. It therefore follows that the utility maximizing bundle remains unchanged under equiproportionate changes in all prices and income. That is, the quantity demanded by a consumer of any good is unchanged by an equal proportionate change in all prices and income.

Technically we say that since $p_1x_1 + p_2x_2 = m$ is the same constraint as $tp_1x_1 + tp_2x_2 = tm$ for all $t > 0$ and since $U(x_1, x_2)$ does not depend on p_1, p_2 , and m , the demand functions $f_1(p_1, p_2, m)$ and $f_2(p_1, p_2, m)$ are the same as the functions $f_1(tp_1, tp_2, tm)$ and $f_2(tp_1, tp_2, tm)$ respectively. Formally, the demand functions $x_1 = f_1(p_1, p_2, m)$ and $x_2 = f_2(p_1, p_2, m)$ fulfil the following:

$$\begin{aligned} x_1 &= f_1(p_1, p_2, m) = f_1(tp_1, tp_2, tm) \\ x_2 &= f_2(p_1, p_2, m) = f_2(tp_1, tp_2, tm) \end{aligned} \quad (3.12)$$

for all $t > 0$. That is, the demand functions are homogeneous of degree zero in prices and income.

This means that individuals' demands will not be affected by a pure inflation during which all prices and incomes rise proportionately. They will continue to demand the same bundle of goods. But if inflation is not pure, that is, if some prices rise more rapidly than others, this will not be the case.

In the next section we examine several demand functions and see that they all verify the homogeneity property.

SOME DEMAND FUNCTIONS

Let us now see how the model of choice we have discussed above can be employed to construct the consumer demand functions under alternative assumptions about preferences.

Perfect Substitutes

When two goods are perfect substitutes, optimal choice may be a corner or an interior solution. Let the slope of the indifference curves be $-a/b$. That is, the utility function is of the form $U(x_1, x_2) = ax_1 + bx_2$, where a and b are positive constants.

If $p_1/p_2 < a/b$, the budget line is flatter than the indifference curves. In this case the consumer will spend his entire income m on the consumption of good 1. On the other hand, if $p_1/p_2 > a/b$, the consumer purchases only good 2. Thus, if $p_1/p_2 \neq a/b$, the optimal choice is a boundary point. This seems quite reasonable. If two goods are perfect substitutes and if one of them is relatively cheaper, then the consumer will spend his total income on the cheaper good. To see this more explicitly, note that $p_1/p_2 < a/b$ means that the market exchange rate (p_1/p_2) is smaller than the individual's psychic exchange rate (MRS_{21}). Since the two goods are perfect substitutes, the individual will not be willing to exchange good 1 for good 2 at all. Thus, the individual will buy only good 1. A similar explanation holds for the case when $p_1/p_2 > a/b$.

If $p_1/p_2 = a/b$, the budget line will coincide with the indifference curve, and therefore any point that lies on the budget line is utility maximizing. In this case an interior solution can as well be optimal. Therefore, the demand function for good 1 will be

$$x_1 = f_1(p_1, p_2, m) = m/p_1 \text{ if } \frac{p_1}{p_2} < \frac{a}{b},$$

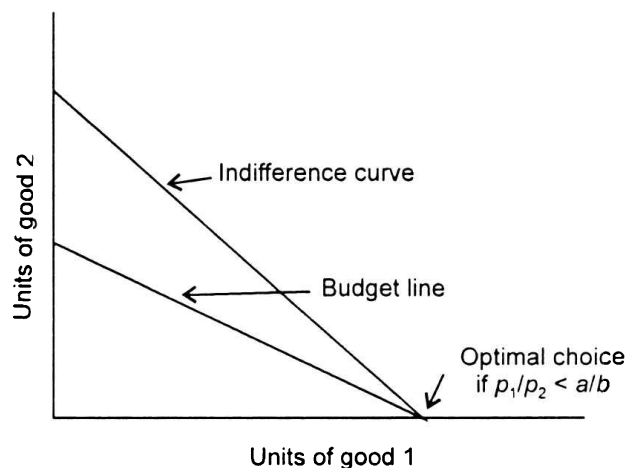


Fig. 3.7 Utility Maximizing Choice with Perfect Substitutes

$$= c \text{ where } 0 \leq c \leq m/p_1 \text{ if } \frac{p_1}{p_2} = \frac{a}{b} \quad (3.13)$$

$$= 0 \text{ if } \frac{p_1}{p_2} > \frac{a}{b}$$

We can derive the demand function for good 2 analogously. The function of Eq. (3.13) is clearly homogeneous of degree zero in its arguments.

Perfect Complements

Suppose that the two goods, which are perfect complements, are purchased in the proportion 1:1. The optimal choices will lie at the vertices of the L-shaped indifference curves, no matter what the prices are. Now, if z is the number of units of good 1 demanded, then z is the demand for good 2. Therefore, it must be the case that $p_1 z + p_2 z = m$, which gives $z = m/(p_1 + p_2)$. Therefore, the demand functions for goods 1 and 2 are

$$\begin{aligned} x_1 &= \frac{m}{p_1 + p_2} \\ x_2 &= \frac{m}{p_1 + p_2} \end{aligned} \quad (3.14)$$

We note homogeneity of degree zero of these demand functions. The formulae in Eq. (3.14) are quite intuitive. Since the two goods are bought together in the proportion 1:1, it is just as if the consumer is spending his income for buying a composite good whose price is $p_1 + p_2$.

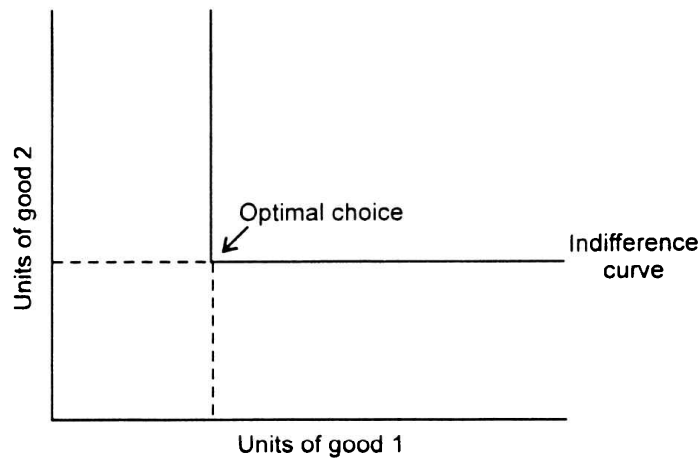


Fig. 3.8 Utility Maximizing Choice with Perfect Complements

Neutrals and Bads

If a good is neutral then the consumer does not buy it. Therefore of the two goods 1 and 2, if good 1 is a neutral and good 2 is liked by the consumer, then he will spend m entirely on the consumption of good 2. Consequently, his demand functions are

$$\begin{aligned} x_1 &= 0, \\ x_2 &= m/p_2. \end{aligned} \quad (3.15)$$

If good 1 is a bad and good 2 is a good, then also his demand functions will be of the type given by Eq. (3.15).

Discrete Goods

Assume that good 2 is available in integer units and good 1 is the amount of money to be spent on other goods. Then the consumer will choose the bundles $(m - p_2, 1)$, $(m - 2p_2, 2)$, $(m - 3p_2, 3)$, ..., $(m - kp_2, k)$, ... given that he chooses 1, 2, 3, ..., k , ... units of good 2. We can compare the utility of each of these combinations to see which has the highest utility.

Strictly Concave Preference Relations

In Fig. 3.9, A is the point of tangency between the budget line and an indifference curve. But the attainable point B lies on a higher indifference curve. Therefore B is the optimal choice here. That is, the consumer spends his entire income m on the consumption of good 2. Hence his demand functions are

$$\begin{aligned} x_1 &= 0, \\ x_2 &= m/p_2 \end{aligned} \quad (3.16)$$

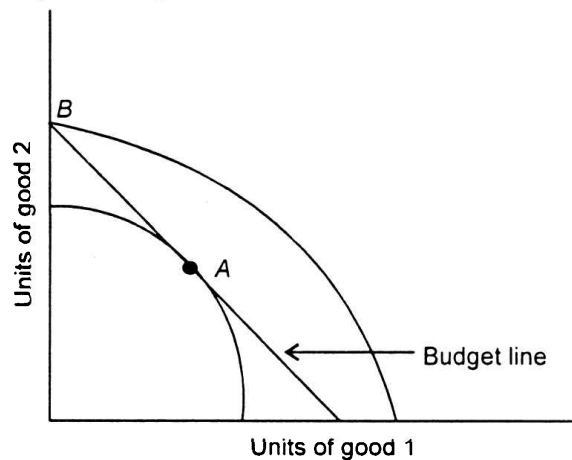


Fig. 3.9 Optimal Choice for Strictly Concave Preference

Stone-Geary Utility Function

A variant of the Cobb-Douglas utility function is the Stone-Geary utility function

$$U(x_1, x_2) = (x_1 - c_1)^{a_1}(x_2 - c_2)^{a_2} \quad (3.17)$$

where $0 < a_1, a_2 < 1$, $a_1 + a_2 = 1$ and $x_1 > c_1$, $x_2 > c_2$. We can rationalize this function by assuming that c_1 and c_2 are the subsistence requirements for x_1 and x_2 . These are met before any structure of preferences comes into play. So $p_1c_1 + p_2c_2$ is the level of the consumer's income spent on subsistence, the rest of income ($m - p_1c_1 - p_2c_2$), which is referred to as the supernumerary income, is divided between above-subsistence expenditures on x_1 and x_2 in the ratio $a_1:a_2$. It follows that the demand function for good i is

$$x_i = c_i + a_i \frac{m - [(p_1c_1 + p_2c_2)]}{p_i} \quad (3.18)$$

where $i = 1, 2$. These demand functions, like the utility function itself, are defined if and only if $x_i > c_i$. The system of demand equations of the form described in Eq. (3.18) is called the *linear expenditure system*, because expenditure p_ix_i on good i is a linear function of m , p_i and the other price, where $i = 1, 2$ (Stone, 1954). To observe this more explicitly note that

$$\begin{aligned} p_1x_1 &= p_1c_1 + a_1[m - (p_1c_1 + p_2c_2)], \\ p_2x_2 &= p_2c_2 + a_2[m - (p_1c_1 + p_2c_2)], \end{aligned} \quad (3.19)$$

we see that p_1x_1 , expenditure on good 1, is linear in p_1, p_2 and m . Similarly, p_2x_2 , is linear in p_1, p_2 and m .

INDIRECT UTILITY FUNCTION

A consumer's demanded bundle gives us optimal values of x_1 and x_2 . These optimal values depend on the prices p_1, p_2 and the consumer's income m . That is, $x_1 = f_1(p_1, p_2, m)$, $x_2 = f_2(p_1, p_2, m)$. These optimal values of x_i 's can be substituted into the original utility function to yield

$$U(x_1, x_2) = U[f_1(p_1, p_2, m), f_2(p_1, p_2, m)] = V(p_1, p_2, m), \text{ say.} \quad (3.20)$$

In other words, because of the consumer's need to maximize preference, given a budget constraint, the level of utility obtainable will depend indi-

rectly on the prices of the goods and the income of the consumer. If any one of these arguments changes, the level of utility that can be obtained also gets affected. V is called the indirect utility function of the consumer.

The following are some of the interesting properties of V :

- a) $V(p_1, p_2, m)$ is decreasing in prices.
- b) $V(p_1, p_2, m)$ is increasing in m .
- c) $V(p_1, p_2, m)$ is homogeneous of degree zero in (p_1, p_2, m) .
- d) $V(p_1, p_2, m)$ is continuous in p_1, p_2, m .

Property (a) means that if p_1 or p_2 , or both increase, then the indirect utility function decreases. This is quite reasonable. When one or both prices go up, the consumer's budget set shrinks and therefore the quantities demanded by the consumer will decrease. Consequently, the indirect utility function decreases. A similar argument shows that an increase in m increases $V(p_1, p_2, m)$. Homogeneity of degree zero of $V(p_1, p_2, m)$ in (p_1, p_2, m) is a consequence of homogeneity of degree zero of the budget line. The continuity property, whose proof is difficult, is also quite reasonable if we want $V(p_1, p_2, m)$ to vary continuously with p_1, p_2, m . In fact, $V(p_1, p_2, m)$ satisfies a fifth property which demands strict convexity of the price indifference curves to the origin. A price indifference curve is a graph of all those prices (p_1, p_2) such that $V(p_1, p_2, m) = k$, where $k > 0$ is a constant.

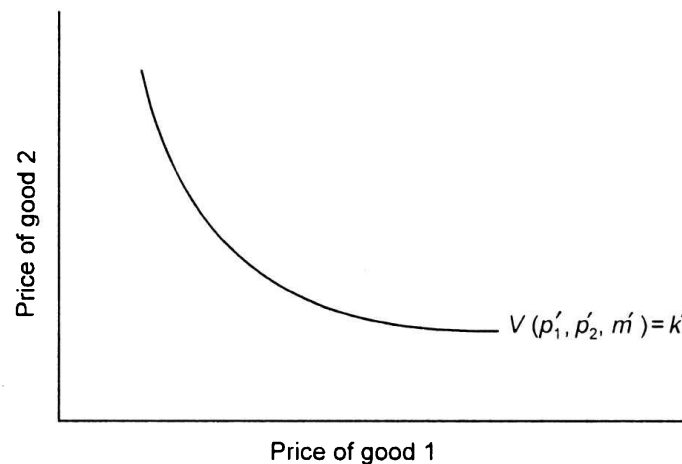


Fig. 3.10 Price Indifference Curve

EXPENDITURE FUNCTION

To a consumer's utility maximization problem, there corresponds an expenditure minimization problem that allocates income in such a way so that a

given utility level is achieved with the minimal expenditure. The problem is clearly analogous to the utility maximization problem, but the goal and constraints have been reversed.

Fig. 3.11 illustrates the expenditure minimization problem. The individual must attain the utility level I_2 (this is now the constraint in the problem).

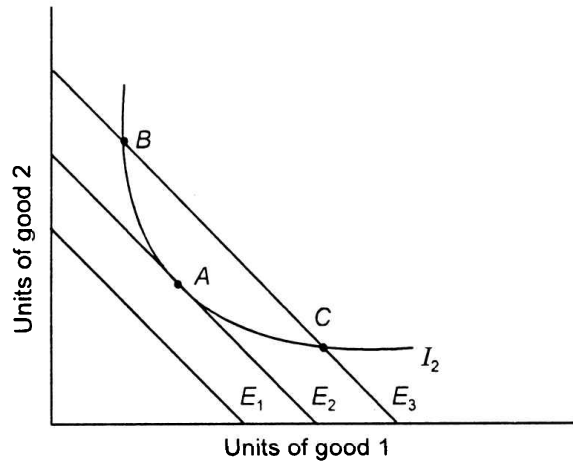


Fig. 3.11 Expenditure Minimization

Three possible expenditure levels E_1 , E_2 and E_3 are shown as three budget lines in the figure. E_1 is too small to achieve I_2 . Hence it cannot solve the problem. With expenditure level E_3 , the individual can reach I_2 (at B or C), but this is not the minimal expenditure required. In fact, E_2 is the expenditure just enough to reach I_2 (at point A). By comparing Figs. 3.4 and 3.11, it should be clear that both the utility maximization problem and the expenditure minimization problem yield the same solution—they are alternative ways of viewing the same process.

Formally, a consumer's expenditure minimization problem is to choose x_1 and x_2 so as to minimize

$$\text{total expenditure} = E = p_1x_1 + p_2x_2 \quad (3.21)$$

subject to the constraint

$$\text{Utility} = I = U(x_1, x_2) \quad (3.22)$$

where I is the given level of utility.

The optimal amounts of x_1 , x_2 chosen here will depend on p_1 , p_2 and I . This dependence can be summarized by an expenditure function:

A consumer's expenditure function shows the minimal expenditure necessary to achieve a given utility level for a particular set of prices. That is,

$$\text{minimal expenditure} = e(p_1, p_2, I). \quad (3.23)$$

The following are some of the important properties of an expenditure function.

- a) $e(p_1, p_2, I)$ is increasing in prices.
- b) $e(p_1, p_2, I)$ is homogeneous of degree 1 in p_1, p_2 .
- c) $e(p_1, p_2, I)$ is continuous in prices.

If we increase p_1 or p_2 or both in the objective function of Eq. (3.21), then obviously $e(p_1, p_2, I)$ increases. Similarly, if p_1 and p_2 are both increased by the same proportion, e also increases by the same proportion. Hence properties (a) and (b) of the expenditure function are quite reasonable. Property (c) demands continuous variation of e with respect to variations in p_1, p_2 .

It may be interesting to note that we can determine the expenditure function directly from the indirect utility function also. Since $V(p_1, p_2, m)$ is increasing in m , we can invert the function and solve it for m as a function of the level of utility and prices. That is, given any level of utility I , we can determine the minimal amount of income necessary to achieve I at prices p_1, p_2 . Clearly, the inverse indirect utility function that relates income and utility in this way is the consumer's expenditure function.

Example 3.1

Suppose that a consumer consumes only two goods 1 and 2 and that his utility function is given by

$$U(x_1, x_2) = \log x_1 + \log x_2.$$

We want to calculate the quantities demanded of goods 1 and 2. We also want to determine the consumer's indirect utility function and the expenditure function.

Let us find the demand functions for goods 1 and 2. The problem we want to solve is

$$\text{Max}_{x_1, x_2} (\log x_1 + \log x_2)$$

such that $p_1 x_1 + p_2 x_2 = m$. The tangency condition for utility maximization gives

$$\frac{\partial U(x_1, x_2) / \partial x_1}{\partial U(x_1, x_2) / \partial x_2} = \frac{MU_1}{MU_2} = \frac{p_1}{p_2}$$

that is,
$$\frac{1/x_1}{1/x_2} = \frac{p_1}{p_2}$$

which gives $p_1x_1 = p_2x_2$. Substituting $p_1x_1 = p_2x_2$ into the budget line equation we get

$$2p_1x_1 = m$$

Hence $x_1 = \frac{m}{2p_1}$. This is the demand function for good 1. Similarly, $x_2 = \frac{m}{2p_2}$, which is the demand function for good 2.

The indirect utility function is

$$\begin{aligned} V(p_1, p_2, m) &= \log \left(\frac{m}{2p_1} \right) + \log \left(\frac{m}{2p_2} \right) \\ &= \log \frac{m^2}{4p_1p_2}. \end{aligned}$$

Let us determine the expenditure function by inverting the indirect utility function. If I denotes some arbitrary given utility level, then

$$I = \log \frac{m^2}{4p_1p_2}$$

gives $m^2/4p_1p_2 = e^I$, from which we have

$$m = 2\sqrt{(p_1p_2)e^I}$$

Therefore, the required expenditure function is

$$e(p_1, p_2, I) = 2\sqrt{(p_1p_2)e^I}$$

Applied Example 3.1 Income Tax Versus Commodity Tax

The consumer theory we have discussed so far will now be employed to make a choice between two types of taxes; quantity tax and income tax. We saw that a quantity tax is a tax on the amount purchased of a good. An income tax is a tax on income.

Here we wish to show that if two taxes, one a tax on a specific commodity and the other an income tax, a lump sum tax on general purchasing power, collect equal revenue, then the lump sum tax will do so at lower cost since it reduces the consumer's utility by a lower amount. The proof is shown in Fig. 3.12.

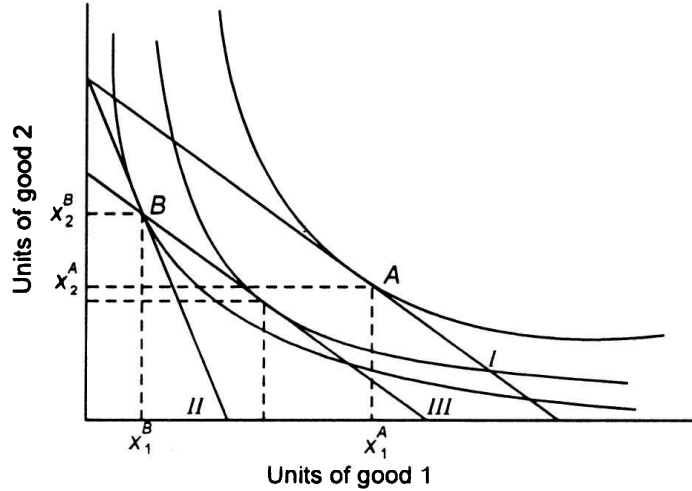


Fig. 3.12 Efficiency of an Income Tax

Initially the consumer has m to spend and wishes to consume (x_1^A, x_2^A) , since this point obeys the tangency condition for a maximum utility and the budget constraint

$$m = x_1 p_1^A + x_2 p_2^A$$

A tax on good 1 of Rs. t per unit will raise its price to $(p_1 + t)$ and the budget line will be

$$(p_1 + t)x_1 + p_2 x_2 = m$$

With this budget line (shown as II in Fig. 3.12), the consumer will have to accept a lower utility level and will consume the combination (x_1^B, x_2^B) . Total tax revenue will be given by $T = tx_1^B$.

An income tax collecting Rs. T in revenue will leave the consumer with Rs. $(m - T)$ to spend and the budget line will be

$$p_1 x_1 + p_2 x_2 = m' = m - T.$$

Since $m = (p_1 + t)x_1^B + p_2 x_2^B$, it must be true that $m' = p_1 x_1^B + p_2 x_2^B$. Therefore, the above budget line's slope is, same as that of the original budget line (shown as I in the figure).

But, can (x_1^B, x_2^B) be an optimal choice in the case of income taxation? The answer is no. At (x_1^B, x_2^B) , $MRS_{21} = (p_1 + t)/p_2$. But since the budget line for the income tax case is parallel to the original budget line, income tax allows the consumer to trade at the market exchange rate p_1/p_2 . Therefore the budget line III will cut the indifference curve for the commodity tax case at (x_1^B, x_2^B) , which means that there exists some point on III that will be preferred to (x_1^B, x_2^B) .

Therefore the income tax is more efficient than the quantity tax in the sense that though they raise the same amount of revenue from the consumer, he becomes better off under the income tax than under the quantity tax. To understand the reasoning behind this result, note that the quantity tax affects an individual's position in two ways. It reduces general purchasing power and it directs consumption away from the taxed good. An income tax incorporates the first effect only. Therefore, the consumer is better off under this.

Appendix 3.1

UTILITY MAXIMIZATION

We are required to maximize $U(x_1, x_2)$ with respect to x_1, x_2 subject to the constraint $p_1x_1 + p_2x_2 = m$. One method for solving this constrained maximization problem is the Lagrangian multiplier method which involves a mathematical trick that also turns out to be economically meaningful. The rationale of this method is quite simple.

We formally state the problem as

$$\text{Max}_{x_1, x_2} U(x_1, x_2)$$

subject to $p_1x_1 + p_2x_2 - m = 0$. The Lagrangian method starts with setting up the expression

$$\mathcal{L} = U(x_1, x_2) - \lambda(p_1x_1 + p_2x_2 - m)$$

where λ is an additional variable that is called the Lagrange multiplier. Later on we shall interpret λ . Note that when the constraint holds, $\mathcal{L}$ and U have

the same value (since $p_1x_1 + p_2x_2 - m = 0$). Therefore, if we restrict attention only to those values of x_1, x_2 that meet the constraint, finding the maximum of U is equivalent to finding the maximum of $\mathcal{L}$, which is referred to as the Lagrangian. Let us then try to do so, treating λ also as a variable.

The first order conditions for a maximum are

$$\frac{\delta \mathcal{L}}{\delta x_1} = \frac{\delta U(x_1, x_2)}{\delta x_1} - \lambda p_1 = 0,$$

$$\frac{\delta \mathcal{L}}{\delta x_2} = \frac{\delta U(x_1, x_2)}{\delta x_2} - \lambda p_2 = 0,$$

$$\frac{\delta \mathcal{L}}{\delta \lambda} = p_1x_1 + p_2x_2 - m = 0.$$

We thus have three equations in three unknowns (x_1, x_2 and λ). These equations can be solved for x_1, x_2 and λ . Such a solution will have the property that x_i 's will obey the constraint.

The above equations are necessary but not sufficient for a maximum. The assumption of a diminishing marginal rate of substitution ensures that any point satisfying the above three equations is a maximum. From now on only the first order conditions will be discussed, under the implicit assumption that the law of diminishing marginal rate of substitution makes these conditions sufficient also for a maximum. Therefore the Lagrangian method provides a way to find a solution for the constrained maximization problem.

Let us now look at the implications of the first order conditions. The first two equations give us

$$\frac{\delta U(x_1, x_2) / \delta x_1}{\delta U(x_1, x_2) / \delta x_2} = \frac{p_1}{p_2}.$$

But it was shown that the ratio of marginal utilities of two goods is identical to the marginal rate of substitution between them. Therefore,

$$MRS_{21} = \frac{p_1}{p_2}.$$

This is exactly the result derived in the main body of the chapter; the individual should equate his own psychic trade-off rate (MRS_{21}) with the market rate of trade-off (p_1/p_2).

Another result is that

$$\lambda = \frac{\delta U(x_1, x_2)/\delta x_1}{p_1} = \frac{\delta U(x_1, x_2)/\delta x_2}{p_2}.$$

This equation says that at the utility maximizing point, each good purchased should give the same marginal utility per rupee spent on that good. That is, each good has an identical marginal benefit (marginal utility) to marginal cost (price) ratio. If this were not the case, one good would yield more marginal enjoyment per rupee than some other good and funds would be allocated non-optimally. Therefore λ can be regarded as the marginal utility of an extra rupee of consumption expenditure (marginal utility of income).

We also have

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

That is, for every good that an individual purchases, the price of the good shows his evaluation of the last unit consumed.

CORNER SOLUTIONS

When boundary or corner solutions (such as illustrated in Fig. 3.6) arise, the first order conditions require slight modifications. In this case the first order conditions become

$$\frac{\delta \mathcal{L}}{\delta x_1} = \frac{\delta U(x_1, x_2)}{\delta x_1} - \lambda p_1 \leq 0,$$

$$\frac{\delta \mathcal{L}}{\delta x_2} = \frac{\delta U(x_1, x_2)}{\delta x_2} - \lambda p_2 \leq 0,$$

$$\frac{\delta \mathcal{L}}{\delta \lambda} = p_1 x_1 + p_2 x_2 - m = 0.$$

and if $\frac{\delta \mathcal{L}}{\delta x_i} < 0$ for any i , then $x_i = 0$, $i = 1, 2$.

To interpret these conditions, suppose $\frac{\delta \mathcal{L}}{\delta x_1} < 0$, that is,

$$p_1 > \frac{\delta U(x_1, x_2)/\delta x_1}{\lambda}.$$

The above inequality means that the price of good 1 exceeds its marginal value to the consumer. Therefore the good should not be purchased, that is, x_1 should be zero.

EXERCISE

Multiple Choice

1. If in a model with only two goods, the marginal rate of substitution of good 2 for good 1 is a constant greater than the price ratio p_1/p_2 , then the consumer will
 - a) Not buy good 1.
 - b) Not buy good 2.
 - c) Be indifferent among all combinations of goods 1 and 2 permitted by his budget line.
 - d) Adjust his preference, for if he is rational, the marginal rate of substitution will not be a constant.
2. In which of the following cases will the budget line move out from the origin without changing its slope?
 - a) The price of good 1 increases by 5% while the price of good 2 decreases by 5%.
 - b) The prices of goods 1 and 2 increase by 5%.
 - c) The prices of goods 1 and 2 decrease by 5% while money income decreases by 2.5%.
 - d) The prices of goods 1 and 2 increase by 10% while money income increases by 5%.
3. When a consumer is maximizing his utility, which of the following is necessarily true?
 - a) The marginal utility of good 1 divided by that of good 2 is equal to p_1/p_2 .
 - b) The marginal rate of substitution of goods for leisure is equal to the wage rate.

- c) The law of diminishing marginal rate of substitution holds.
 - d) None of the above is necessarily true.
4. Suppose goods 1 and 2 are perfect complements and they are bought together in 1:1 ratio. If p_1 and p_2 are the prices of the goods and m is the consumer's income, then the demand function for good 1 is
- a) m/p_1
 - b) $m/(p_1 + p_2)$
 - c) $2m/(p_1 + p_2)$
 - d) $m/(p_1 + p_2)^{1/2}$.
5. If the marginal rate of substitution of good 2 for good 1 exceeds the price ratio p_1/p_2 , it is possible for the individual to gain by giving up
- a) Good 1 in exchange of good 2.
 - b) Good 2 in exchange of good 1.
 - c) Either good 1 or good 2.
 - d) Any one of the above is possible.

Quantitative Problems and Questions

1. Prove or disprove the following.
 - a) If a consumer in a two-good world with the utility function $U(x_1, x_2) = x_1x_2$, faces prices $p_1 = 2$ and $p_2 = 3$, has income $m = 300$, and consumes 60 units of good 1 and 40 units of good 2, then since his marginal rate of substitution equals the price ratio, he is maximizing utility.
 - b) If in a two-good world an individual's utility function is $U(x_1, x_2) = x_1^a x_2^a$, then he will spend half his income on good 1.
 - c) if p_1 , the price of good 1 is Re. 1 and p_2 , the price of good 2 is Rs 2 and an individual with fixed money income consumes 10 units of each good, then if p_1 increases to 1.5 and p_2 increases to 3 while his income increases to 45, his consumption of good 2 will not fall.
 - d) If a consumer's indifference curves are L-shaped, he will be just as well-off facing a quantity tax as an income tax.
 - e) Suppose goods 1 and 2 are perfect substitutes. If the government imposes a quantity tax of Rs. t per unit on the consumption of good 2, then the consumer will buy only good 1.
2. A person likes soft drinks but does not distinguish between brands A and B. He wants to have 14 bottles of soft drinks in a week. He would not drink any more even if they were free.
 - a) Draw his indifference curves showing brands A and B on two axes.

- b) Draw his indifference curves showing soft drinks and other goods on two axes.
3. A person's utility from good 1 and 2 is given by $U(x_1, x_2) = \min\left(\frac{x_1}{2}, x_2\right)$. Determine his demand functions.
4. A person enjoys commodities 1 and 2 according to the utility function $U(x_1, x_2) = \sqrt{x_1^2 + x_2^2}$. Maximize his utility if $p_1 = 3$, $p_2 = 2$ and he has Rs 50 to spend.
5. Suppose that a consumer's utility function is of the form $U(x_1, x_2) = x_1^a x_2^b$, where a and b are positive constants. Determine his: (a) demand functions; (b) indirect utility function; and (c) the expenditure function.
6. Suppose an individual consumes 3 goods, food, clothing and cars. Denote the quantities of these consumed by x_1 , x_2 and x_3 . The individual's utility function is given by $U(x_1, x_2, x_3) = 5 \log x_1 + 4 \log x_2 + \log(1 + x_3)$ and the prices of these goods are $p_1 = 1$, $p_2 = 2$ and $p_3 = 2,000$. Good 3 is a discrete good. If the person's total income is Rs. 9000, how will he allocate income so as to maximize his utility?
7. Show that a given government expenditure on a lump sum transfer program provides greater utility than does the same expenditure on subsidizing purchase of some good.
8. A gasoline tax of Rs 2.00 per litre is to be imposed for conservation of fuel. Since the government does not want the consumers to be worse off under taxation, it wishes to make lumpsum payments such that each consumer will get an amount sufficient to buy the bundle chosen before the tax-subsidy program was started. Given that other things did not change, what can you say about an average consumer's position under the tax-subsidy program by comparing with his original position, that is, is he better off, worse off, or just as well-off?
9. Assume that a consumer is choosing between housing services and consumption of other goods aggregated and measured in rupees.
- Show the utility maximizing position in a diagram.
 - The government agrees to subsidize consumer's by paying 50% of their housing cost. Show the new utility maximizing position.
 - Determine the utility maximizing position under an income supplement program which incurs the same amount of expenditure as the subsidizing program. Compare your results with those in situation (b).

10. A consumer purchases two goods. Under what circumstance does the consumer maximize utility? Move down/up the budget line to the left/right, and at some selected point explain the incentives that operate on the consumer.
11. Find the optimal commodity bundle for a consumer whose utility function and the budget constraint are $U(x_1, x_2) = x_1^{1.5}x_2$ and $3x_1 + 4x_2 = 100$ respectively.
12. A consumer spends Rs. 450 on two goods, 1 and 2. His utility function is $U(x_1, x_2) = 0.5 x_1 x_2^2$, $p_1 = 5$, $p_2 = 3$. What quantities of the two goods does he buy?
13. Consider the utility function $U(x_1, x_2) = g(x_1) + x_2$, where x_1 is the quantity of good 1, x_2 is the amount of money spent on the consumption of other goods and the function g is increasing and strictly concave. Determine the forms of the indirect utility function and the expenditure function.
14. A consumer's expenditure function is given by $e(p_1, p_2, I) = p_1^a p_2^{\frac{1}{2}} I$. Calculate the value of a .
15. A consumer's utility function is of the form $U(x_1, x_2) = -\frac{1}{x_1} - \frac{1}{x_2}$. Determine the ordinary demand functions, the indirect utility function and the expenditure function.
16. Show why a consumer will spend up to the point where the marginal benefit from the last unit of money spent on each commodity is the same.

4

Further Topics on Individual Demand

One major purpose of this chapter is to study the change in the quantity demanded of a particular commodity resulting from certain changed conditions. In particular, we will look at the effect of changes in income, in the price of the commodity and in the price of some other commodity. We will compare the new utility-maximizing choices with those that prevailed before conditions changed. This type of investigation is referred to as comparative statics analysis. Comparative means that we are comparing two optimal positions: before and after change in the economic environment. Statics mean that we are not studying the dynamics of movement from one choice to another, rather we are looking at the optimal choice.

An important issue that should be kept in mind in the context of comparative statics analysis is that the *ceteris paribus* assumption needs a clear recognition. We are changing only one variable at a time; everything else is being held constant. Not only are all other variables in the demand function held constant, but also preferences of the consumer are assumed to remain constant. Graphically, we will keep the consumer's indifference curves fixed and study the effects of shifting the budget constraint to alternative positions.

Another important objective of this chapter is to present the revealed preference approach to consumer choice.

CHANGES IN INCOME

We wish to investigate the impact of income changes on the quantities demanded of various goods when the prices of the goods remain fixed. This situation is illustrated in Fig. 4.1. As income increases (say, from m' to m'') the intercepts along the vertical and the horizontal axes increase. Since the slope of the budget line depends on the price ratio only and since the price ratio has not changed, the new budget line is parallel to the old one.

As income increases from m' to m'' , the quantity of commodity 1 demanded changes from x'_1 to x''_1 . Also the quantity of good 2 demanded changes from x'_2 to x''_2 . Since the price ratio p_1/p_2 stays constant, the utility maximizing condition also requires that the marginal rate of substitution of good 2 for good 1, MRS_{21} , stay constant as the consumer moves to a higher level of satisfaction. Therefore, MRS_{21} is the same at the point (x''_1, x''_2) as it is at (x'_1, x'_2) .

In Fig. 4.1, the amounts demanded by the consumer for both commodities 1 and 2 increase as income increases. But this need not be true for all goods and all consumers. It is well-known that as income increases, people tend to switch from low quality goods to high quality goods. Thus, income increase (with price constant) tends to increase the quantity demanded for some goods and decrease it for others.

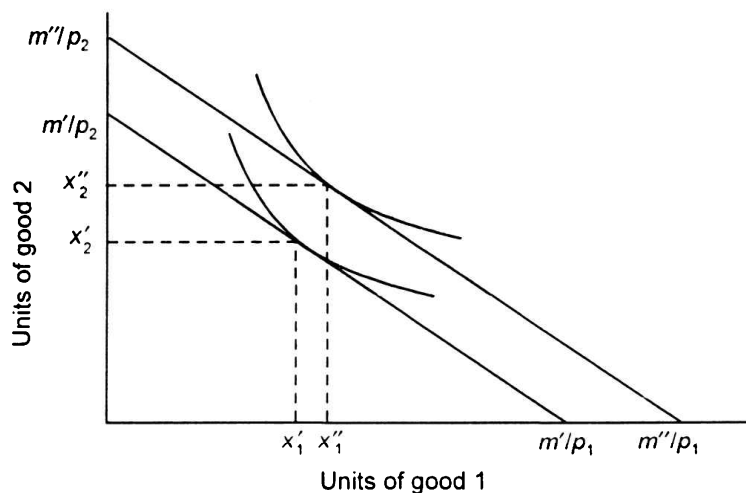


Fig. 4.1 Effect of Income Increase on Demand

Normal and Inferior Goods

A commodity is said to be a normal good if, when income increases with prices held constant, the quantity demanded of the good increases. In Fig. 4.1, both goods 1 and 2 increase as income increases. Therefore in this case both the goods are normal.

As stated earlier, there are some goods for which quantity decreases as income increases. Some examples of these goods are shacks, second hand garments, coarse clothes, low quality cigarettes, or nearly any kind of low quality good. Such a good is called an inferior good. In Fig. 4.2, good 1 is an inferior good and good 2 is a normal good, because an increase in income results in a reduction in consumption of good 1 and augments the consumption of good 2.

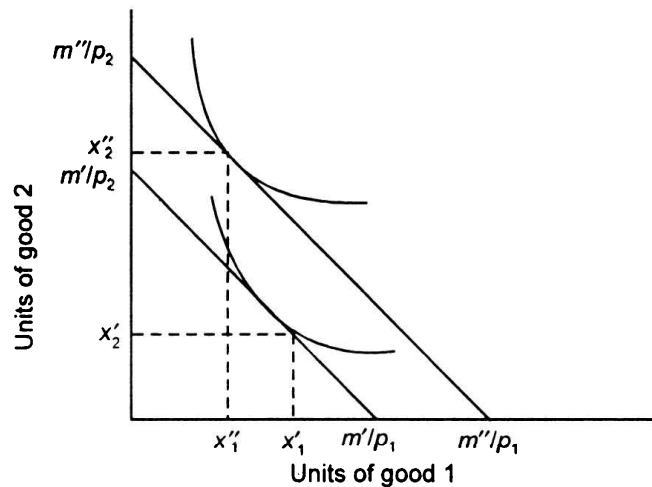


Fig. 4.2 Good 1 is Inferior and Good 2 is Normal

To define a normal good and an inferior good technically, recall that the demand function for good 1 can be written as $x_1 = f_1(p_1, p_2, m)$. Then good 1 is said to be normal or inferior according as $\Delta x_1 / \Delta m$ is positive or negative, where $\Delta m > 0$ is the increase in consumer's income and Δx_1 is the change in the consumption of good 1, keeping p_1 and p_2 constant. If $\Delta x_1 / \Delta m$ is zero, then good 1 is called impartial. An example of an impartial good is salt. Impartial goods are those whose consumption does not increase or decrease as the consumer's income goes up.

Whether a good is inferior or not depends on the consumer's income and his preferences. A good may be normal for one range of income but inferior

for another range. What is inferior for one consumer may be normal for another. The following general conclusion can be made regarding the extent of inferiority of commodities purchased by a consumer.

THEOREM: *For any values of prices and income, it cannot be the case that for a consumer all goods are inferior; at least one good must be normal.*

Proof. Let $f_1(p_1, p_2, m)$ and $f_2(p_1, p_2, m)$ be the consumer's demand functions for goods 1 and 2 respectively. We know that

$$p_1 f_1(p_1, p_2, m) + p_2 f_2(p_1, p_2, m) = m. \quad (4.1)$$

Suppose that the consumer's income increases by $\Delta m > 0$. We denote the changes in the quantities demanded by the consumer because of this income increase by Δx_1 and Δx_2 respectively. Since p_1 and p_2 are assumed to be fixed, we must have

$$p_1 \Delta x_1 + p_2 \Delta x_2 = \Delta m \quad (4.2)$$

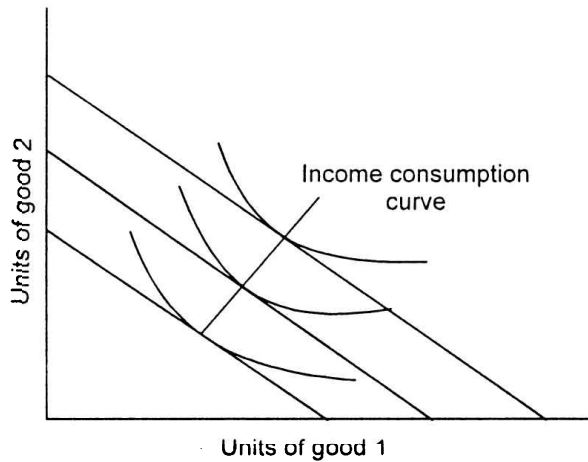
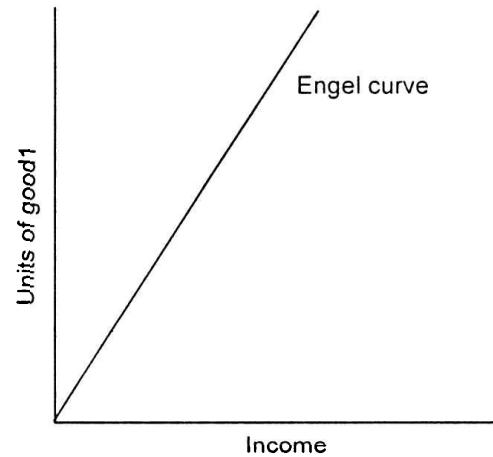
from which it follows that

$$p_1 \frac{\Delta x_1}{\Delta m} + p_2 \frac{\Delta x_2}{\Delta m} = 1. \quad (4.3)$$

Now, if both goods 1 and 2 are inferior, then both $\Delta x_1/\Delta m$ and $\Delta x_2/\Delta m$ are negative. But since p_1 and p_2 are positive, this possibility is ruled out in view of Eq. (4.3). Therefore, for Eq. (4.3) to hold, we require at least one of $\Delta x_1/\Delta m$ and $\Delta x_2/\Delta m$ to be positive. This means that both goods 1 and 2 cannot be inferior—at least one of them is normal.

Income Consumption and Engel Curves

Suppose that for given prices p_1 and p_2 , we allow income to vary over a broad range of values. Then the points of tangency of the budget lines with the indifference curves show the combinations of goods 1 and 2 which will be demanded at various income levels (Fig. 4.3). We connect together these tangency points to generate the consumer's income consumption curve. This curve therefore traces out the responsiveness of the quantities demanded of the two goods to income changes. The income consumption curve is also referred to as the income offer curve or income expansion curve. If both goods are normal then the income consumption curve will have a positive slope. If good 1 is inferior the curve will turn back itself to the left, and if good 2 is inferior it will fall to the right.

**Fig. 4.3** Income Consumption Curve**Fig. 4.4** Engel Curve

Together with the corresponding income levels, the income consumption curve provides the data for the construction of the consumer's Engel curves. The Engel curve for a commodity records the relationship between the quantities demanded of the good and the total income, with all prices being held constant (Fig. 4.4). That is, it is a graphical representation of the demand for the good as a function of income, keeping all prices constant. The Engel curves are not necessarily straight lines. The demand for some luxury goods may increase more proportionately than income, whereas the demand for necessities may grow less proportionately than income. It is clear that for an inferior good, the Engel curve is negatively sloped and for a normal good the curve is positively sloped.

To write the equation for the Engel curve of a good, say good 1, technically consider the demand function $x_1 = f_1(p_1, p_2, m)$ of this good. Since p_1 and p_2 are held constant, say at $\bar{p}_1$ and $\bar{p}_2$, we can rewrite f_1 as $x_1 = f_1(\bar{p}_1, \bar{p}_2, m) = E(m)$, for some real valued function E . The graph of $x_1 = E(m)$ is the Engel curve of good 1.

Some Examples

We will now illustrate the construction of income consumption curves and Engel curves using some examples.

Perfect Substitutes

Goods 1 and 2 are perfect substitutes. Suppose that the slope of the indifference curves is -1 . If $p_1 < p_2$, then the consumer will buy only good 1. Hence, if income increases, he will increase the consumption of good 1. Thus, the income consumption curve for good 1 is the horizontal axis. This is shown in Fig 4.5(a).

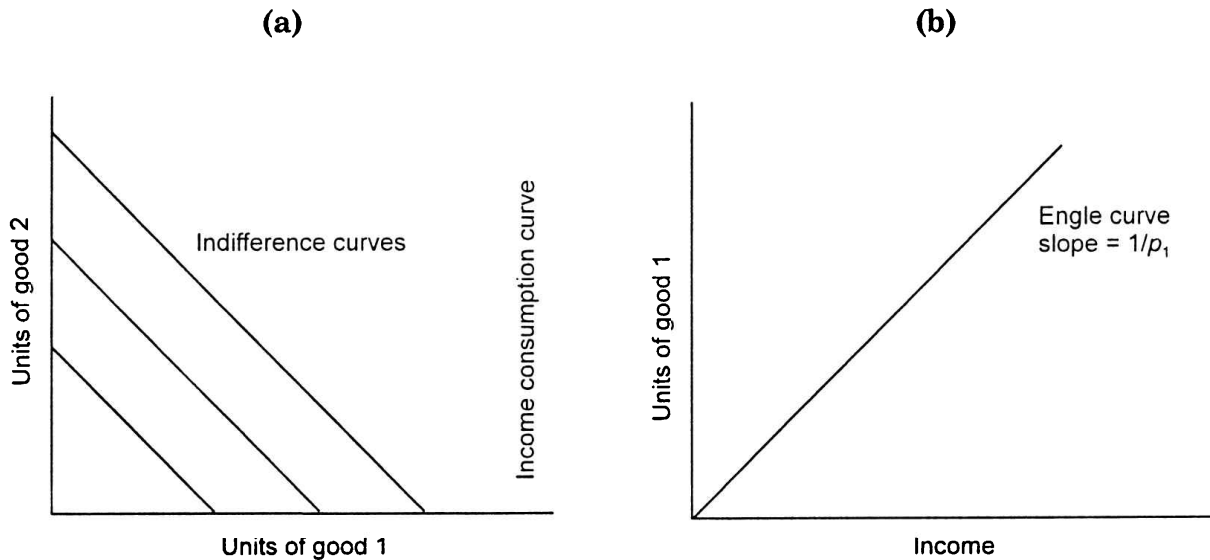


Fig. 4.5 Income Consumption Curve and Engel Curve for Perfect Substitutes

The demand function for good 1 is $x_1 = m/p_1$. The Engel curve is therefore a straight line with slope $1/p_1$. This Engel curve is depicted in Fig. 4.5(b).

Perfect Complements

Assume that goods 1 and 2, which are perfect complements, are purchased in 1:1 ratio. The consumer will always consume the same amount of each good. The income consumption curve will be the diagonal line through the origin. In this case the demand for good 1 is $x_1 = m/(p_1 + p_2)$. Therefore, the Engel curve (Fig. 4.6 (a) and (b)) will be a straight line with a constant slope of $1/(p_1 + p_2)$.

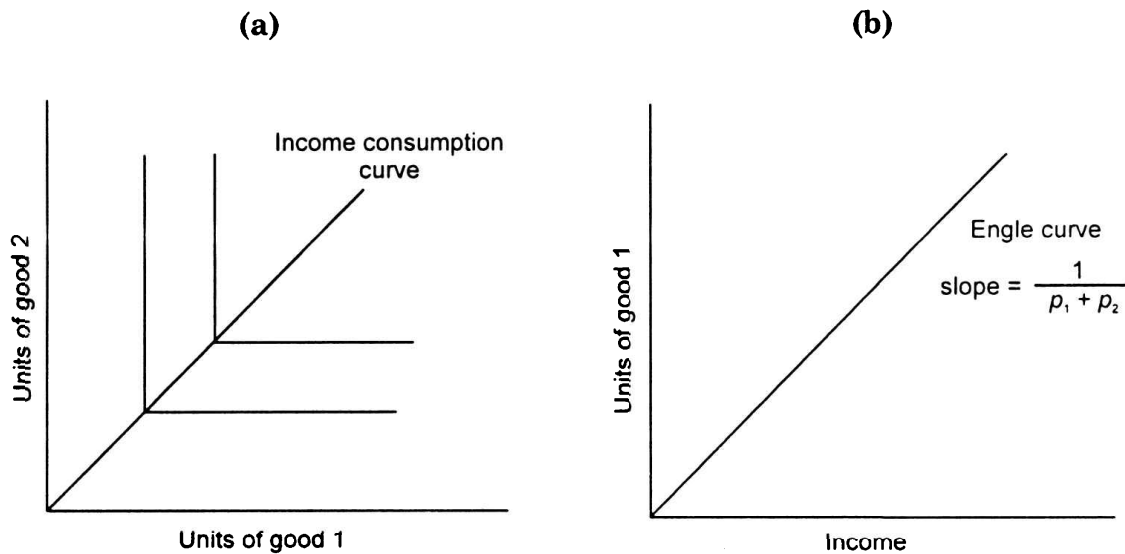


Fig. 4.6 Income Consumption Curve and Engel Curve for Perfect Complements

Homothetic Preferences

Suppose that a consumer's preferences depend on the ratio between quantities of good 1 and 2. That is, if the consumer prefers (x_1, x_2) to (y_1, y_2) , then he prefers (tx_1, tx_2) to (ty_1, ty_2) for all $t > 0$. Preferences satisfying this property are called homothetic; when income is scaled up or down by the scalar $t > 0$, the demanded bundle also gets scaled up or down by the same amount. This follows from the fact that if the indifference curve is tangent to the budget line that has income m at (x_1, x_2) , then the indifference curve through (tx_1, tx_2) is tangent to the budget line that has income tm and the same prices. Thus, for homothetic preferences, the income consumption curve and the Engel curves are straight lines through the origin. Examples of homothetic preferences are perfect substitutes, perfect complements and Cobb-Douglas preferences. If preferences are homothetic, then the entire set of indifference curves can be obtained by radial expansion or contraction of a single indifference curve with the help of rays emanating from the origin.

Impartial Good

Suppose that the consumer's utility function $U(x_1, x_2)$ is given by $U(x_1, x_2) = V(x_1) + x_2$, where x_1 is the quantity of good 1, x_2 is the amount of money spent on the consumption of other goods, and V is increasing and strictly concave. The utility function $U(x_1, x_2) = V(x_1) + x_2$ is linear in x_2 but non-linear in x_1 .

Hence it is called a quasilinear utility function. Clearly, the equation for an indifference curve takes the form $x_2 = I - V(x_1)$, where I is the constant level of utility along the indifference curve. In fact $x_2 = I - V(x_1)$ is the height of the indifference curve along the vertical axis from the origin. Therefore, in this case all indifference curves are shifted versions of one indifference curve. If an indifference curve is tangent to the budget line with income m' at (x'_1, x'_2) , then the budget line with income m'' ($m'' > m'$) will be tangent to a higher indifference curve at $(x'_1, x'_2 + k)$, where $k > 0$ is some constant. Increasing income does not change the demand for good 1, and all the extra income is spent on the consumption of good 2, that is, other goods. Therefore, the budget line with income m'' will touch an indifference curve at $[x'_1, x'_2 + (m'' - m')/p_2]$, given that the budget line with income m' touches a lower indifference curve at (x'_1, x'_2) . In this case good 1 is impartial. The Engel curve for good 1 is a horizontal line.

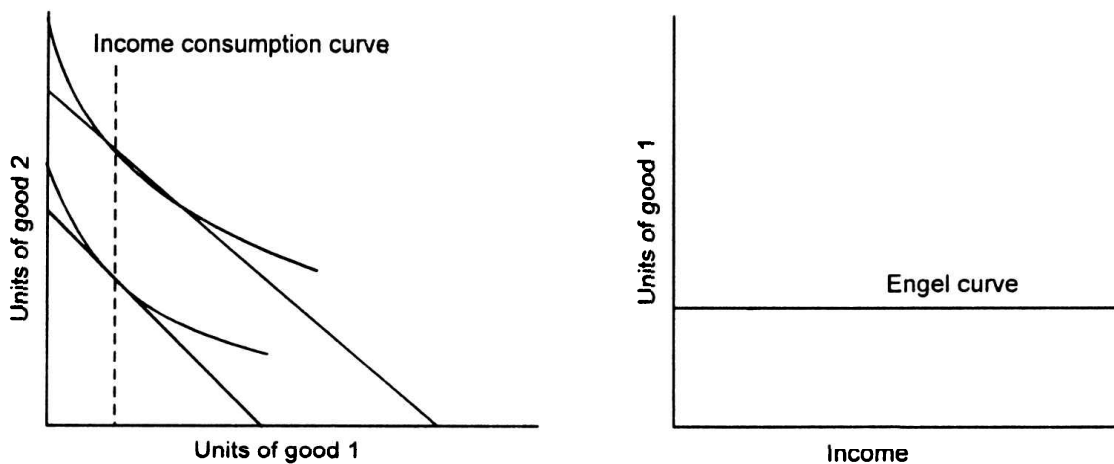


Fig. 4.7 Income Consumption Curve and Engel Curve for an Impartial Good

CHANGES IN A GOOD'S OWN PRICE

Another concern of this chapter is to investigate how a consumer's behavior changes in response to changes in the price of a good. We will first consider how a consumer's choice of a good responds to changes in its price. The analysis of the effect of a price change on the quantity of a good demanded is somewhat more complex than that of the effect of a change in income. Geometrically, this is because a price change involves not only a change in the position of the budget constraint, but also a change in the slope of the budget line. As a result, moving to the new utility maximizing choice entails not only moving

to another indifference curve but also changing the marginal rate of substitution. Therefore, when a price changes, two analytically different effects come into play. One of these is the substitution effect—the change in demand due to change in the rate of exchange between the two goods. The second effect, the income effect, arises because a price change naturally changes a consumer's purchasing power. We begin with a graphical analysis of these effects.

Substitution and Income Effects: Both Goods Normal

In order to give precise definitions of the substitution and income effects we have to analyze the impact of price change in greater detail. For this let us refer to Fig. 4.8. Here we have a situation where the price of good 1 has increased from p'_1 to p''_1 , with the price of good 2 p'_2 and income m' remaining constant. The initial and new budget lines are therefore given by $p'_1x_1 + p'_2x_2 = m'$ and $p''_1x_1 + p'_2x_2 = m'$ respectively, with the former being flatter than the latter. Suppose that the consumer's initial utility maximizing bundle A is on the indifference curve I_2 . An increase in the price of good 1 from p'_1 to p''_1 makes the consumer worse off; he ends up at the lower indifference curve I_1 . His new utility maximizing bundle is given by C . At C , the consumer buys a less quantity of good 1 and a larger quantity of good 2 than at A . (The reason behind this is that good 1 is more expensive at C than at A . However, it may be possible that at the new utility maximizing choice the consumer demands less of both goods than at A).

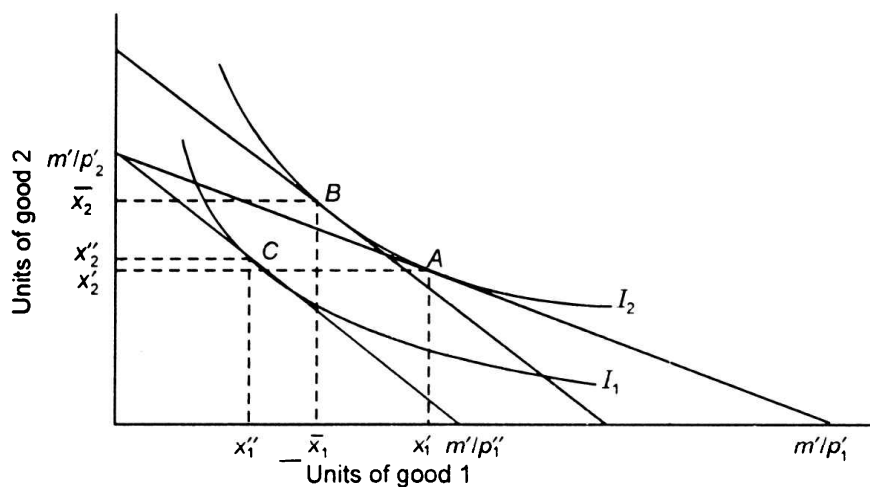


Fig. 4.8 Substitution and Income Effects (Both Goods Normal)

The movement from A to C can be viewed as being composed of two effects. An increase in the price of good 1 makes it relatively more expensive compared to good 2 than before and decreases the purchasing power of income. But suppose that because of this price increase the consumer is given sufficient additional income to remain on the original indifference curve I_2 . The additional purchasing power or the compensating increase in income enables the consumer to remain at the same level of preference as the combination A . But since now good 1 is relatively more expensive than it was, good 2 becomes a more attractive alternative to the consumer. Therefore the consumer is induced to substitute the relatively cheaper good 2 for the relatively more expensive good 1. Consequently, the consumer's utility maximizing position under the income compensation scheme is given by B . Since the consumer faces the same prices (p_1'' , p_2') at the combinations C and B , the budget line through C is parallel to the budget line through B . The movement from A to B is called the substitution effect. This effect indicates how the consumer substitutes one good for the other when the price of the latter changes but the consumer remains on the same indifference curve. Since the income effect of the increase in price of good 1 has been eliminated by the compensating variation in the consumer's income, the substitution effect results solely from the change in the price of good 1 relative to that of good 2.

The income effect can be determined by taking the compensating variation away from the consumer. The movement from B to C is the income effect of the increase in price of good 1. Since at B and C the price levels are the same, and since the consumer's income is lower when he is at C than at B , the move from B to C is a move along the consumer's income consumption curve. Consequently, since goods 1 and 2 are normal, the quantities demanded of both the goods will be lower when the consumer is at C than at B .

It is important to realize that the consumer does not actually move from A to B and then to C . We never observe the point B , because the consumer is not actually compensated by an increase in income when the price of good 1 increases. What happens is that he actually moves from the bundle A to bundle C . However, the notions of income and substitution effects are analytically valuable because they show that a price change affects the quantity of good 1 that is demanded in two conceptually different ways.

To express the substitution and the income effects analytically, let $\Delta p_1 = p_1'' - p_1'$ represent the change in price of good 1, and $\Delta m = m'' - m'$ represent the change in income necessary to keep the consumer on the same indifference curve, that is, $m' + \Delta m = m''$ is the income level that enables the consumer

to buy the bundle B at prices (p_1'', p_2') . Then the substitution effect, Δx_1^s , the change in demand for good 1 when price of good 1 increases to p_1'' , and, at the same time, income increases to m'' is

$$\Delta x_1^s = f_1(p_1'', p_2', m'') - f_1(p_1', p_2', m') \quad (4.4)$$

where f_1 is the demand function for good 1.

On the other hand, the income effect, Δx_1^n , the change in demand for good 1 when we decrease income from m'' to m' , keeping the price of good 1 fixed at p_1'' is

$$\Delta x_1^n = f_1(p_1'', p_2', m') - f_1(p_1'', p_2', m'') \quad (4.5)$$

Since $m' < m''$ and good 1 is normal, it follows that $\Delta x_1^n < 0$. When the price of a good increases, we need to increase income in order to keep the purchasing power constant. If the good is normal, this increase in income will increase the demand.

Now, the total change in demand, Δx_1 , is the change in demand due to a price change, keeping income fixed:

$$\begin{aligned} \Delta x_1 &= f_1(p_1'', p_2', m') - f_1(p_1', p_2', m') \\ &= [f_1(p_1'', p_2', m'') - f_1(p_1', p_2', m')] + \\ &\quad [f_1(p_1'', p_2', m') - f_1(p_1'', p_2', m'')] \\ &= \Delta x_1^s + \Delta x_1^n \end{aligned} \quad (4.6)$$

That is, the total change in demand equals the sum of the substitution effect and the income effect. This equation is called the Slutsky equation, named after the Russian economist Eugen Slutsky (1880–1948). Note that in calculating the substitution effect, since we move along an indifference curve, in Eq. (4.6) Δx_1^s assumes constancy of utility. In the case when goods 1 and 2 are normal, the signs of the substitution and the income effects unambiguously determine the sign of the overall change in demand Δx_1 . That is, in this case $\Delta x_1 < 0$.

The substitution effect always moves opposite to the movement of price. That is, when price goes up, the demand for the good due to the substitution effect reduces. Therefore, the substitution effect Δx_1^s is always negative. In the case considered here, where both goods 1 and 2 are normal, the substitution effect and the income effect therefore operate in the same direction.

The Slutsky equation, Eq. (4.6), has been stated in terms of absolute changes. However, it is more common to express it in terms of rates of change. If we divide each side of Eq. (4.6) by Δp_1 , we get

$$\frac{\Delta x_1}{\Delta p_1} = \frac{\Delta x_1^s}{\Delta p_1} - \frac{\Delta x_1^m}{\Delta p_1} \quad (4.7)$$

where $\Delta x_1^m = -\Delta x_1^n$. (When the Slutsky equation is expressed in terms of rate of change, it becomes convenient to define Δx_1^m as the negative of Δx_1^n .) The first term on the right hand side of Eq. (4.7) is the rate of change of demand when the price changes and income is adjusted to keep the consumer on the same indifference curve—the substitution effect. The second term can be written as

$$\Delta x_1^m / \Delta p_1 = (\Delta x_1^m / \Delta m)(\Delta m / \Delta p_1).$$

Now, $\Delta x_1^m / \Delta m$ is the rate of change of demand when income changes, and $\Delta m / \Delta p_1$ is the rate of change of income when price changes. Thus, the income effect itself can be decomposed into two terms: how demand changes as income changes, and how income should change as price changes if the consumer should remain on the same indifference curve. (We use calculus to decompose total change in demand into income and substitution effects in Appendix 4.1.).

Substitution and Income Effects: One Good Normal and One Good Inferior

If good 1 is inferior, then the substitution effect and the income effect of an increase in its price act in opposite directions. While the substitution effect acts to decrease the quantity demanded of good 1, the income effect acts to increase the quantity demanded.

In Fig. 4.9, the substitution effect increases the consumption of good 2 at the expense of good 1, that is, $\bar{x}_2 > x'_2$ while $\bar{x}_1 < x'_1$. However, the income effect (the move from B to C), results in a higher consumption of good 1 ($\bar{x}_1 < x''_1$), while the consumption of good 2 falls from $\bar{x}_2$ to x''_2 . Thus, $\Delta x_1^n > 0$ in this case. We however note that the substitution effect dominates the income effect in the figure. Thus, the overall change in demand Δx_1 is negative.

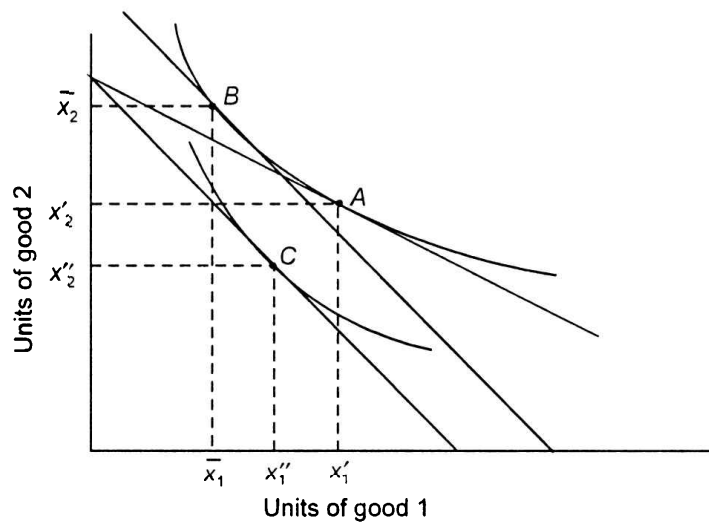


**Fig. 4.9 Substitution and Income Effects
(Good 1 Inferior and Good 2 Normal)**

If the income effect dominates the substitution effect, then the total change in demand Δx_1 is positive. This means that an increase in price generates an increase in demand. Such a good is called a Giffen good after the nineteenth century economist Robert Giffen who first noted this possibility. The basis of Robert Giffen's claim is that in late nineteenth century England, when the price of bread rose, the working class consumed more of it. The increase in price reduces the consumer's purchasing power so much that he increases his consumption of the inferior good. Because of this the consumer is forced to reduce the consumption of the other good.

Given that the income effect is positive for a Giffen good, we can say that a Giffen good is an inferior good. But an inferior good is not necessarily a Giffen good. For a Giffen good we not only require the income effect to be positive, it has also to be large enough to outweigh the negative sign of the substitution effect so that the total change in demand becomes positive.

We illustrate the situation graphically in Fig. 4.10. Here the income effect is so strongly inferior for good 1 that the substitution effect is overwhelmed, and at point C, the quantity demanded of good 1, x_1'' is larger than the original quantity demanded.

Giffen goods are unlikely to be encountered in real-world behavior. In economic literature the Giffen good case is regarded as a theoretical oddity. A considerable amount of empirical evidence should be available to support the behavior that price increase may lead to increase in demand.

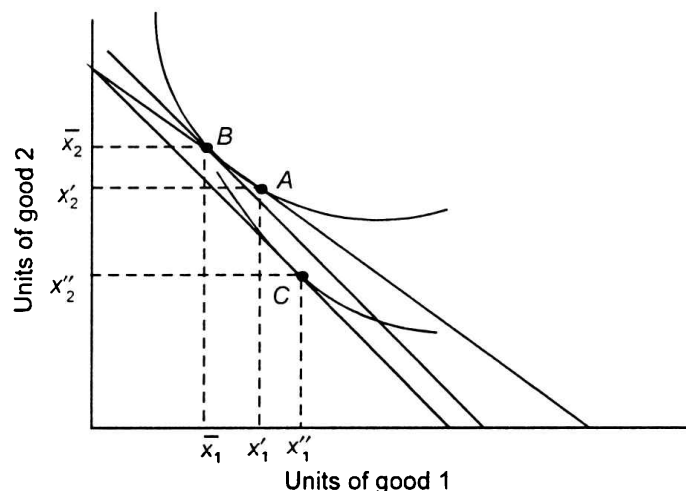


Fig. 4.10 Case of a Giffen good

In one case we have a believable argument for the existence of a Giffen good. Suppose that some outsiders are thinking of buying shares of stock in a firm, but they are not certain whether the firm's dividends will increase in the future. Insiders who have access to such information are likely to buy stock if they receive favorable information. The outsiders then might regard an increase in the price of the stock as a signal that dividends are going to rise, which in turn may generate an increase in the demand for the number of shares. Hence in this case an increase in price leads to an increase in demand.

The notion of the substitution effect we have studied above is called the Hicks substitution effect. There is another definition of the substitution effect. This definition of the substitution effect, which is due to Slutsky, is the change in demand when price changes but a consumer's purchasing power is held constant, so that the original bundle remains affordable. We will discuss the Slutsky substitution effect later in this chapter.

EXAMPLES OF SUBSTITUTION AND INCOME EFFECTS

In this section we consider some examples of preferences and look at the Hicks decomposition of demand change resulting from price changes into the substitution and income effects for these preferences.

Perfect Complements

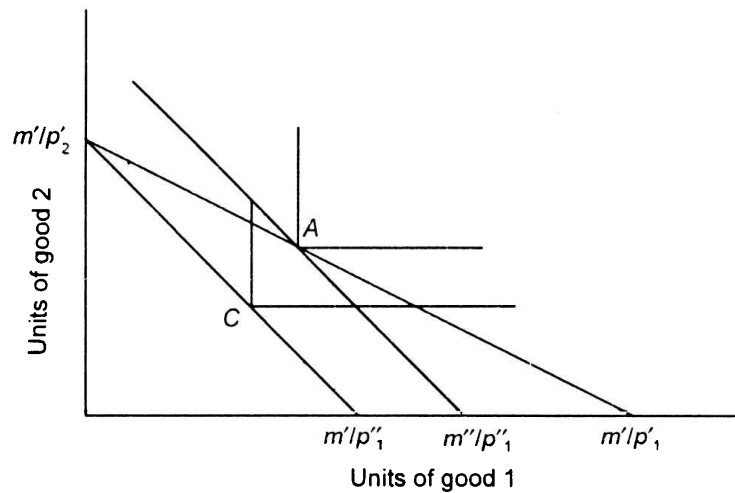


Fig. 4.11 Hicks Decomposition for Perfect Complements

An increase in the price of good 1 from p'_1 to p''_1 shifts the consumer's optimal choice from A to C. The budget line with horizontal intercept m''/p''_1 that parallels the new budget line touches the original indifference curve at the original optimal choice A, where m'' is the income that enables the consumer to remain on the original indifference curve after a price rise. (The consumer's actual income is m' .) Therefore, the substitution effect is zero. The change in demand is due to the income effect only.

Perfect Substitutes

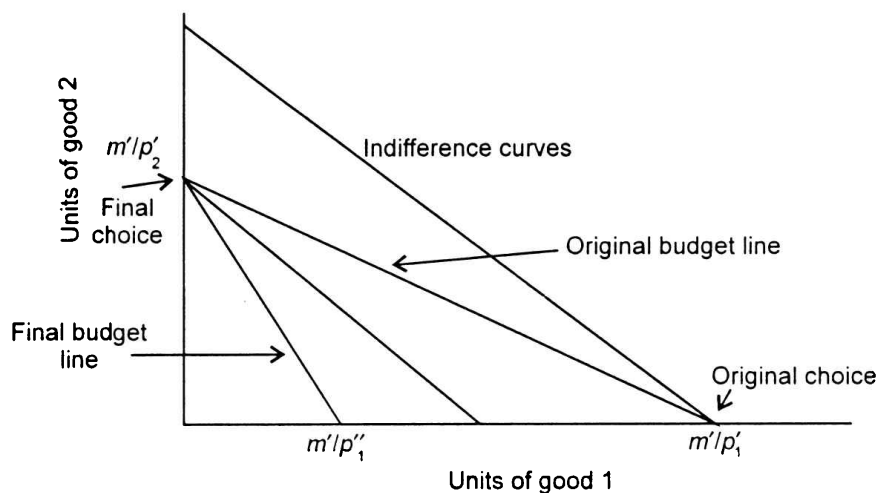


Fig. 4.12 Hicks Decomposition for Perfect Substitutes

Here as p'_1 increases to p''_1 , the demand bundle jumps from the horizontal axis to the vertical axis. The total demand change is due to the substitution effect. The income effect is zero.

Impartial Good

We have already seen that if the utility function $U(x_1, x_2)$ is of the type $V(x_1) + x_2$, then an increase in income level does not change the demand for good 1, that is, good 1 is an impartial good. This will mean that the total change in demand resulting from a rise in the price of good 1 is due to the substitution effect, and the income effect is zero.

THE LAW OF DEMAND

We have seen that the demand for a good may increase or decrease in response to an increase in the consumer's income. An increase in the price of a commodity may also lead to an increase or a decrease in demand. Furthermore, the change in demand resulting from a price change can be decomposed into two components: the income effect, and the substitution effect.

We say we have a normal good if its demand increases when income increases. Now, for a normal good the substitution and the income effect work in the same direction: they are both negative. Therefore, for a normal good an increase in price unambiguously reduces demand.

In view of the above arguments we can state the following Law of Demand:

LAW 4.1: *If the demand for a good increases when income increases, the demand for that good must decrease when its price increases.*

DERIVING CONSUMER'S DEMAND AND PRICE OFFER CURVES

We now show how a consumer's demand curve and price offer curve can be derived from utility maximization. It was shown that a consumer's demand for a good (say, good 1) depends on the shape of the utility function and all prices and income. The Marshallian demand curve, or simply, the demand curve for good 1 shows the quantities demanded of the good as a function of its price, assuming that income and all other prices are held constant. To show the construction of the demand curve for good 1, consider the demand function $x_1 = f_1(p_1, p_2, m)$ for this good.

Fig. 4.13 shows that by varying p_1 , the price of good 1, we can read the utility maximizing quantities of good 1 associated with these prices. The quantities of the good chosen decrease from x'_1 to x''_1 as the good's price increases from p'_1 to p''_1 . This is in accord with our general conclusion that, except in the unusual case of a Giffen good, the quantity demanded of a good falls as price increases. Clearly, $x'_1 = f_1(p'_1, p'_2, m')$ and $x''_1 = f_1(p''_1, p'_2, m')$. The demand curve for good 1 is then defined as the plot of quantities demanded against prices. The demand curve is written as $f_1(p_1, p'_2, m')$, because while it gives the quantities demanded of good 1 as a function of p_1 , still the location of the curve depends on p'_2 and m' . We show the curve in Fig. 4.14.

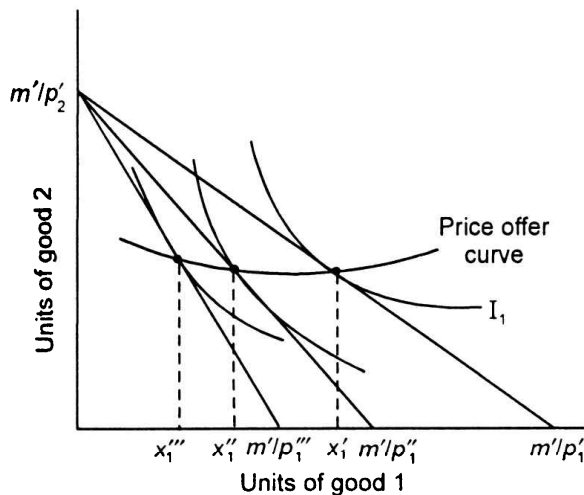


Fig. 4.13 Indifference Map

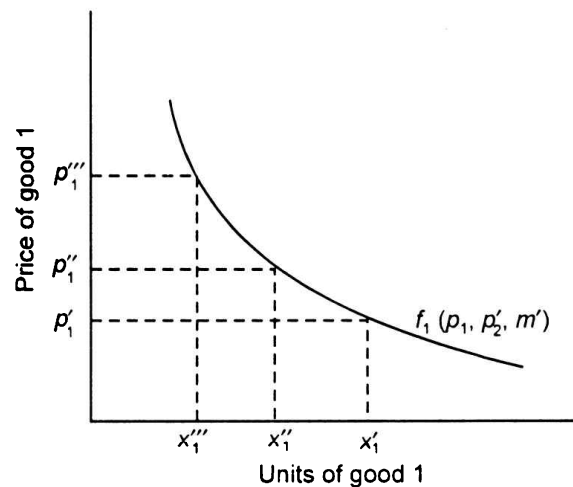


Fig. 4.14 Demand Curve

If p_1 varies, with p_2 and m fixed, we move along a given demand curve. If any of the three basic factors income, price of the other good and the individual's preferences that were held constant in deriving the demand curve in Fig. 4.14 were to change, the entire demand curve would shift to a new position. For example, if m increases the curve will shift outward (provided that the good is normal). More quantity of good 1 will be demanded at each price. If the other price p_2 changes, the curve will shift inward or outward depending on how goods 1 and 2 are related. Later in this chapter we will examine this relationship in more detail. Finally, the curve will shift if the individual's preference for good 1 changes. A sudden improvement in the quality of good 1 may shift the demand curve outward.

Thus, when any one of the determinants of demand changes, the result is a change in demand—an entirely new demand schedule is represented graphically by a shift of the demand curve to a new position. On the other hand, a

change in the price of the product results in a change in the quantity demanded, and is represented graphically by a movement along a stationary demand curve.

The curve we have depicted in Fig. 4.14 is in fact the inverse demand curve that pictorially represents price as a function of quantity. Given that the quantity demanded of a good is a decreasing function of its price, the inverse demand curve measures the same relationship as the direct demand curve that shows quantity as a function of price.

The inverse demand curve can be interpreted economically in an interesting way. Recall that a consumer's optimal choice must satisfy the condition that the marginal rate of substitution of good 2 for good 1, MRS_{21} , equals the price ratio p_1/p_2 :

$$MRS_{21} = \frac{p_1}{p_2} \quad (4.8)$$

We can rewrite the above condition as

$$p_1 = p_2 MRS_{21} \quad (4.9)$$

That is, at the optimal level of demand, p_1 is directly proportional to MRS_{21} .

Now, suppose for simplicity that $p_2 = 1$. Then Eq. (4.9) says that p_1 measures the rate at which the consumer is willing to exchange good 1 for good 2. Thus, in this case the inverse demand curve simply measures MRS_{21} . For any optimal quantity of x_1 , the inverse demand curve shows how much of good 2 the consumer would require to compensate himself for a small reduction in the amount of good 1.

If good 2 is regarded as the amount of money spent on other goods, then MRS_{21} is how much money the consumer would be willing to give up to purchase a little more of good 1. That is, MRS_{21} measures the marginal willingness to pay. Thus, for $p_2 = 1$, p_1 itself is measuring the marginal willingness to pay. The marginal willingness to pay decreases as we increase the consumption of good 1.

Let us now refer to Fig. 4.13. If we connect the optimal points together we get the price offer curve for good 1. This curve represents the bundles that would be chosen by the consumer at different prices for good 1.

SOME EXAMPLES OF DEMAND AND PRICE OFFER CURVES

We will now consider some examples of demand and offer curves using the preferences that we discussed earlier.

Perfect Substitutes

Goods 1 and 2 are perfect substitutes. The slope of the indifference curves is equal to -1 . The demand for good 1 is zero if $p_1 > p_2$, any amount on the budget line if $p_1 = p_2$ and m/p_1 when $p_1 < p_2$. The offer curve traces out these situations.

To derive the demand curve for good 1, we fix the price of good 2 at $\bar{p}_2$ and plot the quantities demanded against different price levels.

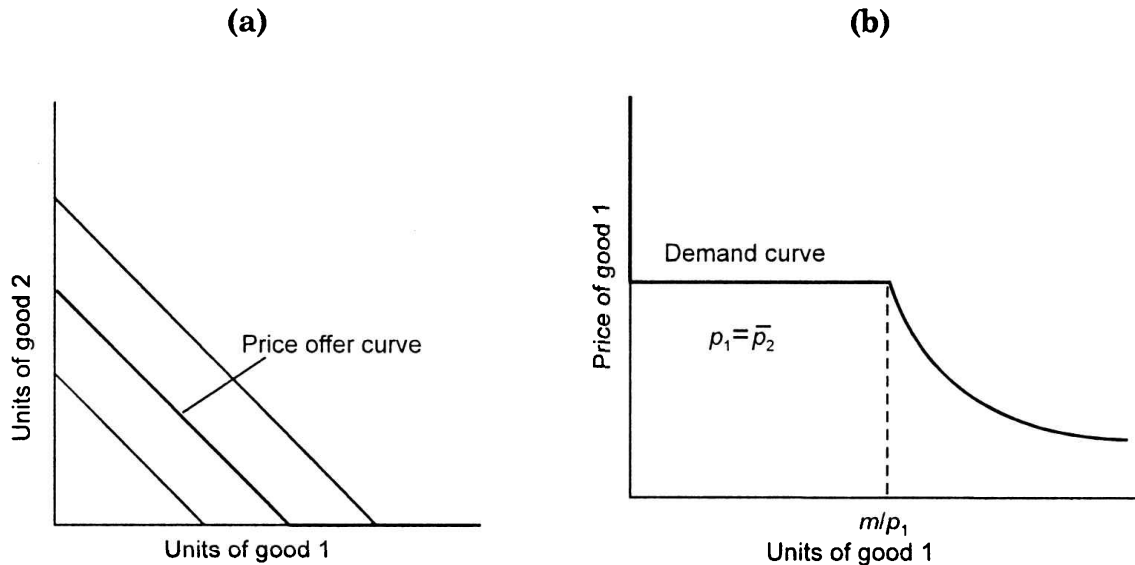


Fig. 4.15 Price Offer Curve and Demand Curve for Perfect Substitutes

Perfect Complements

Suppose that goods 1 and 2, which are perfect complements, are purchased in 1:1 ratio. Then the demand function for good 1 is $x_1 = m/(p_1 + p_2)$. We show this graphically in Fig. 4.16(b). The offer curve will be a diagonal line as shown in Fig. 4.16(a).

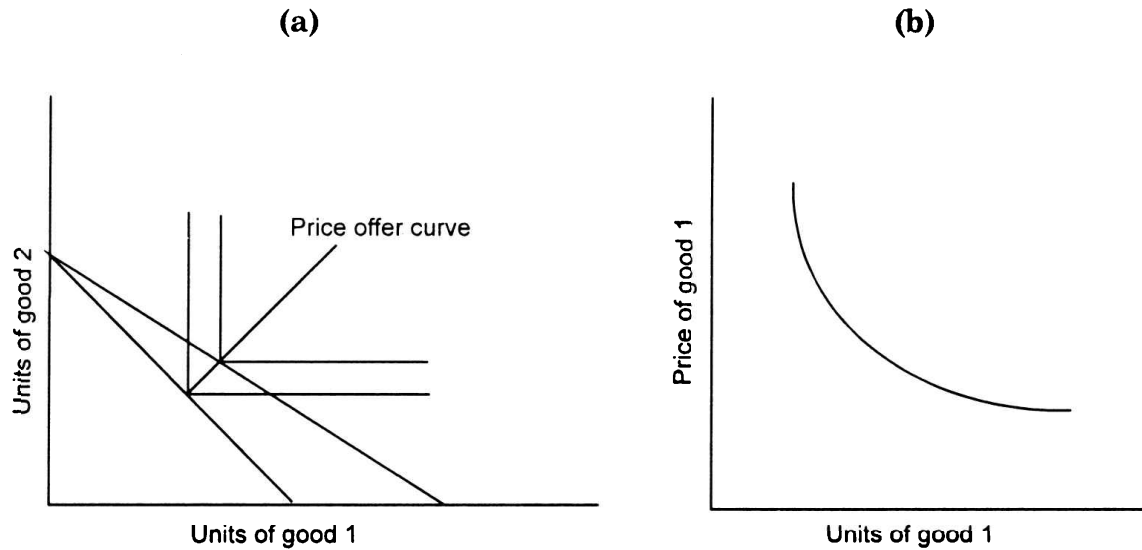


Fig. 4.16 Price Offer Curve and Demand Curve for Perfect Complements

A Discrete Good

Good 1 is assumed to be a discrete good. The consumer will buy zero or one unit of the good according as the price level p_1 is very high or very low. At some price r_1 , he may be indifferent between buying good 1 or not buying it. The price at which the consumer is indifferent between buying and not buying the good is called the reservation price.

At price r_1 , the consumer wishes to buy one unit of the good. If the price decreases to r_2 , he wishes to buy one more unit, and so on. Therefore, the demand behavior can be described by a sequence of reservation prices.

We can define the reservation price sequence using the utility function. For instance, r_1 is the price where the consumer is indifferent between buying one unit of the good or not buying it. Therefore,

$$U(0, m) = U(1, m - r_1) \quad (4.10)$$

Similarly, r_2 is implicitly defined by

$$U(1, m - r_2) = U(2, m - 2r_2) \quad (4.11)$$

The left hand side of Eq. (4.11) gives the utility from consuming one unit of good 1 at price r_2 . The right hand side gives the utility from consuming two units of the good, each of which sells at r_2 .

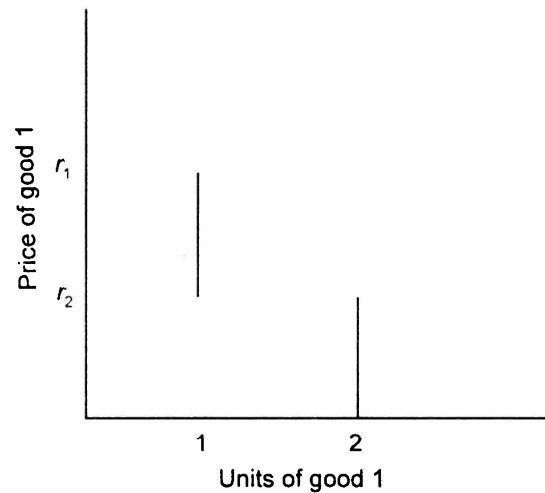


Fig. 4.17 Demand Curve for a Discrete Good

By plotting these reservation prices against the corresponding units of good 1, we get the demand curve for this good.

COMPENSATED DEMAND CURVES

The demand curve we have derived above is the observed demand curve of the consumer. This observed demand curve reflects both the substitution and the income effects of price changes. As the price of the good, say good 1, falls, the individual demands larger quantities of the good, assuming that the nominal income and the other price remain fixed. Although this is the most common way of imposing the *ceteris paribus* assumption in developing a demand curve, it is not the only way. An alternative approach holds the individual's utility and the price of the other good constant while examining reactions to changes in p_1 . The demand curve that we will generate here by plotting the price-quantity values on a graph is called the compensated demand curve because the effects of price changes on purchasing power are compensated so as to constrain the consumer on a given indifference curve. The compensated demand curve is also called the Hicks demand curve.

The derivation of the compensated demand curve is illustrated in Fig. 4.18. In panel (a), we see that if the price of good 2 is held constant at p'_2 and if income of the consumer is m' , then x'_1 units of good 1 is demanded at a price p'_1 . When p'_1 increases to p''_1 , with p'_2 constant, income has to increase to m'' for enabling the consumer to remain on the indifference curve I_1 .

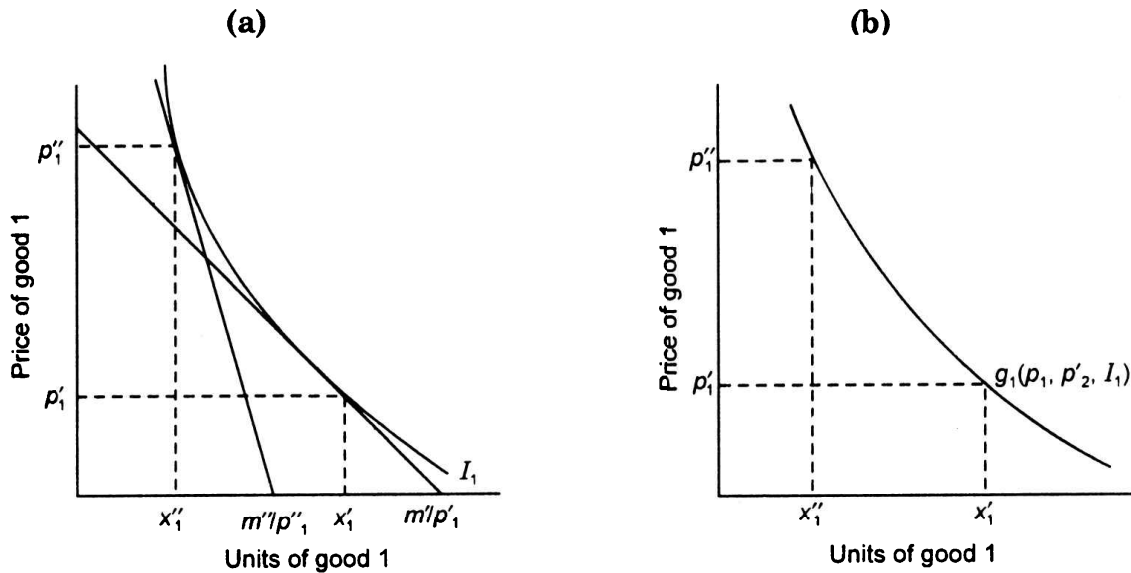


Fig. 4.18 The Compensated Demand Curve

Given m'' , the consumer chooses to consume x_1'' units of good 1 at the price level p_1'' . Transferring these choices to a price-quantity diagram we obtain the compensated demand curve for good 1 that shows the relationship between the price of good 1 and the quantity purchased on the assumption that the other price and utility are held constant. The location of the demand curve which is shown in panel (b) of Fig. 4.18 is indicated by $g_1(p_1, p_2', I_1)$. The price of good 2 is held constant at p_2' along the curve, as is the level of preference at I_1 .

We can also arrive at the compensated demand curve using the consumer's expenditure function. Given an indifference curve I_1 , and given some fixed p_2' , we choose x_1 and x_2 to minimize the cost of reaching the indifference curve I_1 . The minimum expenditure is $p_1x_1 + p_2'x_2$, and this is the income needed to reach the indifference curve.

Since the compensated demand curve is derived under the assumption of a fixed level of utility, there is no income effect operating on demand. Now, in the uncompensated case only a positive income effect (inferior good) may lead to an upward sloping demand curve. Since the compensated demand curve involves the substitution effect only, and since the substitution effect is always negative, we can unambiguously conclude that the compensated demand curve always slopes downward to the right.

The relationship between the compensated and the uncompensated demand curves is shown in Fig. 4.19. At p_1' the curves intersect, since at that price the individual's income is just sufficient to attain the indifference curve I_1 .

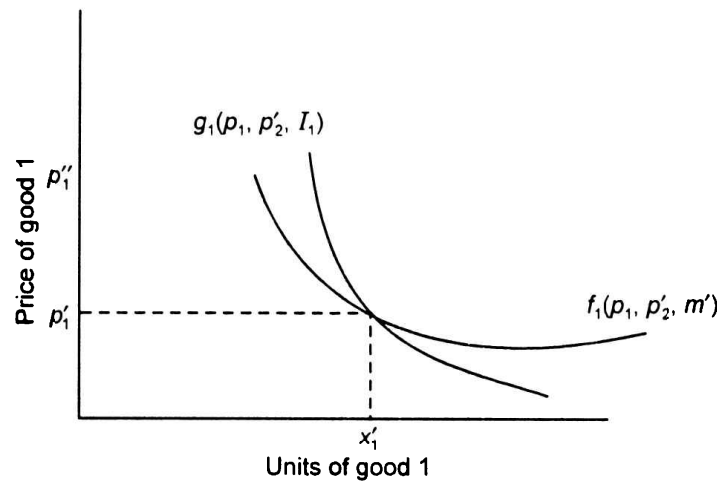


Fig. 4.19 Comparison of Uncompensated and Compensated Demand Curves

(Compare Fig. 4.14 and panel (b) of Fig. 4.18.) Hence x'_1 is demanded under either demand concept. For a price above p'_1 (such as p''_1), the individual requires a positive income compensation since he needs some monetary help to remain on I_1 . Hence assuming that good 1 is normal, at p''_1 more of good 1 is demanded in the compensated case than in the uncompensated case. For prices below p'_1 , the individual suffers a negative compensation to prevent an increase in utility from the lower price. Therefore less of good 1 is demanded with the compensated curve than with the uncompensated curve. In general, the uncompensated demand is flatter than the compensated one because the former incorporates both substitution and income effects whereas the latter reflects only the substitution effect.

The choice between the use of compensated and uncompensated demand curves in economic analysis is a matter of convenience. In most empirical works uncompensated curves are used since the data on income and prices to estimate them are readily available. However, for some theoretical works compensated curves are more appropriate because constancy of utility offers several advantages.

GROSS SUBSTITUTES AND COMPLEMENTS

We have already talked about perfect substitutes and perfect complements. But substitutes and complements need not be perfect always. For instance, a pencil is not a perfect substitute for a pen. This is a case of imperfect substitution. A pen and a pencil are substitutes for each other to some extent.

Similarly, a right shoe and a left shoe are perfect complements because they are always bought together. A pair of socks is another example of perfect complements. A pair of shoes and a pair of socks are usually bought together, but not always. That is why they are not perfect complements.

In order to define substitutes and complements rigorously, let us recall the demand functions $x_1 = f_1(p_1, p_2, m)$ and $x_2 = f_2(p_1, p_2, m)$ for goods 1 and 2 respectively. We say that good 1 is a gross substitute for good 2 if the demand for good 1 goes up as the price of good 2 goes up. In terms of rate of change, good 1 is a gross substitute for good 2 if

$$\frac{\Delta x_1}{\Delta p_2} > 0. \quad (4.12)$$

The idea behind this definition is that as good 2 becomes more expensive, the consumer substitutes the more expensive good by the less expensive one.

Similarly, we say that good 1 is a gross complement to good 2 if the demand for good 1 decreases as the price of good 2 increases. That is,

$$\frac{\Delta x_1}{\Delta p_2} < 0. \quad (4.13)$$

Complements are goods that are consumed together. Therefore if the price of one good increases, the demand for the other good goes down.

Eqs. (4.12) and (4.13) make it clear that the above definitions are gross definitions in the sense that they include both income and substitution effects that arise from price changes. These definitions illustrate perfect substitutes and perfect complements nicely. Note that $\Delta x_1/\Delta p_2$ is positive (negative) in the case of perfect substitutes (perfect complements). If $\Delta x_1/\Delta p_2 = 0$, then the goods are independent.

There are several things that are undesirable about gross definitions of substitutes and complements. The most important is that these definitions are not symmetric. It may be possible for good 1 to be a substitute for good 2, and at the same time for good 2 to be a complement of good 1. The presence of income effects can give rise to such paradoxical results. For example, consider two consumption items: good 1 (food) and good 2 is inferior. If the price of food (p_1) rises, it may be that $\Delta x_2/\Delta p_1 < 0$. Because food is an important item in consumption, a rise in p_1 will substantially reduce the real income and may reduce the demand for good 2 even though there may be a substitution

effect favoring a consumption of good 2 as a replacement of good 1. On the other hand, if p_2 rises, it is quite likely that $\Delta x_1/\Delta p_2$ will not induce an important income effect, and the principal effect here is a substitution effect of good 1 for good 2. Consequently, good 2 would be termed a complement for good 1, whereas good 1 would be regarded as a substitute for good 2. Next, since income is fixed, in the two-good case if we spend more money on the consumption of good 1, we will have to spend less on good 2. This puts some restrictions on the kinds of possible behavior in the two-good case.

NET SUBSTITUTES AND NET COMPLEMENTS

Good 1 is said to be a net substitute for good 2 if, when the consumer remains on a given indifference curve, an increase in the price of good 2 leads to an increase in the demand for good 1. In terms of the rate

$$\frac{\Delta x_1}{\Delta p_2} \Big|_{U = \text{const.}} > 0. \quad (4.14)$$

Good 1 is said to be a net complement of good 2 if, along a given indifference curve, an increase in the price of good 2 generates a decrease in the demand of good 1.

$$\frac{\Delta x_1}{\Delta p_2} \Big|_{U = \text{const.}} < 0. \quad (4.15)$$

These definitions look only at the substitution effect to determine whether two goods are substitutes or complements.

It is clear that if there are two goods, they are always net substitutes. Along a given indifference curve an increase in p_2 leads to less of good 2 and more of good 1 being consumed when utility is maximized. When there are more than two goods, the law of diminishing marginal rate of substitution implies that at least one of the goods must be a net substitute for any given good.

The notions of net substitutes and net complements are symmetric. Good 1 being a net substitute for good 2 implies that good 2 is a net substitute for good 1. The substitution effect of a change in p_1 on good 2 is identical to the substitution effect of a change in p_2 on good 1.

If goods 1 and 2 are normal, net complementarity implies gross complementarity. The income effect of an increase in p_2 leads to a decrease in the demand for good 1 and since the two goods are net complements, the

substitution effect also decreases the quantity of good 1 demanded. Therefore, good 1 is a gross complement of good 2. Similarly, if goods 1 and 2 are inferior, net substitutability implies gross substitutability. In other cases, the relationship between gross and net concepts depends on the strengths of the income and substitution effects.

An advantage of the above net definitions is that they are consistent. Once goods 1 and 2 have been discovered to be substitutes they remain substitutes, no matter in which direction the definition is applied. It can be shown that there can be only a few complementary relations in the net sense. Net substitution is the prevalent relationship among commodities.

ANOTHER SUBSTITUTION EFFECT

The substitution effect we have studied in the section 'Changes in a Good's Own Price' refers to change in demand when prices change but the consumer is given income compensation so that he remains on the original indifference curve. This is the Hicks substitution effect. An alternative to this is the Slutsky substitution effect which is the change in demand due to change in prices but the consumer's purchasing power is kept constant so that the initial bundle remains affordable. We discuss the alternative approach below.

Suppose that good 1 has become costlier: its price has increased from p_1' to p_1'' . The purchasing power of income has gone down. The budget line rotates around the vertical intercept m'/p_2' and becomes steeper. We break this movement of the budget line into two steps: first, pivot the budget line around the original optimal choice, and then shift the pivoted budget line to the new optimal choice. Clearly, the first step is a movement that changes the slope of the budget line but keeps its purchasing power constant. On the other hand, the second step keeps the slope constant but changes the purchasing power.

Let us assume that both goods are normal. The consumer's original choice is (x_1', x_2') , which is denoted by A. Clearly the pivoted budget line has the same slope and thus the same relative price as the final budget line. Hence money income m'' associated with this line is different, since the vertical intercept is different. In fact $m'' > m'$. Since the original bundle A lies on this pivoted budget line, this bundle becomes just affordable. The purchasing power of income has remained unaltered in the sense that the original bundle is just affordable at the pivoted budget line.

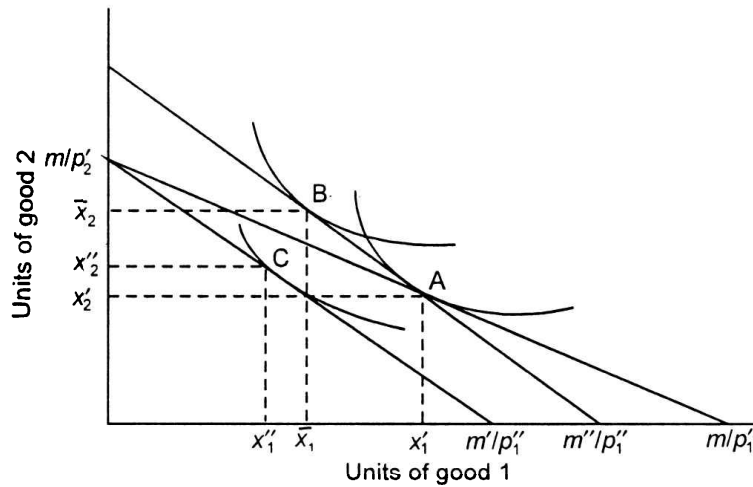


Fig. 4.20 The Slutsky Substitution Effect

Let us now calculate the additional amount of income that the consumer requires so that the original bundle A becomes just affordable when the prices are (p'_1, p'_2) . Since (x'_1, x'_2) is affordable at both (p'_1, p'_2, m') and (p'_1, p'_2, m'') , we have

$$p'_1 x'_1 + p'_2 x'_2 = m', \quad (4.16)$$

$$p'_1 x'_1 + p'_2 x'_2 = m'', \quad (4.17)$$

from which it follows that

$$x'_1 (p'_1 - p_1) = m'' - m' \quad (4.18)$$

We can rewrite the above as

$$x'_1 \Delta p_1 = \Delta m \quad (4.19)$$

Eq. (4.19) shows that change in price and change in income move in the same direction: if price goes up we must increase income to make the original bundle just affordable.

Although the original bundle A is still affordable the consumer will not choose it at the pivoted budget line. The reason behind this is that with increase in price of good 1 from p'_1 to p''_1 , the consumer has a tendency to substitute good 2 for good 1. The optimal choice for the pivoted case is B. This bundle of goods is the optimal bundle when the price-income vector is (p''_1, p'_1, m'') . The movement from A to B is the substitution effect. It shows how the consumer substitutes good 2 for good 1 when the price of good 1 changes but the purchasing power remains intact.

The second stage of price adjustment involves shifting the pivoted budget line parallel to itself to the consumer's new optimal position. This parallel shift changes the consumer's income from m'' to m' , keeping the prices constant at (p_1'', p_2') . In Fig. 4.20, this change moves the consumer from B to C . It is natural to call this movement an income effect, since we are changing income keeping prices constant. Since the goods have been assumed to be normal, the consumption of both goods will be smaller at C than at B .

The total change in demand (A to C) is still equal to the sum of the substitution effect (A to B) and the income effect (B to C), but now it is the Slutsky substitution effect. Clearly, the Slutsky substitution effect can be interpreted in the same way as the Hicks effect. For marginal changes in price, the two substitution effects are virtually identical. Finally, note that using arguments similar to those employed for constructing the Hicks demand curve, we can generate the Slutsky demand curve also.

THEORY OF REVEALED PREFERENCE

We have already mentioned in Chapter 2 that recent work on consumer behavior has followed two approaches; the modern utility approach, which uses information about the consumer's preferences and budget constraint to determine his demand; and the revealed preference approach, which focuses on the consumer's demand to gather information about his preferences. When we say that we wish to determine the consumer's preferences from his observed behavior, it is assumed implicitly that preferences remain unchanged when we observe the behavior.

Idea of Revealed Preference

We begin by assuming that the preferences underlying the consumer's demand are monotone and that these relations are strictly convex also. This ensures that given a budget set, the consumer's demanded bundle will be unique.

Suppose that the consumer with a given income chooses the bundle $A = (x_1, x_2)$ on his budget line. Let $B = (y_1, y_2)$ be another bundle that lies beneath the budget line. The prices of the goods are (p_1, p_2) , and m is the consumer's income. Since (x_1, x_2) is on the budget line it must be true that $p_1x_1 + p_2x_2 = m$. Since B is inside the budget set it is certainly affordable at the price-income vector (p_1, p_2, m) . Therefore, $p_1y_1 + p_2y_2 \leq m$. Hence we have

$$p_1x_1 + p_2x_2 \geq p_1y_1 + p_2y_2. \quad (4.20)$$

If the above inequality holds, then the bundle (x_1, x_2) is said to be directly revealed preferred to B. The choice of (x_1, x_2) when (y_1, y_2) was available reveals that the consumer strictly prefers (x_1, x_2) to (y_1, y_2) . We state this formally as

The Principle of Revealed Preference: Let (x_1, x_2) be the consumer's chosen bundle at prices (p_1, p_2) , and let (y_1, y_2) be another bundle such that $p_1x_1 + p_2x_2 \geq p_1y_1 + p_2y_2$. Now, if the consumer is choosing the most preferred bundle among the bundles which he can afford, then (x_1, x_2) must be strictly preferred to (y_1, y_2) .

Revealed preference is a relation that holds between a chosen bundle with some budget, and a bundle that could be chosen with the same budget. The principle of revealed preference says that if one bundle is chosen when a second one is available, then the former is strictly preferred to the latter.

Now, suppose (y_1, y_2) is a chosen bundle at prices (q_1, q_2) and (y_1, y_2) is revealed preferred to another bundle $C = (z_1, z_2)$. That is, $q_1y_1 + q_2y_2 \geq q_1z_1 + q_2z_2$. Thus, (y_1, y_2) is strictly preferred to (z_1, z_2) . By transitivity of the consumer's preference relation, (x_1, x_2) is strictly preferred to (z_1, z_2) . In this case we say that (x_1, x_2) is indirectly revealed preferred to (z_1, z_2) .

If a bundle is either directly or indirectly revealed preferred to another bundle, then the former is said to be revealed preferred to the latter. From Fig. 4.21 it turns out that since (x_1, x_2) is revealed preferred to any bundle directly or indirectly in the region OKNM, (x_1, x_2) is strictly preferred to all these bundles. Equivalently, the consumer indifference curve that passes through A must lie above the region. Given monotonicity of the preference relation and strict convexity of the same relation, the indifference curve will be strictly convex to the origin. These therefore show how an application of the idea of revealed preference and a few simple assumptions about preferences enable us to trap the indifference curve.

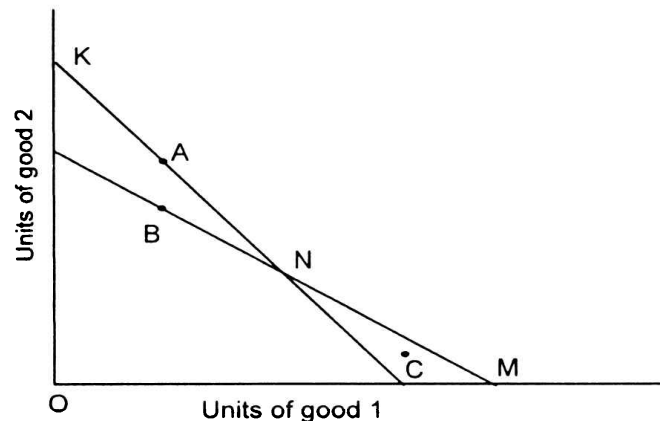


Fig. 4.21 Revealed Preference

Weak Axiom of Revealed Preference

The above notion of revealed preference relies on the assumption that the consumer is choosing the best bundle he can afford. But how can we be sure that the consumer is choosing the best bundle? We have already seen that in order to approximate an indifference curve, the consumer should choose the best bundle he can afford. A simple consistency condition developed independently by Abraham Wald and Paul A. Samuelson shows whether the consumer is following the maximizing model or not. The consistency condition is known as the Weak Axiom of Revealed Preference (WARP) which is stated as follows.

Suppose that A and B are two different commodity bundles. If A is directly revealed preferred to B, then B cannot be directly revealed preferred to A.

To state the axiom algebraically, suppose that $A = (x_1, x_2)$ is directly revealed preferred to $B = (y_1, y_2)$ in the price situation (p_1, p_2) . That is,

$$p_1x_1 + p_2x_2 \geq p_1y_1 + p_2y_2.$$

Let (y_1, y_2) be chosen at prices (q_1, q_2) . Given that A and B are different, the WARP states that A is not available when B is chosen, or in other words, A is too expensive; that is,

$$q_1x_1 + q_2x_2 > q_1y_1 + q_2y_2.$$

Formally, the WARP is stated as follows:

$$p_1x_1 + p_2x_2 \geq p_1y_1 + p_2y_2,$$

which implies that

$$q_1x_1 + q_2x_2 > q_1y_1 + q_2y_2,$$

where $A = (x_1, x_2) \neq B = (y_1, y_2)$.

Fig. 4.22(a) shows the case in which A is chosen when B is available (B lies in A's budget set), and B is chosen when A is not affordable (A lies outside the budget set of B). Hence the WARP is satisfied for the pair in Fig. 4.22(a). In Fig. 4.22(b), A is chosen when B is affordable, and B is chosen when A is affordable. Hence the WARP is violated in this case. Thus, in this case we cannot find indifference curves for which the consumer's behavior is optimal. Clearly, for the case in Fig. 4.22(a) it is possible to find indifference curves for describing optimal behavior of the consumer.

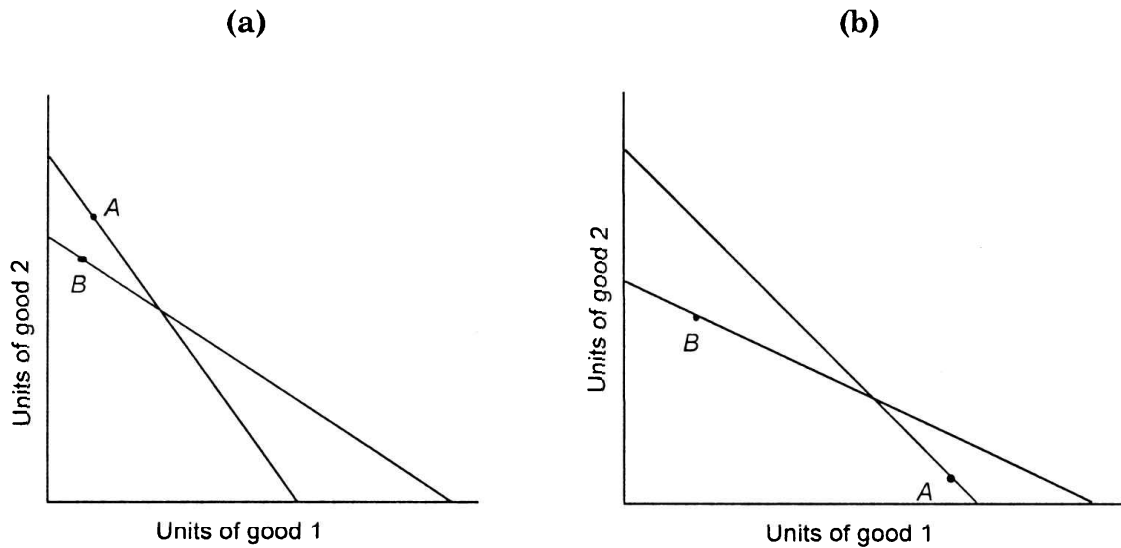


Fig. 4.22 Weak Axiom Satisfied (a); Weak Axiom Violated (b)

There exist interesting relationships between the WARP and the view that the consumer maximizes utility. The following results hold along this line:

- a) Assume that there are only two goods. If the consumer's choice meets the WARP, then they are consistent with the hypothesis that the consumer is a utility maximizer.
- b) If the consumer maximizes utility subject to his budget constraint, and his indifference map is characterized by non-satiation and the law of diminishing marginal rate of substitution, then his choices will fulfil the WARP. The result holds for any number of goods.

We have already indicated how it is possible to recover a consumer's preferences (draw indifference curves) from revealed preference. The argument should convince the reader that if choices meet the WARP, then they should be consistent with the utility maximization hypothesis.

To prove result (b), refer to Fig. 4.23. The consumer maximizes his utility at A. If a different bundle B is affordable when A is chosen, B must lie in the budget set and the indifference curve through B must lie below that through A. Therefore if B is chosen, then A is not affordable. Hence the WARP is satisfied.

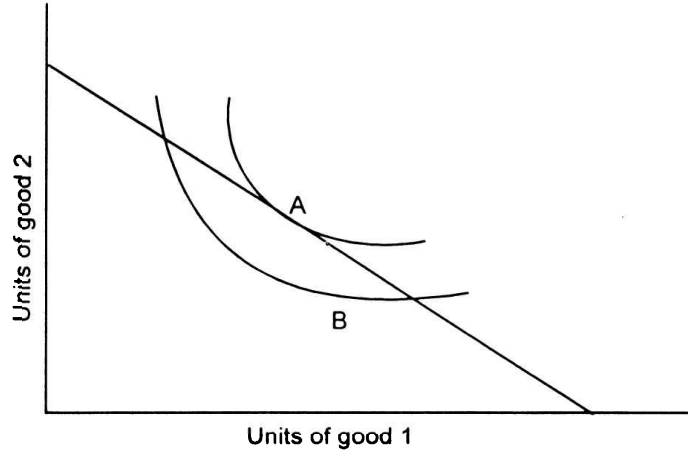


Fig. 4.23 Revealed Preference and Utility Maximization

Weak Axiom of Revealed Preference and Negativity of the Hicks Substitution Effect

The WARP can be employed to show a variety of results on consumer behavior. One interesting situation is to show negativity of the Hicks substitution effect. Suppose that $A = (x_1, x_2)$ is the chosen bundle at the price situation (p_1, p_2) , and let $B = (y_1, y_2)$ be the demanded bundle at some other prices (q_1, q_2) . Suppose also that the consumer's income is such that he is indifferent between A and B . Since the consumer is indifferent between A and B , neither can be revealed preferred to the other.

Therefore, the following inequalities are not true

$$p_1 x_1 + p_2 x_2 > p_1 y_1 + p_2 y_2 \text{ and} \\ q_1 y_1 + q_2 y_2 > q_1 x_1 + q_2 x_2.$$

Hence the following inequalities must hold:

$$p_1 x_1 + p_2 x_2 \leq p_1 y_1 + p_2 y_2 \text{ and} \\ q_1 y_1 + q_2 y_2 \leq q_1 x_1 + q_2 x_2. \quad (4.21)$$

The first inequality of Eq (4.21) means that since the individual is indifferent between A and B , it must be the case that when A is chosen, B must be costlier than A . A similar explanation holds for the second inequality.

We can rearrange the two inequalities in Eq (4.21) as

$$p_1 (x_1 - y_1) + p_2 (x_2 - y_2) \leq 0 \text{ and} \\ q_1 (y_1 - x_1) + q_2 (y_2 - x_2) \leq 0.$$

Adding these, we get

$$(p_1 - q_1)(x_1 - y_1) + (p_2 - q_2)(x_2 - y_2) \leq 0.$$

This is a statement about how demands change when prices change if income is adjusted so that the consumer remains on the same indifference curve. Now suppose that only the price of good 1 changes: assume $p_2 = q_2$. Then we have

$$(p_1 - q_1)(x_1 - y_1) \leq 0. \quad (4.22)$$

This equation means that price and quantity move in the opposite direction when the consumer is on the same indifference curve. This is precisely a statement about the negativity of the Hicks substitution effect. (We ignore the equality sign in Eq. (4.22), because it means that the demand for good 1 is independent of its prices). Thus, we demonstrated the result by a technique that requires neither the existence of a utility function nor the assumption of a diminishing marginal rate of substitution.

Strong Axiom of Revealed Preference

When we turn to the arbitrary number of commodity bundles, we need to impose a somewhat more complicated condition to ensure consistency of choices. This is known as the Strong Axiom of Revealed Preference (SARP). This axiom is formally stated as follows.

Let $A^1, \dots, A^n$ be n commodity bundles of which at least two are distinct. Assume that A^1 is directly revealed preferred to A^2 , A^2 is directly revealed preferred to A^3 , $\dots$, A^{n-1} is directly revealed preferred to A^n . Then A^n cannot be directly revealed preferred to A^1 . This holds for $n = 2, 3, \dots$. Clearly, the WARP is a special case of the SARP when $n = 2$. Further, the SARP ensures transitivity of revealed preferences. We state the relationship between the strong axiom and utility maximization as follows.

1. If a consumer maximizes a utility function subject to a budget constraint, where the utility function satisfies non-satiation and the law of diminishing marginal rate of substitution, then his choices meet the SARP.
2. If a consumer's choices meet the SARP, then they are consistent with the hypothesis that he maximizes utility.

Both of these results hold for any number of goods.

PRICE INDEX NUMBERS

One practical problem for which the study of consumer welfare offers considerable insights is the construction of price indexes by which we can measure

inflation. That is, we wish to investigate first, how the inflation is measured. Second, what impact inflation has on the welfare of a consumer. The Laspeyres price index and the Paasche price index are commonly used to provide partial answers.

The Laspeyres price index is illustrated as follows. The illustration is for one consumer and covers a two-year time series composed of years b and t , where b is the base year and t is some other time.

Suppose in the year b the consumer's income is m^b and the prices of goods 1 and 2 are (p_1^b, p_2^b) . The most preferred position is $B^b = (x_1^b, x_2^b)$. The total expenditure on this bundle of goods is $x_1^b p_1^b + x_2^b p_2^b = m^b$.

Next, consider how much money bundle B^b would cost in year t . If the prices in that year are (p_1^t, p_2^t) and we let $\bar{m}^t$ be the cost of the bundle, then $\bar{m}^t = x_1^b p_1^t + x_2^b p_2^t$. The Laspeyres price index L_p is

$$L_p = \frac{x_1^b p_1^t + x_2^b p_2^t}{x_1^b p_1^b + x_2^b p_2^b}. \quad (4.23)$$

It measures the expenditure that the consumer must incur in year t as a fraction of the expenditure in the base year b in order to obtain in year t the same bundle of goods that maximized satisfaction in year b .

What happens to the consumer's welfare from year b to year t ? Note that $\bar{m}^t$ is not the actual expenditure in year t . Rather, $\bar{m}^t$ is the hypothetical expenditure that would be necessary in the year t to purchase the base year satisfaction maximizing bundle. Let the consumer's actual expenditure in t be $m^t = p_1^t x_1^t + p_2^t x_2^t$, where $B^t = (x_1^t, x_2^t)$ is the consumer's satisfaction maximizing bundle in year t .

The consumer's welfare is greater in year t than in year b if

$$\frac{p_1^t x_1^t + p_2^t x_2^t}{p_1^b x_1^b + p_2^b x_2^b} > \frac{p_1^t x_1^b + p_2^t x_2^b}{p_1^b x_1^b + p_2^b x_2^b} \quad (4.24)$$

This means that the proportional increase in the consumer's income exceeds the estimate of the proportional increase in the cost of living. The above inequality can be rewritten as

$$p_1^t x_1^t + p_2^t x_2^t > p_1^t x_1^b + p_2^t x_2^b. \quad (4.25)$$

The statement means that the bundle chosen in year t is directly revealed preferred to the bundle chosen in year b . This is quite reasonable. If the price

rise is less than the income rise in the movement from b to t , that would tend to make the consumer better off. The revealed preference analysis confirms this. If the inequality in Eq. (4.25) is reversed we cannot be certain whether the consumer's welfare has increased, decreased or remained constant.

A different estimate of the changes in the cost of living over time is provided by the Paasche price index. Whereas the Laspeyres price index shows the relative change in the cost of purchasing a base year bundle of goods between years b and t , the Paasche price index shows the relative change in purchasing a year t bundle of goods between year b and year t .

Suppose the consumer spends his income m^t in year t on the bundle $B^t = (x_1^t, x_2^t)$ at the prevailing prices (p_1^t, p_2^t) . Thus, $p_1^t x_1^t + p_2^t x_2^t = m^t$. The bundle of goods in year t at prices (p_1^b, p_2^b) would cost $\bar{m}^b = p_1^b x_1^t + p_2^b x_2^t$. The Paasche price index P_p then is

$$P_p = \frac{p_1^t x_1^t + p_2^t x_2^t}{p_1^b x_1^t + p_2^b x_2^t}. \quad (4.26)$$

The Paasche price index represents the estimate of the increase in the cost of living from year b to year t . It reflects the increase from $\bar{m}^b$ to m^t in the cost of combination B^t .

Consider the change in the consumer's welfare between years b and t . Let the actual income in year b be m^b where $m^b = p_1^b x_1^b + p_2^b x_2^b$ with (x_1^b, x_2^b) being the utility maximizing bundle in year b . If

$$\frac{p_1^t x_1^t + p_2^t x_2^t}{p_1^b x_1^b + p_2^b x_2^b} < \frac{p_1^t x_1^t + p_2^t x_2^t}{p_1^b x_1^t + p_2^b x_2^t} \quad (4.27)$$

we can be sure that the consumer's welfare decreased between year b and year t , because the increase in the actual income is less than the estimate of the increase in the cost of living. We can rewrite Eq. (4.27) as

$$p_1^b x_1^b + p_2^b x_2^b > p_1^b x_1^t + p_2^b x_2^t \quad (4.28)$$

This statement means that the bundle (x_1^b, x_2^b) is revealed preferred to the bundle (x_1^t, x_2^t) .

QUANTITY INDEX NUMBERS

Quantity index numbers make the comparison between the consumption bundle of a consumer at two different times in almost the same way as price index numbers. In year t the consumer is assumed to choose the bundle (x_1^t, x_2^t)

at the price situation (p_1^t, p_2^t) . In the base year b , the consumer's choice is (x_1^b, x_2^b) when prices are (p_1^b, p_2^b) . Suppose that the bundle (x_1^t, x_2^t) is revealed preferred to (x_1^b, x_2^b) at (p_1^t, p_2^t) . That is,

$$p_1^t x_1^t + p_2^t x_2^t > p_1^t x_1^b + p_2^t x_2^b \quad (4.29)$$

which means that the consumer is better off at t than at b , since he could have consumed the bundle (x_1^b, x_2^b) at t but chooses not to do so.

We can rewrite the inequality, (4.29) as

$$P_q = \frac{p_1^t x_1^t + p_2^t x_2^t}{p_1^t x_1^b + p_2^t x_2^b} > 1. \quad (4.30)$$

The left hand side of inequality Eq (4.30) is the Paasche quantity index. A quantity index number compares the cost of buying distinct commodity bundles in two time periods at constant prices. Therefore, if the Paasche quantity index P_q is larger than one, the consumer is better off at time t than at time b .

We define the Laspeyres quantity index as

$$L_q = \frac{p_1^b x_1^t + p_2^b x_2^t}{p_1^b x_1^b + p_2^b x_2^b}. \quad (4.31)$$

Thus, while in the case of the Paasche quantity index we compare the costs of buying the bundles (x_1^b, x_2^b) and (x_1^t, x_2^t) in the years b and t respectively at the t -year price levels (p_1^t, p_2^t) , the Laspeyres quantity index number compares the costs of buying the same bundles at the b -year price levels (p_1^b, p_2^b) . If the Laspeyres quantity index is less than one, then we can say that (x_1^b, x_2^b) is revealed preferred to (x_1^t, x_2^t) . That is, the consumer is better off in b than in t .

CONSUMER'S SURPLUS

We have already introduced the idea of a consumer's reservation price. This price is regarded as the maximum price that the consumer would be willing to pay per unit of the good to buy the designated quantity. At any quantity, the reservation price for that quantity is the price associated with the quantity demanded on the consumer's demand curve. To understand this more explicitly, let us refer to Fig. 4.24. In panel (a) of the figure the reservation price for one unit is Rs. 10; for two units, Rs. 9; and so on. One way of interpreting this price is that the first unit consumed is worth Rs.10 to the consumer, the second unit is worth Rs. 9, and so on.

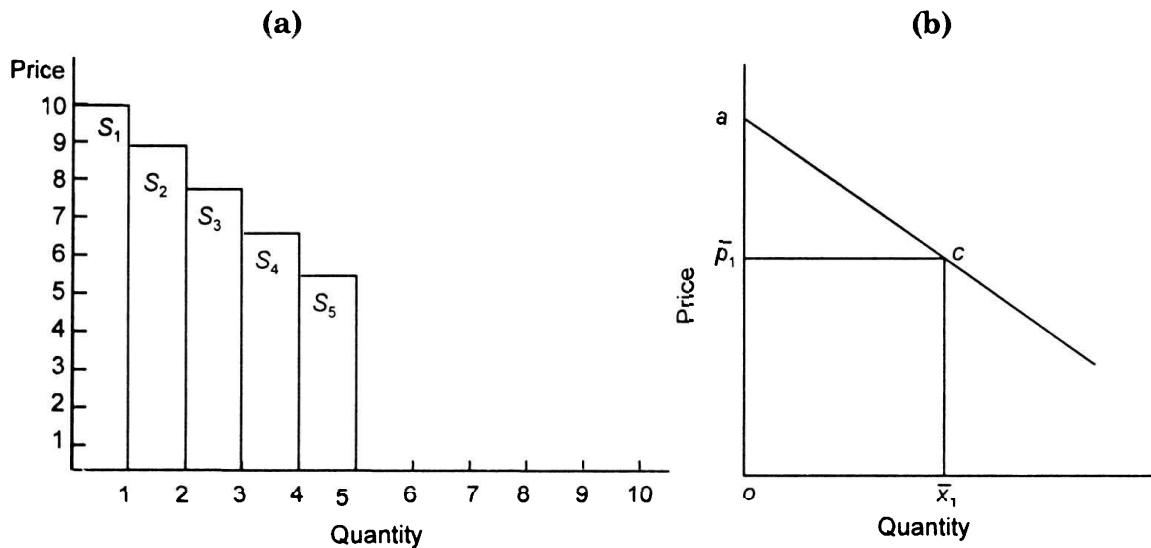


Fig 4.24 Consumer's Surplus

The difference between the total expenditure that must be made to obtain a given quantity of a product and what the consumer would be willing to pay rather than go without it is called the consumer's surplus. In terms of Fig. 4.24(a), at Rs. 10 the consumer would buy one unit. If the price were decreased to Rs. 9 he would buy an additional unit, at Rs. 8 he would buy one more unit, and so on. Let the market price of the product be Rs. 6. The consumer would buy five units of the good, paying Rs. 30 for the entire amount. But he would have been willing to pay Rs. $(10 + 9 + 8 + 7 + 6) = \text{Rs. } 40$. The consumer's surplus is therefore given by Rs. 10.

In general, suppose that the price of the good is p . The reservation price for the first unit is r_1 ; that is, r_1 is the value that the consumer places on the consumption of the first unit of the good. Since he is only paying p for it, he has a surplus of $(r_1 - p)$ on the first unit of consumption. Next, suppose his valuation about the second unit of consumption is r_2 . But since he is paying p for it, this again gives him a surplus of $(r_2 - p)$. Therefore, when we add up the surpluses for all the units he is buying, we get the total consumer's surplus, which is given by

$$\begin{aligned} CS &= r_1 - p + r_2 - p + \dots + r_n - p \\ &= r_1 + r_2 + \dots + r_n - np \end{aligned} \quad (4.32)$$

We can interpret the consumer's surplus in another way. Assume that there exists only one seller for the good in question. This seller is a price discriminator, that is, he can charge different prices on different units of the good sold to a single consumer and can charge different prices from different

consumers. Then, knowing the consumer's demand curve (say, the one shown in Fig. 4.24[a]), this discriminating seller would run down the demand curve and will charge the consumer's reservation prices for different units sold. Thus, he charges Rs. 10 for the first unit, Rs. 9 for the second unit, and so on. In contrast, given a market price, all consumers are charged the same price, and that price holds for all units sold. Consequently, in this case the consumer pays less than the reservation prices on all except the last unit purchased. As a result the consumer has surpluses on all except the marginal unit purchased. As stated earlier, the sum of all such surpluses is the consumer's surplus.

To illustrate the calculation of consumer's surplus, let us refer to Fig. 4.24(a) again. The height of each bar in the figure is the reservation price associated with the corresponding level of demand and the width of each bar is one. The sum of the areas S_i 's is the total area under the demand curve up to the designated quantity. (We have assumed that the market price is Rs. 6 and the consumer buys five units of the good.). Therefore, the consumer's surplus is the area underneath the demand curve up to the designated quantity, minus what the consumer actually pays for that quantity.

Employing the same reasoning, the consumer's surplus in the case of a good which is available in continuous quantities can be easily measured at any given price level. This is shown in panel (b) of Fig. 4.24. At price $\bar{p}_1$, the consumer takes $\bar{x}_1$. He would have been willing to pay the amount represented by $Oac\bar{x}_1$, but he is required by the market to pay only $O\bar{p}_1c\bar{x}_1$. The consumer's surplus at the price $\bar{p}_1$ is measured by the area $\bar{p}_1ac$.

The above calculation of consumer's surplus is based on the compensated demand curve. To understand this, note that the consumer's surplus is the amount of additional income that we would have to pay the consumer to make him as well off with that addition to his present income but with access to a single discriminating seller, as he is with his present income and with access to the market at a given (competitive) price. Thus, the consumer's surplus is derived from a comparison of two situations in which the consumer remains equally well off. Therefore, it will be measured using a demand curve along which the consumer remains on the same indifference curve. Consequently, for measuring the consumer's surplus we should use the compensated demand curve.

The consumer's surplus is underestimated when the ordinary or uncompensated demand curve is used and the good is a normal good. To see this, refer to Fig. 4.19. We want to calculate the consumer's surplus at p'_1 . At this

price level, the quantity demanded under either demand curve is x'_1 . Clearly, the consumer's surplus under the compensated demand curve $g_1(p_1, p'_2, I_1)$ is higher than that under the ordinary demand curve $f_1(p_1, p'_2, m')$. Similarly, the consumer's surplus is overestimated when the ordinary demand curve is used and the good is inferior.

COMPENSATING AND EQUIVALENT VARIATIONS

We often become interested in measuring the change in well-being of an individual between two situations. For this purpose it may be convenient to obtain a monetary representation of the individual's utility function, defined as the amount of money required to attain various utility levels at a set of prices. It should be evident that the expenditure function becomes the appropriate tool here.

As an illustrative example, suppose we ask how much money we would have to give a consumer to compensate him for a change in his consumption pattern. A measure of this type essentially involves a change in utility, but the change is measured in monetary units.

Let us consider the following situation. Suppose the consumer with income m' initially faces some prices (p'_1, p'_2) and consumes some bundle (x'_1, x'_2) . The price of good 1 then increases from p'_1 to p''_1 , and the consumer changes his consumption to (x''_1, x''_2) . We wish to determine the amount of money we would have to give the consumer after the price change to make him as well off as he was before the price change took place. The change in income required to keep the consumer on his original indifference curve is referred to as the compensating variation. The compensating variation gives the amount of extra money that the consumer requires as compensation for remaining on the original indifference curve after the price change took place.

Let us denote the consumer's initial utility level by I_1 . The compensating variation is the income needed to just achieve I_1 after prices have changed from their level (p'_1, p'_2) to their level (p''_1, p'_2) minus the actual income. But in terms of expenditure function the actual income is $e(p'_1, p'_2, I_1)$, and the new amount of income needed is $e(p''_1, p'_2, I_1)$. So the compensating variation (CV) is

$$CV = e(p''_1, p'_2, I_1) - e(p'_1, p'_2, I_1) \quad (4.33)$$

An alternative way of measuring the impact of price change in terms of money is to investigate how much money would have to be taken away from the consumer before the price change to make him as well off as he would

become after the price change takes place. This income change is known as the equivalent variation because of its equivalence with the price change in terms of change in utility. The equivalent variation is the maximum amount of income that the consumer would be willing to pay for avoiding the price change.

Let us denote the consumer's utility level after the price change by I'_1 . Then the equivalent variation (EV) is the difference between the amount of money needed to achieve I'_1 when prices are at the level (p''_1, p'_2) and the amount of money needed to achieve I'_1 at (p'_1, p'_2) , that is,

$$EV = e(p''_1, p'_2, I'_1) - e(p'_1, p'_2, I'_1). \quad (4.34)$$

In general, the amount of money that the consumer requires as compensation after a price change is different from the amount of money that the consumer would be willing to pay to avoid a price change. The source of difference between the two concepts is the utility level at which the expenditure difference due to a price change is measured. The CV is concerned with the original utility level and the EV deals with the final utility level. In terms of geometry, the EV and CV are just two different ways of measuring the distance between the two indifference curves. Since for measuring this distance we in fact see how far apart their tangent lines are, the distance measured will depend on the slopes of the tangent lines, that is, on price levels.

Example 4.1

Suppose that a consumer's utility function is given by $U(x_1, x_2) = x_1x_2$. We wish to answer the following questions:

- a) Assume that $p_1 = 1$, and $p_2 = 1$. Verify that goods 1 and 2 are normal goods.
- b) Let $p_1 = 1$, $p_2 = 1$, and $m = 20$. Determine the utility maximizing choices of goods 1 and 2.
- c) Suppose that $p_1 = 2$, $p_2 = 1$ and $m = 20$. Determine the new utility maximizing choice of goods 1 and 2. Identify the Hicks substitution effect and the income effect of the increase in the price of good 1.
- d) Suppose that $p_1 = 2$, $p_2 = 1$ and $m = 20$. Determine the Slutsky substitution effect and the income effect of the change in the price of good 1.
- e) Show that the weak axiom of revealed preference is satisfied in a comparison between choices under (b) and under (c).

- f) Determine two points on the compensated demand curve for good 1. Also identify two points on the Slutsky demand curve.
- g) What is the relationship between the two goods? Are they independent? If not, is good 2 a gross complement for good 1, or, a gross substitute for good 1?
- h) Compare the choices under (b) and (c) and determine the equivalent and compensating variations.

Answers

- a) At the utility maximizing point, the marginal rate of substitution of good 2 for good 1 equals the price ratio p_1/p_2 . Now, the marginal rate of substitution is equal to the ratio between the marginal utility of good 1 and that of good 2. The marginal utility of good 1 is

$$MU_1 = \frac{U(x_1 + \Delta x_1, x_2) - U(x_1, x_2)}{\Delta x_1} \quad (4.35)$$

$$= \frac{(x_1 + \Delta x_1) x_2 - x_1 x_2}{\Delta x_1} \quad (4.36)$$

$$= x_2.$$

Similarly $MU_2 = x_1$. Thus,

$$\frac{x_2}{x_1} = \frac{p_1}{p_2} = \frac{1}{1} = 1. \quad (4.37)$$

The consumer's budget line is

$$p_1 x_1 + p_2 x_2 = m, \quad (4.38)$$

that is,

$$x_1 + x_2 = m. \quad (4.39)$$

From Eq. (4.37) we have $x_1 = x_2$, which when substituted in Eq. (4.39) gives

$$2x_2 = m$$

and hence $x_2 = m/2$. Similarly, $x_1 = m/2$. These are the utility maximizing choices of goods 1 and 2 as functions of the income m . Clearly, as m increases both x_1 and x_2 increase. Consequently, both goods 1 and 2 are normal.

- b) In (a) we have already established that the utility maximizing choices of goods 1 and 2 are $x_1 = m/2$, $x_2 = m/2$. Given $m = 20$, these choices become $x_1 = 10$, $x_2 = 10$.
- c) In the general case when p_1 and p_2 are the prices, and m is the income, the utility maximizing condition is

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

or,
$$\frac{x_2}{x_1} = \frac{p_1}{p_2}. \quad (4.40)$$

The consumer's budget line is

$$p_1x_1 + p_2x_2 = m. \quad (4.41)$$

From Eqs. (4.40) and (4.41)

$$x_1 = \frac{m}{2p_1}, \quad (4.42)$$

$$x_2 = \frac{m}{2p_2}. \quad (4.43)$$

These are the demand functions for goods 1 and 2.

We know that for

$$p_1 = 1, p_2 = 1, m = 20, x_1 = 10, x_2 = 10.$$

These are the initial choices of the goods. For $p_1 = 2$, $p_2 = 1$, $m = 20$, the choices become $x_1 = 5$, $x_2 = 10$. These are the final choices.

The initial utility level is

$$U(x_1x_2) = x_1x_2 = 10.10 = 100.$$

Let Δm be the income compensation that the consumer requires to remain on the initial indifference curve. Then the consumer's budget line is

$$2x_1 + x_2 = 20 + \Delta m, \quad (4.44)$$

Further, given $p_1 = 2$, $p_2 = 1$, from Eqs (4.42) and (4.43) we have $x_2 = 2x_1$. Since in this case the consumer remains on the initial indifference curve we must have

$$x_1.2x_1 = 100,$$

that is,

$$x_1 = 5\sqrt{2}.$$

Therefore $x_2 = 10\sqrt{2}$, and from Eq. (4.44) it follows that $\Delta m = 20(\sqrt{2} - 1)$. Thus, the choices of the two goods when the consumer gets income compensation to remain on the initial indifference curve are $x_1 = 5\sqrt{2}$, $x_2 = 10\sqrt{2}$.

Given that the consumer's initial and final choices of good 1 are $x_1 = 10$ and $x_1 = 5$, the Hicks substitution effect is $5\sqrt{2} - 10$, and the income effect is $5 - 5\sqrt{2}$.

- d) Recall from (c) that for $p_1 = 1$, $p_2 = 1$, $m = 20$ we have $x_1 = 10$, $x_2 = 10$. Also for $p_1 = 2$, $p_2 = 1$, $m = 20$, we have $x_1 = 5$, $x_2 = 10$.

The increase in the income necessary to keep purchasing power constant when the price of good 1 increases from $p_1 = 1$ to $p_1 = 2$ is

$$\Delta m = x_1 \Delta p_1 = 10.1 = 10$$

(see Eq. [4.19]). Therefore, the income level that keeps purchasing power constant is $m = 20 + 10 = 30$. For $m = 30$ and $p_1 = 2$, the quantity demanded of good 1 is $30/4 = 7.5$ (see the demand function for good 1 in Eq [4.42]). Therefore, given that the consumer's initial choice of good 1 is $x_1 = 10$, the Slutsky substitution effect is $7.5 - 10 = -2.5$; since the consumer's final choice of good 1 is $x_1 = 5$, the income effect is $5 - 7.5 = -2.5$.

- e) At $p_1 = 1$, $p_2 = 1$, the bundle $(x_1 = 5, x_2 = 10)$ costs $1.5 + 1.10 = 15$. This is less than the cost of the bundle $(x_1 = 10, x_2 = 10)$ which was chosen at $(p_1 = 1, p_2 = 1)$. The cost of this latter bundle is 20. By the weak axiom of revealed preference, at $(p_1 = 2, p_2 = 1)$, when $(x_1 = 5, x_2 = 10)$ was chosen, $(x_1 = 10, x_2 = 10)$ is not available because it is too expensive. At these prices, the cost of $(x_1 = 5, x_2 = 10)$ is $2.5 + 1.10 = 20$, while the cost of $(x_1 = 10, x_2 = 10)$ is $2.10 + 1.10 = 30$. Thus, the weak axiom holds.
- f) It should be clear from our calculations in (c) that $(p_1 = 1, x_1 = 10)$ and $(p_1 = 2, x_1 = 5\sqrt{2})$ are two points on the compensated demand curve for good 1. (In this case for each increase in the price of good 1, we need to calculate the income compensation required so that the consumer's utility level remains fixed at 100.) Similarly, $(p_1 = 1, x_1 = 10)$ and $(p_1 = 2, x_1 = 7.5)$ are two points on the Slutsky demand curve.
- g) When p_1 increases from 1 to 2, the quantity demanded of good 2 did not change ($x_2 = 10$). Hence goods 1 and 2 are independent.

- h) Our calculations in (c) show that when p_1 increases from 1 to 2, the amount of income compensation that the consumer requires for being on the initial indifference curve is given by $20(\sqrt{2} - 1)$. Therefore, the compensating variation is $20(\sqrt{2} - 1)$.

In order to calculate the equivalent variation, we ask how much money would be necessary at prices $p_1 = 1$, $p_2 = 1$ to make the consumer as well-off as he would be consuming the final bundle $x_1 = 5$, $x_2 = 10$. Let m be the consumer's money income. Then putting $p_1 = 1$, $p_2 = 1$ in the demand functions given by Eq. (4.42) and (4.43), we have $x_1 x_2 = m^2/4$. But by assumption $x_1 x_2 = 5 \cdot 10 = 50$. Therefore, $m = 10\sqrt{2}$.

Thus, if the consumer had an income of $10\sqrt{2}$ at the original prices $p_1 = 1$, $p_2 = 1$, he would be just as well-off as he would be facing the new prices $p_1 = 2$, $p_2 = 1$ and having an income of 20. Therefore, the equivalent variation in income is $20 - 10\sqrt{2}$.

Applied Example 4.1 Consumer Price Index

Perhaps the most important conclusion about price indexes is that they are, as their name implies, measures of changes in the average prices of particular consumer goods and services.

As stated earlier, the most common procedure to develop such indexes is to choose a market basket of commodities in a base year, and then examine how the cost of this basket changes. For example suppose that the market basket for commodities in the base year 0 is denoted by x_i^0 and that prices p_i^0 prevail in the base year. Then the cost of the basket in year 0 is $\sum_{i=1}^n p_i^0 x_i^0$.

If in year 1 prices change to p_i^1 , the cost of the original basket is $\sum_{i=1}^n p_i^1 x_i^0$ and the consumer price index is

$$CPI = \frac{\sum_{i=1}^n p_i^1 x_i^0}{\sum_{i=1}^n p_i^0 x_i^0},$$

the Laspeyres price index.

CPI computed in this way may overstate the welfare loss from inflation, since the use of a fixed market basket does not permit commodity substitution in response to price changes. Suppose that in year 0 a representative person obtained utility I_0 from consuming x_i^0 . Then the minimum expenditure in year 0 is $e(p_i^0, I_0)$. The minimum expenditure in year 1 is $e(p_i^1, I_0) \leq$

$\sum_{i=1}^n p_i^0 x_i^0$, since the individual can always choose to obtain I_0 more cheaply by consuming some other bundle. Since the construction of the CPI does not permit such substitution, it may overstate the effect of inflation on individual well-being.

Economic Survey published annually by the Government of India, Ministry of Finance (Economic Division), presents CPIs for industrial workers, urban non-manual employees and agricultural laborers for different years by varying base years. (For the lists of items included in the market basket, see the Economic Surveys and references cited therein.)

Economic Survey, 1991–92, Part II—Sectoral Developments, shows that while for industrial workers the CPI rose 80.3 percent between the years 1960 and 1988–89, the same rose 72.4 percent for urban non-manual employees. For agricultural laborers the percentage increase in the CPI between 1960–61 and 1988–89 turned out to be 70.8.

In order to investigate whether the CPI is overestimating the effect of price rise on individual welfare, it becomes necessary to estimate the true cost of living index that takes account of substitution in consumption. A comparison between the two will determine the overstatement of inflation.

Appendix 4.1

SLUTSKY EQUATION

To show the separation of the Hicks substitution effect and the income effect for a good, say good 1, mathematically we return to the utility maximizing first order conditions derived in Appendix 3.1. These conditions are

$$\frac{\delta U(x_1, x_2)}{\delta x_1} - \lambda p_1 = 0,$$

$$\frac{\delta U(x_1, x_2)}{\delta x_2} - \lambda p_2 = 0,$$

$$p_1 x_1 + p_2 x_2 - m = 0.$$

We rewrite these conditions as

$$U_1 - \lambda p_1 = 0,$$

$$U_2 - \lambda p_2 = 0.$$

$$p_1 x_1 + p_2 x_2 - m = 0.$$

where $U_i = \delta U(x_1, x_2) / \delta x_i$, $i = 1, 2$. The total differentials of these equations are respectively

$$U_{11} dx_1 + U_{12} dx_2 - \lambda dp_1 - p_1 d\lambda = 0$$

$$U_{21} dx_1 + U_{22} dx_2 - \lambda dp_2 - p_2 d\lambda = 0$$

$$dm - x_1 dp_1 - p_1 dx_1 - x_2 dp_2 - p_2 dx_2 = 0$$

where $U_{ij} = \delta^2 U / \delta x_i \delta x_j$, $i, j = 1, 2$, and so on. We can rearrange the above equations as

$$U_{11} dx_1 + U_{12} dx_2 - p_1 d\lambda = \lambda dp_1, \quad (\text{A.1})$$

$$U_{21} dx_1 + U_{22} dx_2 - p_2 d\lambda = \lambda dp_2, \quad (\text{A.2})$$

$$-p_1 dx_1 - p_2 dx_2 = -dm + x_1 dp_1 + x_2 dp_2. \quad (\text{A.3})$$

The terms on the right hand side of the above rearranged equations can be regarded as constants. Then we have three equations in three unknowns dx_1 , dx_2 and $d\lambda$.

To find the solution of dx_1 we will first eliminate $d\lambda$ from equations (A.1) and (A.2). For this, multiply Eq (A.1) by p_2 and Eq (A.2) by p_1 to get

$$U_{11} p_2 dx_1 + U_{12} p_2 dx_2 - p_1 p_2 d\lambda = \lambda p_2 dp_1, \quad (\text{A.4})$$

$$U_{21} p_1 dx_1 + U_{22} p_1 dx_2 - p_1 p_2 d\lambda = \lambda p_1 dp_2. \quad (\text{A.5})$$

Subtracting Eq. (A.5) from Eq (A.4) we have

$$(U_{11} p_2 - U_{21} p_1) dx_1 + (U_{12} p_2 - U_{22} p_1) dx_2 = \lambda (p_2 dp_1 - p_1 dp_2). \quad (\text{A.6})$$

We will now eliminate dx_2 from Eqs (A.3) and (A.6). For this, multiply Eq. (A.3) by $(U_{12} p_2 - U_{22} p_1)$ and Eq (A.6) by p_2 to get

$$\begin{aligned} -p_1 (U_{12} p_2 - U_{22} p_1) dx_1 - p_2 (U_{12} p_2 - U_{22} p_1) dx_2 = \\ (-dm + x_1 dp_1 + x_2 dp_2) (U_{12} p_2 - U_{22} p_1), \end{aligned} \quad (\text{A.7})$$

and

$$p_2 (U_{11} p_2 - U_{21} p_1) dx_1 + p_2 (U_{12} p_2 - U_{22} p_1) dx_2 = \lambda p_2 (p_2 dp_1 - p_1 dp_2). \quad (\text{A.8})$$

Adding Eq (A.7) with Eq (A.8) we have

$$(U_{11}p_2^2 - U_{12}p_1p_2 - U_{21}p_1p_2 + U_{22}p_1^2)dx_1 = \lambda p_2(p_2dp_1 - p_1dp_2) + (U_{12}p_2 - U_{22}p_1)(-dm + x_1dp_1 + x_2dp_2).$$

Therefore,

$$dx_1 = \frac{-\lambda p_2^2 dp_1 + \lambda p_1 p_2 dp_2 + (U_{22}p_1 - U_{12}p_2)(-dm + x_1 dp_1 + x_2 dp_2)}{D} \quad (\text{A.9})$$

where $D = -U_{11}p_2^2 + U_{12}p_1p_2 + U_{21}p_1p_2 - U_{22}p_1^2$.

Now, from Eq (A.9), assuming p_2 and m remain constant, and dividing through by dp_1 , we obtain the total response in the quantity taken of good 1 to a change in its price:

$$\frac{\delta x_1}{\delta p_1} = \frac{-\lambda p_2^2}{D} + x_1 \frac{(U_{22}p_1 - U_{12}p_2)}{D}. \quad (\text{A.10})$$

If in Eq (A.9) p_1 and p_2 are held constant, then

$$\frac{\delta x_1}{\delta m} \Big|_{dp_1=dp_2=0} = - \frac{(U_{22}p_1 - U_{12}p_2)}{D}. \quad (\text{A.11})$$

It is now assumed that change in the price of good 1, the price of good 2 remaining fixed, is compensated by an income change such that the consumer remains on the same indifference curve. That is, when p_1 changes $dU = 0$. Thus, $U_1 dx_1 + U_2 dx_2 = 0$, which gives $U_1/U_2 = -dx_2/dx_1$. Now, for the consumer to maximize utility we require

$$\frac{U_1}{U_2} = \frac{p_1}{p_2}.$$

Therefore, $p_1 dx_1 + p_2 dx_2 = 0$. This expression is the left hand side of Eq. (A.3). Hence, the equation becomes $-dm + x_1 dp_1 + x_2 dp_2 = 0$. Then, from Eq. (A.9)

$$\frac{\delta x_1}{\delta p_1} \Big|_{U=\text{const.}} = \frac{-\lambda p_2^2}{D}. \quad (\text{A.12})$$

Substituting Eqs. (A.11) and (A.12) into Eq. (A.10), we get

$$\overline{\frac{\delta x_1}{\delta p_1}} = \frac{\delta x_1}{\delta p_1} \Big|_{U=\text{const.}} - x_1 \frac{\delta x_1}{\delta m} \Big|_{dp_1=dp_2=0}.$$

That is,

The total effect of change in p_1 = Substitution effect + Income effect.

The procedure we have employed above for solving dx_1 becomes gradually tedious as the number of commodities increases. In such a case we need to use Cramer's rule, which relies on the theory of determinants. For a discussion on Cramer's rule, see Henderson and Quandt (1971). In the two-commodity case, our procedure appears to be much simpler than Cramer's rule.

Alternative Derivation

The Slutsky equation can also be derived using the expenditure function. We write $f_1(p_1, p_2, m)$ for the ordinary demand function and $g_1(p_1, p_2, I)$ for the compensated demand function. Let $e(p_1, p_2, I)$ be the expenditure function. Then by definition

$$g_1(p_1, p_2, I) = f_1[p_1, p_2, e(p_1, p_2, I)].$$

This identity means that the compensated demand function—the solution to the expenditure minimization problem—is equal to the ordinary demand function at an appropriate level of income, namely, the minimum income required to achieve the desired level of utility at given prices. Hence

$$\frac{\delta g_1}{\delta p_1} = \frac{\delta f_1}{\delta p_1} + \frac{\delta f_1}{\delta e} \cdot \frac{\delta e}{\delta p_1}.$$

Rearranging terms

$$\frac{\delta f_1}{\delta p_1} = \frac{\delta g_1}{\delta p_1} - \frac{\delta f_1}{\delta e} \cdot \frac{\delta e}{\delta p_1}.$$

The first term on the right side of the above expression is the slope of the compensated demand curve, that is, the substitution effect. Now,

$$\frac{\delta f_1}{\delta e} \cdot \frac{\delta e}{\delta p_1} = \frac{\delta f_1}{\delta m} \cdot \frac{\delta e}{\delta p_1}$$

since changes in income or expenditure amount to the same thing in the function f_1 . But by Shephard's Lemma (proved below), $\delta e / \delta p_1 = x_1$. (Intuitively, an increase in p_1 by Re. 1 raises necessary expenditure by Rs. x_1 since Re. 1 extra must be paid for each unit of x_1 purchased).

Therefore

$$\frac{\delta f_1}{\delta p_1} = \frac{\delta g_1}{\delta p_1} - x_1 \frac{\delta f_1}{\delta m}.$$

That is, the

Total effect of a change in p_1 = substitution effect + income effect.

Shephard's Lemma

Finally, we show that the partial derivative of the expenditure function with respect to the price of a good gives the compensated demand function for the good. This is known as Shephard's Lemma.

The expenditure function is

$$e(p_1, p_2, I) = \min_{x_1, x_2} (p_1 x_1 + p_2 x_2)$$

such that $U(x_1, x_2) = I$.

We need to show that $\frac{\delta e}{\delta p_i} = g_i(p_1, p_2, I)$, $i = 1, 2$. The proof relies on the envelope theorem.

To state the envelope theorem formally, consider the following problem

$$\min_{x_1, x_2} F(x_1, x_2, a)$$

such that $h(x_1, x_2, a) = 0$, where we have explicit dependence of the function F and h on some parameter a .

Let $\mathcal{L} = F(x_1, x_2, a) - \lambda h(x_1, x_2, a)$

be the associated Lagrangian, where λ is the Lagrange multiplier. Denote the optimal values of x_1 and x_2 by $\bar{x}_1$ and $\bar{x}_2$ respectively, and the constrained minimum of $F(x_1, x_2, a) = y$, say, by $\bar{y}$. Then the envelope theorem states that

$$\frac{d\bar{y}}{da} = \frac{\delta \mathcal{L}}{\delta a} = \frac{\delta F(\bar{x}_1, \bar{x}_2, a)}{\delta a} - \lambda \frac{\delta h(\bar{x}_1, \bar{x}_2, a)}{\delta a}$$

Thus, the envelope theorem concerns calculating how the optimal value of a function changes when a parameter of that function changes.

If we write $\mathcal{L}_1$ for the Lagrangian associated with the expenditure function, that is, if $\mathcal{L}_1 = p_1x_1 + p_2x_2 - \lambda_1(U(x_1, x_2) - I)$, where λ_1 is the Lagrange multiplier, then by the envelope theorem

$$\frac{\delta e(p_1, p_2, I)}{\delta p_i} = \frac{\delta \mathcal{L}_1}{\delta p_i} = x_i, \quad i = 1, 2.$$

Since x_i generated in this way depends on the utility level I , it should be regarded as the compensated demand function.

EXERCISES

Multiple Choice

1. In a two-good model, if the price of good 1 falls and the substitution effect on the quantity of good 1 consumed is absolutely greater than the income effect on the quantity of good 1 consumed, then we can be sure that
 - a) Good 1 is normal.
 - b) Good 1 is inferior.
 - c) Good 1 is not Giffen.
 - d) Good 2 is normal.
2. If in a two-good model, the income consumption curve approaches the axis that represents the quantities of good 1 as income increases, then we can be sure that
 - a) Good 1 is normal.
 - b) Good 1 is inferior.
 - c) Good 2 is normal.
 - d) Good 2 is inferior but not Giffen.
3. If in a model with only two goods, where the quantity of good 1 is plotted on the horizontal axis and the quantity of good 2 on the vertical axis, and if the demand curve for good 1 is a downward sloping straight line, we know that the price offer curve generated by changing the price of good 1 must
 - a) Fall continuously.
 - b) Rise continuously.
 - c) Be a horizontal straight line.
 - d) Fall at high prices and rise at low prices.

4. If the consumer price index rises by 10 percent and an individual's income rises by 10 percent, we can say that
 - a) He is equally well-off.
 - b) He is now better off.
 - c) He is now worse off.
 - d) We cannot infer any of the above.
5. If the demand curve for apples shifts to the left, this could be due to
 - a) An increase in the price of apples.
 - b) An increase in consumer income.
 - c) An increase in the prices of other fruits.
 - d) A shift in taste and preferences to other fruits.
6. If the price of tea goes up, the probable result will be
 - a) A decrease in the demand for tea.
 - b) An increase in the price of substitutes for tea.
 - c) A decrease in the price of substitutes for tea.
 - d) None of the above.
7. If the prices of automobiles increase and all other factors remain constant, it will be reasonable to expect
 - a) An increase in the demand for automobiles.
 - b) A reduction in the demand for automobiles.
 - c) An increase in the quantity of automobiles demanded.
 - d) A decrease in the quantity of automobiles demanded.
8. In deriving the Slutsky substitution effect, the purchasing power of income is kept constant when the price of a commodity falls by
 - a) Keeping the consumer on the same indifference curve.
 - b) Relegating the consumer to a lower indifference curve.
 - c) Enabling the consumer to purchase the same bundle of goods as before the price change.
 - d) Enabling the consumer to buy more of both commodities than before the price change.
9. Which of the following is correct about the theory of revealed preference?
 - a) It is possible to recover a consumer's preferences from his choices.
 - b) The consumer's demand curve can be derived.

- c) The consumer's indifference curve can be derived.
 - d) All of the above.
10. If we use an ordinary demand curve in place of a compensated demand curve for a normal good, then the consumer's surplus is
- a) Overestimated.
 - b) Underestimated.
 - c) Exactly estimated.
 - d) Any of the above is possible.
11. Consider the linear demand curve $x = 10 - 2p$. If the price is 2, the consumer's surplus is
- a) 7.
 - b) 8.
 - c) 9.
 - d) 6.

Quantitative Problems and Questions

1. Prove or disprove the following.
- a) If a consumer's demand curve for a good is upward sloping, then an increase in his income would shift it to the left.
 - b) In a two-good world, if the price offer curve for good 1 approaches the axis that represents its quantities as the price of good 1 is reduced, then since the quantity of good 2 falls, we know that good 2 is inferior.
 - c) If over the relevant range the price offer curve for good 1 is parallel to the axis that represents its quantities, then the demand curve for good 1 is a straight line.
 - d) There must be at least one good that a consumer buys for which the demand curve is downward sloping.
 - e) Perfect substitutes can be regarded as an example of homothetic preferences.
 - f) Suppose that in a two-good world the price offer curve showing the consumer's response to a change in the price of good 1 approaches the axis that indicates the quantities of good 2 as the price of good 1 falls. We can then infer that the price offer curve showing the consumer's response to a change in the price of good 2 will approach the axis that indicates its quantities as the price of good 2 falls.
 - g) Let goods 1 and 2 be inferior. If they are net substitutes, then they are gross substitutes as well.
 - h) An inferior good does not generate any consumer's surplus.

2. Show that the Engel curves for the utility function $U(x_1, x_2) = x_1^a x_2^a$, where $a > 0$, are straight lines for both goods 1 and 2.
3. Suppose that goods 1 and 2 are perfect complements and they are purchased in 1:1 ratio. Identify the form of the inverse demand function for good 1.
4. Suppose that a consumer's utility function for three goods is separable, that is,

$$U(x_1, x_2, x_3) = f_1(x_1) + f_2(x_2) + f_3(x_3)$$

where each f_i is increasing and strictly concave, $i = 1, 2, 3$. Show that

- a) None of the goods can be inferior.
 - b) $\delta f_i / \delta p_1$, $i = 2, 3$, can be either positive or negative, but they must have the same sign, where f_i is the demand function for good i and p_1 is the price of good 1.
5. Suppose that there are only two goods 1 and 2, and that the consumer has as the only source of income an initial endowment of each good given by w_1, w_2 . In attempting to maximize utility, he may exchange some of one good in the market for the other at prices p_1 and p_2 respectively.
 - a) What is the consumer's budget constraint?
 - b) Use a graph to show the individual's initial endowment and the budget line. Also show the utility maximizing choices of goods 1 and 2 on the same graph.
 - c) How should the Slutsky equation be revised to represent price changes in this situation?
 - d) Show the consumer's price offer curves and demand curves.
 - e) Can you think of any real world consumption decisions for which the initial endowment model becomes appropriate?
 6. A consumer's income is Rs. 120 and he spends it on two goods 1 and 2. His utility function is $U(x_1, x_2) = x_1 x_2$. The price of good 1 is Rs. 2 and that of good 2 is Rs. 3.
 - a) What are the utility maximizing choices of goods 1 and 2?
 - b) If the price of good 1 increases by 44%, with no change in the price of good 2, by how much would his income have to increase for enabling him to maintain the original utility level?
 - c) Calculate the Laspeyres price index when the price of good 1 rose.
 - d) What are the equivalent and compensating variations?
 - e) Identify the Hicks and the Slutsky substitution effects.

7. Employ the weak axiom of revealed preference to generate a consumer's demand curve for a good.
8. Suppose that with prices (p'_1, p'_2) and income m' , a consumer's choice is the bundle (x'_1, x'_2) . Now assume that good 1 is taxed so that p'_1 rises to p''_1 , and that p'_2 remains unchanged. Since the government does not want the consumer to suffer, m' is increased by providing a subsidy such that the consumer can still buy the bundle (x'_1, x'_2) . Is the consumer better or worse off? Employ the weak axiom of revealed preference to justify your argument.
9. Check whether the weak axiom of revealed preference is satisfied in each of the following cases:
 - a) With income $m = 20$, and $(p_1, p_2) = (1, 1)$ the choice is $(5, 15)$; with income $m = 20$ and $(p_1, p_2) = (2, .50)$ the choice is $(8, 8)$.
 - b) With income $m = 5$, and $(p_1, p_2) = (1, 2)$, the choice is $(1, 2)$; with income $m = 5$ and $(p_1, p_2) = (2, 1)$, the choice is $(2, 1)$.
 - c) With income $m = 20$ and $(p_1, p_2) = (1, 1)$, the choice is $(10, 10)$; with income $m = 20$ and $(p_1, p_2) = (0.5, 3)$, the choice is $20, (10/3)$.
 - d) With income $m = 4$ and $(p_1, p_2) = (2, 1)$, the choice is $(1, 2)$; with income $m = 4$ and $(p_1, p_2) = (1, 2)$, the choice is $(2, 1)$.
10. A consumer's demand function is given by $x = 10 - p$. Determine the gain in the consumer's surplus when p reduces from 3 to 2.
11. Suppose that a consumer's utility function is of the form $U(x_1, x_2) = v(x_1) + x_2$, where x_1 is the quantity of good 1 consumed and x_2 is the amount of money spent on the consumption of other goods, and v is increasing and strictly concave. Suppose the price of good 1 increases from p'_1 to p''_1 . Show that the compensating and the equivalent variations are the same. Also show that $v(x_1)$ is simply the area under the demand curve for good 1.
12. Suppose that an individual is initially maximizing utility by consuming the combination (x'_1, x'_2) . His income is m' and the prices are (p'_1, p'_2) . Now assume that p'_1 reduces to p''_1 .
 - a) Identify the Hicks substitution effect when
 - i) Good 1 is normal.
 - ii) Good 1 is inferior.
 - iii) Good 1 is Giffen.
 - b) Repeat the exercise in (a) for the Slutsky substitution effect.

- c) Suppose p_1' instead of decreasing increases to p_1'' . Identify the Slutsky substitution effect when
- Good 1 is inferior.
 - Good 1 is Giffen.
13. Clearly distinguish between the
- Income consumption curve and the Engel curve.
 - Demand curve and the price offer curve.
 - Gross substitutes and net substitutes.
 - Hicks substitution effect and the Slutsky substitution effect.
 - Gross complements and net complements.
 - Normal goods and inferior goods.
 - Slutsky demand curve and the Hicks demand curve.
14. Show that inferiority of a good is necessary for its being Giffen.
15. Give a graphical exposition of the equivalent and the compensating variations.
16. Draw and explain a consumer's indifference map for two goods, 1 and 2. Explain and illustrate the construction of the consumer's demand curve for good 1 and the Engel curve for good 2.
17. Why must a Giffen good be an inferior good always, while an inferior good need not be Giffen? Justify your answer with a diagram.
18. Suppose $U(x_1, x_2) = x_1 x_2$. Find the expenditure function if $p_2 = 1$. Suppose income $m = 100$. Derive the compensated demand curve for good 1, and compare its slope with that of the ordinary demand curve. Suppose p_1 falls from 1 to 0.25. Evaluate the compensating and equivalent variations. Obtain the Laspeyres price index. (If you differentiate the expenditure function $e(p_1, p_2, I)$ partially with respect to p_i , you get the compensated demand function for good i , $i = 1, 2$. Why?)
19. The true cost of living index is defined by
- $$C = e(p_1^1, p_2^1, I_0) / (p_1^0 x_1^0 + p_2^0 x_2^0)$$
- where 0 and 1 refer to the base and final periods, and I_0 is the utility level in the base period. Show that $C \leq$ the Laspeyres price index. —
20. In a world of two commodities, 1 and 2, the following observations have been made for a consumer in two years:

Base year: $p_1 = 100$, $p_2 = 100$, $x_1 = 100$, $x_2 = 100$;

Final year: $p_1 = 125$, $p_2 = 100$, $x_1 = 120$.

Determine the range of values for the quantity of good 2 consumed in the final year that would enable us to infer each of the following:

- a) The consumer is better off in the base year.
 - b) The consumer's behavior is inconsistent.
 - c) Good 1 is inferior.
 - d) There is insufficient evidence to conclude (a), (b), or (c).
 - e) Good 2 is inferior.
21. A consumer's utility function is of the form $U(x_1, x_2) = x_1^{.02} x_2^{.78}$, where x_1 is the number of kilograms of wheat consumed and x_2 is the amount of money spent on other goods. The consumer's annual income is Rs. 30,240.
- a) The price of wheat is initially Rs. 20 per kilogram. How many kilograms does the consumer buy per year?
 - b) The price of wheat has now increased to Rs. 60 per kilogram. How many kilograms will the consumer consume per year?
 - c) If the government subsidizes wheat, maintaining the price of Rs. 20 per kilogram by providing a subsidy of Rs. 40 per kilogram, and financing the subsidy by a proportional income tax, how many kilograms of wheat will the consumer buy then?
 - d) Compare situations (b) and (c) and show in which situation the consumer is better off.
 - e) Suppose that the consumer can buy wheat at Rs. 40 per kilogram if he joins a club. Compare this with situation (b) to determine the maximum amount of income that the consumer could be willing to pay as club membership.
22. A person has moved from place A to place B. Different prices exist in the two places, and the person's income has also changed as a result of the move. The person is better off in B than in A. Using the original prices as base prices, what relationships do you expect among Laspeyres, Paasche and expenditure price indexes? (The expenditure price index is the ratio between final year and base year expenditures.)
23. We write L_p , P_p and E_p for the Laspeyres, Paasche and expenditure price indexes. Which of the following four conclusions, that
- a) The consumer is better off in the base period.

- b) The consumer is better off in the final period,
 c) The consumer's behavior is inconsistent,
 d) There is insufficient evidence to conclude (a), (b) or (c),

can you draw in each of the following cases?

- a) $P_p < E_p < L_p$. b) $P_p > L_p > E_p$.
 c) $P_p < E_p, L_p < E_p$. d) $P_p > E_p, L_p < E_p$.
 e) $L_p = E_p = P_p$. f) $P_p = E_p < L_p$.

24. The following are the prices and the quantities bought by a consumer of three goods in two different years, b and t .

	Year b		Year t	
	Price	Quantities bought	Price	Quantities bought
Good 1	1	30	2	40
Good 2	2	40	3	50
Good 3	2	50	5	10

- a) Calculate
 i) The Laspeyres price index,
 ii) The Paasche price index,
 iii) The expenditure index.
- b) Which of the following can you conclude?
 i) He is better off in the base year.
 ii) He is better off in the final year.
 iii) He is equally well-off in both periods.
 iv) His behavior is not consistent.
 v) It is not possible to conclude any of the above.
25. What is a Giffen good? Draw an indifference map showing potatoes on one axis and expenditure on all other goods on the other axis, where potatoes are assumed to be a Giffen good. Plot the price offer curve in this case.
26. Draw an indifference map illustrating the Engel curve for an inferior good.
27. Consider the utility function $U(x_1, x_2) = 2x_1x_2 + x_2$. Construct the ordinary and compensated demand functions for good 1.

28. Do you think all goods for a consumer can be luxury goods? Justify your answer.
29. Show that a Slutsky demand curve is never positively sloped.
30. The slope of the indifference curve of a consumer is given by $-x_2/x_1$. What can you say about his
- a) Demand curves,
 - b) Price elasticities of demand,
 - c) Income elasticities, and
 - d) The Engel curves?
31. Under an income supplement program a person receives a cash grant of Rs. G but faces the prices p_1 and p_2 for food and other goods. On the other hand, under a food stamp program, the person facing the same prices, is given a subsidy of Rs. s per unit of food purchased. Show that the cash grant program is preferred to the food stamp program costing the same amount of money.

5

Market Demand

In Chapter 4 we described the construction of the individual demand curve for a good by looking at changes in the utility maximizing choices for the good in response to changes in its price. In this chapter we will be concerned with adding up of individual demand curves for a good to generate the market demand curve. We will examine how the positions of the market demand curve change in response to changing conditions. Various elasticity measures of the market demand curve will also be defined in the chapter.

MARKET DEMAND CURVES

For expositional ease let us assume that there are only two goods (1 and 2). Let $x_{1i} = f_{1i}(p_1, p_2, m_i)$ be the quantity of good 1 demanded by person i , and $x_{2i} = f_{2i}(p_1, p_2, m_i)$ be his demand for good 2. Two features should be emphasized. First, all individuals are assumed to face the same price levels p_1 and p_2 . Each person is price taker. Second, each person's demand depends on his own income.

Now if there are n consumers, the market demand or the aggregate demand for a good, say good 1, is the sum of the amounts of good 1 demanded by the consumers. Formally, if the market demand for good 1 is denoted by X_1 , then

$$\begin{aligned}
 X_1 &= \sum_{i=1}^n f_{1i} (p_1, p_2, m_i) \\
 &= F_1 (p_1, p_2, m_1, \dots, m_n) \text{ (say)}
 \end{aligned}
 \tag{5.1}$$

Note that the market demand depends on the price p_1, p_2 , and on the income distribution, $(m_1, \dots, m_n)$.

To construct the market demand curve for good 1, p_1 is allowed to vary while $p_2, m_1, \dots, m_n$ are held constant. If we assume that each individual's demand curve is downward sloping, the market demand curve will be as well. Fig. 5.1 shows the construction of the market demand curve of good 1. Let us now assume that $n = 2$, that is, there are only two persons. For each price, the point on the market demand curve is found by summing the quantities demanded by each individual. For example, at price p'_1 , consumer 1 is willing to buy x'_{11} units, while consumer 2 wants to buy x'_{12} units. The total quantity demanded in the market at p'_1 is therefore the sum of these two amounts: $X'_1 = x'_{11} + x'_{12}$; consequently, the point (X'_1, p'_1) is a point on the market demand curve. Additional points can be generated similarly and the market demand curve is drawn through them. The market demand curve for a commodity, then, is the horizontal summation of the individual consumer demand curves for that commodity.

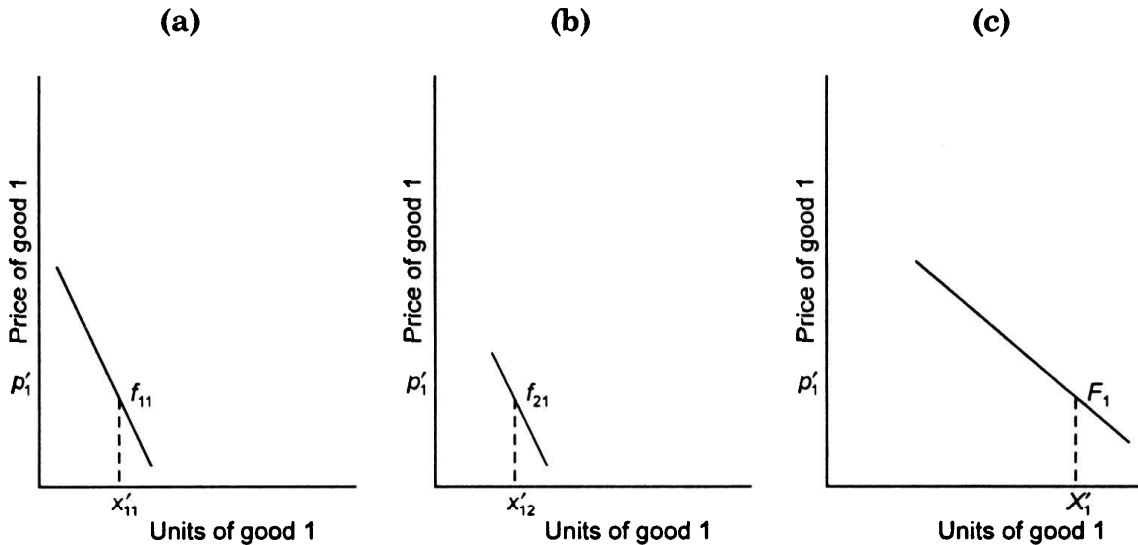


Fig. 5.1 Market Demand Curve

The market demand curve summarizes the *ceteris paribus* assumption. If the factors underlying the construction of the market demand curve do not

change, the position of the curve remains fixed. If we wish to analyze the shifts of a market demand curve, we must first examine how individual demand curves shift and then see how the horizontal sum of these new demand curves compares with the old market demand. In some cases it is possible to conclude unambiguously in which direction the market demand curve will shift. For instance, if both m_1 and m_2 decrease and if both persons regard good 1 as a normal good, then the individual demand curves and hence the market demand curve for good 1 will shift downward.

Some cases may be more complex. For instance, if one person's demand curve shifts upward while the other's shifts downward, the resultant effect will be ambiguous. This situation may arise if m_1 increases and m_2 decreases. The total effect on the location of the market demand curve will depend on the relative shift of the individual demand curves.

Similar comments apply if we wish to examine the impact of a change in p_2 on X_1 . For instance, if p_2 rises and goods 1 and 2 are regarded as gross substitutes by the persons, then the curve for X_1 will shift outward. For instance, a rise in the price of mutton will shift the market demand curve for chicken outward. On the other hand, if two good are regarded as gross complements, an increase in p_2 will shift the curve for X_1 inward. An increase in the price of sugar may shift the tea demand downward.

It should be mentioned that individual demand functions were constructed on the assumption that preferences stay fixed. If preferences change, individual and hence market demand curves would change.

The curve we have drawn in panel (c) of Fig. 5.1 is in fact the inverse market demand curve. This curve gives us the market prices for good 1 that have to be in order for the consumers to demand different units of good 1 as a whole. This inverse demand curve measures the marginal rate of substitution, or the marginal willingness to pay, of every consumer who is purchasing the commodity.

Sometimes it may be convenient to consider the average of individuals' demands. This may then be regarded as the demand of some representative individual who has income that is just the average of all individual incomes. The average demand for good 1 may be written as $A_1(p_1, p_2, m)$, where m is the average of the incomes of individual consumers. The average demand is the demand of some representative individual who faces prices (p_1, p_2) and has income m .

ELASTICITY

Quite often economists become interested in summarizing the way in which changes in one variable affect some other variable. For instance, we may be interested in measuring how the change in the price of a good affects the quantity demanded of the good or some other good, or how change in income affects total expenditure.

We will first have a measure of how responsive the demand for a good is with respect to change in its own price. We use the symbol X to denote the quantity of the good demanded and p to denote its price. One crude measure of degree of responsiveness of demand with respect to price change is the slope of the demand function. The slope of the demand function is $\Delta X/\Delta p$, the change in the quantity demanded divided by the change in price. But there are certain disadvantages of the slope measure. The slope of a demand function depends on the units in which price and quantity are measured. For example, if a decrease of 10 paise in the price of potatoes causes a 100-kilogram increase in the quantity taken, the slope of that portion of the curve is $-100/10$, or -10 . However, if we redraw the demand curve measuring the price in rupees instead of paise, the slope of the same segment of the demand curve becomes $-100/10/100 = -1,000$. The shift from paise to rupees in measuring the price causes a drastic change in the measurement of the downward slope of the demand curve, even though there has been no real change in the demand curve itself. If we draw the demand curve again measuring the price in rupees and the quantity taken in grams, the slope of the same segment of the demand curve becomes $-100 \text{ times } 1000/10/100 = -10^6$. Thus, the slope of the demand curve is a very unreliable indicator of how responsive quantity taken is to changes in price.

The comparative slopes of demand curves for different goods are also not appropriate as measures of the comparative responsiveness of the quantities demanded to price changes. In comparing the demand curve for wheat with the demand curve for cars, suppose we want to know in which case quantity demanded is more responsive to price change. The comparative slopes of the two demand curves tell us nothing. A 1-rupee decrease in the price of wheat may increase the quantity demanded by 10 million kilograms per month. A 1-rupee decrease in the price of cars may increase the quantity demanded by 5 cars per month. This situation does not tell us that the quantity taken of wheat is more responsive to change in its price than is the quantity demanded of cars to changes in car prices. Further, a unit of wheat and a unit of car are completely different concepts, and there is no basis for comparison between the two.

In view of the above discussion it is clear that a measure of degree of responsiveness of the quantity demanded of a product to a change in its price that does not depend on the units in which price and quantity are measured should be developed. The great British economist Alfred Marshall devised a measure known as the price elasticity of demand that fulfils this requirement. The price elasticity of demand, e , is defined as the percentage change in the quantity demanded divided by the percentage change in price. Symbolically,

$$e = \frac{\Delta X/X}{\Delta p/p}. \quad (5.2)$$

The number or coefficient denoting elasticity is obtained by dividing a percentage $(\Delta X/X)$ 100 by another percentage $(\Delta p/p)$ 100. It is therefore a pure number independent of units of measurement of price and quantity. The elasticity will be the same regardless of whether the price is measured in paise or rupees, and of whether quantity is measured in kilograms or grams.

Note that we can rewrite Eq. (5.2) as

$$e = \frac{p\Delta X}{X\Delta p}. \quad (5.3)$$

That is, the elasticity coefficient can be expressed as the ratio of price to quantity multiplied by the slope of the demand function.

Since demand curves are generally negatively sloped (except in the case of a Giffen good), e will usually be negative. For example a value of e of -1 would mean that a 1 percent rise in price leads to a 1 percent decline in quantity whereas a value of e of -2 would record the fact that a 1 percent rise in price causes quantity to decline by 2 percent. When we compare two demand curves with respect to their elasticities, say e_1 and e_2 , demand with elasticity e_1 is more elastic than the one with e_2 if $e_1 < e_2$. (Recall that both e_1 and e_2 are negative.) Therefore, though from an algebraic point of view e_1 is smaller than e_2 , economists prefer to call the demand function with a lower elasticity number as more elastic because it shows a higher degree of responsiveness of the quantity demanded of a good to change in its price. Thus, in the example considered above the demand function with $e = -2$ is more elastic than the one with $e = -1$.

A distinction is often made among values of e that are less than, equal to or greater than -1 . If the elasticity is less than -1 , demand is said to be

elastic. For an elastic demand curve, a given percentage change in price brings about a larger percentage change in the quantity demanded. When the elasticity coefficient is greater than -1 , demand is *inelastic*. In this case, a given percentage change in price leads to a smaller percentage change in the quantity demanded. If the elasticity value is exactly -1 , *unitary* elasticity prevails. For a unit elastic curve, a given percentage change in price results in an identical percentage change in the quantity demanded. If a curve is elastic, price affects quantity a lot; if a curve is inelastic price does not have as much of an effect on the quantity demanded. Elasticity can vary from minus infinity to zero, where zero represents the least elastic demand imaginable, and minus infinity represents the most elastic demand imaginable.

If the coefficient of elasticity is zero, a change in price will not generate any change in the quantity demanded. Demand is described in this case as *perfectly inelastic*, and the demand curve is a vertical straight line indicating that the quantity demanded is not responsive to price. Over some price range, the demand for life saving drugs such as insulin may be perfectly inelastic. Another example can be the demand for dialysis treatment by those suffering from kidney failure.

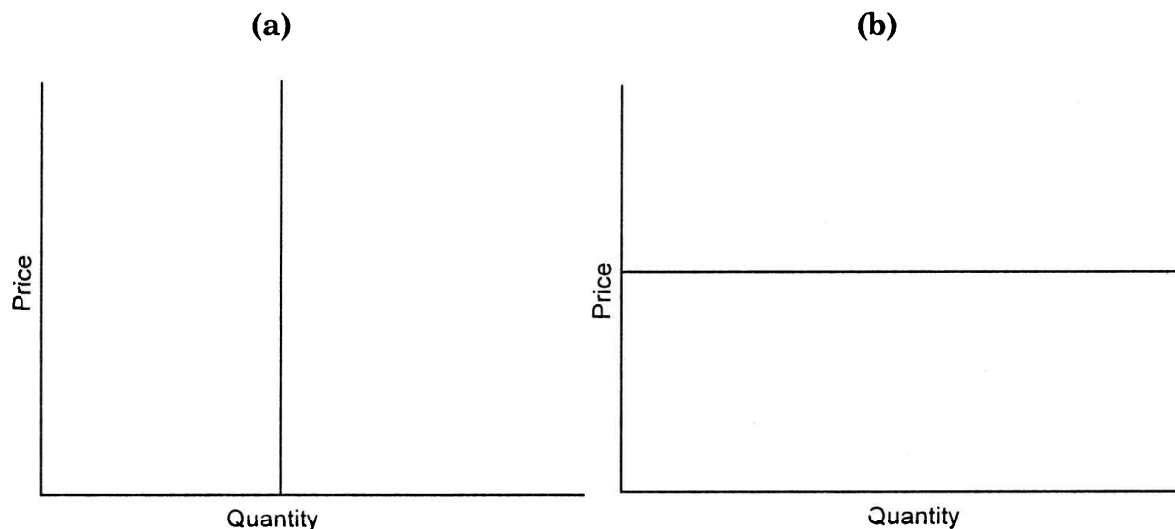


Fig. 5.2 Perfectly Inelastic (a); and Perfectly Elastic (b) Demand Curves

If the elasticity value approaches minus infinity, a very small change in price will lead to an enormous change in the quantity demanded. Demand is referred to as *perfectly elastic* in this case and is graphed as horizontal straight line. An individual potato farmer faces a situation that illustrates perfectly elastic demand. If the market price for potato is Rs. 5 a kilogram, the farmer can sell as much of his production as he wants at that price. But

if he attempts to charge more than Rs. 5, he will sell nothing; consumers will simply buy their potatoes from someone else. This type of situation we represent with a perfectly elastic demand curve. We have much more to discuss about perfectly elastic demand curves in Chapter 10.

ELASTICITY AND REVENUE

One of the most important aspects of price elasticity of demand is the relationship that exists among price changes, elasticity, and the total amount of money spent on the good. The total amount spent on the good can be regarded as the buyer's total outlay or the seller's total revenue or total receipts. The amount is found by multiplying the quantity sold by the price per unit at which it is sold.

If demand decreases significantly when the price increases, revenue will fall. On the other hand, if demand reduction is insignificant when price increases, revenue is expected to increase. Therefore, the direction of change in revenue is closely related to price elasticity. We present this relationship more formally as follows.

THEOREM 5.1: *An increase in the price of a good leads to an increase or a decrease in the total amount spent on purchases of the good according, as the demand for the good is inelastic or elastic. If demand is of unitary elasticity, an increase in the price leads to no change in the amount spent on purchases of the good.*

Proof: Let A denote the expenditure on the good so that $A = pX$. If we let the price change to $p + \Delta p$, where $\Delta p > 0$, and the quantity change to $X + \Delta X$, we have a new expenditure of

$$\begin{aligned} A' &= (p + \Delta p)(X + \Delta X) \\ &= pX + X\Delta p + p\Delta X + \Delta p\Delta X. \end{aligned} \quad (5.4)$$

Subtracting A from A' we get

$$\Delta A = X\Delta p + p\Delta X + \Delta p\Delta X. \quad (5.5)$$

For small values of Δp and ΔX , the last term on the right hand side of ΔA can be neglected. Hence

$$\Delta A \simeq X\Delta p + p\Delta X. \quad (5.6)$$

This means that the change in total expenditure is approximately equal to the quantity times change in price plus price times change in quantity. We can rewrite ΔA in Eq. (5.6) as

$$\begin{aligned}\Delta A &\simeq X\Delta p \left(1 + \frac{p\Delta X}{X\Delta p}\right) \\ &= X\Delta p (1 + e).\end{aligned}\tag{5.7}$$

Now by assumption $\Delta p > 0$. Therefore when demand is inelastic ($-1 < e \leq 0$), an increase (Δp) in p is accompanied by an increase (ΔA) in the amount spent, A . If demand is elastic ($e < -1$), $\Delta p > 0$ implies that $\Delta A < 0$. When demand is of unitary elasticity ($e = 1$), $\Delta p > 0$ implies that $\Delta A = 0$. This completes the proof of the theorem.

The intuitive reasoning behind this result is quite clear. When demand is relatively unresponsive to price (inelastic demand), a price increase does not change the demand very much and total expenditure rises. When demand is highly responsive to price (elastic demand), a price increase reduces demand substantially, and total expenditure falls. Unitary elasticity is the point of demarcation between the two. At this point, the price increase and the quantity decrease are of identical proportional magnitudes, so that overall expenditure does not alter at all.

We have examined above how expenditure or revenue changes when the price of the good under consideration changes. Another interesting issue is to investigate how revenue changes when quantity changes. This is particularly useful when we are considering production decisions by firms.

We saw that for small changes in price and quantity, the change in revenue is given by

$$\Delta A = p\Delta X + X\Delta p.\tag{5.8}$$

If we divide both sides of this expression by ΔX , we get marginal revenue, MR . Thus, MR is the change in total revenue when the quantity changes by 1 unit.

Now

$$\begin{aligned}MR &= \frac{\Delta A}{\Delta X} = p + X \frac{\Delta p}{\Delta X} \\ &= p \left[1 + \frac{X}{\Delta X} \frac{\Delta p}{p}\right] \\ &= p \left[1 + \frac{1}{e}\right].\end{aligned}\tag{5.9}$$

This shows that if demand is of unitary elasticity, then MR is zero—total revenue does not change when output increases. If demand is inelastic, that is, if $e > -1$, then MR is negative. Thus, in this case total revenue will decrease when output increases. On the other hand, if demand is elastic, total revenue increases if output increases.

This result seems quite plausible. If demand is not highly responsive to price, a substantial price reduction is necessary for increasing output, so total revenue decreases. On the other hand, when demand is highly responsive to price we can increase output by a small price reduction and hence the total revenue goes up. We should note here the consistency between this discussion and the earlier discussion about how revenue changes as price increases. This consistency is a consequence of the decreasing relationship between price and quantity.

CONSTANT ELASTICITY DEMAND CURVES

In general we expect price elasticity of demand to vary along a demand curve. However, in many empirical works, for simplifying measurement and estimation it is often assumed that the demand curve has a constant elasticity. That is, the elasticity of demand remains the same at each price.

Suppose that the elasticity of demand is -1 . Then an increase in price level does not change the total expenditure on the product. Let us denote this fixed total expenditure by $\bar{A}$. Then

$$pX = \bar{A} \quad (5.10)$$

which means that

$$X = \bar{A}p^{-1} \quad (5.11)$$

This demand curve is a rectangular hyperbola with a constant area under the curve. Total expenditure is constant along this curve.

The general formula for a constant elasticity demand curve is

$$X = Bp^{-r} \quad (5.12)$$

where $B > 0$ is a constant and $-r$ is the constant elasticity ($r > 0$). The value of the constant B depends on other prices, income and other variables affecting demand. We can rewrite the above equation as

$$\log X = \log B - r \log p. \quad (5.13)$$

This formulation shows that the logarithm of the quantity demanded is a linear function of the logarithm of prices.

DETERMINANTS OF ELASTICITY

Sellers need to know whether the demand for their goods and services is elastic or inelastic before they can make intelligent pricing decisions. But how can sellers know? Quite often they can gather insight into the elasticity of demand by considering some important factors that dictate the elasticity of demand. The major factors influencing elasticity are : (i) the availability of good substitutes for the product under consideration; (ii) the importance of the product in the consumers' budgets, that is, importance of its price relative to buyer's purchasing power; (iii) the number of uses of the product; and (iv) whether the price established is towards the upper end of the demand curve or the lower end.

The **most** important factor in determining the price elasticity of demand is the **number** of good substitutes available. When a large number of good substitutes are available, demand for a good tends to be more elastic because the consumers have alternatives—they can buy a different good if the price of the good becomes quite high. On the other hand, if a product has few good substitutes, demand tends to be inelastic because the consumers do not have many options; they will have to buy the product even at the higher price. The demand for a good also tends to be more elastic in the long run than in the short run, because if the period of adjustment to a price rise is long, substitutes may occur in a higher number of ways.

If consumers spend a higher portion of their income on a particular good, a price rise for that good will force a vigorous search for less expensive substitutes. Demand will tend to be more elastic. On the other hand, if expenditure on a good is relatively small, consumers are likely to ignore the price rise. Demand will tend to be more inelastic.

The wider the range of uses of a product, the more elastic the demand will be. The higher the number of uses, the higher will be the possibility for variation in quantity demanded as price varies. Increases in price reduce, and decreases in price augment the economically desirable uses of a good. These possibilities tend to make the demand for the good more elastic.

If the existing price is towards the upper end of the demand curve, demand is likely to be more elastic than if it were towards the lower end. The reason behind this is that at high price consumers find a wider range of

substitutes for the good, and hence the demand for the good tends to be elastic. In contrast, at low prices the number of substitutes appear to be relatively small, and consequently demand will tend to be inelastic.

CROSS-ELASTICITY OF DEMAND

Another type of elasticity frequently encountered in economics is the cross-elasticity of demand. It is a measure of the extent to which commodities are related to one another. If we consider commodities 1 and 2, the cross-elasticity of commodity 1 with respect to commodity 2 is defined as the percentage change in the quantity demanded of commodity 1 divided by the percentage change in the price of commodity 2. Symbolically,

$$e_{12} = \frac{\Delta X_1 / X_1}{\Delta p_2 / p_2} \quad (5.14)$$

We know that if goods 1 and 2 are gross substitutes $\Delta X_1 / \Delta p_2$ will be positive, as will be e_{12} . When goods 1 and 2 are gross complements, $\Delta X_1 / \Delta p_2$ and e_{12} will be negative. Cross-elasticity of demand is frequently used in attempts to define the boundaries of an industry. High cross-elasticities indicate close relationships or goods in the same industry; low cross-elasticities indicate goods in different industries. For instance, we have high cross-elasticities for various kinds of shoes, but low cross-elasticities with regard to other articles of dress. Thus, we have a basis for separating a shoe industry. For further discussions along this line, see Chakravarty (1996, Chapter 1).

INCOME ELASTICITY OF DEMAND

The final concept of elasticity we introduce in this chapter is the income elasticity of demand. It records the relationship between income changes and quantity changes. The income elasticity of demand for a good is defined as the percentage change in the quantity demanded of the good divided by the percentage change in income. Symbolically,

$$e_m = \frac{\Delta X / X}{\Delta m / m} \quad (5.15)$$

For a normal good, e_m is positive since $\Delta X / \Delta m$ is positive. For an inferior good, e_m is negative.

Among normal goods there is considerable interest about whether e_m is greater than or less than unity. Goods for which e_m is greater than unity are called luxury goods in the sense that purchases of these goods increase more proportionately than income. For instance, if income elasticity of demand for cars is 3, then the implication is that a 10 percent rise in income leads to a 30 percent increase in car purchase. If $e_m < 1$, then the corresponding good is a necessity. An example is food. In this case a 10 percent increase in income will lead to less than 10 percent increase in food purchase.

EMPIRICAL ESTIMATION OF MARKET DEMAND

For many years researchers have been interested in studying the demand for various goods. Some of the earlier studies looked at the expenditure patterns of a small sample of families, attempting to generalize from those empirical regularities that were observed. (Stigler [1954] surveyed some of the early empirical work in demand analysis.) One of the most widely quoted results of the early studies is Engel's law, which states that the percentage of income that families spend on food decreases as their incomes increase.

More recent studies examine a wide variety of goods and attempt to estimate both income and price elasticities. Although it is not possible to discuss here the detailed statistical techniques used in such studies, it may be useful to suggest in an intuitive way how economists have proceeded.

The first important problem that arises in any empirical investigation is how to implement the *ceteris paribus* assumption. In examining the relationship between the price of a good and the quantity demanded, it is assumed that income, other prices, and all other factors are held constant. If income and other prices are not assumed to be constant, observed price-quantity combinations will lie on many different demand curves.

Conceptually, suppose it is possible to find several individuals identical in every respect. We can then plot on a graph the prices the individuals face, together with the per capita quantities demanded at these prices. If it is possible to guarantee that the individuals considered are identical in all respects, the points in the figure will certainly reflect the *ceteris paribus* assumption on the price-quantity relationship.

But in practice it is impossible to impose the *ceteris paribus* principle in this way. Identification of a group of persons who are identical in every respect is certainly impossible. For estimation purposes we must rely on the real world for the data. Economists, therefore, have adopted statistical techniques

for imposing the *ceteris paribus* assumption. The most widely used technique is multiple regression analysis. For instance, in our two-commodity model, multiple regression analysis for commodity 1 attempts to estimate a relationship of the form

$$X_1 = a + bp_1 + cp_2 + dm, \quad (5.16)$$

where a, b, c and d are constants. Once this relationship is estimated, all the terms other than p_1 can be held constant, while the partial relationship between X_1 and p_1 is examined. This is precisely what is required by the *ceteris paribus* assumption. In the linear demand equation stated above, the variables p_1, p_2 and m are called explanatory (exogenous) variables and X_1 is referred to as the dependent (endogenous) variable.

It is important to note that the demand function we have considered in Eq. 5.16 is the market demand function. That is, it has been obtained by aggregating/averaging the individual demand functions.

To illustrate the issue of aggregation, assume again that there are three explanatory variables. The linear demand equation for an individual household h , designated by a superscript, is

$$x_1^h = a^h + b^h p_1 + c^h p_2 + d^h m^h \quad (5.17)$$

where $h = 1, \dots, H$. That is, we are assuming that the number of households that are the buyers of good 1 is H . Under what conditions can these H demand equations be aggregated/averaged into a demand equation for the market as a whole, that is, into a demand function of the form of Eq. 5.16?

Taking the average over H households,

$$\frac{1}{H} \sum_{h=1}^H x_1^h = \frac{1}{H} \sum_{h=1}^H a^h + \frac{p_1}{H} \sum_{h=1}^H b^h + \frac{p_2}{H} \sum_{h=1}^H c^h + \frac{1}{H} \sum_{h=1}^H d^h m^h. \quad (5.18)$$

The left hand side is the average quantity demanded of good 1. On the right hand side the individual intercepts are averaged to a market intercept: $a = \sum_{h=1}^H a^h / H$. The prices p_1 and p_2 are common to all, so $bp_1 = p_1 \sum_{h=1}^H b^h / H$ and $cp_2 = p_2 \sum_{h=1}^H c^h / H$, where $b = \sum_{h=1}^H b^h / H$ and $c = \sum_{h=1}^H c^h / H$. The term that creates an aggregation problem, however, is that involving income, since both d^h and m^h will vary across households. If we assume that d^h is the same for all households, that is, $d^h = d$ for all h , then $dm = \sum_{h=1}^H d m^h / H$ where $m = \sum_{h=1}^H m^h / H$. Thus, individual demand equations given by Eq. (5.17) can be averaged into the market demand equation of the form of Eq. (5.16). Here.

the assumption that the response of demand to income is the same for all households is the aggregation condition for the problem. The condition means that the Engel curves are all linear, with identical slopes across households.

In general, aggregation conditions need to be imposed for generating average demand equations from individual household demand equations. Such conditions usually require that if an explanatory variable changes across households (e.g. income), but the coefficients are the same for each household (for instance, $d^h = d$ for all h), then the average relationship is of the same form, where the explanatory variable is the average of micro-variables (such as $m = \sum_{h=1}^H m^h/H$). They also require that if some explanatory variable is the same for all households (e.g. price), then the average relationship is of the same linear form, but the coefficient in the average relationship is the average of coefficients for micro-relationships (such as $b = \sum_{h=1}^H b^h/H$). However, these statements are true if individual household demand functions are linear. This type of simple averaging principle will not hold for non-linear demand equations. For instance, if the fourth term on the right hand side of Eq. (5.17) is $d^h \log m^h$, then assuming that $d^h = d$, we cannot approximate $\sum_{h=1}^H d^h \log m^h/H$ by $d \sum_{h=1}^H \log [(m^1 + \dots + m^h)/H]$. In practice, however, such problems are usually ignored, with the totality of various individual households treated as if they were a single representative household.

It is also important to note that there is usually an aggregation over commodities as well as households. Empirical studies generally refer to a group of individuals consuming not specific commodities but groups of commodities. In fact, most empirical works consider less than ten commodities, which clearly necessitates aggregation of commodities into broad commodity groups.

To give a brief idea about the estimation of a linear demand function we consider the simple linear regression.

$$X_1 = \alpha + \beta p_1 \quad (5.19)$$

where α and β are constants. Thus, here we are assuming that the demand for commodity 1 depends only on its own price in a linear fashion. Now, one purpose of estimation is to predict the amount of commodity 1 that would be demanded for a given price. Let y_1 be the predicted value of X_1 . For predicting the value of X_1 , we are required to determine the values of parameters α and β from the observed data. Determination of the values of α and β is the problem of estimation associated with Eq. (5.19).

Let us suppose that there are n given pairs of observations on (p_1, X_1) , the i th pair being (p_1^i, X_1^i) . The predicted values of X_1 are therefore $y_1^i = \alpha + \beta p_1^i$,

$y_1^2 = \alpha + \beta p_1^2, \dots, y_1^n = \alpha + \beta p_1^n$, $X_1^i - y_1^i$ is thus the error of estimate for the i th pair, $i = 1, 2, \dots, n$. Given the purpose in mind it is reasonable to require that α and β should be such that these errors of estimate are as small as possible. However, minimization of the sum of errors will not be enough, because the errors, which may be positive or negative, may add up to zero. A satisfactory method of determining α and β , which is known as the method of least squares, consists of minimizing the sum of squares of errors of estimates. Thus, the problem is to minimize

$$\begin{aligned} S &= \sum_{i=1}^n (X_1^i - y_1^i)^2 \\ &= \sum_{i=1}^n (X_1^i - \alpha - \beta p_1^i)^2 \end{aligned} \quad (5.20)$$

with respect to α and β .

For S to be minimum, we require that the partial derivatives of S with respect to α and β to be equal to zero, that is, $\delta S / \delta \alpha = 0$ and $\delta S / \delta \beta = 0$. Now, $\delta S / \delta \alpha = 0$ means that

$$\sum_{i=1}^n (X_1^i - \alpha - \beta p_1^i) = 0. \quad (5.21)$$

On the other hand, $\delta S / \delta \beta = 0$ means that

$$\sum_{i=1}^n p_1^i (X_1^i - \alpha - \beta p_1^i) = 0. \quad (5.22)$$

We can rewrite Eqs. (5.21) and (5.22), which are referred to as normal equations, as

$$\sum_{i=1}^n X_1^i = n\alpha + \beta \sum_{i=1}^n p_1^i \quad (5.23)$$

$$\text{and} \quad \sum_{i=1}^n X_1^i p_1^i = \bar{\alpha} \sum_{i=1}^n p_1^i + \beta \sum_{i=1}^n (p_1^i)^2 \quad (5.24)$$

By solving the above equations we get

$$\beta = \frac{n \sum_{i=1}^n p_1^i X_1^i - (\sum_{i=1}^n X_1^i)(\sum_{i=1}^n p_1^i)}{n \sum_{i=1}^n (p_1^i)^2 - (\sum_{i=1}^n p_1^i)^2} \quad (5.25)$$

$$\alpha = \frac{1}{n} \sum_{i=1}^n X_1^i - \beta \frac{1}{n} \sum_{i=1}^n p_1^i \quad (5.26)$$

Evidently the estimate of β which we determine from Eq. (5.25) can be substituted into Eq. (5.26) for getting the estimate of α . These values of α and β can therefore be used to obtain the desired prediction formula.

The method of estimation we have described above clearly generalizes to more than one explanatory variable. We do not wish to discuss them in detail. Instead we will discuss some non-linear market demand functions. One standard non-linear demand function is the semilogarithmic demand function.

$$X_1 = a + b \log p_1 + c \log p_2 + d \log m \quad (5.27)$$

A third functional form is the log-linear or constant elasticity form. It specifies the demand function as

$$X_1 = A p_1^{b_1} p_2^{b_2} m^c. \quad (5.28)$$

Taking logarithms lead to the linear representation

$$\log X_1 = \log A + b_1 \log p_1 + b_2 \log p_2 + c \log m \quad (5.29)$$

This demand curve will be a line if plotted on a double-log graph paper. All the coefficients in Eq. (5.29) are elasticities. That is, b_1 is the (own) price elasticity of demand for commodity 1, b_2 is the cross elasticity of demand for commodity 1 with respect to commodity 2, and c is the income elasticity of demand for commodity 1.

Sometimes the estimation of the complete system of demand functions is important for identifying the interdependence among the goods, specifically the effects of changes in prices of certain goods on the demand for other goods. One such example is the linear expenditure system discussed in Chapter 3. Other examples can also be constructed.

The discussion of the estimation procedures of above non-linear specifications and the different systems of demand functions are beyond the scope of this book. For such issues the reader is referred to Intriligator (1980).

Example 5.1

A consumer's utility function is $U(x_1, x_2) = x_1^5 x_2^5$. We want to determine the following:

- The equation for his demand curve for good 2, and its elasticity when income is 656 and the prices of goods 1 and 2 are two and one.
- The cross elasticity of good 2 with respect to good 1 if income is 656 and prices of goods 1 and 2 are five and six.
- The income elasticity of demand for good 2 when income is 656 and the prices of goods 1 and 2 are eight and nine.

The utility maximizing conditions are

$$\frac{\delta U(x_1, x_2) / \delta x_1}{\delta U(x_1, x_2) / \delta x_2} = \frac{p_1}{p_2} \quad (5.30)$$

and

$$p_1 x_1 + p_2 x_2 = m \quad (5.31)$$

where p_1 and p_2 are the prices of goods 1 and 2, and m is income. That is,

$$\frac{x_2}{x_1} = \frac{p_1}{p_2} \quad (5.32)$$

and $p_1 x_1 + p_2 x_2 = m$, from which we have

$$x_2 = m / 2p_2. \quad (5.33)$$

This is the demand function for good 2. This function has a constant price elasticity of -1 . Since x_2 is independent of p_1 , the cross elasticity is zero.

The income elasticity of demand for good 2 is

$$e_{2m} = \frac{\delta x_2 m}{\delta m x_2} = \frac{1}{2p_2} \frac{m}{m / 2p_2} = 1. \quad (5.34)$$

Applied Example 5.1 Elasticity of Demand

Majumder (1986) estimated price and expenditure elasticities for the linear expenditure system using Indian consumer expenditure data for the period October 1953–March 1954 to October 1973–June 1974 for the rural and urban

sectors respectively. The groups of items for which expenditure patterns have been analyzed are: (i) cereals and cereal substitutes, (ii) milk and milk products, (iii) edible oils, (iv) meat, fish and eggs, (v) sugar, etc., (vi) other food items (that is, spices, salt, beverages, prepared food, etc.), (vii) clothing, (viii) fuel and light, and finally (ix) other non-food items (which include items like medicines, personal care, education, transport, recreation, rents and taxes, etc.).

Price elasticities and expenditure elasticities for three percentile groups (poorest 30 percent, middle 40 percent and richest 30 percent) have been calculated. As one moves from the lower expenditure class, one would normally expect the magnitude of price and expenditure elasticities to decline for all commodities. Regarding price elasticities, for both urban and rural data this expected trend has been observed for milk and milk products, clothing, and other non-food items. For expenditure elasticities in both urban and rural sectors, an increasing trend was observed for all necessary items, but a decreasing trend was found for the luxury items.

Appendix 5.1

ELASTICITY OF DEMAND

In this appendix we define different elasticity formulae using calculus. The demand function for good 1 is

$$X_1 = A_1(p_1, p_2, m). \quad (5.35)$$

Then the elasticity of demand with respect to the price of good 1 is

$$e_1 = \frac{\delta X_1}{\delta p_1} \frac{p_1}{X_1}. \quad (5.36)$$

The cross elasticity of demand for commodity 1 with respect to commodity 2 is

$$e_{12} = \frac{\delta X_1}{\delta p_2} \frac{p_2}{X_1}. \quad (5.37)$$

The income elasticity of demand for commodity 1 is

$$e_m = \frac{\delta X_1}{\delta m} \frac{m}{X_1}. \quad (5.38)$$

We can rewrite the above elasticity formulae as

$$\begin{aligned} e_1 &= \frac{\delta \log X_1}{\delta \log p_1} \\ e_{12} &= \frac{\delta \log X_1}{\delta \log p_2} \\ e_{1m} &= \frac{\delta \log X_1}{\delta \log m} \end{aligned} \quad (5.39)$$

To see this for e_1 , note that

$$\begin{aligned} \frac{\delta \log X_1}{\delta \log p_1} &= \frac{\delta \log X_1}{\delta X_1} \frac{\delta X_1}{\delta p_1} \frac{\delta p_1}{\delta \log p_1} \\ &= \frac{d \log X_1}{d X_1} \frac{\delta X_1}{\delta p_1} \frac{1}{d \log p_1 / d p_1} \\ &= \frac{1}{X_1} \frac{\delta X_1}{\delta p_1} \frac{1}{1/p_1} \\ &= \frac{p_1}{X_1} \frac{\delta X_1}{\delta p_1} = e_1 \end{aligned} \quad (5.40)$$

Similar demonstrations follow for the rewritten expressions for e_{12} and e_{1m} .

EXERCISES

Multiple Choice

1. Which of the following cases indicates a Giffen good?
 - a) Income elasticity of demand and price elasticity of demand are both positive.

- b) Income elasticity of demand and price elasticity of demand are both negative.
 - c) Income elasticity of demand is positive but price elasticity of demand is negative.
 - d) Income elasticity of demand is negative but price elasticity of demand is positive.
2. For a downward sloping market demand curve we can say that
- a) It is perfectly elastic.
 - b) It is perfectly inelastic.
 - c) Its elasticity falls as price rises.
 - d) Its elasticity falls as quantity rises.
3. On a downward sloping demand curve, demand is more elastic
- a) At the upper end.
 - b) At the lower end.
 - c) In the middle.
 - d) Any of the above is possible.
4. In a model with only two goods, 1 and 2, the cross-elasticity of demand for good 1 with respect to good 2 is positive. Then we can say that
- a) The price elasticity of demand for good 1 is elastic.
 - b) The price elasticity of demand for good 1 is inelastic.
 - c) The price elasticity of demand for good 2 is elastic.
 - d) The price elasticity of demand for good 2 is inelastic.
5. If an increase in price causes total revenue to fall, what can be inferred?
- a) Demand is elastic.
 - b) Demand is inelastic.
 - c) Demand is perfectly inelastic.
 - d) Unitary elasticity prevails
6. Which of the following cases is not possible?
- a) Income elasticity of demand and price elasticity of demand are both positive.
 - b) Income elasticity of demand and price elasticity of demand are both negative.

- c) Income elasticity of demand is positive but price elasticity of demand is negative.
 - d) Income elasticity of demand is negative but price elasticity of demand is positive.
7. In general, demand for a product is more elastic
- a) The smaller the number of substitutes and the larger the fraction of income spent on that product.
 - b) The greater the number of substitutes and the larger the fraction of income spent on that product.
 - c) The smaller the number of substitutes and the smaller the fraction of income spent on that product.
 - d) The greater the number of substitutes and the smaller the fraction of income spent on that product.
8. If the demand curve for a good is a vertical straight line, the elasticity would be
- a) Zero.
 - b) Minus one.
 - c) Minus infinity.
 - d) Different between any two points on the curve.
9. If the price of a product is reduced from Rs. 10 to Rs. 8, and the quantity demanded increases from 5,000 to 10,000, then the price elasticity of demand is
- a) -0.4 .
 - b) -1 .
 - c) -4 .
 - d) None of the above.
10. Which of the following elasticity measures a movement along a curve?
- a) The price elasticity of demand.
 - b) The income elasticity of demand.
 - c) The cross elasticity of demand.
 - d) None of the above.
11. A reduction in the price of a commodity whose demand curve is a rectangular hyperbola causes total expenditure on the commodity to

- a) Increase.
- b) Decrease.
- c) Remain unchanged.
- d) Any of the above is possible.

Quantitative Problems and Questions

1. Prove or disprove the following.
 - a) A consumer purchases two goods, 1 and 2, with a fixed income. His cross-elasticity of demand for good 1 with respect to the price of good 2 is positive. Then the demand for good 2 is inelastic.
 - b) Suppose that the price offer curve of good 1 is parallel to the axis that represents its quantities. Make a similar assumption about the price offer curve of good 2. Then the income elasticities of demand for the two goods are equal.
 - c) Every straight-line demand curve that is downward sloping shows unitary elasticity at its midpoint, elastic demand above it, and inelastic demand below it.
 - d) An individual's income elasticity of demand for a good and price elasticity of demand for the same good cannot be simultaneously positive.
 - e) An individual spends one half of his income on good 1 and the rest on other goods. If his income elasticity of demand for good 1 is four, then he consumes at least one inferior good.
 - f) A consumer spends his income on goods 1 and 2. If his demand for good 1 is unit-elastic with respect to the price of good 1, then the cross-elasticity of demand for good 2 with respect to good 1 is zero.
2. Total revenue from the sale of a good is given by $R = 100X - 2X^2$. Calculate the price elasticity of demand when marginal revenue is 40.
3. A consumer's demand curve for a good is given by $p = 100 - \sqrt{x}$. Calculate the price elasticity of demand when the price of the good is 40.
4. Suppose that the price elasticity of demand for water is -3 and that the government wants to reduce the quantity of water demanded by 35%. By how much must it raise the price of water for achieving this objective?
5. Which would tend to be more elastic, the demand for cars or the demand for Maruti cars? Why?

6. If the government increases the price of petrol as a method of increasing total revenue, what assumption is it making about the elasticity of demand for petrol?

7. Consider the Stone-Geary utility function

$$U(x_1, x_2) = (x_1 - C_1)^{a_1} (x_2 - C_2)^{a_2}.$$

Can there be any inferior good? What distinguishes luxuries from necessities?

8. Show that in a two-good world,

$$s_1 e_1 + s_2 e_{21} = -s_1,$$

where s_i is the proportion of income spent on the consumption of good i .

9. The expenditure elasticity for a good is defined as the proportional change in total expenditure on the good in response to a one percent change in income. Show that the expenditure elasticity for a good is the same as the income elasticity of demand for the good. Can you establish a relation between the expenditure elasticity for a good and its price elasticity of demand?
10. $U(x_1, x_2) = x_1^a x_2^b$ is a person's utility function for two goods, 1 and 2. Determine his demands for the two goods as functions of p_1 , p_2 and m , and the own price and income elasticities of demand for each good.
11. Suppose that the utility from two goods 1 and 2 is given by $U(x_1, x_2) = \log x_1 + x_2$.
- Calculate the demand function for good 1. What are the income, own price and cross price elasticities of demand for good 1?
 - What is the relationship between uncompensated and compensated elasticities of demand for good 1? Interpret your result.
12. Suppose that the quantity demanded of a good, say good 1, demanded by person 1 is given by

$$x_{11} = 10 - 2p_1 + 0.01m_1 + 0.4p_2$$

and that the quantity of good 1 demanded by person 2 is

$$x_{12} = 5 - p_1 + 0.02m_2 - 0.2p_2,$$

where $m_1(m_2)$ is the income of person 1(2).

- a) What is the market demand function for good 1?
 - b) Graph the demand functions of the two individuals assuming that $m_1 = 1,000$, $m_2 = 1,000$ and $p_2 = 10$.
 - c) Using these individual demand curves, determine the market demand curve for good 1.
 - d) Suppose that m_1 increases by 100 and m_2 decreases by 100. How would the individual demand curves shift? How would the market demand curve shift? Graph these new curves.
 - e) Suppose that p_2 increases to 15. Graph the new individual and market demand curves.
13. If for a consumer all income elasticities are constant and equal, show that they must be equal to unity.
 14. Identify the form of the demand function for a good if its income elasticity of demand is constant.
 15. Four persons, A, B, C and D, have the same demand function for a good, say good 1:

$$x_1 = \frac{\sqrt{(mp_2)}}{2p_1},$$

where m represents income, $p_1(p_2)$ is the price of good 1(2).

- a) What is the market demand function for good 1? Determine the elasticities of (market) demand for good 1 with respect to the price of good 1 and the price of good 2. Determine the value of market demand if $p_1 = p_2 = 1$, and the incomes of the four persons are $m_A = m_B = 16$, $m_C = 25$ and $m_D = 100$.
 - b) If $m_A = m_B = m_C = m_D = 25$, what would be the total demand for good 1?
 - c) If p_1 is doubled, what will be the new level of demand for good 1? If m_A is reduced to $m_A/2$, how would the market demand for good 1 be affected?
16. Show that for a two-good world

$$s_1 e_{1m} + s_2 e_{2m} = 1,$$

where e_{im} is the income elasticity of demand for good i , and s_i is the proportion of income spent on good i , $i = 1, 2$. (This result is sometimes referred to as the generalized Engel's Law. It shows that for every good (or group of goods) that has an income elasticity of demand less than 1,

there must exist good (goods) that has (have) income elasticity (elasticities) greater than 1. In the two-good case, knowledge of one good's income elasticity and share of income spent on that good enables us to calculate the income elasticity for the other good.)

17. Show that in a two-good world demand elasticities for good 1 with respect to all prices and income sum to zero. Do you think this is true for good 2 also? Does this result generalize to any number of goods?
18. Show that the Slutsky equation for good 1 can be written as

$$e_1 = e_1^c - s_1 e_{1m},$$

where e_1 is the price elasticity of demand for good 1, e_1^c is the compensated price elasticity of demand for this good, s_1 is the fraction of income spent on it and e_{1m} is its income elasticity of demand. Hence or otherwise show that if good 1 is normal then the ordinary demand curve is more elastic than the compensated demand curve.

19. A person spends his total income on three goods, 1, 2 and 3. He buys 550 units of good 1 at Re. 1 per unit, 425 units of good 2 at Rs. 2 per unit and 200 units of good 3 at Rs. 3 per unit. The price of good 1 increases by 10%, but there is no change in other prices. He now buys 440 units of good 2 and 190 units of good 3. What is the price elasticity of demand for good 1?
20. Total revenue from the sale of a good is given by $R = 60X - X^2$. Calculate the value of marginal revenue when the elasticity of demand is -2 .
21. The relationship between a consumer's income and the quantity x of a good he consumes is given by $m = 1,000x^2$. Calculate the elasticity of demand when his income is 64,000.
22. A consumer's utility function is given by $U(x_1, x_2) = 50x_1 - 0.5x_1^2 + 100x_2 - x_2^2$. The price of good 1 is 4 and his income 672.
 - a) Derive his demand function for good 2.
 - b) If the price of good 2 is 14, how much of good 1 does he buy?
 - c) Calculate his income elasticity of demand for good 1 at the utility maximizing point.
23. Suppose that 1,200 persons buy 1,600 units of a good, and their elasticity of demand is -1.5 . Also suppose that 800 persons buy 800 units of the good and these person's elasticity of demand is -3 . The price of the good is Rs.1.25. Calculate the elasticity of market demand.
24. Explain why demand for any commodity tends to be more elastic in the long run than in the short run.

25. Relating demand to income, show that no demand curve can be perfectly inelastic for every price.
26. Demonstrate rigorously that the price elasticity of demand for a good is constant if and only if it is of the form $X = \beta p^{-r}$, where $\beta > 0$ is a constant, and $-r < 0$ is the constant elasticity of demand.
27. Suppose that the demand curve is given by $p = 30 - 2X$, and there is an increase in demand, so that the new demand curve becomes $p = 50 - 2X$. Does the elasticity at each price change after the shift? Does the elasticity at each quantity change?
28. As income rises, what do you expect to happen to the demand for: potatoes, food, rice and cigarettes?
29. On the same diagram, represent the demand function and the constant expenditure function for a good graphically. What happens to the value of sales if there is a fall (rise) in price given that the demand function cuts the constant expenditure function at a very high (low) price level?
30. The price elasticity of a good is -1 , its income elasticity is 3, and the income share spent on it is $1/4$. What is its compensated price elasticity?

6

Intertemporal Issues

So far we have ignored the intertemporal aspects of decisions by consumers, that is, decisions that take into consideration the consequences for the current period as well those for future periods. The theory that has been presented up to this point can be regarded as being appropriate to a world in which each period is a repeat performance of the previous period, so that decisions made in the current period are determined by current period incomes. In this chapter we first examine consumer behavior by considering the choices involved in saving and consumption over time. Then we discuss the functioning of asset markets under perfect information about the future flow of services provided by the assets.

RATE OF INTEREST

The rate of interest is the rate of return per year that is earned by lending money for use today to borrowers (firms or individuals) for repayment at some future time. Thus, the rate of interest is the rate of return earned by the lender for waiting to get his money back from the borrowers. Alternatively, the interest rate can be regarded as a measure of the terms at which the funds today may be turned into funds in a future period. In other words,

it measures the rate of return that must be paid to obtain money today rather than at some future date.

When money is lent, creditors obtain in exchange for it an obligation to pay (e.g. a promissory note or a bond) that is generally fixed in nominal terms; conversely, the borrower who issues the note or bond acknowledges indebtedness equal to the fixed money value. With assets, liabilities and interest specified in money terms, real values are clearly reduced by any rise in the price level: the real value of the debtor's liability decreases by the same extent to which the real worth of the asset held by the creditor falls. When the general price level is increasing, lenders will want to be compensated for such changes in purchasing power. That is, during periods of inflation, interest rates should cover expected increases in prices. This means that the lenders require the interest rates to be high enough to cover expected price increases, and a return for services of waiting. To see this rigorously, suppose overall prices are expected to increase by the rate d between two periods. Then we would expect the nominal interest rate i to be given by

$$1 + i = (1 + r)(1 + d), \quad (6.1)$$

since a lender requires compensation for both the opportunity cost of not investing (r) and for the general rise in prices (d). Expansion of the above equation gives

$$1 + i = 1 + r + d + rd, \quad (6.2)$$

and assuming $r.d$ is small, we have the approximation

$$i \simeq r + d. \quad (6.3)$$

If there is uncertainty about repayment of a loan, the rate of interest includes a risk premium to compensate the lender for uncertainty of repayment. Therefore, in general, observed interest rates on loans incorporate an interest charge for waiting, an inflation premium to cover expected rates of price increase over the period for which a loan is given, and a risk premium to cover costs related to uncertainty of repayment of the loan.

For our purposes we will assume in the remainder of this chapter a world of certainty. We also assume that an equilibrium rate of return r has been established by a process (such as that shown in Fig. 6.10). This rate of return provides a common price for all investors, telling them how much they can expect on their investment. This rate of return also provides information to savers to enable them to decide how consumption in one period may be traded for consumption in another period.

INTERTEMPORAL BUDGET CONSTRAINT

We imagine a consumer who has preference ranking over commodity bundles consisting of amounts of goods to be consumed in period 1 and amounts to be consumed in period 2. For simplicity we adopt the two dimensional framework and represent the bundles of goods available in periods 1 and 2 by composite goods. Let C_1 denote the consumption in period 1. (C_1 is the number of units of a composite good consumed in period 1.) Similarly C_2 stands for the consumption in period 2.

From his decision on supplying the services of capital goods and labor, a consumer knows with certainty that his income from supplying factor services in periods 1 and 2 are given by M_1 and M_2 respectively. Let us assume that the consumer faces a perfect lending market, in which he can borrow and lend money at the market interest rate r . Suppose S stands for the consumer's saving out of period 1's income. (If S is negative, then the consumer is a borrower and it represents his borrowings in period 1.) We denote the per unit prices of the composite good in periods 1 and 2 by p_1 and p_2 respectively.

The consumer faces two possible types of choices. He can just consume his income in each period, or he can consume less than his income during the first period. In the latter case he is accumulating some money in the form of saving out of his first period consumption for the second period. We need to consider two equations. The first one is

$$S = M_1 - p_1 C_1. \quad (6.4)$$

This equation simply defines the consumer's savings as the difference between his income in period 1 and expenditure on consumption in this period. To determine the consumer's expenditure in period 2, we must take into consideration the facts that he must pay off in period 2 (with interest) any debts incurred in period 1, and he will spend in period 2 any saving (and interest) that he accumulates in period 1. Therefore, we have the second equation

$$\begin{aligned} p_2 C_2 &= M_2 + S + rS \\ &= M_2 + S(1 + r) \end{aligned} \quad (6.5)$$

Eq. (6.5) says that the amount the consumer spends on consumption in period 2 equals his income in this period plus the amount saved from period 1, plus the interest he earned on his savings. If the consumer is a borrower, his first period consumption is greater than his income, that is, S in Eq. 6.4

is negative ($M_1 < p_1 C_1$). In this case $S(1+r)$ represents the amount that is paid in period 2 to repay borrowing S plus interest rS on S .

Combining Eqs. 6.4 and 6.5 we get the consumer's budget constraint :

$$p_2 C_2 = M_2 + (M_1 - p_1 C_1)(1+r) \quad (6.6)$$

$$\text{or} \quad p_1(1+r)C_1 + p_2 C_2 = M_1(1+r) + M_2. \quad (6.7)$$

From Eq. (6.6) we note once again that the consumer pays or earns interest on his borrowings or savings according as $M_1 - p_1 C_1$ is negative or positive. If $M_1 = p_1 C_1$, then the consumer is neither a borrower nor a lender.

Recalling that d stands for the rate of inflation or the proportionate increase in the prices of the consumption of goods between periods 1 and 2, we have

$$d = \frac{p_2 - p_1}{p_1}, \quad (6.8)$$

which implies that $p_2 = p_1(1+d)$. We can then rewrite Eq. (6.7) as

$$p_1(1+r)C_1 + p_1(1+d)C_2 = M_1(1+r) + M_2. \quad (6.9)$$

Note that the budget constraint given by Eq. (6.7) can be alternatively expressed as

$$p_1 C_1 + \frac{p_2}{1+r} C_2 = M_1 + \frac{M_2}{(1+r)} \quad (6.10)$$

We can use the budget constraint in Eq. (6.7) to measure period 1 price relative to the period 2 price, while that in Eq. (6.10) to do the reverse. Now, the maximum amount of first period consumption is $[M_1(1+r) + M_2]/p_1(1+r)$. (We get this by putting $C_2 = 0$ in Eq. [6.10] and then dividing the resulting expression by p_1 .) This is also the intercept of the budget constraint along the C_1 axis. Similarly, the intercept along the C_2 axis is the maximum amount of second period consumption, which occurs when $C_1 = 0$. From Eq. (6.9), this amount turns out to be $[M_1(1+r) + M_2]/p_1(1+d)$.

Letting $W(r) = M_1(1+r) + M_2$, we present the consumer's budget line in Fig. 6.1. For any interest rate prevailing in the market, one choice available to the consumer is to set savings S equal to zero. Therefore, the bundle $[M_1/p_1, M_2/p_1(1+d)]$ will always lie on the budget line. Note that, $W(r)/p_1(1+r)$ becomes

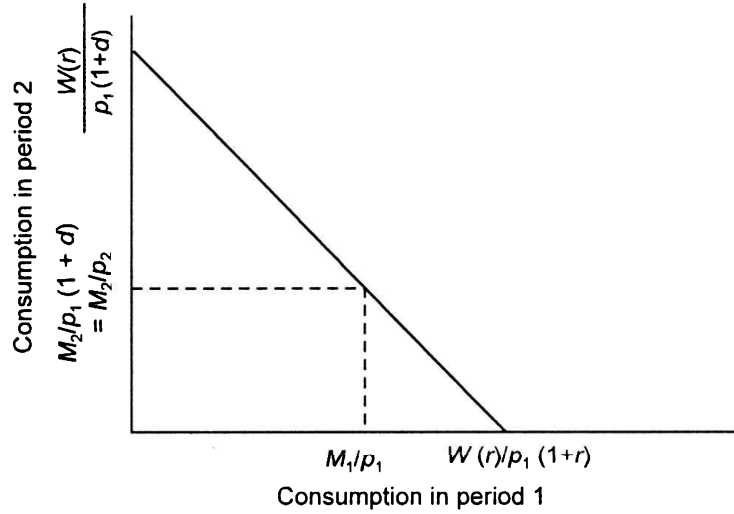


Fig. 6.1 A Consumer's Budget Line with respect to Timing of Consumption

$$\frac{M_1}{p_1} + \frac{M_2}{p_1(1+r)}, \quad (6.11)$$

the present value of the real or deflated endowment $(M_1/p_1, M_2/p_2)$. The present value of an endowment of money in two periods is the amount of money in period 1 that would generate the same budget set as the endowment, where both the endowment and the amount of money in period 1 are expressed in deflated terms. Similarly, $W(r)/p_1(1+d)$ is the future value of the deflated endowment $(M_1/p_1, M_2/p_2)$. Note that the Eqs. (6.7) and (6.10) express budget constraints in terms of future and present values respectively. The form given in Eq. (6.10) is especially important since it measures the future relative to the present.

The slope of the budget line becomes

$$-\frac{C_2 \text{ intercept}}{C_1 \text{ intercept}} = -\frac{1+r}{1+d}. \quad (6.12)$$

To interpret Eq. (6.12), note that if the consumer increases consumption in period 1 by one unit, then his expenditure in period 1 increases by p_1 . This means that his expenditure in period 2 will decrease by $p_1(1+r)$. Therefore, the number of units of consumption the consumer has to give up in period 2 is $p_1(1+r)/p_2$. Since $p_2 = (1+d)/p_1$, $(1+r)/(1+d)$ is the real cost of giving up consumption in period 2 at the expense of increasing consumption in period 1 by one unit.

Finally, note that if the rate of interest is zero and no borrowing is allowed, then C_2 intercept of the budget line becomes $(M_1/p_2 + M_2/p_2)$ and the line does not come down beyond the deflated endowment $(M_1/p_1, M_2/p_2)$. We represent this situation in Fig. 6.2.

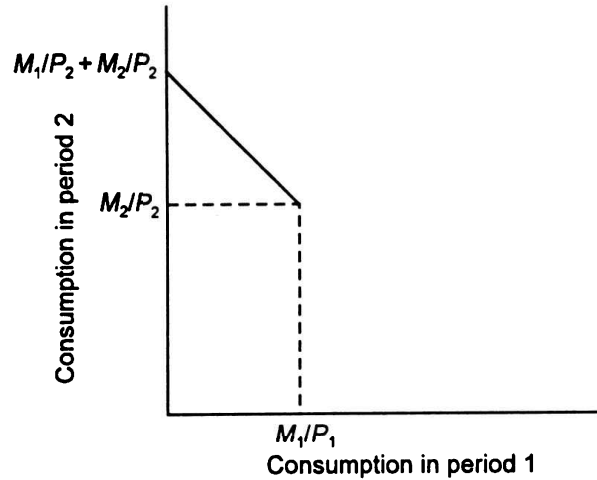


**Fig. 6.2 Consumer's Intertemporal Budget Line
(Rate of Interest Zero, no Borrowing Allowed)**

TIME PREFERENCES

The consumer's indifference map specifies his preferences with respect to timing of consumption. The shape of an indifference curve will indicate his tastes at different times. To characterize the indifference curves we will use the term marginal rate of time preference (MRTP) to refer to the MRS between C_1 and C_2 . Along any indifference curve, MRTP is the marginal rate of substitution defined by

$$MRTP = - \frac{\Delta C_2}{\Delta C_1}. \quad (6.13)$$

Diminishing MRTP means that along any indifference curve, consumption in period 2 becomes a better substitute for consumption in period 1, as the first period consumption increases. This is equivalent to the requirement that the indifference curves be strictly convex to the origin. That is, the consumer wishes to have an average amount of consumption in each period, rather than having a lot in one period and nothing in the other period. In terms of Eq. (6.13) we can say that MRTP indicates the consumer's willingness

to sacrifice consumption in period 2 to increase period 1 consumption. This willingness decreases as consumption in period 1 increases.

If we consider indifference curves with a constant slope of say, -1 , then consumption in period 2 is a perfect substitute of period 1 consumption. That is, the consumer becomes totally indifferent between consumption in period 1 and period 2. In this case MRTP becomes 1. On the other hand, if the consumer is unwilling to substitute consumption from one time period to another, then the indifference curves will correspond to the ones drawn for perfect complements. As usual, the intermediate case with diminishing MRTP is the more reasonable situation. Further, between two or more non-intersecting indifference curves, the highest one represents the maximum level of satisfaction.

UTILITY MAXIMIZATION AND CHOICE BETWEEN TWO PERIODS

Given a consumer's budget constraint and his preferences for consumption in each of the two periods, we can determine the optimal consumption choice. The optimal or the utility maximizing bundle in the figure below is given by $(\bar{C}_1, \bar{C}_2)$. Optimality occurs at the point of tangency between the budget line and the indifference curve. At the utility maximizing point, we have the condition that MRTP equals the real cost (in terms of consumption in period 2) of increasing period 1 consumption by one unit. Now, with a perfect lending market every consumer faces the same rate of interest. Given that the price vector (p_1, p_2) is also the same for all consumers, it must be the case that every consumer will adjust his expenditure pattern to satisfy

$$MRTP = - \frac{\Delta C_2}{\Delta C_1} = \frac{1+r}{1+d}. \quad (6.14)$$

Thus, the MRTP will be the same across consumers.

In panel (a)([b]) of the above figure the consumer allocates a lower (higher) fraction of the income to the current consumption. That is, in some sense in panel (b) the consumer tends to value the current consumption more highly than the future consumption. Such consumers are referred to as impatient or myopic. We also note that in panel (b), $M_1/p_1 \leq \bar{C}_1$, that is, the consumer is a borrower. While the sign of the expression $(\bar{C}_1 - M_1/p_1)$ will depend on the exact positions of the real endowment $(M_1/p_1, M_2/p_2)$ and the consumer's optimal choice, it is reasonable to expect that an impatient consumer is

usually a borrower. On the other hand, in panel (a) the consumer values the future consumption more highly. In this case the consumer is said to possess a patient preference. From this figure, we also note that $\bar{C}_1 < M_1/p_1$, that is, the consumer is a lender. Again, while this need not hold unambiguously, it is quite reasonable to expect that a patient consumer is a lender. Formally, for a two-period planning horizon with one consumer good, a consumer is said to be patient if when $A > B$, he prefers the consumption pattern A in period 2 and B in period 1, to B in period 2 and A in period 1. Therefore, a consumer with relatively patient preference will tend to save more (see panel [a] in Fig. 6.3) or borrow less than he would do with impatient preference.

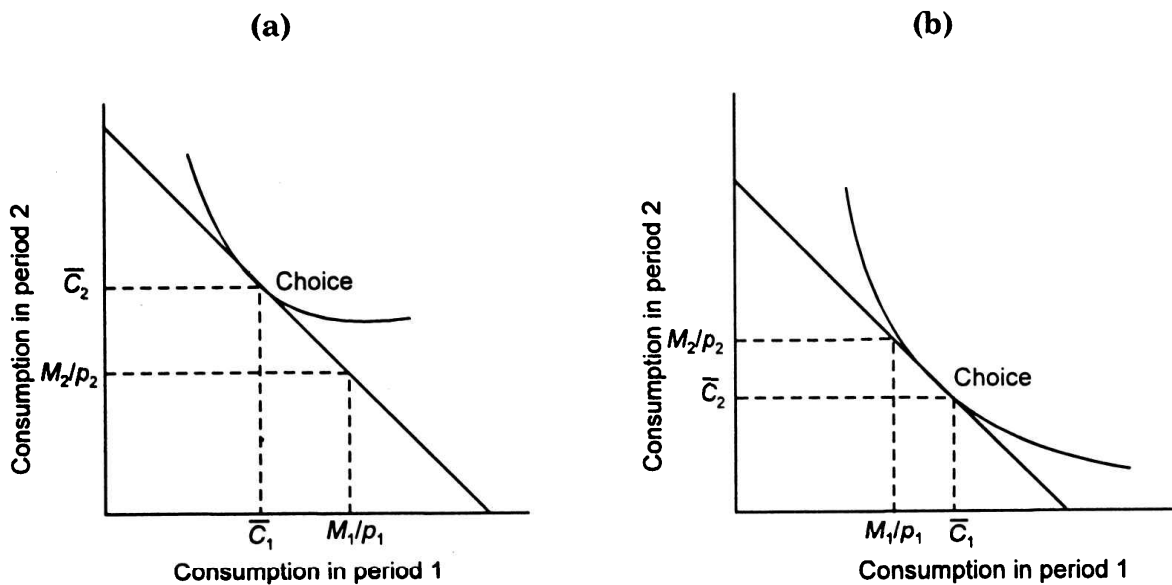


Fig. 6.3 Consumer's Intertemporal Choice

COMPARATIVE STATICS: EFFECTS OF CHANGE IN RATE OF INTEREST ON SAVINGS

We will now examine how the consumer would react to a change in the rate of interest. From Eq. (6.6) it appears that increasing the rate of interest will rotate the budget line around the real endowment $(M_1/p_1, M_2/p_2)$ to a steeper position. That is, for a given reduction in C_1 , the consumer will enjoy more consumption in period 2 if the rate of interest is higher. Since at any interest rate the real endowment is always available, the budget line must pivot around the endowment as the rate of interest goes up.

The result of an increase in the rate of interest is shown in Fig. 6.4. At the initial interest rate r_0 , the bundle $A = (C_1^0, C_2^0)$ is chosen. An increase in the

interest rate from r_0 to r_1 will tilt the budget line in the clockwise direction around the real endowment. The bundle chosen for $r = r_1$ is $D = (C_1^1, C_2^1)$. We can use the Slutsky equation to decompose the change in demand due to a change in the interest rate into substitution and income effects. The substitution effect is illustrated by the movement from A to B. To understand this, note that in terms of future value budget constraint, raising interest rate is analogous to raising the price of consumption in period 1 as compared to consumption in period 2. To put it another way, since the alternative to having one more unit of consumption in period 1 is to save the expenditure and have it available for consumption in period 2 plus interest, we say that the opportunity cost of consumption in period 1 increases as interest rate increases. That is, with increase in interest rate, consumption in period 1 becomes more expensive than before in terms of foregone interest income. Therefore, an increase in interest rate leads to the substitution of period 2 consumption for period 1 consumption along the indifference curve I_0 . That is, the substitution effect says that the consumer should consume less in period 1.

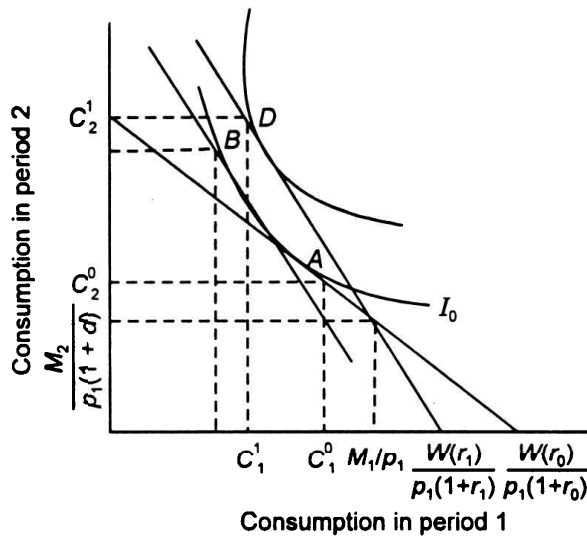


Fig. 6.4 Income and Substitution Effects of Increase in Interest Rate: Positive Savings

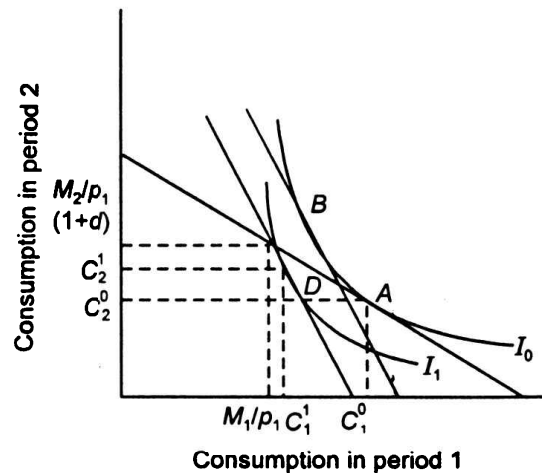


Fig. 6.5 Substitution and Income Effects of Increase in Interest Rate: Negative Savings

We assume that both C_1 and C_2 are normal goods. If the consumer is a lender in period 1, then $S = (M_1 - p_1 C_1) > 0$. For such a person an increase in interest rate increases income. Hence under the assumption that both C_1 and C_2 are normal goods, the income effect induces the individual to consume more of both C_1 and C_2 . This is shown by the movement from B to D.

Next, if the person is a borrower ($S < 0$), then from Eq. (6.6) it follows that an increase in interest rate decreases his income. Now, the substitution effect leads him to reduce the consumption in period 1 and increase the consumption in period 2. This is shown by the movement from A to B in Fig. 6.5. Since both C_1 and C_2 have been assumed to be normal, the income effect (due to reduction in income) decreases both C_1 and C_2 . This is illustrated by movement from B to D .

So far we have examined the effect of an increased interest rate on consumption. Let us now look at the implication of the increased rate of interest on saving. When the person is a lender, the substitution effect increases saving, whereas the income effect acts to decrease it. If the substitution effect dominates the income effect, then the increased rate of interest reduces consumption in period 1 which in turn increases saving. This case is shown in Fig. 6.4. Since $C_1^1 < C_1^0$, we have $(M_1/p_1) - C_1^1 > (M_1/p_1) - C_1^0$. That is, savings are higher at r_1 than at r_0 . On the other hand, if the income effect dominates, then saving decreases with increase in interest.

If the person is a borrower, that is, if $M_1/p_1 < C_1^0$, then an increase in rate of interest implies that $C_1^1 < C_1^0$. This means that borrowings decrease as interest rate increases.

Finally, we will say something about how the choice of being a lender or a borrower changes with changes in interest rate. Suppose that the consumer is a lender initially ($C_1^0 < M_1/p_1$). Then C_1^0 must lie to the left of the endowment point. Now, suppose interest rate increases. Then the original optimal bundle is still available at the new budget line and the axiom of revealed preference requires that the new optimal bundle cannot lie inside the original budget set. Therefore, the new optimal bundle must lie to the left of the endowment. That is, the consumer remains a lender as the rate of interest goes up. We show this in Fig. 6.6.

Similarly, we can check that if the consumer is initially a borrower, then as the rate of interest decreases, he will remain a borrower. However, if a person is a lender (borrower) initially, then as the rate of interest decreases (increases), he may decide to become a borrower (lender). We can also see how a consumer's welfare changes with respect to changes in the interest rate. For instance, if a consumer is a borrower initially and the increased interest rate induces him to remain a borrower, then he must be worse off. This can be seen by drawing appropriate indifference curves in Fig. 6.6.

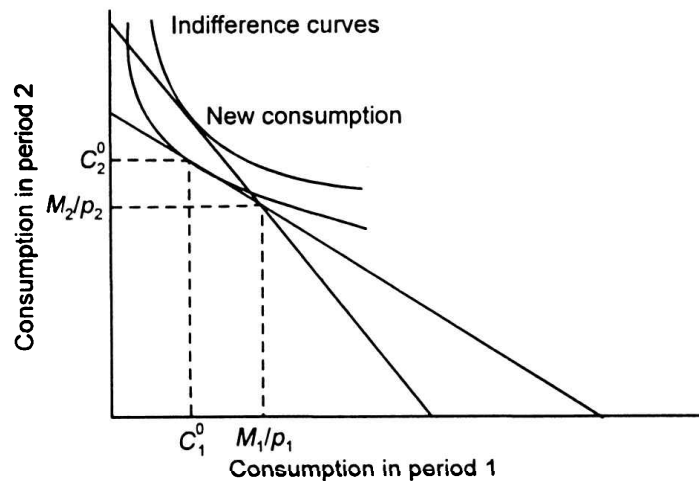


Fig. 6.6 Lending and Rate of Interest

SAVING AND RATE OF INTEREST

A consumer's saving $S(r)$ is a function of the interest rate r . In the figure below we plot a consumer's inverse saving curve. For very low values of r , the consumer will actually be a borrower. That is, the consumer will prefer immediate consumption to consumption in the future. But if the amount of future consumption made possible through saving is sufficiently greater than the sacrifice of the present consumption, the consumer will be willing to defer consumption today in favor of increased future consumption. This will usually occur at a higher rate of interest. Now, since the rate of interest is the opportunity cost of consumption in the current period, it is reasonable to expect that current consumption falls (saving increases), as the rate of interest increases. That is, saving as a function of the interest rate is monotonically increasing.

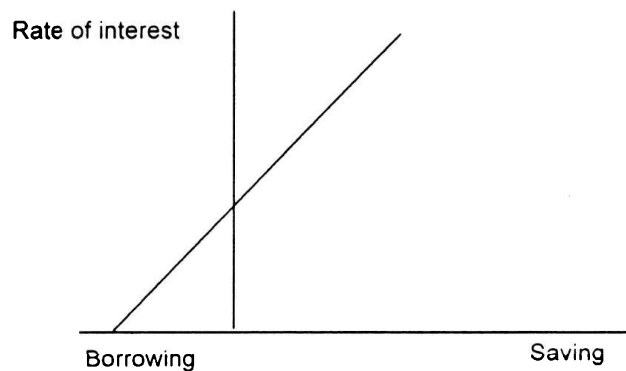


Fig. 6.7 Saving and Rate of Saving Interest

Note that positive saving by a consumer means addition to the stock of his assets, whereas negative saving means liquidation of some of his assets. Hence saving gives rise to changes in a consumer's equity, which is given by assets less liabilities. Therefore, given changes in the equity position of a consumer, it may be necessary to adjust his asset portfolio.

ASSET MARKETS

An asset can be regarded as a good which provides a flow of services over time. In the context of a market for assets, our primary interests are to examine how the consumer decides what kinds and amounts of assets to hold and how the prices of such assets are related to the incomes they earn. The major assets that are available for ownership include titles to capital goods such as machines, buildings, inventories and other manufactured non-labor resources; titles to natural resources such as land and mineral deposits; shares of stock in business firms, bonds issued by governments and corporations; and ownership of human capital in the form of skills acquired by the consumer through education and training. Some assets provide a flow of consumption services. An example of an asset of this type is housing. Some assets, like bonds and shares, are financial assets—they provide a flow of money which can be used for purchasing goods and services. In order to analyze the relationship between the price of an asset and the income stream it generates, we need to consider the discounted present value of the stream.

DISCOUNTED PRESENT VALUE CRITERION

The analysis of the discounted present value criterion is based on the procedure that should be used to add up sums of money that are to be paid in different periods. The starting point is the observation that Re. 1 today is worth more than Re. 1 that is not received until some later date. If one rupee is available today, it can be invested and some interest can be earned at the current interest rate r . More specifically, Re. 1 will grow into Rs. $(1 + r)$ next year. Thus, Rs. $(1 + r)$ is the future equivalent of Re. 1. In other words Rs. $(1 + r)$ next year is equivalent to Re. 1 today since that is how much one will have to pay to borrow Re. 1 today. The value $(1 + r)$ is just the price of Re. 1 today, relative to Re. 1 next year. Alternatively, if one rupee is not available until next year, the amount of interest foregone is r .

The calculation of present value is based on the idea that everything is measured in terms of today's rupees. More precisely, the discounted present value (DPV) of Re. 1 payable next year is that amount which if invested at

the rate of interest r would grow to be exactly Re. 1 at the end of one year. That is,

$$DPV (\text{Re. } 1) = \frac{\text{Re. } 1}{(1 + r)} \quad (6.15)$$

This is because Rs. $1/(1 + r)$ can be turned into one rupee next year by saving it at the interest rate r . Using the definition of Eq. (6.15), future rupees can be converted into present rupees.

Using the present value we can express the budget constraint of Eq. (6.10) for a two-period consumption problem as follows: a consumption plan can be afforded if the present value of consumption equals the present value of income. A similar statement can be made in terms of the future value using the budget constraint of Eq. (6.7). Note that the consumer's budget set gets expanded as the present (hence the future) value of the endowment increases. This follows from the fact that with an increase in the present value of the endowment, the budget line moves out from the origin and the new budget set contains the old one.

It is easy to generalize the idea of present value to any number of periods. For example, if Re. 1 is invested for two years, it will grow into Rs. $1(1 + r)(1 + r) = \text{Rs. } (1 + r)^2$. The present value of Re. 1 payable in two years would be Rs. $1/(1 + r)^2$. This is the amount that would have to be invested today in order to obtain Re. 1 after two years. In general, the present value of Re. 1 payable in n years from now is Rs. $1/(1 + r)^n$. Clearly, the present value of Re. 1 available at some date in the future decreases as the interest rate increases; for a fixed interest rate, rupees payable in the distant future are worth less than those payable in the near future.

Consider now a stream of monetary payments that pay Rs. N_1 one year from now, Rs. N_2 two years from now, and in general Rs. N_i i years from now. If the present rate of interest is r and if the stream makes payments upto n years, how much is such a stream of income worth to a person today? The answer is the *DPV* of the stream $N_1, N_2, \dots, N_n$, which is given by

$$DPV = \frac{N_1}{(1 + r)} + \frac{N_2}{(1 + r)^2} + \dots + \frac{N_n}{(1 + r)^n} \quad (6.16)$$

DPV is the amount that an individual would be willing to pay in return for a promise to receive the stream $N_1, N_2, \dots, N_n$. It represents the amount that would have to be invested now if one wishes to duplicate the payment stream.

ANNUITIES AND PERPETUITIES

Assets that provide monetary returns are called financial assets. In the financial markets people get opportunities to buy/sell assets that generate different types of monetary flows over time. Securities are examples of financial assets that promise to make certain types of payments. The particular kind of security that we will examine here is an annuity. An annuity is a promise to pay Rs. N in each period for n periods, starting next period. The *DPV* of such a contract is

$$DPV = \frac{N}{(1+r)} + \frac{N}{(1+r)^2} + \dots + \frac{N}{(1+r)^n}. \quad (6.17)$$

Let $D = 1/(1+r)$, then

$$\begin{aligned} DPV &= N(D + D^2 + \dots + D^n) \\ &= ND(1 + D + \dots + D^{n-1}) \\ &= ND \left(\frac{1-D^n}{1-D} \right). \end{aligned} \quad (6.18)$$

In particular, suppose that an annuity promises an income of Rs. N per year into the indefinite future. The case of an infinite period annuity is called a perpetuity or a consol. Notice that

$$\lim_{n \rightarrow \infty} D^n = 0. \quad (6.19)$$

Therefore, for a consol,

$$DPV \text{ of infinite annuity} = \lim_{n \rightarrow \infty} DPV = ND \frac{1}{1-D}, \quad (6.20)$$

which by the definition of D ,

$$\begin{aligned} &= N \left(\frac{1}{1+r} \right) \left[\frac{1}{1-1/(1+r)} \right] \\ &= N \left(\frac{1}{1+r} \right) \left(\frac{1+r}{r} \right) \\ &= \frac{N}{r}. \end{aligned} \quad (6.21)$$

The formula simply says that the amount that must be invested if one is to obtain Rs. N per year forever is simply Rs. N/r , since this amount of money would earn Rs. N in interest in each period ($r \cdot N/r = N$). From Eq. (6.21) it follows directly that increasing the interest rate reduces the value of the consol.

SPECIAL CASE OF BONDS

Bonds are issued by governments and corporations to borrow money. The borrower—the agent who issues the bond—promises to pay a fixed amount of money, say, Rs. N (the coupon) each year until a certain period n (the maturity date). On maturity the borrower also promises to return the principal (face) value to the holder of the bond. If the principal value of the bond is Rs. P , then the *DPV* of such a promise is

$$DPV = \frac{N}{(1+r)} + \frac{N}{(1+r)^2} + \cdots + \frac{N}{(1+r)^n} + \frac{P}{(1+r)^n}. \quad (6.22)$$

Again, let $D = 1/(1+r)$, then

$$DPV = ND + ND^2 + \cdots + ND^n + PD^n. \quad (6.23)$$

Eq. (6.23) can be used in another way. Suppose we know the price at which the bond is currently trading. Denote this price by B . Then we can ask the following question: what value of r gives the bond a *DPV* equal to B ? For this, we set

$$B = DPV = ND + ND^2 + \cdots + (N + P)D^n. \quad (6.24)$$

Since, B , P and N are known, we can solve Eq. (6.24) for D and hence for r . The value of r that solves Eq. (6.24) is referred to as the yield on the bond and is the best measure of the return actually available from the bond. The yield of a bond represents the return available both from direct interest payments and from any price differential between the initial price (B) and the principal value P . Note that as r increases, *DPV* decreases. This is a precise way of formulating the well-known concept that bond prices (*DPVs*) and interest rate (yields) are inversely related.

APPLICATIONS OF DISCOUNTED PRESENT VALUES

An interesting application of *DPV* arises in the context of a consumer's holding of different assets. In choosing which assets to hold, it is natural to assume

that the consumer should be motivated solely by the size and the timing of the income streams that assets generate. The following remarkable result of Irving Fisher (1930) indicates a general rule that governs a consumer's choice: 'Whatever his time preferences about consumption are, the consumer facing a perfect lending market chooses among assets solely on the basis of their discounted present values, using the market interest rate to discount income.'

We can illustrate the argument underlying Fisher's proposition as follows. In Fig. 6.8, $(M'_1/p_1, M'_2/p_2)$ is a worse consumption bundle than the consumer's original endowment $(M_1/p_1, M_2/p_2)$ since $(M'_1/p_1, M'_2/p_2)$ lies below the indifference curve passing through $(M_1/p_1, M_2/p_2)$. But if the consumer is able to borrow and lend at the interest rate r , then he would prefer $(M'_1/p_1, M'_2/p_2)$ to $(M_1/p_1, M_2/p_2)$. This follows from the fact that with the real endowment $(M'_1/p_1, M'_2/p_2)$ the consumer can consume a bundle, say, (C'_1, C'_2) which lies on a higher indifference curve than the indifference curve passing through his current consumption bundle $(M_1/p_1, M_2/p_2)$.

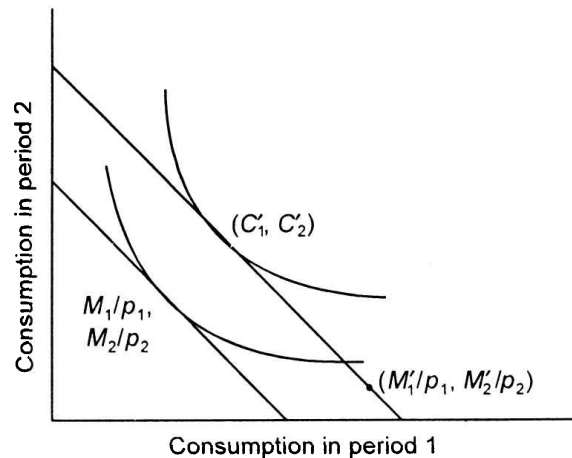


Fig. 6.8 Consumer Choice of Endowment with Higher *DPV*

A very useful application of *DPV* criterion in Eq. (6.16) is to evaluate income streams offered by different types of investment. If we wish to compare two alternative investments that yield different income streams, then to choose between the two we simply calculate their *DPVs* and select the one that generates the higher *DPV*.

Since a consumer will be indifferent between any two assets that generate the same *DPV*, these assets have the same price. In fact, Fisher demonstrated the following stronger proposition: 'Given a perfect lending market, the price of any asset equals the asset's *DPV*, using the market rate of interest as the discounted rate.'

Fisher's proposition enables us to arrive at a straightforward answer to the question of which assets a consumer should hold. Assuming that a perfect lending market extends the notion of perfect competition from commodity markets to loan markets, in equilibrium each asset must sell at a price equal to its *DPV*. Consequently, the consumer becomes indifferent about the types of assets to hold regardless of his time preferences about consumption. Equivalently, in equilibrium in the market for assets, investment on any asset yields a rate of return equal to the market rate of interest.

This analysis applies to all asset markets in the economy. For capital goods, the point of interest is the evaluation of the stream of rental incomes. For a share of stock in a business firm, the stream of earnings per share is to be evaluated, and so on. Only the evaluation of human capital requires a separate treatment and we will return to this issue after a while.

Sometimes it may be necessary to purchase an income stream by making a stream of payments. For instance, suppose that a consumer is in the process of deciding whether or not to buy a particular house. The house will last n years and give its owner a stream of payments in each of n years. Let the return in year i be N_i . N_i can be regarded as the consumer's rental income in period i if the house has been given to someone else on rent. In case the consumer himself is living in the house, then N_i stands for the implicit rental rate. The implicit rental rate is the rate at which the consumer could rent a similar house. Or, equivalently, it is the rate at which the consumer could rent the house to someone else. By deciding to rent his house to himself, the consumer is foregoing the opportunity of earning rental payment from someone else and hence incurring an opportunity cost. Further, the value of the house may increase over time. This increase in the value of an asset is called capital appreciation. Therefore, if the rent is positively related with the value of the house, then the consumer's rental income from the house or the opportunity cost may increase over time. That is, it may be the case that $N_1 \leq N_2 \leq \dots \leq N_n$.

The house can be bought by making a stream of payments $p_1, p_2, \dots, p_n$. If the *DPV* of the income stream $N_1, \dots, N_n$ exceeds that of the payment stream $p_1, \dots, p_n$, the consumer should make the purchase. That is, the consumer purchases the house if

$$\frac{N_1}{(1+r)} + \frac{N_2}{(1+r)^2} + \dots + \frac{N_n}{(1+r)^n} > \frac{p_1}{(1+r)} + \frac{p_2}{(1+r)^2} + \dots + \frac{p_n}{(1+r)^n}, \quad (6.25)$$

which we rewrite as

$$\frac{N_1 - p_1}{(1+r)} + \frac{N_2 - p_2}{(1+r)^2} + \dots + \frac{N_n - p_n}{(1+r)^n} > 0. \quad (6.26)$$

The term on the left hand side of Eq. (6.26) is the net present value. If the inequality in Eq. (6.26) is reversed, the consumer's expenditure exceeds his income and the consumer as a utility maximizing agent will not buy the house. Therefore, the consumer should buy the house if and only if the net present value is positive.

Finally, we come to the problem of maximization of a firm's net cash flows (profits). (A firm is a production unit that transforms inputs into outputs. See Chapter 8 for details.). In a single period set-up a firm's objective is to maximize its current net cash flows. Extending this idea to the intertemporal case we can say that the firms will operate to maximize the *DPV* of their net cash flows. This in fact has the basis that maximization of *DPV* of a firm's net cash flows will maximize the market value of each stockholder's share. This along with a perfect lending market enables each stockholder to attain the most preferred time stream of consumption from his holdings. Note that when each period is a repetition of the previous period, then as expected, maximization of *DPV* of net cash flows reduces to the problem of maximization of net cash flows for the current period.

DISCOUNTED PRESENT VALUE AND RENTALS OF ASSETS

When a consumer holding title to some capital good enters the market today, he signs contracts with some firm(s) for the rentals of the services of the capital good over future periods. Suppose the good lasts n years. That is, there is a natural aging of the good. (There is depreciation independently of the use of the good.) Thus, at the end of n years, the good has no further productive value. Let R_i be the rental in period i . These rentals are net of depreciation due to use of the good in production. The *DPV* of the income stream is then given by

$$DPV = \frac{R_1}{(1+r)} + \frac{R_2}{(1+r)^2} + \dots + \frac{R_n}{(1+r)^n}. \quad (6.27)$$

In a single period set-up the rental earned by a unit of a capital good equals the marginal revenue product of the capital good (See Chapter 10). Analogously, the current market value of the capital good is given by

$$DPV \text{ in Eq. (6.27)} = \text{price of capital good.} \quad (6.28)$$

The marginal revenue product enters into the evaluation of the good through *DPV*. Thus, given a perfect lending market, each asset sells at a price equal to the *DPV* of the asset's income stream.

MARKET FOR NEW CAPITAL GOODS

A firm uses capital services to produce output. These services are a flow magnitude. The factor relevant to the productive process is the number of machine hours (just as the number of labor hours) not the number of machines per se. It is often assumed that the flow of capital services is proportional to the stock of machines. For instance, 10 machines, if fully employed for 1 hour, can provide 10 machine-hours of service. Therefore, these two concepts are used synonymously. If during a particular period of time, a firm desires a certain number of machine hours, this is usually taken to mean that the firm requires a certain number of machines. That is, the firm's demand for capital services is its demand for capital.

Suppose that a firm obeying the profit maximizing rule finds that it requires more capital services than can be provided by its currently existing stock of machines. The firm has two choices here. First, it may hire the additional machines from the rental market. This is identical to the firm's decision to hire additional labor. Second, the firm can buy new machines to meet its requirements. The second alternative, that is, the firm's decision to buy new equipment is its investment. It is obvious that the demand for capital services and investment are not the same concept. Investment occurs only if the firm's demand for capital services exceeds that level of services which can be provided by its existing stock of machines.

So far we have discussed the demand side of the newly produced capital goods. Let us now turn to the supply side. In the short run capital goods are produced under the condition that the production function satisfies the law of variable proportions. This in turn implies that subject to the usual assumptions the concerned firm's average variable cost curve is downward sloping initially and it is increasing beyond some output level. Hence, the firm's marginal cost curve is rising as it passes through the minimum point of average variable cost curve. (See Chapter 9). Therefore, the firm's supply curve of the capital good is the upward rising portion of its marginal cost curve with the portion below the minimum average variable cost curve being wiped out. The market supply for the capital good is obtained by summing the supply curves of the firms producing the good. (See Chapter 10). Let us write $\bar{I}$ for the number of units of a newly produced capital good (an invest-

ment good) and S_I for the industry supply curve of this good. (Note that $S_I = \sum_i s_i^i$, where s_i^i is firm i 's supply curve of the capital good.) In panel (a) of Fig. 6.9 we show the supply curve of a typical firm engaged in manufacturing an investment good and in panel (b) we show the (inverse) supply curve of the good for the entire industry.

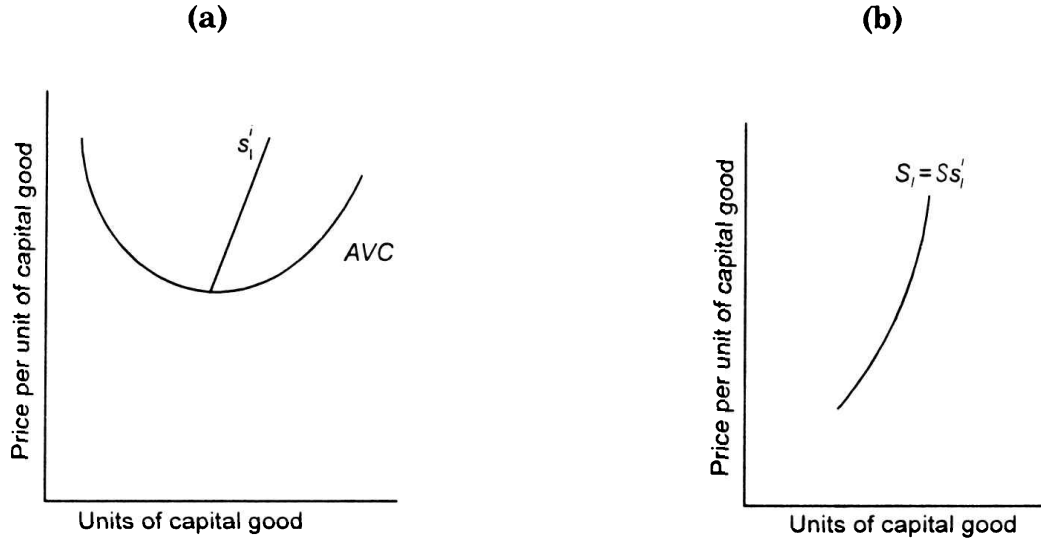


Fig. 6.9 Supply Curves of Investment Goods

There are different types of capital goods, each with its own industry supply curve. With equilibrium on the asset market, each newly produced capital good will sell for a price equal to the DPV of the net rentals associated with the good. (See Eq. [6.28].) Let us assume for simplicity that only one kind of capital good is produced and that it earns a constant rental N per year for an indefinite period. In such a case DPV is given by $DPV = N/r$ (see Eq. [6.21]). Also by Eq. (6.28), $p_I = N/r$, where p_I = price of the newly produced capital good.

At equilibrium in the market for investment goods, the price of an investment good p_I equals DPV , and it also equals the marginal cost of producing the investment good. (See Chapter 10.) Let V stand for the volume of the investment good produced, that is, $V = p_I \cdot \dot{I}$. Thus, V is the money amount of investment that occurs at price p_I of the investment good.

We can rewrite $p_I = N/r$ as

$$r = \frac{N}{p_I}, \quad (6.29)$$

which simply says that in equilibrium an infinitely long-lived capital good is equivalent to a perpetual bond and hence must yield the market rate of return. Note also that N/p_I is analogous to the concept of a perpetual rate of return.

Since $p_I = N/r$, a decrease in r increases p_I . Now, I as a function of p_I is the supply curve of the investment good and this curve slopes upward to the right. Given that $V = p_I \cdot I$, it follows from the behavior of p_I and I with respect to r and p_I respectively that the V curve slopes downward to the right as r increases. (We demonstrate this rigorously in appendix (6.1.)

In Fig. 6.10 we plot the saving and investment curves. At the interest rate r_e , saving equals investment. This is the equilibrium rate of interest. Once the interest rate has been determined, we can determine the market prices of assets, the output of investment goods, the amount of savings in the society and so on.

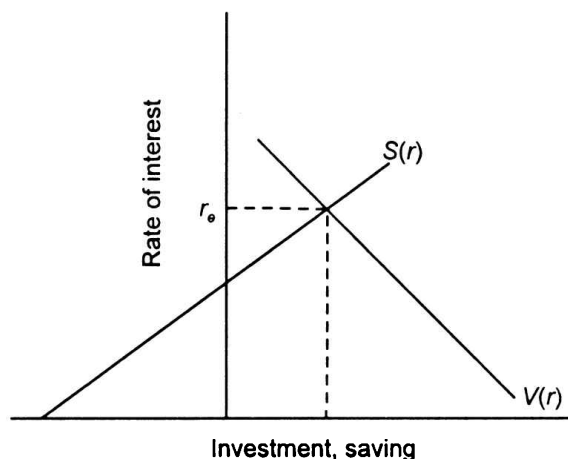


Fig. 6.10 Equilibrium in Market for Newly Produced Capital Goods

A rightward shift of the $V(r)$ curve, which may result from a reduction in the marginal cost of production of investment goods, increases the rate of interest and output of investment goods. Investment becomes more attractive than before, and so the demand for investment increases. Similarly, an increase in the patience on the part of the consumers shifts the $S(r)$ curve to the right, which results in a decrease in the rate of interest and an increase in the volume of saving (and investment).

HUMAN CAPITAL

A consumer's decision to acquire skills by investing in himself is a special kind of investment decision. An individual can invest in himself in a number of ways. A person can acquire a formal education, may accept apprenticeships

and learn skills on the job, etc. All such activities can be regarded as investments in the sense that both time and money are sacrificed currently with the expectation that this somehow will pay off in the future. By acquiring skills an individual is adding to his stock of human capital in much the same way as a firm adds to its stock of physical capital by purchasing machines. It has been verified empirically that most such additions to human capital are quite productive in the sense that they tend to increase an individual's future income. In fact, many investments in human capital generate higher return than investments in physical capital. (Becker [1964] provides a thorough analysis of the education aspects of human capital.)

Human capital has some unique characteristics that distinguish it from other assets. Contrary to other assets, human capital cannot be sold. Although an individual may rent out his investment to employers, he may not sell it in the way in which one can sell a machine. Human capital is totally lost upon the death of its owner. However, a broken machine may be repaired. The acquisition of human capital may take substantial time. Individuals seeking to improve their skills have to invest considerable time in doing so. During that time earnings must be foregone, and they should be regarded as a part of the cost of acquisition. Therefore, a consumer will invest in human capital so long as the *DPV* of the income stream associated with the investment exceeds the *DPV* of the cost of training, using the market rate of interest as the discount rate.

DEMAND FOR DURABLE GOODS

Goods that last for several time periods are called durable goods. Examples of durable goods include light bulbs, automobiles, washing machines, TV sets.

Producers of durable goods must decide how long their products should last. By spending more money at the initial stage, a manufacturer may be able to make the product last longer. When purchasing a durable good, a consumer considers how long it lasts and its resale value in future periods. For instance, producers of expensive automobiles often argue that consumers are better off buying better automobiles than less expensive ones because the better automobiles last longer and can be resold for a higher percent of their initial purchase price in any future period.

When resale is possible, the price a consumer is willing to pay for a durable good depends on both the value of the services of the durable good during the period the consumer owns it and the resale value at the end of that period. That is, consumers' expectations about the resale price may affect

the initial price of the durable good. For instance, if you are going to buy a house, the amount you are willing to pay may depend on how much you think you will receive when you sell it after some years. However, issues which analyze the effects of consumers' price expectations on a manufacturer are not the matters we want to discuss here. Our aim is to discuss the issues relating a consumer's demand for durable goods.

In the area of demand for durables, a number of really very different approaches has been suggested and agreement is far from being uniform. To understand how complex the problem is, let us try to look at some special features of the demand for durables. All of them have been emphasized by researchers working in this area. One feature is that a distinction should be drawn between new demand and replacement demand. Another feature is that discrete choice may cause an aggregation problem. It is not possible for a person to buy 1.3 TV sets. Since such goods usually last for more than one period, past decisions affect present decisions, and present decisions impose constraint on future actions. This is the third feature of the demand for durables. A fourth feature is that a non-durable purchase is one where choice is exercised over a continuous range, but the decision whether or not to have a durable good is a choice between two discrete alternatives. Next, many durable goods may be subject to a high level of technical change. This may affect the consumer's choice to a large extent, since he may be willing to buy a more sophisticated form of the good which may not be available right now but will be available soon. Also note that markets for durable goods are often imperfect. Finally, the consumer may be required to keep some provision for adjustment costs. These are some of the issues that affect a consumer's demand for durables. (For further discussions, see Deaton and Muellbauer, [1980].)

We will now discuss a two-period model in which there is a single durable good whose service flow is proportional to the stock. With its linear budget constraint and assumption of perfect markets, the model is a simple extension of the consumer behavior model to the case of durable goods. In this model there is one durable good and one non-durable good. The stock of durable goods measured at the end of period t is D_t . Utility is derived from a flow of services derived from the good. Here we assume that service flow is proportional to D_t , so that D_t will appear in the utility function. Hence in our two-period model preferences are represented by the utility function

$$U(C_1, C_2, D_1) \tag{6.30}$$

where C_1 and C_2 are non-durable consumptions in periods 1 and 2.

The stock is assumed to deteriorate physically at a rate d per period, so that D_t at the end of period t becomes $(1 - d) D_t$. This is the stock at the beginning of period $(t + 1)$. We are simply assuming here that a fraction d of the stock “evaporates” in each period leaving the fraction $(1 - d)$ of identical stock for use in the next period. For instance, if we measure stock in efficiency units, and if a new refrigerator is one efficiency unit, then a one year old refrigerator represents $(1 - d)$ efficiency units. Let us denote the consumer’s purchase of the durable good by d_t . Then

$$D_t = d_t + (1 - d)D_{t-1}. \quad (6.31)$$

This should be incorporated into the consumer’s budget constraint.

The durable good can be bought and sold at the same price irrespective of who is buying and selling. In the second-hand market the durable good is sold at the same price per efficiency unit as new goods. Given the type of deterioration of the durable good, this is not unreasonable in the absence of informational problems. It is assumed again that the consumer can borrow and lend freely at a fixed interest rate.

We will now formulate the consumer’s budget constraint. If as before, M_1 stands for the consumer’s income in period 1, p_1 denotes the price of the non-durable good in this period, then the consumer’s saving in period 1 is

$$S = M_1 - p_1 C_1 - v_1 d_1 \quad (6.32)$$

where v_1 is the price of a unit of the (new) durable good in period 1. Since the consumer can borrow and lend at the market interest rate r , and since the amount he spends on consumption in period 2 equals his income in this period plus his saving from period 1 and interest on saving, it must be the case that

$$p_2 C_2 + v_2 d_2 = M_2 + S(1 + r), \quad (6.33)$$

where p_2 and v_2 are respectively the prices of the non-durable and durable goods in period 2, and M_2 is the consumer’s income in this period. Since preferences are defined over stocks, we use Eq. (6.31) to eliminate d_1 and d_2 from Eqs. (6.32) and (6.33). Then substituting Eq. (6.32) in Eq. (6.33) we have the single budget constraint

$$p_1 C_1 + \frac{p_2}{(1+r)} C_2 + \left[v_1 - v_2 \frac{(1-d)}{(1+r)} \right] D_1 = M_1 + \frac{M_2}{1+r} + v_1 (1-d) D_0 \quad (6.34)$$

We rewrite Eq. (6.34) as

$$p_1 C_1 + p_2^* C_2 + v_1^* D_1 = W_1, \quad (6.35)$$

where $p_2^* = p_2/(1+r)$ is the discounted future price, $v_1^* = [v_1 - v_2(1-d)/(1+r)]$ and W_1 is the right hand side of Eq. (6.34). Clearly, W_1 can be regarded as the consumer's lifetime income. Note that W_1 includes the valuation of the starting stock of durables $v_1(1-d) D_0$.

Clearly, the utility function in Eq. (6.30) and the budget constraint Eq. (6.35) define a standard problem within the well-known model of consumer behavior and hence demand functions can be determined.

The above formulation enables us to incorporate durables into the standard analysis of non-durables. In particular, v_1^* in Eq. (6.35) is a price that represents the cost of holding one unit of durable good for one period. This is why v_1^* is also referred to as the rental equivalent price or user cost. Evidently, user cost is a function of price of the durable good, interest rate, and deterioration. If price remains constant, that is, if $v_1 = v_2 = v$, then $v^* = v((r+d)/(1+r))$. In the extreme case if $d = 0$, $v^* = v_1$. The case of capital gain or loss arises according as $v_2 > v_1$ or $v_2 < v_1$. Therefore, user cost, unlike a normal price, can be negative. If this happens consumers may wish to hold an infinite quantity of the durable good to reap infinite capital gains.

Example 6.1

Suppose a person faces two income streams A and B which promise to pay according to the following schedules:

<i>Stream</i>	<i>Payment</i>	
	<i>Current period</i>	<i>After one year</i>
A	1	1.10
B	1.1	0.99

We want to determine the interest rate at which the person will be indifferent between the two streams.

Suppose that the required interest rate is r . The present value of A is $1 + \frac{1.1}{1+r}$, while the present value of B is $1.1 + \frac{0.99}{1+r}$. Equating these two present values we have $r = 0.10$. Thus, the interest rate is 10 percent. Note that at any interest rate higher (lower) than 10 percent the present value of $B(A)$ is higher than that of $A(B)$.

Applied Example 6.1 When to Harvest a Tree

As an example of the *DPV* criterion consider the case of a landowner who has to decide when to cut down a tree and sell it. Suppose that the value of the tree at any time, t , is given by $f(t)$. The *DPV* of the tree harvested at time t is then $f(t)/(1 + r)^t$.

Now, suppose we allow continuous compounding. If the interest rate r is paid n times a year, then the recipient gets $(1 + r/n)^{nt}$ after t years. If the interest is paid continuously, then the amount the recipient gets is obtained by taking the limit of $(1 + r/n)^{nt}$ as n goes to infinity. This limit becomes e^{rt} , where e is the base of natural logarithms. The *DPV* of the tree harvested at time t is then given by $e^{-rt}f(t)$.

If the landowner allows the tree to continue growing, he will forego the potential interest that could be earned by investing the market value of the tree in some other use. On the other hand, harvesting the tree will mean that the tree's future growth must be foregone. The profit maximizing decision will be to harvest the tree at the time at which these two costs balance each other.

In order to maximize the *DPV*, we differentiate $e^{-rt} f(t)$ with respect to t and set the resulting expression at zero. This gives

$$e^{-rt} f'(t) - re^{-rt} f(t) = 0, \quad (6.36)$$

which on simplification becomes

$$f'(t) - rf(t) = 0. \quad (6.37)$$

Therefore,

$$r = \frac{f'(t)}{f(t)} \quad (6.38)$$

Hence the tree should be harvested when the rate of interest is equal to the proportional rate of growth of the tree. This result is quite reasonable intuitively. If the tree is growing more rapidly than the interest rate, then the optimal policy for the owner would be to wait. On the other hand, if the tree is growing less rapidly than the rate of interest, the tree should be cut and the funds available from its sale should be invested elsewhere at the interest rate r .

Now, if t given by Eq. (6.38) is to represent a maximum, the second order condition should be verified. That is, at the chosen value of t we must have

$$f''(t) - rf'(t) < 0. \quad (6.39)$$

If we assume that $f'(t) > 0$ (the tree is always growing), then the second order condition will hold if $f''(t) < 0$, that is, if the growth of the tree decreases over time.

To see how the optimal choice of t changes with respect to a change in r , we differentiate equation Eq. (6.37) totally:

$$f''(t) \frac{\delta t}{\delta r} - rf'(t) \frac{\delta t}{\delta r} - f(t) = 0, \quad (6.40)$$

from which it follows that

$$\frac{\delta t}{\delta r} = \frac{f(t)}{f''(t) - rf'(t)}. \quad (6.41)$$

Given $f(t) > 0$, this expression becomes negative in view of the second order condition. Therefore, an increase in r will decrease t . That is, the tree will be cut sooner if the rate of interest rises.

Applied Example 6.2 Effects of Tax Policies

In this applied example we consider the quantitative effect of taxation on investment return and the problem of adjustment for differences among returns.

i) Taxation of Investment Returns

Suppose a person invests Rs. X in an asset and receives an interest payment of Rs. rX after one year. But he has to pay a tax of Rs. trX ($0 < t < 1$) on the interest income. Consequently, he is left with Rs. $(1 - t)rX$ as after-tax income. The expression $(1 - t)r$ is the after-tax interest rate.

Now, suppose that the person borrows Rs. X in the lending market. Then he has to make an interest payment of Rs. rX . Assume that the person enjoys some tax deduction on interest payment. That is, the interest payment can be subtracted from other incomes and taxes are to be paid only on what is left. In India, some interest payments are of this type and some are not. For instance, one gets tax deductions on house building loan interest payments, but not on ordinary consumption loans. An interest payment of Rs. rX will reduce taxes by Rs. trX . Therefore, the total cost of Rs. X borrowed becomes Rs. $(rX - trX) = Rs. (1 - t)rX$. Thus, after-tax interest payment is the same whether the person is a borrower or a lender.

ii) Adjustment for Differences among Asset Returns

Suppose that an asset pays a before-tax rate of return r_1 , and another asset pays a tax-free return r_2 . If an asset holder holds both the assets and pays taxes on income at rate t , then it must be the case that $(1 - t)r_1 = r_2$. That is, after-tax return on each asset must be the same. Otherwise, the individual will hold the asset that generates the higher after-tax return.

Applied Example 6.3 Use of Exhaustible Resources

Recent concern with rising energy prices has motivated economists to reexamine theories of the optimal use of natural resource stocks. Of particular interest is the question of whether the markets in which decisions are based on supply-demand considerations are sufficient to meet the needs of future consumers. That is, the question necessarily involves the examination of the optimal time pattern for the depletion of a fixed stock of some resource (e.g. oil, coal).

It turns out that the total quantity consumed during a period be chosen so that market price rises at the market rate of interest. (It is assumed that the extraction cost of resource under consideration is zero.) Any slower rate of price increase would prompt investors to put their funds in some other form of capital, whereas any faster rate would motivate the investors to put all available funds in the resource. Therefore, if we write p_t for the price of the resource at time t , then $p_{t+1} = p_t(1 + r)$. If the resource stock is to be fully depleted, then the final period price, say, p_T , be such that demand becomes zero at that price. Such a price can be found by choosing it to be high enough so that the perfect substitute for the resource in question totally dominates the market. For example, suppose it is known that solar power, which can be produced at a constant cost of C , would totally replace petroleum energy in T years from now. Clearly, after T years from now the price of petroleum must be C , the price of solar power. This means that the current price p_0 of petroleum must grow at the interest rate r over the next T years to be equal to C . Thus, we have

$$p_0(1 + r)^T = C, \quad (6.42)$$

$$\text{or} \quad p_0 = \frac{C}{(1 + r)^T}. \quad (6.43)$$

Clearly, an increase in r or T decreases p_0 and a reduction in C also decreases p_0 . An increase in the value of T means that petroleum supply has

gone up. So its current price p_o must decrease. Similarly, a reduction in the cost of production of solar power, a perfect substitute of petroleum, also decreases p_o .

Appendix 6.1

INVESTMENT FUNCTION

Investment as a Function of the Rate of Interest Slopes Downward to the Right.

The money amount of investment V is given by the product of the number of units I of the investment good and its price p_I . Now, I is a function of p_I and p_I depends on the net rental N and the interest rate r . Since N is a constant, we can write V as

$$V(r) = p_I(r)I[p_I(r)].$$

Then
$$V'(r) = p'_I(r)I[p_I(r)] + p_I(r) I'[p_I(r)] \cdot p'_I(r).$$

Both $I[p_I(r)]$ and $p_I(r)$ are positive. Since I slopes upward to the right as a function of p_I , $I'[p_I(r)]$ is positive. Therefore, given that $p'_I(r)$ is negative (since variations in r produce variations in the opposite directions in p_I), $V'(r)$ is negative. This demonstrates rigorously that $V(r)$ is decreasing in r .

EXERCISES

Multiple Choice

1. A consumer's choices about the timing of consumption reflect his
 - a) Tastes at different times.
 - b) Current and expected future incomes and the interest rate.
 - c) The current and expected future prices.
 - d) All of the above.
2. The absolute value of the slope of the intertemporal budget line for a consumer in a two-period horizon is given by

- a) The ratio between period 1 and period 2 prices.
 - b) The real cost of giving up consumption in period 2 at the expense of increasing consumption in the first period by one unit.
 - c) The interest rate.
 - d) Proportional rate of change of the price of the consumption good between the two periods.
3. A consumer's marginal rate of time preference between consumptions in period 1 and period 2 indicates
- a) His willingness to save in period 2 if the interest rate is high.
 - b) His willingness to borrow in period 1 if the interest rate is low.
 - c) His willingness to sacrifice consumption in period 2 to increase consumption in period 1.
 - d) His willingness to maximize saving in period 1 as a function of the interest rate.
4. Suppose that the market rate of interest is 10 percent. Then the present value for an asset that will earn Rs. 100 per year for two years and then Rs. 50 per year indefinitely into the future is
- a) Rs. 586.78.
 - b) Rs. 568.78.
 - c) Rs. 658.78.
 - d) Rs. 685.78.
5. Suppose that the market rate of interest is r . Then the amount that must be invested if one wishes to obtain Rs. N per year forever is
- a) Rs. N/r .
 - b) Rs. Nr .
 - c) Rs. $N(1 + r)$.
 - d) Rs. $N\left(1 - \frac{1}{r}\right)$.
6. Suppose that the present value of a consumer's endowment increases. Then the consumer's intertemporal budget set
- a) Expands.
 - b) Contracts.
 - c) Remains unaltered.
 - d) Any one of the above is possible.

Quantitative Problems and Questions

1. An individual has a fixed amount of money I to allocate between consumption in two periods C_1 and C_2 . The individual's utility function is given by $U(C_1, C_2) = C_1 C_2$, and the budget constraint is $I = C_1 + C_2/(1 + r)$, where r is the one period interest rate. (Both period 1 and period 2 prices are assumed to be 1.)
 - a) Show that in order to maximize utility, given his budget constraint, the individual should choose C_1 and C_2 so that the marginal rate of time preference (of C_1 for C_2) is $(1 + r)$.
 - b) Show that consumption in period 2 increases as r increases.
2. Suppose an individual has an income of Rs. 100 in each period and his utility function is $U(C_1, C_2) = \log C_1 + \log C_2$, where C_1 and C_2 are consumptions in period 1 and 2 respectively. The one-period interest rate is 10 percent. Both period 1 and period 2 prices are 1.
 - a) Determine the values of C_1 and C_2 .
 - b) How much will he borrow/lend in the first period and repay/be repaid in period 2?
 - c) Determine the present values of consumption and income.
3. A person faces two income streams A and B which generate incomes according to the following schedules:

<i>Stream</i>	<i>Incomes</i>	
	<i>Current period</i>	<i>Future period</i>
A	10	15
B	12	13

Which stream will be adopted if the one-period interest rate is (i) 10 percent, (ii) 20 percent?

4. Suppose you have decided to spend Rs. 100,000 for buying a house. Since you could be earning 10 percent on that money in a bank, you will lose Rs. 30,000 in interest over the next 3 years. On the other hand, if you take a house building loan of Rs. 100,000, you will pay Rs. 3,500 per month for 3 years. Overall then you pay back Rs. 126,000. Which is the better option for you?
5. Prove or disprove the following.
 - a) As the interest rate rises, the intertemporal budget constraint becomes flatter.

- b) Saving as a function of the rate of interest is monotonically increasing.
 - c) For a consumer in a two-period horizon who lends, an increase in the interest rate increases or decreases saving according as the substitution effect or the income effect dominates.
 - d) Given a perfect lending market, the sale price of an asset is equal to the discounted present value of the asset's income stream.
 - e) When a consumer wishes to have an average amount of consumption in each period, rather than having a lot in one period and nothing in the other period, his intertemporal preferences must satisfy the diminishing marginal rate of time preference condition.
6. In what sense can education be regarded as an asset? What is the opportunity cost of money spent on education? Under what conditions is it worth foregoing present consumption to invest in education?
 7. How does the interest rate affect demand for future consumption? Why is the opportunity cost of current consumption reflected in the rate of interest?
 8. Suppose a person has Rs. 200 to invest. There are two options available to him: (i) investing at 20.6 percent interest compounded annually, (ii) investing at 20 percent interest compounded quarterly. Which is the better option if the person decides to invest for one year?
 9. Suppose that the market rate of interest is 10 percent. A three-period bond has a face value of Rs. 2,000. The bond pays Rs. 200 interest per year at the end of each of the three years. What is the discounted present value of the bond?
 10. An individual has spent Rs. 1,000 for buying a machine which will last two years. The person will get an annual return of Rs. 1,200 from the machine. The operating cost of the machine is Rs. 600 per year. What is the internal rate of return on the purchase of the machine? (The internal rate of return on an investment is the largest number, which when used as a discount rate, makes the discounted present value of incomes from the investment equal to its cost. Recall that in the case of a bond we have referred to this rate of return as the yield on the bond. If we assume a perfect lending market, then the internal rate of return equals the market rate of interest.)
 11. Analyze graphically the impact of a change in the rate of inflation on the aggregate amount of saving.

12. An asset is currently earning Rs. 20 per year. Suppose earnings are expected to increase at a 5 percent rate indefinitely into the future. If the market rate of interest is 15 percent, what is the discounted present value?
13. An individual expects to work for 30 years. His income rises at a rate of 4 percent per year and the market rate of interest is also 4 percent. What constant fraction of the income must the individual save in each working year to finance a level of retirement income equal to 60 percent of earnings in the year just prior to retirement?
14. Discuss the issues that affect a consumer's demand for durable goods.

7

Uncertainty

In previous chapters it has been implicitly assumed that decision making by different individuals takes place in environments characterized by certainty. In such environments when individuals purchase goods they know how much utility will be yielded by the goods. Similarly, in such a world a profit making decision by a firm will give exact information about the firm's profit.

In many real world situations the certainty assumption cannot be regarded as appropriate. Uncertainty plays an important role in a variety of economic activities. First, for some goods buying can take the form of games or lotteries for which the outcome is not known with certainty: stock market transactions and insurance purchase fit well into this category. A second way in which an individual can be confronted with uncertainty is through lack of requisite information to convert uncertainty into certainty. We might not possess the scientific knowhow to obtain the correct information, or the market for information as a commodity is not able to provide the correct information. Since information is often costly to collect, uncertainty may persist in some situations if the value of information is less than the cost of obtaining it. In such cases, certainty exists from deliberate choices by decision makers. A final way in which uncertainty affects individual behavior is

in dealings with others. Many encounters between individuals are in the form of adversary proceedings in which an individual's reward will depend on what others do. This type of uncertainty arises in many situations ranging from poker games to the conduct of diplomacy.

Of these three types of uncertainty the first will occupy the major portion of this chapter. We will discuss the remaining ones, particularly, the one concerning information, in Chapter 25.

UNCERTAINTY AND PROBABILITY

The study of individual behavior under uncertainty and that of probability have a common origin in attempts to understand games of chance. For instance, the study of coin tossing games has been quite productive in the mathematics that has developed from these games and in illuminating certain characteristics of human behavior they exhibit.

The probability of an event happening is, loosely speaking, the relative frequency with which the event occurs. For instance, when a fair coin is tossed the probability of occurrence of a head being half means that if a fair coin is tossed a large number of times, a head would appear in approximately one-half of the trials.

Suppose that a lottery offers n prizes $X_1, X_2, \dots, X_n$, (some of which may be zero), and that the probabilities of winning these prizes are $p_1, p_2, \dots, p_n$. These probabilities are non-negative numbers between zero and one. We say that events $X_1, \dots, X_n$ are mutually exclusive if the occurrence of one of them prevents the occurrence of any of the others; they are collectively exclusive if one of the outcomes *has* to occur. For mutually exclusive and collectively exhaustive events we have

$$\sum_{i=1}^n p_i = 1 \quad (7.1)$$

For a lottery (X) with prizes $X_1, X_2, \dots, X_n$ and probabilities of winning $p_1, p_2, \dots, p_n$, the expected value $E(X)$ of the lottery is

$$E(X) = \sum_{i=1}^n p_i X_i. \quad (7.2)$$

The expected value of the lottery is a weighted average of the prizes, where the weights are the respective probabilities. More precisely, it is the size of the prize that the player will win, on the average.

In economic worlds characterized by uncertainty, a consumer gets concerned with the probability distribution of alternative consumption bundles of goods. A probability distribution here gives a list of different consumption bundles and the probability of each outcome. When a consumer has decided how much fire insurance to buy or how much to invest in the share market, he has in effect taken a decision concerning the probability distributions across different amounts of consumptions.

We can refer to different outcomes of some random event as being different states of nature. Thus, in the lottery example considered above there are n states of nature. The prizes $X_1, \dots, X_n$ can be regarded as contingent consumption plans, since they specify what will be consumed in each different state of nature. We use the word contingent to mean that a consumption plan depends on something not yet certain; more precisely, it depends on the outcome of some event.

A consumer has preferences over different plans of consumption, just like he has preferences over actual consumptions. If the consumer has preferences about consumptions in different circumstances, we can use a utility function to describe these preferences. Strictly speaking, given that a consumer's beliefs about uncertainty can be formulated in terms of probabilities, the problem of decision making under uncertainty can be regarded as one of choosing from among alternative probability distributions over pay-offs. The approach to this problem followed in economic literature is referred to as the 'measurable utility' approach.

MEASURABLE UTILITY

In the context of decision making under uncertainty, two dimensions of the problem need to be considered: first, the decision maker's preferences with respect to possible pay-offs from his decisions, and; second, his attitude towards risk. The notion of measurable utility captures both dimensions of the decision making problem in a single concept.

The concept of measurable utility, in the context of uncertainty, was introduced by von Neumann and Morgenstern (1944) in connection with their pioneering work on game theory, for axiomatizing the notion of rational behavior under uncertainty. Their axioms represent an attempt to generalize

some of the foundations of the theory of individual choice to cover uncertain situations. Therefore, utility numbers are assigned to alternative probability distributions in a manner such that individual's preference ranking over such distributions are preserved. Thus, the rational rule of behavior turns out to be the problem of choosing the probability distribution that has the highest measurable utility associated with it.

We will now illustrate the way in which measurable utility numbers are determined in a simple set-up characterized by uncertainty. Assume that there are n possible prizes that a consumer might win by participating in a game. Denote these prizes by $X_1, X_2, \dots, X_n$, and assume without loss of generality that they have been arranged so that $X_1 < X_2 < \dots < X_n$. Consequently, X_1 is the least preferred prize and X_n is the most preferred prize for the individual. For concreteness, let us assign the utility number 0 to X_1 and a utility of 1 to X_n , that is,

$$U(X_1) = 0, \quad (7.3)$$

$$U(X_n) = 1. \quad (7.4)$$

The point of the von Neumann-Morgenstern theorem is that using these two values of utility it should be possible to show that a reasonable way exists to assign specific utility numbers to the other prizes available. Suppose that we choose some other prize, say X_i . We then ask the individual with what probability, say p_i , he will be indifferent between receiving X_i with certainty and participating in the gambling offering prizes of X_n with probability p_i and X_1 with probability $(1 - p_i)$. An individual will always be indifferent between a gamble and a sure thing, if in the gamble a sufficiently high chance of winning the best prize is offered. Therefore, it is reasonable to assume that p_i exists. Note that p_i also represents the desirability of the prize X_i , since the better X_i is the better a chance of winning X_n must be given in order to get the individual involved in gambling. In fact, according to the von Neumann-Morgenstern formulation the number p_i becomes the measurable utility associated with the money pay-off X_i . The von Neumann-Morgenstern technique is to define the utility of X_i as $p_i U(X_n) + (1 - p_i) U(X_1)$, the expected value of utility of the pay-offs from the gamble, that is, the weighted average of the utility numbers associated with X_n and X_1 , where the weights are the probabilities p_i and $(1 - p_i)$. This expected value is considered to be equally desirable to X_i . More precisely,

$$U(X_i) = p_i U(X_n) + (1 - p_i) U(X_1). \quad (7.5)$$

Because of our choice of scale in Eqs. (7.3) and (7.4), we have

$$U(X_i) = p_i \cdot 1 + (1 - p_i) \cdot 0 = p_i. \quad (7.6)$$

Thus, by sensibly choosing the utility numbers to the worst and best prizes we show that the utility number assigned to any other prize is simply the probability of indifference defined above. It should be clear that the choice of utility numbers is quite arbitrary. We can choose any two numbers in Eqs. (7.3) and (7.4), and hence $U(X_i)$ gets redefined.

EXPECTED UTILITY MAXIMIZATION

In view of the above discussion we can now say that it is possible to assign utility numbers to every prize. Suppose that some probability p_i has been assigned to represent the utility of X_i . We have $p_1 = 0$, $p_n = 1$ and other utility values lie between these two extremes. Using these utility numbers we can show that a 'rational' individual's basis of choice among probability distributions/gambles is the expected utility, that is, the expected value of these von Neumann-Morgenstern numbers.

To see this, consider two games. The first game offers X_2 with probability q and X_3 with probability $(1 - q)$. The second one offers X_4 with probability r and X_5 with probability $(1 - r)$. We need to show that the individual will choose game i if and only if the expected utility of i exceeds that of j ($i, j = 1, 2, i \neq j$). Now, for the games:

$$E_1 = \text{expected utility of game 1} = qU(X_2) + (1 - q)U(X_3) \quad (7.7)$$

$$E_2 = \text{expected utility of game 2} = rU(X_4) + (1 - r)U(X_5) \quad (7.8)$$

Substituting p_i for $U(X_i)$, $i = 2, 3, 4, 5$, we get

$$E_1 = qp_2 + (1 - q)p_3 \quad (7.9)$$

$$E_2 = rp_4 + (1 - r)p_5. \quad (7.10)$$

We need to show that the individual will prefer game 1 to game 2 if and only if

$$E_1 > E_2. \quad (7.11)$$

To demonstrate this recall the definition of the utility numbers. The individual is indifferent between X_i with certainty, and a game consisting of X_n with probability p_i and X_1 with probability $(1 - p_i)$, where $i = 2, 3, 4, 5$. Consequently,

$$\begin{aligned}
E_1 &= q[p_2 U(X_n) + (1 - p_2) U(X_1)] + (1 - q) [p_3 U(X_n) + (1 - p_3) U(X_1)] \\
&= [qp_2 + (1 - q)p_3] U(X_n) + [q(1 - p_2) + (1 - q)(1 - p_3)] U(X_1), \\
E_2 &= r[p_4 U(X_n) + (1 - p_4) U(X_1)] + (1 - r) [p_5 U(X_n) + (1 - p_5) U(X_1)] \\
&= [rp_4 + (1 - r)p_5] U(X_n) + [r(1 - p_4) + (1 - r)(1 - p_5)] U(X_1).
\end{aligned}$$

From the above expressions it turns out that game 1 is equivalent to a game promising X_n with probability $qp_2 + (1 - q)p_3$ and X_1 with probability $q(1 - p_2) + (1 - q)(1 - p_3)$. Similarly game 2 coincides with a game promising X_n with probability $rp_4 + (1 - r)p_5$ and X_1 with probability $r(1 - p_4) + (1 - r)(1 - p_5)$. Clearly, a rational consumer will prefer the game that has a higher probability of winning the better prize. Consequently, he will opt for game 1 if and only if

$$E_1 = qp_2 + (1 - q)p_3 > rp_4 + (1 - r)p_5 = E_2,$$

which is what we wanted to show. Thus, we have proved that under uncertainty an individual will choose among the options that one which provides the highest level of expected (or von Neumann-Morgenstern) utility.

We conclude this section by making some important observations. First, the entire idea of the expected utility maximization hypothesis has been presented in terms of games. Since any situation involving choices under uncertainty can be regarded as one in which each alternative under consideration is in essence a list of payments together with their probabilities, we can apply the notion of measurable utility to any choice under uncertainty. Thus, we can regard $X_1, X_2, \dots, X_n$ as consumptions in different states of nature, and the corresponding probabilities $(p_1, p_2, \dots, p_n)$ as the probabilities of materialization of these states. Then in analogy with Eq. (7.9), the expected utility function takes the form

$$v(X_1, X_2, \dots, X_n; p_1, p_2, \dots, p_n) = \sum_{i=1}^n p_i U(X_i). \quad (7.12)$$

When we say that under uncertainty an individual's preferences can be represented by an expected utility function, we mean that we can choose a utility function that can be written in the form described in Eq. (7.12). Any monotonically increasing transformation of the expected utility function will represent the same preferences, but the expected utility property may not be preserved under an increasing transformation. For instance, the utility functions $\sqrt{(p_i X_i^{1/2} + p_j X_j^{1/2})}$ and $(p_i X_i^{1/2} + p_j X_j^{1/2})$ with the states of nature

i and j represent the same preferences, but while the latter possesses the expected utility property, the former does not. In fact, an expected utility property remains preserved under an affine transformation only. A transformation $f(v)$ is said to be affine if it is of the form $f(v) = a + bv$, where $b > 0$. An affine transformation not only preserves the expected utility property, it also represents the same preferences (since it is monotonically increasing). This is why it is often told in the literature that an expected utility function is unique only up to an affine transformation. Technically, a von Neumann-Morgenstern utility function is unique only up to a choice of scale and origin. (In the transformation $a + bv$, 'a' allows to choose the origin and 'b' enables to choose the scale.) An analogy with temperature measurement will be useful. The important thing to notice here is that the two scales of temperature measurement (Celsius and Fahrenheit) are related by the affine transform, degrees Fahrenheit = 32 + (9/5) degrees Celsius.

Finally, note that in choices under uncertainty, there is an implicit assumption of independence between outcomes in different states of nature. That is, the choice that people wish to make in one state is independent of the choices they plan to make in other states. In fact, this assumption ensures that the utility function is additive across the different contingent consumption bundles, that is, it is of the form of Eq. (7.12). This in turn implies that the marginal rate of substitution between any two goods is independent of the consumption of any other good. For instance, from (7.12) we note that the marginal rate of substitution between goods i and j takes the form

$$MRS_{ij} = \frac{p_i \Delta U(X_i) / \Delta X_i}{p_j \Delta U(X_j) / \Delta X_j} \quad (7.13)$$

which does not depend on the consumption of good $k \neq i, j$.

RISK AVERSION

Two games may have the same expected monetary pay-off but they may differ with respect to riskiness. To explain this, suppose a person participates in a coin tossing game in which he wins or loses (pays) Re. 1 according as a head or a tail appears. The expected pay-off of this game is zero. Consider another coin flipping game where the person wins or pays Rs. 5,000 depending on whether a head or a tail comes up. This game also has an expected value of zero. However, the latter is in some sense more risky than the former and fewer people would participate in the game in which there is a 50-50 chance

of winning or losing Rs. 5000. In this section we will explain the term 'risky', and hence discuss people's attitude towards risk.

The term risk refers to the variability of the outcomes of an uncertain activity. (A standard measure of variability is the standard deviation which we explain later in this chapter.) If the variability is higher, the outcome of the uncertain activity must be regarded as risky. On the other hand, if the variability is low, the outcome may be considered as a sure thing. Therefore it is reasonable to expect that when an individual faces a choice between two games/gambles with the same expected value, he will usually choose the one with the lower variability of outcome. The intuitive reasoning behind this is as follows. The marginal utility from an extra unit of prize money declines as prizes increase. A coin tossing game for a prize of Rs. 5,000 promises a relatively low utility gain if the person wins, but the loss of utility is large if he loses. In contrast, for the other game the gain in utility from a win is almost counterbalanced by the decline in utility from a loss.

To represent the above idea graphically, let W_c be an individual's current level of wealth. Suppose that the individual is offered two fair gambles: a 50-50 chance of winning or losing an amount M , or a 50-50 chance of winning or losing an amount $2M$. The expected utilities from these gambles are given respectively by

$$U^1(W_c) = \frac{1}{2} [U(W_c + M) + U(W_c - M)], \quad (7.14)$$

$$U^2(W_c) = \frac{1}{2} [U(W_c + 2M) + U(W_c - 2M)], \quad (7.15)$$

where U is the von Neumann-Morgenstern utility function. In the figure below we draw U as a strictly concave function of its argument to reflect the assumption of diminishing marginal utility. From the figure it follows that

$$U(W_c) > U^1(W_c) > U^2(W_c) \quad (7.16)$$

Note that we can interpret $U(W_c)$ as the utility of the expected wealth: $U(W_c) = U((W_c - M)/2 + (W_c + M)/2) = U[(W_c - 2M)/2 + (W_c + 2M)/2]$. Thus, in the figure the *expected utility* of wealth is less than the utility of the *expected wealth*. In this case we say that the consumer is risk averse, since he prefers the expected value of his wealth to the outcomes of the gambles. Further, of the two gambles the expected utility is lower for the one with higher variability.

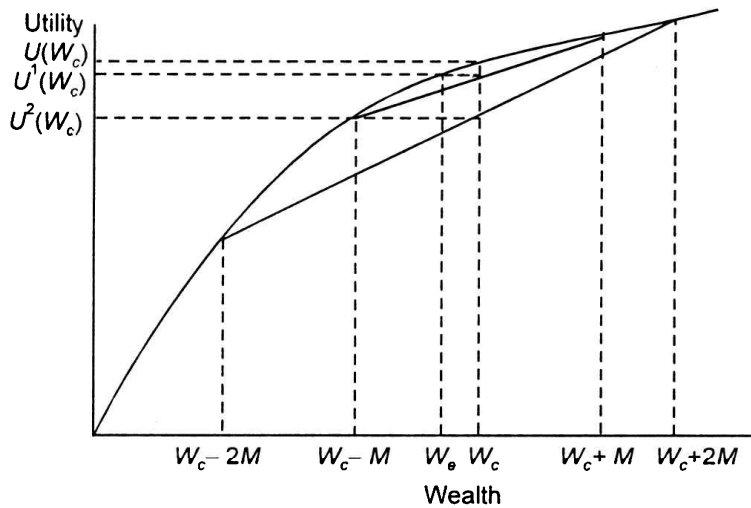


Fig. 7.1 Risk Aversion

It may happen that a consumer prefers a random distribution of wealth to its expected value. In this case we say that the consumer is a risk lover. A risk loving individual has a strictly convex utility function. For such a utility function the slope becomes steeper as wealth increases. In contrast, the slope of strictly concave utility function gets flatter with increase in wealth. Therefore, the curvature of the utility function, that is, the slope of the marginal utility function, measures the consumer's attitude towards risk. In general, the more concave (convex) the utility function is, the more risk averse (loving) the individual will be. The intermediate situation between these two cases is risk neutrality. In this case, the expected utility of wealth is the same as the utility of its expected value. Thus, the consumer does not care about the risks involved in gambles: the only consideration is their expected values. This situation arises when the consumer has a linear utility function.

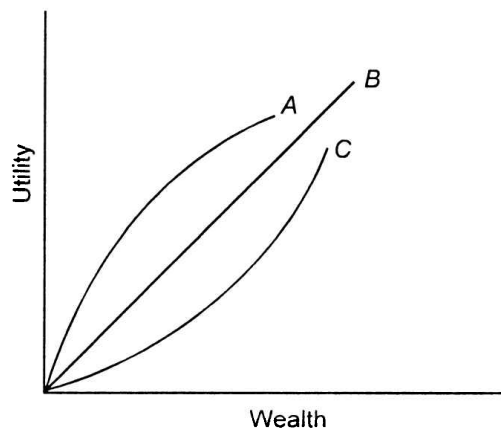


Fig. 7.2 Attitudes Towards Risk

In Fig. 7.2, curve *A* indicates a risk averse attitude, curve *B* shows risk neutrality and curve *C* represents a risk loving mentality. In the mathematical appendix to this chapter (Appendix 7.1) we examine a theoretical measure of risk aversion and some of its practical applications.

RISK AVERSION AND INSURANCE

A risk averse individual will prefer his current wealth W_c to W_c combined with a fair gamble, and will prefer a gamble with lower variability to one with a higher variability. Although the expected values of wealth are equal in money terms, in utility terms the loss is more serious. In fact, an individual may be willing to pay some amount to avoid participating in a gamble. Note that in Fig. 7.1 the certain wealth W_e provides the same utility as does participating in gamble 1. We refer to W_e as the certainty equivalent level of wealth associated with gamble 1, since this amount of wealth when enjoyed by the individual with certainty generates the same level of utility as gamble 1 does. The individual may be willing to pay anything up to $W_c - W_e$ to avoid participating in the gamble. This therefore explains why people buy insurance: they are willing to give up a certain amount (the premium) to avoid the risky outcome that is insured against.

We will explain this by an example. The idea of the example has been taken from Varian (1993). Suppose that an individual's current level of wealth is W_c , but there is a possibility that he may lose $W_l < W_c$. For instance, a fire may damage his house, or his car may be stolen. Let the probability of occurrence of the loss be p . Then $(1 - p)$ is the probability that the loss will not occur. Suppose that if the person purchases an insurance of worth W_m , then he has to pay a premium rW_m , $0 < r < 1$. Note that the premium has to be paid whether or not the loss occurs. Then the individual faces the gamble:

probability p of getting $W_c - W_l + W_m - rW_m$

and probability $(1 - p)$ of getting $W_c - rW_m$.

The individual's endowment of contingent consumption is $W_c - W_l$ if the loss occurs (state 2), and W_c if the loss does not occur (state 1). If he purchases an insurance of W_m , he gives up rW_m in state 1 in exchange for $W_m - rW_m$ in state 2. Therefore, the consumption he loses in state 1, divided by the additional consumption he gains in state 2 is

$$-rW_m/(W_m - rW_m) = -r/(1 - r). \quad (7.17)$$

This is the slope of the budget line (Fig. 7.3) through his endowment. It shows that r and $(1 - r)$ can be regarded as the prices of consumption in states 1 and 2 respectively.

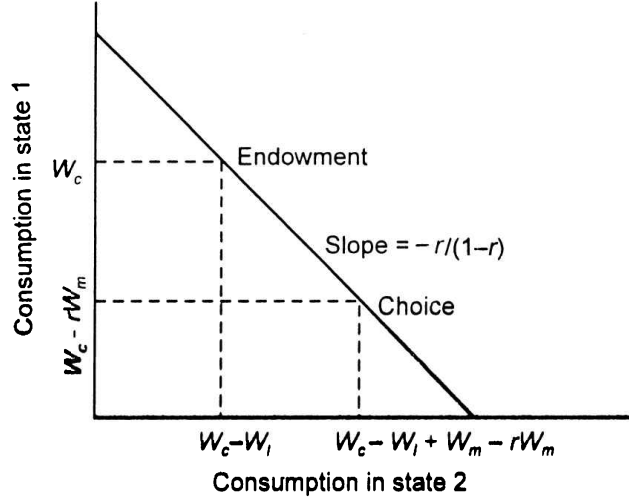


Fig. 7.3 Budget Line Associated with Demand for Insurance

Denoting the individual's wealth in states 1 and 2 by $W_1 (= W_c - rW_m)$ and $W_2 (= W_c - W_l + W_m - rW_m)$ respectively, his optimal choice of insurance is determined by the condition that the MRS between consumptions in the two outcomes equals the price ratio:

$$MRS = \frac{p \Delta U(W_2) / \Delta W_2}{(1 - p) \Delta U(W_1) / \Delta W_1} = \frac{r}{1 - r} \quad (7.18)$$

Now, let us assume that the insurance market is 'fair'. By a 'fair market' for contingent commodities, we simply mean that the ratio of prices of the contingent commodities is the same as the ratio of probabilities of the states of the world. That is,

$$\frac{r}{1 - r} = \frac{p}{1 - p} \quad (7.19)$$

which in turn implies that $r = p$. Eq. (7.19) is one way of stating the assumption that buying contingent commodities is a fair game. This may be true if there are a large number of buyers and sellers and the probabilities are known to all. Substituting $r = p$ in Eq. (7.18) we can rewrite the optimality condition as

$$\frac{\Delta U(W_2)}{\Delta W_2} = \frac{\Delta U(W_1)}{\Delta W_1}, \quad (7.20)$$

which means that the marginal utility of an extra unit of income if the loss does not occur is the same as the marginal utility of an extra unit of income if the loss occurs.

Now, a risk averse person's marginal utility of income decreases as income increases. Therefore, the equality Eq. (7.20) will hold only if $W_1 = W_2$, which we rewrite as

$$W_c - rW_m = W_c - W_l + W_m - rW_m. \quad (7.21)$$

From Eq. (7.21) it now follows that $W_m = W_l$. This shows that if a risk averse consumer is offered a fair insurance against a loss, then his optimal choice will be to fully insure the loss. The intuitive reasoning behind this is that the risk averse individual is willing to give up something in the form of expected income in order to reduce the variability of income.

Reduction in variability of income can be achieved through some alternative ways also. For instance, by diversifying investment in two or more companies a person may be able to reduce the overall risk of investment, while keeping the expected income constant. A person may be able to spread his risk over other consumers and reduce the amount of his risk bearing. For instance, suppose n persons decide to insure one another. If anybody incurs a loss of W_l , each of the n consumers will contribute W_l/n to the loser. Thus, in this case we have a cooperative of consumers. The cooperative may collect some subscriptions from its members annually to build up a cash reserve fund that can be used for premium payment. It may also invest its cash reserve and earn interest on its assets. Investment in the stock market may also enable a person to spread his risk. By issuing shares of a company, the original owners can spread their risk over a large number of share holders.

So far we have described the expected utility approach to choices under uncertainty. Two alternative approaches to the problem are to describe the underlying probability distributions through stochastic dominance, and mean-variance analysis. To explain these we need some more discussions on probability distributions.

MORE ON PROBABILITY DISTRIBUTIONS

Suppose that a given variable X has outcomes $X_1, X_2, \dots, X_n$ with probabilities $p_1, p_2, \dots, p_n$. We can associate with this distribution its average or expected value $\mu(X)$ (also called the mean of the distribution), defined by

$$\mu(X) = \sum_{i=1}^n X_i p_i. \quad (7.22)$$

A measure of the dispersion of outcomes about the mean is the variance, $\sigma^2(X)$, defined by

$$\sigma^2(X) = \mu[X - \mu(X)]^2 = \sum_{i=1}^n p_i [(X_i - \mu(X))]^2. \quad (7.23)$$

The variance is therefore the weighted average of squared deviation of outcomes from the mean, where the weights are the probabilities of the outcomes. As evident from the definition, the variance measures the spread of the distribution and therefore can be regarded as a measure of riskiness involved. Another related measure of dispersion is the standard deviation, $\sigma(X)$, defined as the positive square root of the variance.

It is sometimes convenient to plot the probability distribution over outcomes as a graph. When there is a continuous variation in outcomes, the probability distribution becomes a smooth curve. An increase in the variance (standard deviation) will spread the distribution out, and an increase in the average value moves the distribution to the right.

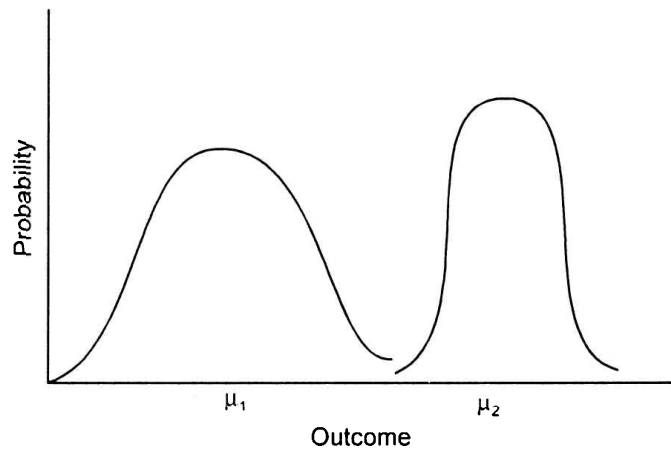


Fig. 7.4 Mean and Variance

The probability distribution depicted in panel (a) of Fig. 7.4 has a higher variance but a lower mean than that depicted in panel (b).

The stochastic dominance approach to uncertainty relies on the cumulative probability distribution. A cumulative probability distribution gives us for any income, the probability that the individual will receive the amount of income or less. We determine the cumulative probability at any income level by summing up the probabilities associated with incomes less than or equal to the given level of income. To illustrate this, assume that three coins

are independently tossed and an individual is paid Re. 1 for each head that appears. Events, pay-offs and the probabilities of pay-offs are given in the following table.

Table 7.1 Probability Distribution over Pay-Offs

<i>Event</i>	<i>Pay-off X</i>	<i>Probability</i>
1. No head appears	0	1/8
2. One head appears	1	3/8
3. Two heads appear	2	3/8
4. Three heads appear	3	1/8

For this example, $\mu(X) = 1.50$, $\sigma(X) = \sqrt{3/2}$. If we denote the cumulative probability distribution by F , then $F(0) = 1/8$, $F(1) = 4/8$, $F(2) = 7/8$ and $F(3) = 1$.

STOCHASTIC DOMINANCE AND DECISIONS UNDER UNCERTAINTY

In Fig. 7.5 we plot two cumulative probability distributions over income. The curve labelled 1 has the property that the cumulative probability associated with every level of income is less than or equal to that for curve 2, with strict inequality for at least one income level. When this occurs we say that the probability distribution 1 first order stochastically dominates the probability distribution 2. From the figure it is clear that for any income x , the probability of getting more than x is higher (or not smaller) under 1 than it is under 2. Therefore, if an individual prefers more money to less and his preferences

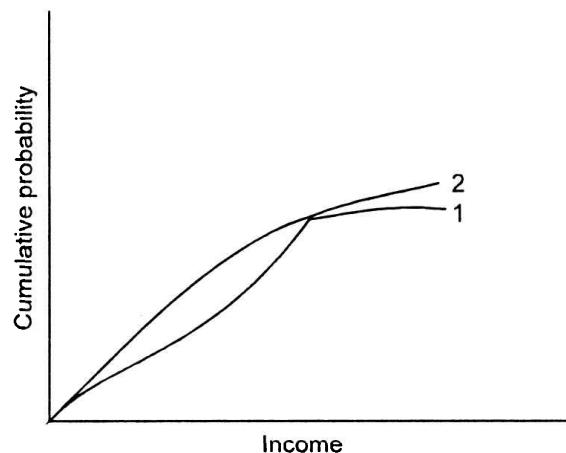


Fig. 7.5 Stochastic Dominance and Cumulative Probability Distributions

follow the measurable utility approach, then he will opt for probability distribution 1.

In fact it has been shown that one probability distribution is preferred to another probability distribution by all individuals who prefer more money with certainty to less, if and only if the former distribution first order stochastically dominates the latter. This gives us a general statement about preferences among probability distributions without making any assumption regarding an individual's attitudes towards risk. The only assumption involved is that individuals prefer more money to less, with certainty.

MEAN VARIANCE ANALYSIS

In the stochastic dominance approach to uncertainty an individual's preferences depend on the entire probability distribution of his wealth over different possible outcomes. An alternative possibility is to suppose that his preferences can be described by some summary statistics about the probability distribution of his wealth. One reasonable formulation along this line is to assume that the individual's utility v can be written as a function of the mean and standard deviation of the wealth distribution, $v[\mu(W), \sigma(W)]$. Under this assumption a utility function will be able to rank distributions in the same way an expected utility function will rank them. A natural supposition here is that given other things, a higher mean is better and that a higher standard deviation or variance is worse.

We will use the mean-variance model to analyze a simple portfolio problem. The analysis here follows that introduced by Tobin (1958) and Varian (1993). An individual has decided to invest his wealth in two different assets. One of them is a risk-free asset F which generates a fixed rate of return $R(F)$. This can be something like earning a fixed rate of interest on some amount of money deposited in a bank. The other asset is a risky asset B . We can think of this asset as being an investment in a share market. The return on this risky asset will depend on the performance of the share market. The key decision faced by the investor is the division of his wealth W between the risk-free asset F and the risky asset B . Let B_j be the return on the risky asset if situation j occurs, $j = 1, 2, \dots, n$. (A situation describes the performance of the share market, for instance, good, moderate, bad, etc.) Suppose that the probability of occurring situation j is p_j .

We denote the proportions of wealth held as risk-free and risky assets by A_1 and A_2 respectively, ($A_1 = F/W$, $A_2 = B/W$). Then the expected return on the investor's portfolio is given by

$$\mu(W) = \sum_{j=1}^n [A_1 R(F) + A_2 B_j] p_j \quad (7.24)$$

$$= A_1 R(F) \sum_{j=1}^n p_j + A_2 \sum_{j=1}^n B_j p_j \quad (7.25)$$

$$= A_1 R(F) + A_2 \mu(B), \quad (7.26)$$

where $\mu(B) = \sum_{j=1}^n B_j p_j$ is the expected return of the risky asset and the first term of Eq. (7.26) follows from the fact that $\sum_{j=1}^n p_j = 1$. Thus, the expected return on the portfolio is a weighted average of returns from two types of assets.

The risk associated with the portfolio is expressed in terms of the variance of W :

$$\sigma^2(W) = \sum_{j=1}^n [B_j A_2 + A_1 R(F) - \mu(W)]^2 p_j, \quad (7.27)$$

which on simplification becomes

$$\sigma^2(W) = A_2^2 \sigma^2(B), \quad (7.28)$$

where $\sigma^2(B) = \sum_{j=1}^n p_j [B_j - \mu(B)]^2$ is the variance of the return of the risky asset. If the investor believes that holding risky assets involves the likelihood of large gains or losses, $\sigma^2(B)$ (hence $\sigma(B)$) will be large. Alternatively, a low standard deviation $\sigma(B)$ implies that variations in return around its mean value are expected to fall within a narrow range and the associated risk is correspondingly low.

Observe that the mean return and the risk attached to the portfolio depend on the proportion of wealth held as risky assets. For $A_1 = 0$, the vector of expected return and standard deviation of the portfolio turns out to be $[\mu(B), \sigma(B)]$. If $A_2 = 0$, the corresponding vector becomes $[R(F), 0]$. For $0 < A_1 < 1$, the resulting vector will be somewhere in the middle of the line joining these two extreme points. This line is the investor's constraint or budget line describing the trade-off between risk and return. That is, this line defines the terms on which a greater expected return can be obtained through assuming more risk.

To explain this note that since a risk averse individual would never hold a risky asset if it had a smaller expected return than the risk free asset, it is reasonable to assume that $R(F) < \mu(B)$. This in turn implies that if an investor decides to devote a higher fraction of his wealth to the risky asset (A_2 increases and A_1 decreases), then his expected return $\mu(W)$ will increase, but he will also incur higher risk $\sigma(W) = A_2\sigma(B)$. We represent this in Fig. 7.6. From the figure, by inspection the slope of the constraint line turns out to be $[\mu(B) - R(F)]/\sigma(B)$.

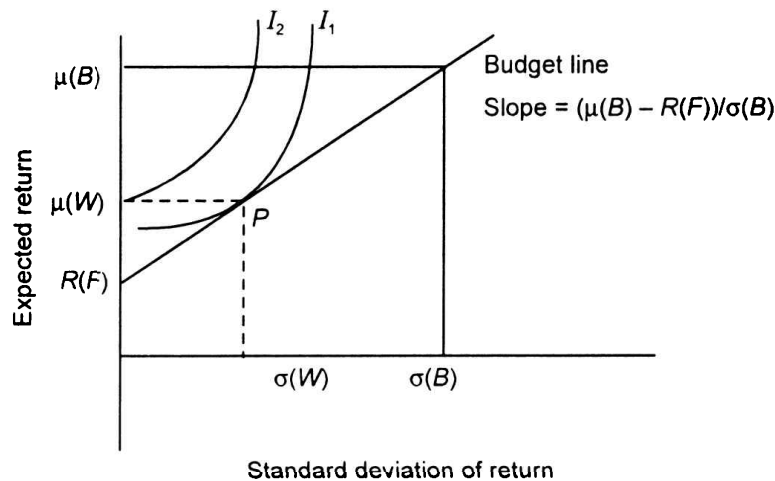


Fig. 7.6 Portfolio Choice and Uncertainty

To establish a demand for money to be invested in the risky asset, it is necessary to make specific assumptions about the investor's preferences (that is, the shape of the indifference curves). Now, for a risk averse individual, that is, an individual for whom expected return is a desirable portfolio characteristic and risk is an undesirable characteristic, more risk will be acceptable only if he is compensated by greater expected return. Therefore, the indifference curves will be positively sloped. We can say that expected return is a good commodity and standard deviation is a bad commodity here. Next, if the investor is assumed to be subject to increasing risk aversion (that is, additional units of risk will be accepted only if the investor is provided with larger and larger increments of return), the indifference curves will be strictly concave upward. Indifference curves for an investor with these characteristics are illustrated in Fig. 7.6. This individual makes no distinction between all combinations of expected return and risk that lie on a given indifference curve; all return-risk pairs on I_2 are preferred to those on I_1 .

Given the trade-off between expected return and risk, the constraint line and the investor's subjective preferences between risk and return, the utility maximizing individual adjusts the amount of investment in a risky asset

relative to that in a risk free asset in order to reach the highest attainable indifference curve. This will occur at P , where there is a tangency between an indifference curve and the constraint line. The utility maximizing return-risk combination requires that the proportion of portfolio wealth to be held as a risky asset be $A_2 = \frac{\sigma(W)}{\sigma(B)} < 1$, hence some wealth is held in the form of a risk-free asset ($0 < A_1 < 1$). A utility maximizing outcome of this type involves portfolio diversification (as long as $A_2 < 1$), so that the investor holds risky and risk free assets simultaneously.

Since the slope of the constraint line measures how risk and return can be adjusted in making portfolio choices, we can call it the price of risk. Now, for a risk averse person the utility function $v[\mu(W), \sigma(W)]$ is increasing in $\mu(W)$ and decreasing in $\sigma(W)$. The marginal rate of substitution between risk and return then becomes $[-\Delta v / \Delta \sigma(W)] / (\Delta v / \Delta \mu(W))$. Therefore, optimal portfolio diversification can be characterized by the condition:

$$MRS_{\sigma\mu} = \frac{-\Delta v / \Delta \sigma(W)}{\Delta v / \Delta \mu(W)} = \frac{\mu(B) - R(F)}{\sigma(B)} \quad (7.29)$$

If there are many persons who are required to choose between these two assets, then given the price of risk the optimality condition in Eq. (7.29) implies that the marginal rate of substitutions will be equal across the persons.

We can easily discuss the impact of changes in the parameters of the problem on the portfolio choice and hence on the risky asset demand for money. For instance, suppose that the return rate from a risk free asset has increased from $R(F)$ to $\bar{R}(F)$. This rotates the risk-return trade-off constraint upwards as shown in Fig. 7.7. As usual the effect of a price change on the optimal choice of an individual may be separated into a substitution effect and an income effect. The higher rate of return, working through the substitution effect, induces the investor to increase A_1 and hence less of the risky asset is demanded. If the individual regards risk as an inferior good, with decrease in the price of risk (as $R[F]$ increases), the income effect further reduces the demand for risky assets. The two positions of optimal choice are given by P and Q respectively. The exact position of Q depends on the strengths of the substitution and income effects.

It should now be clear that all of the analyses carried out in the context of a consumer behavior problem under certainty can be brought to bear on the above mean variance model. The model can also be extended to incorporate other assets in addition to the two assets considered here.

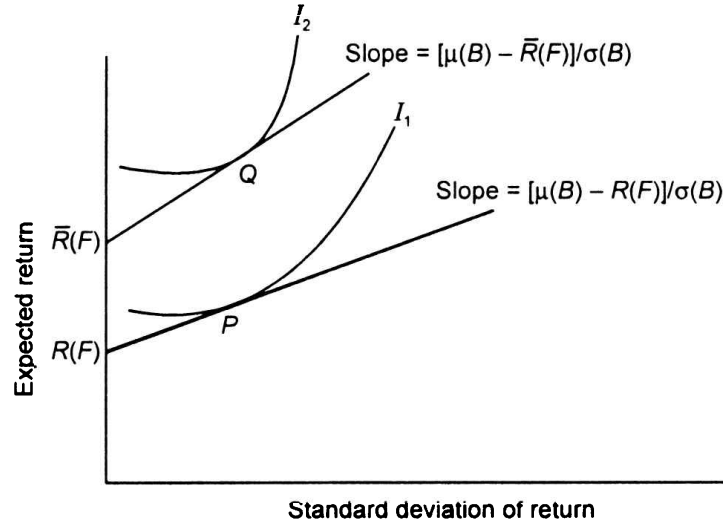


Fig. 7.7 Effect of an Increased Rate of Return on Risk Free Asset

Example 7.1

Suppose an individual's utility from his wealth W is given by $U(W) = A - Be^{-cW}$, where B and c are positive. The individual plans on investing his total wealth in a project. The probability that the project will generate a gain of 100 units of wealth is p . On the other hand, it may give rise to a loss of 90 units of wealth with probability $(1 - p)$. The individual will evaluate the project on the basis of his resulting total wealth $W + 100$ or $W - 90$. We want to show that the person's decision to do the project or not will not be affected by the level of initial wealth.

The person will do the project if his expected utility is greater than the utility from his initial wealth. The expected utility is $pU(W + 100) + (1 - p)U(W - 90)$, which under the above specification of U becomes $p(A - Be^{-c(W + 100)}) + (1 - p)(A - Be^{-c(W - 90)}) = A - Bpe^{-c(W + 100)} - (1 - p)Be^{-c(W - 90)}$. His utility from the initial wealth W is $A - Be^{-cW}$. Therefore, the person will invest in the project if

$$A - Bpe^{-cW - c100} - (1 - p)Be^{-cW + c90} > A - Be^{-cW},$$

which can be rewritten as

$$-Bpe^{-cW - 100c} - B(1 - p)e^{-cW + 90c} > -Be^{-cW}.$$

Multiplying both sides of the above inequality by $B^{-1}e^{cW}$, we get

$$-pe^{-100c} - (1 - p)e^{+90c} > -1.$$

Thus, his decision to invest in the project does not depend on the initial wealth holding

Applied Example 7.1 Uncertainty in Crop Production

In one sense all individual behavior in uncertain situations can be regarded as being motivated by the deficiency of correct information. Many economists have been interested in the relationship between individuals' behavior under uncertainty and the value of information they receive. For instance, obtaining accurate weather forecasts is important in a variety of situations where the decisions of firms/individuals depend on the weather. Perhaps the most important examples of the impact of weather on production occur in agriculture, where anticipated rainfalls affect a number of decisions such as which crop to plant, when to harvest, and so on. Better weather forecasts undoubtedly reduce the uncertainty associated with crop cultivation. Therefore, uncertainty affects economic calculations in such situations.

In a recent study Chattopadhyay, Neogi and Maity (1993) attempted to examine the stability of the pattern of growth of crop production in three states of Eastern India (namely, Bihar, Orissa and West Bengal) during the period 1950 to 1988. The growth of agriculture in these states has been found to be unstable. One of the factors which have been held responsible for instability in agricultural production in these states is uncertain natural conditions. Hence even for a relatively small segment of the economy, the value of information can be substantial.

Appendix 7.1

MEASUREMENT OF RISK AVERSION

The most commonly used measure of risk aversion was suggested by Arrow (1971) and Pratt (1964). The risk aversion measure, $r(W)$, as a function of wealth W , is defined as

$$r(W) = - \frac{U''(W)}{U'(W)}. \quad (7.30)$$

This measure, which is popularly known as the Arrow-Pratt absolute risk aversion measure, is positive, zero or negative accordingly an individual is risk averse, neutral or a risk lover. It is easy to check that the measure is invariant under affine transformations of the utility function. Hence it is not affected by which particular von Neumann-Morgenstern ordering is used.

Risk Aversion and Insurance Premiums

An interesting feature of the Arrow-Pratt measure is that it is proportional to the amount an individual will pay for insurance against participating in a fair gamble. Suppose that the winnings from such a fair gamble are denoted by the random variable h (h may take positive or negative values). Since the gamble is fair, its expected value $\mu(h) = 0$. Now, let p be the size of the insurance premium that would make the individual exactly indifferent between participating in the fair gamble, and paying p with certainty to avoid the gamble. Therefore,

$$\mu[U(W + h)] = U(W - p),$$

where W is the individual's current wealth. Expanding both sides of the above equation by Taylor's series we have

$$\begin{aligned} \mu[U(W) + hU'(W) + \frac{h^2}{2} U''(W) + \text{higher order terms}] \\ = U(W) - pU'(W) + \text{higher order terms.} \end{aligned}$$

Taking the expectation on the left hand side of the above we have

$$\begin{aligned} U(W) + \mu(h)U'(W) + \mu\left(\frac{h^2}{2}\right) U''(W) + \text{higher order terms} \\ = U(W) - pU'(W) + \text{higher order terms.} \end{aligned}$$

Now, remembering that $\mu(h) = 0$, using the constant k to represent $\mu(h^2)/2$ and ignoring higher order terms on both sides of the above, we rewrite it as

$$U(W) - pU'(W) = U(W) + kU''(W),$$

from which it follows that

$$p = -k \frac{U''(W)}{U'(W)} = kr(W).$$

Thus, the amount that a risk averse individual would be willing to pay to avoid a fair gamble is proportional to the Arrow-Pratt absolute measure. Since insurance premiums are observable in the real world, they can be used to calculate an individual's risk aversion measure. Thus, we can make use of market information to learn about attitudes toward risk.

STOCHASTIC DOMINANCE

Consider two income distribution functions F and G defined on $[0, \infty]$. F and G are non-decreasing, $F(0) = G(0) = 0$, $F(\infty) = G(\infty) = 1$. F is said to dominate G by the n th degree stochastic dominance criterion if

$$F_n(x) \leq G_n(x)$$

for all $x \in [0, \infty]$, with $<$ for at least one x , where

$$F_n(x) = \int_0^x F_{n-1}(t) dt$$

and

$$F_1(x) = F(x).$$

Thus, F first order stochastically dominates G if $F(x) \leq G(x)$ for all x , with $<$ for at least one x . This is equivalent to the condition that $\int_0^\infty U(x) dF(x) > \int_0^\infty U(x) dG(x)$, for all increasing utility functions U , that is, the expected utility value under F is greater than that under G for all increasing U . Clearly, in the case of first order stochastic dominance, higher income is the criterion for preferring one distribution to another.

For F to dominate G by the second degree stochastic dominance criterion we require that $\int_0^x F(t) dt \leq \int_0^x G(t) dt$ for all x , with $<$ for at least one x . This holds if and only if $\int_0^\infty U(x) dF(x) > \int_0^\infty U(x) dG(x)$ for all increasing and strictly concave U . Therefore, under second degree stochastic dominance criterion F is preferred to G by all risk averse persons who prefer higher income. For further discussions along this line, see Fishburn (1980) and Chakravarty (1990).

EXERCISES

Multiple Choice

1. A risk averter's utility function over money pay-offs is
 - a) Strictly concave.
 - b) Strictly convex.
 - c) Concave.
 - d) Convex.

2. Let C_1 and C_2 represent consumptions in states 1 and 2 respectively. Suppose $p_1(p_2)$ is the probability that state 1(2) will actually occur. Then which of the following does not have the expected utility property?
- $\frac{p_1}{p_2}C_1 + C_2$
 - $p_1 \log C_1 + p_2 \log C_2$
 - $p_1 \sqrt{C_1} + p_2 C_2$
 - $\frac{p_1}{p_2} \log C_1 + \log C_2$
3. Which of the following property (properties) of the utility function describes (describe) a consumer's attitude towards risk?
- Continuity.
 - Differentiability.
 - Increasingness.
 - Curvature.
4. An individual's utility function over money pay-offs W is of the form $U(W) = A\sqrt{W}$ where $A > 0$. He faces a situation in which he receives Rs. 784 with probability $1/3$ and Rs. 100 with probability $2/3$. The certainty equivalent here is
- 16A.
 - 16.
 - $16 + 16A$.
 - None of these.
5. The expected utility property remains preserved under
- Any increasing transformation.
 - Affine transformation only.
 - Identity mapping only.
 - Logarithmic transformations only.

Quantitative Problems and Questions

- Prove or disprove the following:
 - For a risk averse individual the indifference curves between expected return and variability of return for the expected utility function are negatively sloped.
 - Let U and V be two expected utility functions. Suppose U is more risk averse than V according to the Arrow-Pratt absolute risk aversion measure. Then U is more concave than V , that is, U can be expressed

as an increasing, strictly concave function of V . (For instance, if $V[W] = \sqrt{W}$ and $U(W) = W^{1/4}$, then $U[W] = g[V(W)]$ where $g(Z) = Z^{1/2}$. Thus, U is more concave than V .)

- c) Of two persons A and B , A is willing to pay a higher amount of premium than B to avoid a fair gamble. Then A is more risk averse than B according to the Arrow-Pratt absolute risk aversion measure.
 - d) Suppose you are given the choice of selecting and keeping the contents of one of three envelopes from group A or group B , where the three envelopes in group A contain Rs. 400, Rs. 500 and Rs. 600, and those in group B contain Rs. 50, Rs. 500 and Rs. 950. Since you are a risk lover you will select group B .
 - e) A consumer's expected utility function is strictly concave. He is offered a bet with probability p of winning a very small amount t and the probability $1 - p$ of losing t , where $p > 1 - p$. Then the consumer will refuse the bet.
2. Suppose that an individual's utility from wealth W is given by $U(W) = a + bW + cW^2$, where $b > 0$, $c < 0$ and for all relevant values of W , $b + 2cW > 0$.
- a) Show that $U(W) > 0$.
 - b) If W is a random variable, show that the expected utility $E[U(W)]$ depends on the mean μ and the variance σ^2 of W .
 - c) Show that the individual will accept a higher σ^2 only if he gets a higher mean μ . What can you say about the slope of the indifference curve between μ and σ^2 for the expected utility function?
 - d) Does the indifference curve calculated in (c) have a diminishing marginal rate of substitution?
3. The measure of risk aversion $r(W)$ we have considered in Appendix 7.1 is called the Arrow-Pratt measure of absolute risk aversion. An alternative to $r(W)$ is the Arrow-Pratt relative risk aversion measure $m(W)$, where

$$m(W) = - \frac{WU''(W)}{U'(W)}$$

Consider the following utility functions

- a) $U(W) = A - Be^{-cW}$, $c > 0$, $B > 0$.
- b) $U(W) = A + B \frac{W^r}{r}$, $r < 1$, $r \neq 0$, $B > 0$,
 $= A + B \log W$, $r = 0$, $B > 0$.

One of these two functions is the only utility function exhibiting constant absolute risk aversion, and the other is the only function displaying constant relative risk aversion. Identify them.

4. An individual has a utility function over money pay-offs Z of the form $U(Z) = \log Z$. He faces a situation in which he gets Rs. 2 with probability $1/2$ and Rs. 100 with probability $1/2$. What pay-off with certainty is indifferent to this situation?
5. A person's utility function from wealth W is given by $U(W) = A\sqrt{W}$, where $A > 0$ is a constant. The person's initial wealth holding is Rs. 25. The person can invest this money in a gamble from which he can gain Rs. 11 with probability $1/2$ and lose Rs. 9 with probability $1/2$. What is his expected utility?
6. Let W_1 and W_2 be two independent random variables representing the pay-offs from two investments. Assume that the mean of W_1 is same as that of W_2 and the variances are also the same. Then show that a risk averse consumer will divide his wealth holding between the two assets equally. How does the result change if the consumer is risk loving?
7. Suppose a tourist wishes to visit important places of a country. The utility from his trip is a function of how much he spends on it (W), and is given by $U(W) = \sqrt{W}$. He has Rs. 10,000 to spend on the trip.
 - a) If there is a 25 percent probability that the person will lose Rs. 1,000 on the trip, what is his expected utility?
 - b) Suppose that the person can buy insurance against losing Rs. 1,000. What is the maximum amount that the person would be willing to pay to insure his Rs. 1,000?
8. Suppose that an individual's utility from wealth (W) is given by $U(W) = a + bW + cW^2$ where $b > 0$, $c < 0$. Show that the Arrow-Pratt absolute risk aversion measure is an increasing function W . How does the measure change with respect to changes in W if the utility function is of the type $U(W) = a \log W$, where $a > 0$?
9. Consider the utility function $U(W) = W(1 - e^{-W})$.
 - a) Show that this function exhibits risk loving behavior for small values of W and risk averse behavior for high values of W .
 - b) What is the value of W above which the function shows risk averse behavior? (A utility function of the type $U[W] = W[1 - e^{-W}]$, which is

strictly convex for low values of W and strictly concave for high values of W is called a Friedman-Savage utility function [Friedman and Savage (1948)].

10. Suppose a person's responses when he has to make a choice from among gambles are as follows: He is indifferent between receiving (Rs. 20 with probability $1/2$, Rs. 200 with probability $1/2$), and Rs. 100 with certainty. He is indifferent between receiving (Rs. 20 with probability $1/3$, Rs. 200 with probability $2/3$) and Rs. 120 with certainty. What will be his choices in the following situations?
- a) Between (Rs. 100 with probability $1/4$, Rs. 200 with probability $3/4$ and (Rs. 120 with probability $1/2$ and Rs. 200 with probability $1/2$),
 - b) Between (Rs. 100 with probability $1/2$, Rs. 200 with probability $1/2$) and Rs. 120 with certainty.

8

Technology

Production means transformation of inputs into outputs. A firm is an organization that engages in production. A firm allocates resources that it has bought in order to produce and sell its output. In this chapter we examine the motivations of a firm and its technological constraints, which show that there are only certain feasible ways to produce outputs from inputs.

PROFIT MAXIMIZATION AND THE FIRM

In the economists' model a firm is a decision making unit that mainly operates to maximize profits for its owners through its choice of input-output combinations. Profits are a residual, being the difference between what the firm receives as income (revenue) and what it spends for the resources (costs). (We shall discuss conceptual problems associated with the definition of profit in Chapter 10.) The profit maximization postulate is a direct consequence of the motivation of consumers as owners of firms.

Firms can be classified into proprietorships, partnerships and corporations. A proprietorship is a firm which is owned by a single individual only. A partnership is a firm with multiple owners. A partnership will exist as long

as all partners are alive and agreeable in maintaining its existence. If one member of a partnership leaves, the entire partnership is automatically dissolved. If the business is to continue, a new partnership must be formed. On the other hand, a corporation, which is also usually owned by several owners, can last longer than the lifetime of any of the owners. Thus, a corporate firm has an existence separate from that of its owners. Most large firms are organized as corporations.

In a proprietorship or partnership, owners of the firm usually exercise direct control over the day-to-day operations of the firm. Generally, the owners become interested in profit maximization. In large corporate firms, ownership is often separated from the control of the firm. Direct communication among owners is expensive. The managers operate simply as agents of the owners, which amounts to saying that the job of the managers is to run the firm as an entity and do their best to see that they pursue the goals of the owners. Profit maximization can be a common goal of the owners.

In a world of perfect certainty, the present value of the stream of future profits of a firm is the present value of the firm. This is the value that a person would be willing to pay for buying the firm.

A corporation issues stock certificates to represent ownership of shares in the corporation. A shareholder is entitled to receive dividends which come out of the corporation's profits. Dividends are one way stockholders earn return on their investments. Even if a corporation does not pay any dividend, shareholders can get return on their investments. If the price of the stock increases above the price which the shareholder paid, the shareholder can sell it and make a profit. The price of a share represents the present value of the future flow of dividends that people expect to get from the corporate firm. The market in which the shares of ownership of a corporation are traded is called the stock market. In a world of certainty, maximization of the stock market value is the same as maximization of the present value of flow of future profits.

The benefit of a corporation is that shareholders have limited liability. This means that if a corporation fails, a stockholder's losses are limited to the price paid for the stock. The worst situation that a shareholder can face is that the stock becomes worthless. The value of limited liability is that it makes individuals willing to buy shares.

In a situation of uncertainty, if managers attempt to make the value of a corporation's shares as high as possible, the shareholders are as well-off as possible. Therefore we can talk about stock market value in a situation of

uncertainty also. On the other hand, it is difficult to say something about the present value of flow of future profits in such a case.

In this book we will assume that a firm's motivation is to maximize profit and analyze several models under alternative assumptions about markets in Chapters 10–12. This will give significant insights into the firms' behavior and develop intuition to study more general models. At the end of this chapter we briefly discuss some other objectives which will give an idea about the alternative motivations a firm owner may have in mind.

PRODUCTION FUNCTION

As a production unit, a firm transforms inputs into outputs. Inputs are factors of production. These factors can be machine (that is, capital) usage, hours of labor input, raw materials used, land usage, building usage and so on. For any firm only certain combinations of inputs are feasible ways to produce a given amount of output. That is, there are only certain kinds of technological choices that are possible. A firm must limit itself to technologically feasible production plans. The set of all input-output combinations that are technologically feasible is called a production set.

For example, suppose that we have only one input L , the hours of labor input, and one output, measured by x . Then we say that a point (L, x) is in the production set if it is technologically possible to produce x amount of output when we have L amount of labor. The production set gives us the possible technological choices that a firm faces. A production set may have the shape shown in Fig. 8.1.

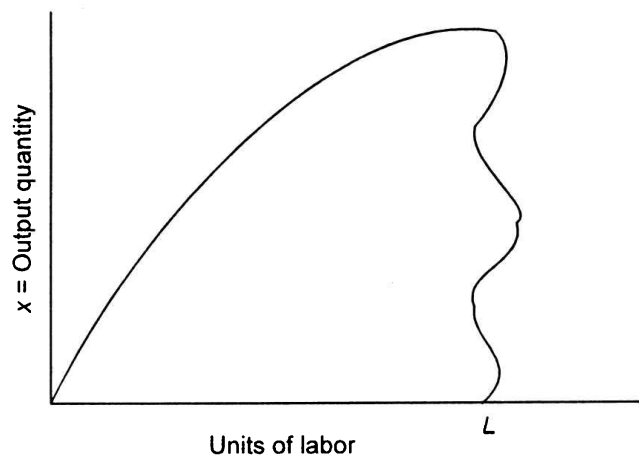


Fig. 8.1 A Production Set

If the firm wastes input (produces less than the maximum output possible from a given input level), it cannot be a profit maximizing entity. Therefore, as long as the inputs are costly we should be interested only in the maximum output that can be obtained from a given input level. This is the boundary of the production set shown in Fig. 8.1. This boundary is in fact the graph of the firm's production function that associates with each input level the firm employs, the maximum output the firm can obtain from that level of input.

The above notion of the production function applies equally well if we have several inputs. In the general case of an arbitrary number of inputs, we write the production function as $x = f(K, L, M, \dots)$. Here x is the maximum amount of output that we can get if we use K units of capital services, L units of labor services, M amount of raw material, and so on. (The notation indicates the possibility of other inputs affecting the technology.) Since the function f transforms inputs K, L, M , etc., into output x , f is the production function. It is important to note that we have assumed that inputs and outputs are measured in flow units: a certain number of machine hours per unit of time, a certain amount of labor per unit of time, a certain amount of raw materials used per unit of time, and so on, will produce a certain amount of output per unit of time.

PRODUCTION PROCESS

Essential to the notion of a production function is the production process. A production process is a combination of inputs in a certain fixed ratio required for the production of a given amount of output. For instance one process (say process 1, P1 for short) may involve using capital and labor services in 1:1.5 ratio, another process (say, P2) may involve 1:1 ratio, and so on. Since P1 operates with a 1:1.5 capital-labor ratio and P2 operates with a 1:1 capital-labor ratio, P1 is said to be more labor intensive than P2. Alternatively, P2 is said to be more capital intensive than P1.

A production process A is technically inefficient if there is another process B (or a combination of processes) which, for any given output, uses no more of any input and less of at least one input. For instance, suppose that one unit of a good can be produced by two processes: process A that uses three units of capital and one unit of labor and process B which involves one unit of capital and one unit of labor. Clearly, A is an inefficient method of producing one unit of output since it uses more capital than B . All processes that are not inefficient are called efficient. Theories of production concentrate only on efficient methods.

If for producing a given output level, a process A uses less of some input(s) and more of some other(s) than another process B , then A and B cannot be directly compared in terms of technical efficiency. Both processes are included in the technology. The input prices determine which particular process will be chosen. The theory of production describes the law of production. The choice of a particular process form among the set of technically efficient processes is an economic one, based on input prices and not a technical one. The problem of choice of a specific technique by a firm is described in the next chapter. A technically efficient process need not be economically efficient.

THE SHORT RUN AND LONG RUN

We may be required to distinguish between production plans from short run and long run viewpoints. A planning period so short that the firm is unable to vary the quantities of some of the factors it uses is called the short run. That is, in the short run the firm is unable to vary the quantities of some of its factors. In other words, in the short run, there will be some factors of production that are fixed at predetermined levels. Examples are land, building, top management personnel, heavy machinery, etc. Land and building may be leased by the firm for given time periods, or they may be owned outright. It may take some time to get additional amounts or to dispose off a part of the quantities already owned. Similarly, the number of top management personnel is not ordinarily readily variable: quantities of heavy machinery especially designed for the firm's use cannot be quickly decreased or increased. Typically, in the short run variation in the quantities of resources such as labor, light machinery, power, transportation and raw materials can take place. These are the firm's variable factors. On the other hand, factors such as land, building, top management, heavy machinery are the firm's short run fixed inputs. The firm does not have time to vary quantities of such resources in the short run.

The quantities of fixed factors used by a firm determine the size of the firm's plant. The size of the plant sets the upper limit to the amount of output that the firm can produce per unit of time. The firm can vary its output within that upper limit by increasing or decreasing the quantities of the variable inputs used in the fixed sized plant.

A planning period which is long enough to vary the quantities of all inputs used is called long run. Thus, all inputs are variable in the long run. There is no problem of classifying factors as variable or fixed. The firm can change its size of the plant as it desires.

The distinction between short run and long run situations depends on what we are investigating. There is no specific time interval implied here. In the short run some factor(s) is (are) fixed at certain level(s), but in the long run all the factors can be changed.

For simplicity of exposition, let us suppose that we are dealing with a technology involving three inputs—capital services K , labor services L and plant services P . Output x is then given by the production function $x = F(K, L, P)$. In the short run P is fixed, while K and L are variable. In the long run K , L and P are variable inputs. Since in the short run P is fixed, we can rewrite the production function $x = F(K, L, P)$ as $x = f(K, L)$, say.

VARIATIONS IN ONE INPUT

In this section we will examine the impact of a change in one input on the output. The earliest economists who studied production were mainly concerned with how increasing labor utilization on a fixed plot of land might affect output. Hence there are historical reasons for looking at variations in a single input.

To examine the impact of variations in one input on output levels, let us concentrate on the production function $x = f(K, L)$. Assume that we are operating at some point (K, L) . Suppose we use a little more of L , say ΔL , while keeping K constant. How much additional output do we get per additional unit of L ? This is given by

$$MP_L = \frac{\Delta x}{\Delta L} = \frac{f(K, L + \Delta L) - f(K, L)}{\Delta L} \quad (8.1)$$

Since the additional amount of labor employed (ΔL) is marginal, MP_L is called the marginal product of labor. MP_L is the increase in output from adding an additional unit of labor, with all other inputs held constant. It is important to observe that MP_L is a rate: it is the extra amount of output per unit of extra input.

We can similarly, define the marginal product of capital as

$$MP_K = \frac{\Delta x}{\Delta K} = \frac{f(K + \Delta K, L) - f(K, L)}{\Delta K} \quad (8.2)$$

The notion of marginal product is analogous to that of marginal utility. However, the marginal product of a factor is a specific real number, which can be observed, whereas marginal utility represents a psychological

phenomenon. This difference arises because in the case of production we are concerned with physical output levels.

To illustrate the behavior of marginal productivity, let us consider the case of a farmer hiring one more laborer to harvest the crop, but holding all other inputs constant. The extra output that this laborer produces is the marginal product, measured in physical quantities, say bushels of wheat. For instance, suppose that 10 workers on a farm are able to produce 100 bushels of wheat per unit of time, whereas 11 workers can produce 103 bushels of wheat per unit of time. Then the marginal product of the eleventh worker is 3 bushels.

Now, labor cannot be added indefinitely to a given field, keeping the amounts of fertilizer, equipments, etc., fixed, without eventually exhibiting some deterioration in its productivity. This is because as we add more and more units of the variable input, we decrease the number of units of the fixed inputs available per unit of the variable input. This reduction hampers the productivity of the variable input and output increases at a decreasing rate, that is, the marginal product of a factor is a positive but decreasing function of the factor. In the example we have considered, after 4 to 5 workers are added the additional output per worker will drop to 2, 1.5, 1 or even fewer bushels. That is, if a firm adds additional units of a variable factor to production process, the fixed inputs become over-utilized and a decline in productivity will set in.

This principle is known as the law of diminishing marginal productivity. We state this law as follows: As more and more units of a variable factor are added to the constant amounts of other factors, a point is reached beyond which the marginal product of the variable factor starts decreasing. Equivalently, if the amount of a variable factor is increased while keeping the amounts of other factors fixed, total output will increase, but beyond some point, the resulting output increases will be smaller and smaller.

Given fixed amounts of other inputs, the graphical relationship between a particular input (say, labor) and total output is called the total product curve of the input. That is, the TP_L curve, the total product of the labor curve, is obtained by plotting output x against labor employed, L , given that other inputs are fixed.

The TP_L curve drawn in panel (a) of Fig. 8.2 reflects the law of diminishing marginal productivity. For small amounts of L , output increases at an increasing rate as L increases, that is, MP_L increases for small values of L .

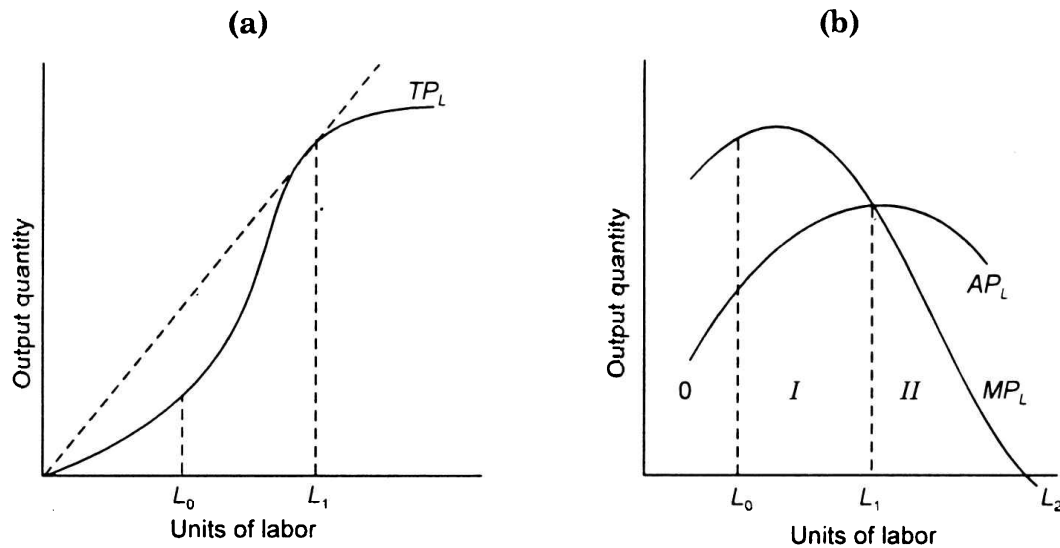


Fig. 8.2 Total Product Curve (a), Marginal and Average Product Curves (b)

Now, MP_L is the slope of the TP_L curve. The slope of the TP_L curve just shows how output expands as additional labor is added. Therefore the MP_L curve can be constructed from the TP_L curve. The MP_L curve is shown in panel (b) of Fig. 8.2. Notice that MP_L attains a maximum at L_0 . We say that L_0 is the point of diminishing returns because MP_L starts declining as more labor input is added beyond L_0 . This is a reflection of the assumption of diminishing marginal product of labor. MP_L is equal to zero at the point L_2 at which TP_L reaches a maximum. Beyond L_2 , further addition of labor input reduces output. The point L_2 is sometimes called the intensive margin of production. Production will not take place beyond L_2 since using more labor will result in less output.

We define the average product of labor, AP_L , as the total product per unit of labor. That is,

$$AP_L = \frac{\text{output}}{\text{labor input}} = \frac{x}{L}. \quad (8.3)$$

Geometrically, for any quantity of labor input, AP_L is the slope of the line drawn from the origin to the relevant point on the TP_L curve. This follows since the slope of such a line is x/L . By drawing a series of lines from the origin to the various points on the TP_L curve, the AP_L curve can be constructed. This curve is shown in panel (b) of Fig. 8.2.

Observe that AP_L increases up to L_1 and decreases after that. At L_1 , $MP_L = AP_L$. For the level of labor input L_1 , the line through the origin is just

tangent to the TP_L curve. Hence MP_L and AP_L are equal at L_1 . Also at L_1 , AP_L is at its maximum value.

The relationship between MP_L and AP_L is summarized in the following theorem.

Theorem 8.1

- a) So long as AP_L is rising ($L < L_1$), MP_L is greater than AP_L .
- b) So long as AP_L is falling ($L > L_1$), MP_L is less than AP_L .
- c) $MP_L = AP_L$ at the maximum of the AP_L curve ($L = L_1$).

A rigorous proof of this theorem is given in Appendix 8.1. A verbal reasoning is provided here. If AP_L rises, output x increases faster than labor input L . But this can occur only if the increment to output from hiring one more worker exceeds the average product of workers who have been hired previously. Hence $MP_L > AP_L$. Similarly, if AP_L is declining, x increases more slowly than L , which implies that $MP_L < AP_L$.

We have shown that to the left of L_1 , the AP_L curve is rising; to the right of L_1 , it is falling. Therefore, the average product of labor is maximum at L_1 . The point L_1 is often referred to as the extensive margin of production, since labor added beyond this point will result in a reduction of the average product of labor.

We identify three regions in Fig. 8.2. In region 0, MP_L increases because of the relative abundance of fixed inputs compared to labor input. In region I, diminishing marginal productivity starts working, but AP_L is still rising, with $AP_L > MP_L$. In region II, both MP_L and AP_L are declining. It is important to note that if both output and factor markets are competitive, hiring of labor input occurs in region II only. In this region as larger quantities of labor per unit of fixed inputs are used, the production per worker decreases. On the other hand, regions 0 and I are characterized by increases in the average production of labor as more labor per unit of fixed inputs is used. We can draw the same kind of figures for any other input, but under competitive assumptions the hiring of the input will occur in the region in which both average and marginal products of the input are declining. These issues are discussed in greater detail in Chapter 14.

It is possible for diminishing marginal productivity or diminishing increases in total output to occur for all increments of the variable input. That is, the law of diminishing marginal productivity may apply immediately. This is frequently the case with the application of fertilizer to given complexes of

seed, land, labor and machinery. In this case in Fig. 8.2(b), we will have only two regions I and II. Region 0 will be absent because the diminishing marginal productivity condition sets in immediately.

THE LAW OF VARIABLE PROPORTION

The law of diminishing marginal productivity relies on variations in only one input. In some cases we may allow variability of more than one input, keeping one or more inputs constant. For instance, suppose plant capacity is held fixed, but other inputs are variable. Then it seems reasonable to expect that beyond a certain point equiproportional increases in all variable inputs will lead to less than proportionate increases in output. This generalized version of the law of diminishing marginal productivity is known as the law of variable proportion. We state this formally as: If one or more inputs are held constant, a point is reached after which equal proportionate increases in all other inputs lead to a less than proportionate increase in output.

To state the law mathematically, we consider the production function $F(K, L, P)$, where K and L are variable and plant capacity P is fixed. Then the law of variable proportion says that

$$F(tK, tL, P) < tF(K, L, P) \quad (8.4)$$

beyond a certain capital-labor combination (K_0, L_0) , where $t > 1$ is arbitrary. However, if the law applies immediately, then the law states that

$$F(tK, tL, P) < tF(K, L, P) \quad (8.5)$$

for all $t > 1$.

ISOQUANT MAP

One convenient way to depict a production function for two inputs is by using its isoquant map. We use the production function $x = f(K, L)$, with the understanding that capital (K) and labor (L) are two examples of inputs that might be of interest. An isoquant is the locus of all possible combinations of capital and labor that are just sufficient to produce a given level of output. An isoquant records the fact that there are alternative ways of producing the given level of output. Suppose that x_0 is the given level of output. Mathematically, an isoquant shows the set of K and L that satisfies

$$f(K, L) = x_0. \quad (8.6)$$

A graph of family of isoquants, each of which connects capital-labor combinations that can produce a given level of output, is called an isoquant map.

Production isoquants are level curves of a firm's production function. Each isoquant represents a different level of output. Isoquants record successively higher levels of output as we move to the right. This reflects the assumption that using more of each of the inputs will permit output to increase. Clearly, isoquants cannot intersect, by their construction.

Now, as stated earlier, we concentrate on ranges of output over which marginal products of factors are diminishing but positive. The locus of points of the isoquants where the marginal products of the factors are zero constitute the ridge lines.

In panel (a) of Fig. 8.3, the upper ridge line OA implies that the marginal product of capital is zero, while zero marginal product of labor is indicated by the lower ridge line OB . Outside the ridge lines, that is, where isoquants are positively sloped, the production techniques are (technically) inefficient, since they require more quantities of both factors for producing a given level of output. That is, outside ridge lines or, equivalently, in the positively sloped region of the isoquants, marginal products of factors are negative. Such inefficient methods are ruled out from the purview of the theory of production. Production techniques are efficient only between the ridge lines. Between the ridge lines isoquants are negatively sloped. This means that if the firm wants to use less of K , it must use more of L to produce the same level of output, for remaining on the same isoquant. In this relevant range, isoquants are not only negatively sloped but also strictly convex. The firm's isoquant map is shown in panel (b) of Fig. 8.3.

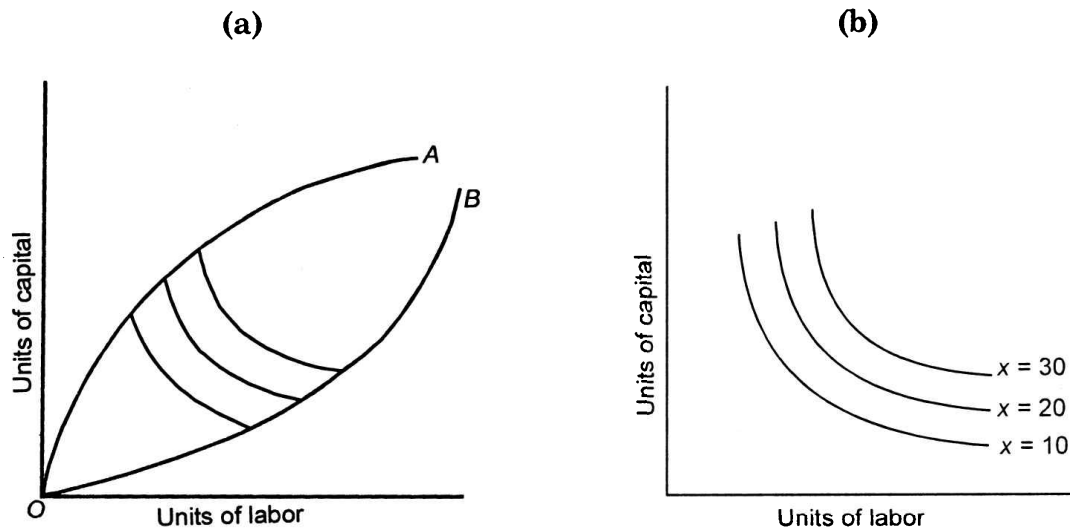


Fig. 8.3 Isoquant Map

There are some parallels between an isoquant map and an indifference map. An indifference map graphically represents the consumer's tastes, whereas a production isoquant map graphically represents the firm's technology. An indifference curve is a curve along which utility of a consumer is constant, whereas along a production isoquant a firm's output is constant. Indifference curves are level curves of the consumer's utility function, and isoquants are level curves of the firm's production function. However, there is one important difference. For isoquants, the labeling of curves is measurable since the numbers associated with isoquants are output levels of the firm. On the other hand, the numbers associated with indifference curves are arbitrary so long as they represent the consumer's preferences.

INPUT SUBSTITUTION AND MARGINAL RATE OF TECHNICAL SUBSTITUTION

Suppose that a firm is operating with some capital-labor combination (K, L) and it now wants to give up a little bit of L and use just enough more of K to produce the same amount of output. How much extra of K , ΔK , does the firm need if it gives up ΔL amount of L ? The marginal rate of technical substitution (MRTS) measures this kind of trade-off between two factors of production. The MRTS of capital for labor refers to the rate at which capital can be substituted for labor while holding output constant. In terms of the notation introduced

$$\text{MRTS (K for L)} = - \frac{\Delta K}{\Delta L}$$

along a given isoquant. Since $\Delta L < 0$ and $\Delta K > 0$, $\text{MRTS} > 0$. MRTS is the absolute value of the slope of the isoquant (see Fig. 8.4). Its value depends on the point on the isoquant at which the slope is measured. In other words, the particular value of the trade-off rate indicated by the MRTS will depend not only on the level of output but also on the quantities of labor and capital being used.

In order to examine the shape of the isoquant, it is useful to show that the MRTS (of K for L) is equal to the ratio of the marginal product of labor (MP_L) to the marginal product of capital (MP_K). The demonstration proceeds as follows. Let Δx be the change in output brought about by ΔL change in labor and ΔK change in capital. Then

$$\Delta x = MP_L \Delta L + MP_K \Delta K. \quad (8.7)$$

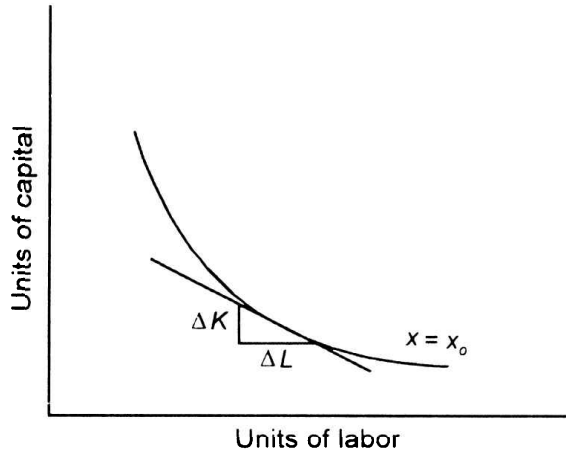


Fig. 8.4 Marginal Rate of Technical Substitution

That is, the total change in output Δx equals the change in output due to one unit change in labor (MP_L) times the change in labor input (ΔL) plus the change in output that is due to one unit change in capital (MP_K) times the change in capital (ΔK). But along an isoquant $\Delta x = 0$, so:

$$MP_L \Delta L = - MP_K \Delta K. \quad (8.8)$$

This says that along an isoquant the gain in output from increasing K slightly is exactly offset by the loss in output from suitably decreasing L . Rearranging terms, we have

$$-\frac{\Delta K}{\Delta L} = \frac{MP_L}{MP_K} \quad (8.9)$$

or, equivalently,

$$\text{MRTS of } K \text{ for } L = \frac{MP_L}{MP_K} \quad (8.10)$$

as was to be shown.

Since both MP_L and MP_K are positive (no firm would choose to produce where the marginal product is negative), MRTS is positive, an observation which we have made earlier.

In Fig. 8.3 the isoquants are drawn not only with a negative slope (as they should be) but also as strictly convex curves. Along any isoquant the MRTS is diminishing. This is known as the law of diminishing marginal rate of technical substitution and it is formally stated as: Along any production

isoquant, the marginal rate of technical substitution of capital for labor decreases as the amount of labor increases.

To understand this law, note that for high ratios of K to L , the MRTS is a large positive number, indicating that a low amount of labor is to be given up if one more unit of capital is to be employed along an isoquant. On the other hand, when a lot of a labor is already in use, the MRTS is low, signifying that only a high amount of labor can be traded for an extra unit of capital, if output is held fixed. This seems intuitively reasonable. The more labor (relative to capital) is used, the more difficult it is to substitute capital for labor. Loosely speaking, this means that labor becomes less potent as a substitute as more of it is used. Technically the assumption of diminishing MRTS says that the slope of an isoquant must decrease in absolute value as we move along an isoquant to the direction of increasing L and it must increase as we move in the direction of increasing K . Equivalently, we say that the isoquants are strictly convex to the origin.

The assumption of diminishing MRTS has some close relationship with that of diminishing marginal productivities, as Eq. (8.10) implies. But the two assumptions are not identical. It is not possible to derive a diminishing MRTS from the assumption of diminishing marginal productivities alone. We show this rigorously in the Appendix. 8.1.

INPUT SUBSTITUTION: TWO POLAR CASES

Suppose that the production function is given by

$$x = f(K, L) = aK + bL, \quad a, b > 0 \quad (8.11)$$

All isoquants associated with this production function are straight lines with a constant slope of $-b/a$. Such an isoquant map is pictured in panel (a) of Fig. 8.5. Clearly, the constant MRTS (of L for K) here is b/a . Although this linear production function is a nice example, it is rarely encountered in practice since there are very few production processes characterized by constant MRTS. In this case capital and labor can be regarded as perfect substitutes for each other. A firm characterized by such a production function can use only capital or only labor.

Quite often two inputs are used in a fixed proportion for producing a good. For instance, many machines require a certain number of people to run them, any excess labor is superfluous. Consider combining capital (a typewriter) and a worker (a typist) to do some typing job. Only one typist can use a

typewriter at a time, and either input without the other is not able to produce any output at all. In this case capital and labor are combined in 1:1 ratio. Suppose a typist can type one page per unit of time. Then the total number of pages that can be typed will be the minimum of the number of typists and the number of typewriters available. Therefore, we can write the production function as $f(K, L) = \min(K, L)$.

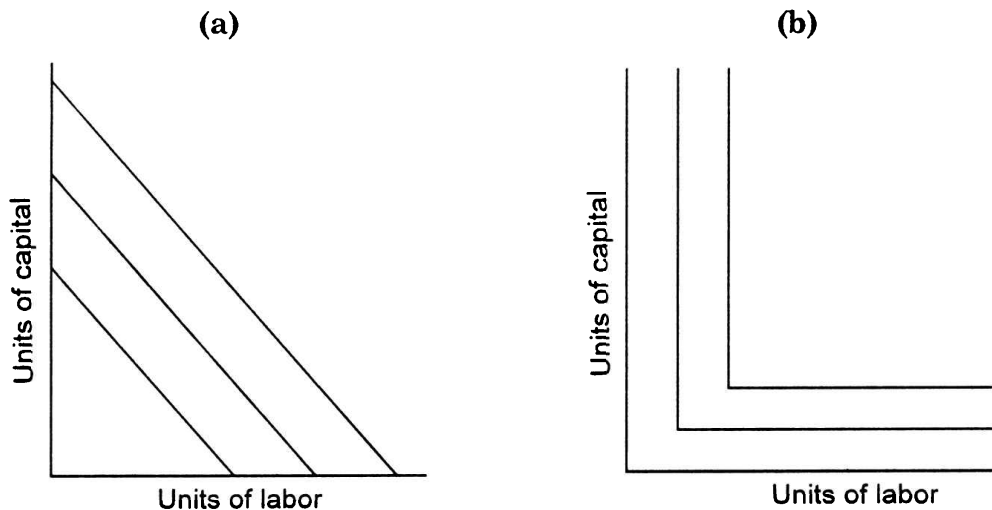


Fig. 8.5 Isoquants: Perfect Substitutes (a); and Fixed Proportions (b)

In general, where capital and labor are used in a fixed ratio, say b/a , the production function will be of the form

$$x = \min(aK, bL), \quad a, b > 0. \quad (8.12)$$

This production function is known as the Leontief production function. The isoquants for this production function are L-shaped and are shown in panel (b) of Fig. 8.5. A firm characterized by this production function will operate along the ray where K/L is b/a . To operate at some other point would mean inefficiency in production. To see this, suppose $aK < bL$. Then $x = aK$ and capital is the binding constraint here. The employment of additional labor does not increase x , and hence the marginal product of labor is zero. The additional labor is superfluous here. Similarly, if $aK > bL$, labor is the binding constraint here and additional capital becomes superfluous. When $aK = bL$, both inputs are fully utilized and the inputs are combined in the proportion $K/L = b/a$. The production takes place at the vertex of the isoquant. Clearly, in the case of fixed proportion production function MRTS is zero, indicating no possibility of substitution.

RETURNS TO SCALE

Since production functions represent tangible production processes, a lot of attention has been paid to the actual form of these functions. Knowledge of the shape and properties of a firm's production function is important from a variety of perspectives.

An important question that may be asked about production functions is how it responds to increases or decreases in all inputs. For instance, suppose that all inputs are doubled. Then how does the output change? This is a question of the returns to scale exhibited by the production function.

If a proportional change in all inputs increases output by the same proportion, the production function is said to exhibit constant returns to scale. That is, if, when all inputs are increased (decreased) by the same percent, output increases (decreases) by the same percent, there are constant returns to scale. Mathematically, constant returns to scale means that

$$F(tK, tL, tP) = tF(K, L, P) \quad (8.13)$$

for all $t > 0$. Thus, in this case if the firm has twice as much of each input, it can simply set up two plants side by side and thereby get twice as much output. Hence constant returns to scale is just a case of replication: if the quantities of different inputs that the firm are using are replicated $t > 1$ times, then output also gets replicated t times.

Constant returns to scale are not the only consequences of equiproportionate variations in inputs. It may happen that when all inputs are increased by some percent, output increases by a greater percent. This is called the case of increasing returns to scale. Mathematically,

$$F(tK, tL, tP) > tF(K, L, P) \quad (8.14)$$

for all $t > 1$. Thus, in this case a proportionate increase in inputs leads to a more than proportionate increase in output. The increasing returns to scale may take place due to technical efficiency.

To illustrate increasing returns to scale, consider the classic example of an oil pipeline. A pipeline's carrying capacity is approximately equal to the area of the cross section of the pipeline times the speed of flow of oil. The area of cross section of a pipeline is πr^2 , where r is the radius of the pipeline. The amount of material required to construct the pipeline is approximately proportional to $2\pi r$, the circumference of the pipeline. Let r_0 be the original radius. Suppose we double the radius. Then the new circumference is $4\pi r_0$, but the new cross sectional area is $4\pi r_0^2$. Therefore, in this case doubling of

input (circumference) quadruples the output (as measured by the cross sectional area). Hence this is a case of increasing returns to scale. Consequently, the firm with the largest pipeline has important cost advantages.

Sometimes for a proportional increase in inputs, output increases less than proportionately. This is the case of decreasing returns to scale. Mathematically,

$$F(tK, tL, tP) < tF(K, L, P) \quad (8.15)$$

for all $t > 1$. Thus, when inputs increase by a certain percent, in this case, output increases by a smaller percent. The most common cause for decreasing returns to scale are 'diminishing returns to management'. As output grows, management becomes overburdened and hence may become inefficient in coordinating the productive activities of the firm. In other words, it becomes difficult to scale management input upward in production without losing productivity. Although all firms face the scaling problems of management (tending towards decreasing returns), in some firms these are more than offset by scaling characteristics that give rise to increasing returns to scale.

It may be noted that a constant return to scale is equivalent to the statement that the production function is linear homogeneous or homogeneous of degree one. If a production function is homogeneous of degree $k > 1 (< 1)$, then it exhibits increasing (decreasing) returns to scale. For instance, suppose that the production function $F(K, L, P)$ is of the Cobb-Douglas type, that is, $F(K, L, P) = aK^{a_1}L^{a_2}P^{a_3}$, where, $a, a_1, a_2, a_3 > 0$. If $a_1 + a_2 + a_3 = 1$, then F exhibits constant returns to scale. On the other hand, if $a_1 + a_2 + a_3 > 1 (< 1)$ we have increasing (decreasing) returns to scale.

It will probably be interesting to represent different notions of returns to scale graphically. For this let us assume that the long run production function involves only two inputs, K and L . We then write the production function as $x = G(K, L)$. In Fig. 8.6 returns to scale are shown in the firm's production map in terms of the spacing of isoquants representing different output levels. For constant returns to scale, doubling the inputs achieves double the level of the initial output, and so the distance between any two successive multiple-isoquants remains constant. That is, the isoquants are equally spaced as output expands. For decreasing returns to scale, the distance between consecutive multiple-isoquants increases. By doubling the inputs, output increases by less than twice its original level. Finally, for increasing returns to scale the distance between consecutive multiple-isoquants reduces.

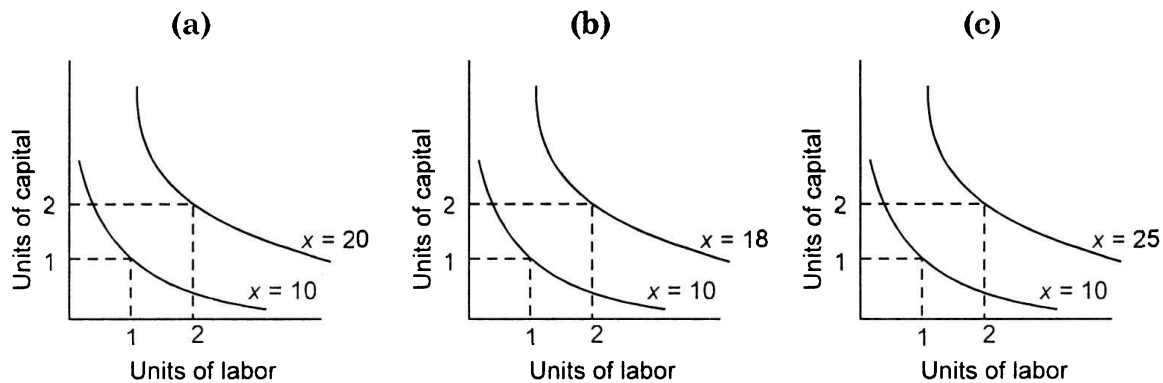


Fig. 8.6 Constant (a); decreasing (b); and increasing (c) returns to scale

In the constant returns to scale case all the isoquants are radial blow-ups of the unit isoquant. Along any ray through the origin (a line along which K/L remains constant) the slope of the isoquant is the same. That is, along any process line the marginal rate of technical substitution between capital and labor is the same. In other words, along any process line the ratio of marginal products of capital and labor is the same, independent of the output level. This property is, in fact, enjoyed by any homothetic production function. A production function is said to be homothetic if it can be written as an increasing transformation of a linearly homogeneous production function. That is, if $G(K, L)$ is linearly homogeneous, then $h[G(K, L)]$ is homothetic, where h is increasing. An example of a homothetic production function is $K + L + b$, where $b > 0$. This non-homogeneous function is homothetic because it is an increasing transformation of the linearly homogeneous production function $G(K, L) = K + L$.

Finally, we may point out that a technology may exhibit different kinds of returns to scale at different stages of production. For instance, it may happen that at low levels of production the technology exhibits increasing returns to scale; over a certain range of outputs there may be constant returns to scale, and; for any output beyond that, decreasing returns to scale may appear.

ELASTICITY OF SUBSTITUTION

An important characteristic of a production function is how easy it is to substitute one input for another. For instance, is it relatively simple to substitute capital for labor along an isoquant? We know that if the production function is linear, input substitution is very easy. In contrast, if a technology

uses inputs in a fixed proportion, input substitution is not possible at all. In the intermediate case where isoquants are strictly convex to the origin, substitution is possible, but cannot be made so easily, since minor variations in the input mix will have a major impact on the inputs' relative productivities.

It seems worthwhile to have a scale-free measure that will indicate this type of responsiveness. As expected, the measure should take on the maximum value (may be plus infinity) for a linear production function, minimal value (zero) if the isoquant is of the fixed proportion type, and some finite positive value for other types of technologies. One such measure is the elasticity of substitution σ defined as

$$\begin{aligned}\sigma &= \frac{\text{Percent change in } (K/L)}{\text{Percent change in } MRTS} \\ &= \frac{\Delta(K/L) \text{ } MRTS}{\Delta(MRTS)(K/L)}\end{aligned}\quad (8.16)$$

That is, for a production function $x = f(K, L)$, the elasticity of substitution (σ) measures the proportionate change in the capital-labor ratio relative to the proportionate change in the marginal rate of technical substitution of capital for labor.

Since in the usual case along an isoquant, $MRTS$ and K/L move in the same direction, so $\sigma > 0$. Graphically we explain this concept in Fig. 8.7. Let us consider a movement from point B to A on the isoquant. In this movement both $MRTS$ and K/L will change and we are interested in the relative magnitude of these changes. If the isoquant is relatively flat, $MRTS$ will not

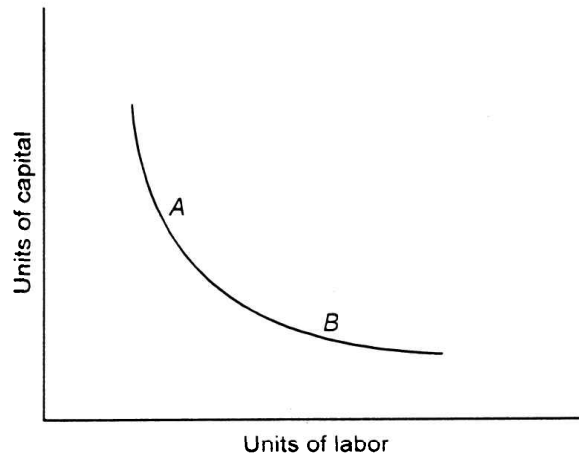


Fig. 8.7 Elasticity of Substitution

change much relative to K/L and σ is high. In the extreme case, when the isoquant is linear, change in $MRTS$, $\Delta(MRTS)$ is zero and hence $\sigma = \infty$ in this case. If the isoquant is sharply curved, $MRTS$ will change by a substantial amount as K/L changes. In this case σ is low. When the isoquant is L-shaped, $MRTS$ is zero, and σ is also zero.

In general, it is possible that the elasticity of substitution will change as one moves along the isoquant. However, if constant returns to scale are assumed, then since all isoquants are merely radial blow-ups of each other, σ will be the same along all isoquants. Many empirical investigations have centered on constant returns to scale, and the constant elasticity of substitution (CES) production function developed by Arrow, Chenery, Minhas and Solow (1961). The mathematical form of this production function is given by

$$x = a[bK^{-\theta} + (1 - b)L^{-\theta}]^{-1/\theta}, \quad (8.17)$$

where $a > 0$, $0 \leq b \leq 1$, $\theta \geq -1$. For $a = 1$, the production function in Eq. (8.17) is the weighted mean of order $-\theta$ of K and L , where the weights are b and $(1 - b)$ respectively. We get the weighted arithmetic, geometric and harmonic means of K and L as special cases of this mean for $\theta = -1$, 0 and 1 respectively. The parameters have been given the following interpretation; a is an efficiency parameter since it can shift the whole production process; b is a distribution parameter that shows the relative importance of K and L in the aggregation; and θ is a substitution parameter because it can be shown that $\sigma = 1/(1 + \theta)$. When $\theta = -1$, the CES production function becomes linear and in this case $\sigma = \infty$. When $\theta = 0$, the CES production function becomes the Cobb-Douglas form $x = aK^b L^{1-b}$, and the corresponding elasticity of substitution is 1. On the other hand, when $\theta \rightarrow \infty$, $x \rightarrow a$ times $\min [bK, (1 - b)L]$, the Leontief production function and $\sigma = 0$.

An important point that should be pointed out here is that in Eq. (8.17) we have defined σ involving the K/L ratio and $MRTS$ of K for L . We can also define σ in terms of L/K and $MRTS$ of L for K . However, the two definitions are identical.

ALTERNATIVE MODELS OF A FIRM: A BRIEF DISCUSSION

In situations of imperfect information about the demand curve, firms may opt out for output decision rules that require less knowledge than profit maximization. A particularly simple alternative is that of sales maximization. This idea was put forward by Baumol (1967). When firms are uncertain

about the demand curve they actually face or when they have no highly reliable notion of marginal costs of their output (as may be true for multi-product firms), a decision to try to maximize sales may be a reasonable alternative. Indeed a number of consulting firms stress the importance of maximizing their market shares.

A strictly revenue maximizing firm will choose to produce that level of output for which the marginal revenue is zero. Selling further units will cause total revenue to fall. We know from Chapter 5 that marginal revenue

$$MR = p(1 + \frac{1}{e}), \quad (8.18)$$

where p is the price and e is the elasticity of demand. Therefore, $MR = 0$ implies that $e = -1$. That is, demand will be unit elastic at the sales maximizing output level.

A firm that chooses to maximize revenues does not give any attention to its costs or to the profitability of its sales. It may be possible that the sales maximizing output level generates negative profit for the firm. Since no firm can survive forever with extremely low profit, it may be more realistic to assume that firms must meet some minimum profitability level for their activities. Now, for positive profit we require positivity of marginal revenue. A positive marginal revenue means that the demand must be elastic ($e < -1$) at the corresponding output level. Therefore, if a firm maximizes sales subject to the condition that it will meet some profitability level, it will produce an output level somewhere in the output region whose extreme points are profit maximizing and sales maximizing output levels.

Firms' contracts with workers may be quite complex. They may incorporate provisions that will encourage workers to stay with firms. Such provisions (for instance, compensation allowance) may affect firms' output decisions.

Workers' contracts may also contain risk-sharing arrangements. For instance, if workers are risk averse but firms are not, firms may be required to absorb risks arising from uncertain demand in the product market.

Contracts with managers will also reflect incentive issues. Since managers act as agents of firm owners, provisions may be required for bringing profit-maximizing behavior more close to managers' utility maximizing behavior. Otherwise managers may not take decisions that are consistent with profit maximization.

Many firms are organized on a non-profit basis. Examples of such firms include public schools, most hospitals and philanthropic foundations. For such organizations the hypothesis of profit maximization is inappropriate, and some other objectives (e.g. welfare maximization) may be required.

Example 8.1

Let us consider the Cobb-Douglals production function $x = aK^bL^{1-b}$, where $a > 0$, $0 < b < 1$. We show that the law of diminishing MRTS holds. The marginal products of labor and capital are $f_L = \delta x / \delta L = a(1-b)K^bL^{-b} > 0$ and $f_K = \delta x / \delta K = abK^{b-1}L^{1-b} > 0$. Since $f_{LL} = \delta f_L / \delta L = a(1-b)(-b)K^bL^{-b-1} < 0$ and $f_{KK} = \delta f_K / \delta K = ab(b-1)K^{b-2}L^{1-b} < 0$, the law of diminishing marginal productivity holds. Now, $f_{KL} = \delta f_K / \delta L = ab(1-b)K^{b-1}L^{-b} > 0$. Therefore, we have $f_K > 0$, $f_L > 0$, $f_{KK} < 0$, $f_{LL} < 0$ and $f_{KL} > 0$. This shows that the law of diminishing MRTS holds. (See Appendix 8.1)

Let us now interpret the parameters b and $(1-b)$. Note that $Lf_L/x = (1-b)$ and $Kf_K/x = b$. Thus, the powers of L and K are simply the elasticities of output with respect to L and K respectively.

Applied Example 8.1 Returns to Scale in Indian Agriculture

Economists have conducted many empirical studies of real world production relationships. Here we look at the study made by Saini (1969) to evaluate the efficiency with which farmers in two Indian states (Uttar Pradesh and the Punjab) use their resources for crop production. The following form of agricultural production function was chosen for this purpose:

$$Q = CA^{b_1}L^{b_2}B^{b_3}F^{b_4}I^{b_5},$$

where Q = Gross value of output (in rupees), A = land (acres), L = human labor (adult men-days), B = bullock labor (pair days), F = farm manures and fertilizers (in rupees), I = irrigation expenditure (in rupees) and C is a constant (which may be regarded as the efficiency parameter).

The function has been estimated by the ordinary least squares method for both Uttar Pradesh and the Punjab using farm management data for the years 1955–56 and 1956–57. The returns to scale are increasing, constant or decreasing accordingly $b_1 + b_2 + b_3 + b_4 + b_5$ is greater than, equal to or less than one. Saini's finding confirmed results of some earlier investigations indicating constant returns to scale in Indian agriculture.

Marginal products of land in both states turned out to be very close to each other in both the years. Marginal products of human labor (except for Uttar Pradesh in 1955–56) moved in line with the level of market wages for casual labor in the respective region. Marginal products of other inputs turned out to be positive in all cases with the exception of manures and fertilizers in Uttar Pradesh for 1955–56. The negative figure in this case was a consequence of excessive rainfall in the region in 1955–56.

Irrigation gave the highest return per unit of cost in 1955–56. Farmers' response to this economic opportunity has been reflected in higher expenditure on irrigation in the following year. In general, farmers' behavior was found to be rational in terms of their response to other economic opportunities of this type. This rationality however does not imply that farmers always succeeded in operating their farms at economically optimum levels. Existence of an excess of marginal products over factor costs in the two states suggest that farmers were not always efficient allocators of resources. It should however be kept in mind that while farmers make *ex ante* decisions on farming, evaluation of their efficiency is based on an *ex post* examination of their decisions. Realized results may not fully reflect the results desired at the time of decision making.

Appendix 8.1

PROOF OF THEOREM

In this appendix we will first give a formal proof of theorem 8.1. Differentiating $AP_L = x/L = f(K, L)/L$ partially with respect to L we get

$$\begin{aligned}\frac{\delta(AP_L)}{\delta L} &= \frac{L \frac{\delta f(K, L)}{\delta L} - f(K, L)}{L^2} \\ &= \frac{\frac{\delta f(K, L)}{\delta L} - \frac{f(K, L)}{L}}{L} \\ &= \frac{MP_L - AP_L}{L}.\end{aligned}$$

Therefore, if AP_L is increasing, then $\delta(AP_L)/\delta L > 0$, that is, $MP_L > AP_L$. Similarly, if AP_L is maximum, $MP_L = AP_L$, and if AP_L is decreasing, $MP_L < AP_L$. This completes the proof.

DIMINISHING MRTS

We now show that it is not possible to derive diminishing MRTS from the assumption of diminishing marginal productivities only. Let f_L be the marginal product of labor, that is, $f_L = \delta f(K, L)/\delta L$. Similarly, we write $f_K = \delta f(K, L)/\delta K$ for the marginal product of capital. Both f_L and f_K are positive. MRTS (of K for L) = f_L/f_K . Diminishing MRTS requires that $d(\text{MRTS})/dL < 0$. Now,

$$\frac{d(\text{MRTS})}{dL} = \frac{\{f_K[f_{LL} + f_{LK}(dK/dL)] - f_L[f_{KL} + f_{KK}(dK/dL)]\}}{(f_K)^2}, \text{ where } f_{LL} = \delta f_L /$$

$\delta L, f_{LK} = \delta f_L / \delta K$, and so on. Using the fact that $dK/dL = -f_L/f_K$ along an isoquant and the fact that $f_{KL} = f_{LK}$, we have

$$\frac{d(\text{MRTS})}{dL} = \frac{f_K^2 f_{LL} - 2f_K f_L f_{KL} + f_L^2 f_{KK}}{(f_K)^3}$$

Since we have assumed that $f_K > 0$, the denominator of $d(\text{MRTS})/dL$ is positive. Hence $d(\text{MRTS})/dL$ is negative if the numerator of the expression is negative. Now, the diminishing marginal productivity of capital (labor) means that $f_{KK}(f_{LL}) < 0$. It clearly follows that diminishing marginal productivities, that is, $f_{KK} < 0, f_{LL} < 0$, are not sufficient alone to guarantee that $d(\text{MRTS})/dL < 0$. However, if we assume that $f_{KL} > 0$, then $d(\text{MRTS})/dL < 0$. Therefore, given positivity of marginal products and the law of diminishing marginal productivity, a sufficient condition for diminishing MRTS to hold is that $f_{KL} > 0$.

CES AND COBB-DOUGLAS PRODUCTION FUNCTIONS

Finally, we show that the CES production function

$$x = a[bK^{-\theta} + (1-b)L^{-\theta}]^{-\frac{1}{\theta}}$$

becomes the Cobb-Douglas production function for $\theta = 0$. To see this note that

$$\begin{aligned}\log x &= \log a - \frac{1}{\theta} \log [bK^{-\theta} + (1-b)L^{-\theta}] \\ &= \log a - H(K, L), \quad \text{say.}\end{aligned}$$

For $\theta = 0$, $H(K, L)$ becomes $0/0$. Hence L' Hospital's rule is applicable. According to L' Hospital's rule, for two functions g and h , if $g(a) = h(a) = 0$ and if h' , the derivative of h around a is not zero, then

$$\lim_{x \rightarrow a} \frac{g(x)}{h(x)} = \lim_{x \rightarrow a} \frac{g'(x)}{h'(x)},$$

where g' is the derivative of g .

Now, differentiating the numerator and denominator of $H(K, L)$ with respect to θ we get

$$-\frac{bK^{-\theta} \log K + (1-b)L^{-\theta} \log L}{bK^{-\theta} + (1-b)L^{-\theta}},$$

which for $\theta = 0$ becomes $-\log [K^b L^{1-b}]$. Therefore, for $\theta = 0$, $\log x$ becomes $\log a + \log [K^b L^{1-b}]$, which in turn means that $x = a K^b L^{1-b}$ for $\theta = 0$.

EXERCISES

Multiple Choice

- Suppose that by increasing the quantity of labor used by 1 unit, a firm can give up 3 units of capital and can still produce the same amount of output. Then the marginal rate of technical substitution of labor for capital is
 - 1/3.
 - 3.
 - 1/3.
 - 3.
- When the average product of labor is positive but declining, the marginal product of labor is
 - Decreasing.
 - Zero.
 - Negative.
 - Any of the above.
- If inputs cannot be scaled downward without loss of productivity,
 - Increasing returns to scale occur.
 - Decreasing returns to scale occur.

- c) The law of variable proportions occur.
 - d) Any of the above may occur.
4. If inputs cannot be scaled upward without loss of productivity,
- a) Increasing returns to scale occur.
 - b) Decreasing returns to scale occur.
 - c) Constant returns to scale occur.
 - d) None of the above occurs.
5. How many of the following production functions do not satisfy the law of diminishing returns to labor? (i) $f(K, L) = K + L$, (ii) $f(K, L) = \min(K, L)$, (iii) $f(K, L) = KL$, (iv) $f(K, L) = \max(K, L)$, where $K(L)$ stands for capital (labor) input?
- a) One.
 - b) Two.
 - c) Three.
 - d) None.

Quantitative Problems and Questions

(For solving some parts of some problems you may need to consult Chapters 9 and 10.)

1. Prove or disprove the following.
- a) The law of diminishing marginal productivity does not apply to a production function that exhibits constant returns to scale.
 - b) Consider a two-factor production function $x = f(K, L)$ that exhibits constant returns to scale. Then if the marginal productivity of labor (L) is greater than the average productivity of labor, the marginal productivity of capital (K) is negative.
 - c) Constant returns to scale occur when inputs can be scaled upward or downward with no productivity loss.
 - d) It is never possible to have decreasing marginal product of an input and increasing returns to scale in a production process.
 - e) The two-factor production function $x = f(K, L)$ exhibits constant returns to scale. Then the law of diminishing marginal productivity ensures that $f_{KL} > 0$, where $f_{KL} = \partial^2 f(K, L) / \partial K \partial L$.
 - f) A two-factor production function $f(K, L)$ satisfies strict concavity and the condition $f(0, 0) = 0$. Then f exhibits decreasing returns to scale.
 - g) Any homothetic production function is homogeneous.

2. Draw an isoquant for a firm that uses two inputs, K and L to manufacture a product. Explain:
 - a) The downward slope of an isoquant.
 - b) The non-intersection of any two isoquants.
 - c) Strict convexity of an isoquant to the origin.
 - d) The ridge lines.
3. A firm uses one input (L) to produce two outputs using the relation $L = A(x_1^a + x_2^b)$, where x_1 and x_2 are quantities of outputs, and A , a and b are positive constants. The firm buys the inputs and sells the outputs at fixed prices. Express the firm's profit maximizing outputs as functions of prices.
4. The production function of a firm is given by $x = -L^3 + 24L^2 + 240L$, where L is the number of workers employed and x is the firm's output. Suppose that the wage rate is Rs. 12 and each worker works for 40 hours.
 - a) What is the minimum product price at which the firm will operate?
 - b) Suppose that the firm's profit is Rs. 1,096. (The firm has no control over product price.) In order to achieve this level of profit, the firm employs 16 workers. What is the firm's fixed cost?
5. Iron is produced using iron ore in blast furnaces. The larger the blast furnace, the more iron ore is produced. What factors should determine whether a firm increases the size or the number of furnaces when increasing output?
6. A firm's production function $x = -0.1L^3 + 6L^2 + 12L$, where x is output and L is the number of workers employed.
 - a) How many workers are employed if the average product of labor is maximized?
 - b) How many workers are employed if the marginal product of labor is maximized?
 - c) What quantity of x is produced if the average variable cost is minimized (wage rate is assumed to be given)?
 - d) If the wage bill is Rs. 360 and output price is Rs. 30, how much output should be produced to maximize profit?
 - e) If the wage bill Rs. 510, at what output price will the firm be indifferent between producing and closing down?
 - f) Suppose that the output price is Rs. 10 and that the total fixed cost is Rs. 15,000. What will the firm's profit be if it hires 36 workers?

7. Clearly distinguish between diminishing returns to scale and diminishing returns to a factor. Can a production function exhibit diminishing returns to scale but not diminishing returns to a factor?
8. A firm's production function is $x = -0.2L^3 + 18L^2 + 1620L$, where L is the number of persons employed and x is the output.
 - a) At what output does marginal cost equal the average variable cost?
 - b) If the firm is currently maximizing the average product of labor, by what proportion would it have to increase employment if it wishes to maximize total product?
 - c) By what proportion would output increase if the total product, instead of average product, is maximized?
 - d) If the price of output is Re. 0.20, what is the maximum wage bill that the firm can pay?
9. What is the law of diminishing marginal rate of substitution? Give an argument justifying this law.
10. A firm uses two factors of production. Its production function is $x = -(A_1^3 A_2^3)^{1/2} + 22A_1A_2 + 15(A_1 A_2)^{1/2}$, where A_1 and A_2 are the quantities of the factors 1 and 2, and x is output.
 - a) In the short run $A_2 = 100$.
 - i) Derive the equations for the average and marginal products for the other factor.
 - ii) Over what ranges of output does the firm experience diminishing average and marginal products of factor 1?
 - b) Repeat question (a) treating A_2 as an arbitrarily given constant.
 - c) Both factors are now variable. Derive the equation for the marginal rate of technical substitution of factor 1 for factor 2.
 - d) If the price of factor 1 is 1, and the price of factor 2 is 4, derive an equation in the form $A_1 = f(A_2)$ for the firm's expansion path.
 - e) Along the expansion path of (d), calculate the ranges of output for which the firm experiences increasing, decreasing and constant returns to scale.
 - f) Are the answers to (e) dependent on input prices?
 - g) Derive the equation in the form $A_1 = f(A_2)$ when $x = 1,350$.
 - h) Calculate the firm's maximum possible output.

- i) Assume that the prices of factors 1 and 2 are 4.44 and 9.99 respectively and the output price is 0.12. The firm is a profit maximizer. Calculate (i) the amounts of the factors used, (ii) the amount of output produced, (iii) the firm's profit.
 - j) Let $x = 4y$, $A_1 = 4a_1$ and $A_2 = 4a_2$. Then write the production function in terms of y , a_1 and a_2 .
 - k) Return to the original production function. Assume that the price of both the factors is 10. The firm's expenditure on the two factors is 320. What is the maximum possible output?
 - l) With prices and expenditure as in (k), suppose that the firm has to buy 22 units of factor 1, what is then the maximum output?
11. What is meant by elasticity of substitution? In the production of soap, stearic acid and sodium hydroxide are used in fixed proportions. Capital and labor are also used. What will be the elasticity of substitution between (i) stearic acid and sodium hydroxide (ii) capital and labor, (iii) labor and sodium hydroxide, and (iv) capital and stearic acid?
12. A firm uses one variable factor, labor (L), together with fixed plant to produce a single product. The production function is $x = -0.002L^3 + 0.5L^2 + 29.6L$, where x is the output. The firm is considering investing in new capital equipment which, if acquired, will increase fixed cost. If the new equipment is acquired the production function takes the form $x = -0.002L^3 + 0.6L^2 + 36L$. The wage rate is Rs. 540 and the output price is Rs. 15. The cost of new equipment would be Rs. 65,000. Would it be profitable to acquire the new equipment?
- Suppose that the wage rate increases to Rs. 622.50 and the output price falls to Rs. 10. Would the firm continue to produce if it had not bought the new equipment? Would it continue to produce if it had bought the new equipment?
13. Show that for a constant returns to scale production function the marginal rate of technical substitution depends on the input ratio.
14. Consider the Cobb-Douglas production function $x = L^{a_1} K^{a_2}$, $a_1, a_2 > 0$, $a_1 + a_2 = 1$. Determine the elasticities of output with respect to labor and capital.
15. A firm faces the following production processes involving capital and labor: Process 1 has $K/L = 1$, Process 2 has $K/L = 2$, Process 3 has $K/L = 1/2$. The production function for the processes are $x = K$, $x = 2L$ and $x = 2K$ respectively, where x is the output. Identify the efficient processes.

16. Let $f(K, L)$ be a production function that exhibits constant returns to scale. Show that the marginal products MP_K and MP_L are homogeneous of degree zero. Hence or otherwise show that along any ray the slopes of all the isoquants are identical.
17. Suppose that a firm seeks to maximize total revenue subject to a minimum profit constraint. Show graphically that the firm will choose an output level in the interval $[\underline{x}, \bar{x}]$, where $\underline{x}$ and $\bar{x}$ are the profit and revenue maximizing outputs respectively.
18. A firm can produce some good at a constant average (and marginal cost) of 4. The demand for the good is given by $x = 100 - 10p$. Assume that the firm owner requires a profit of 80 from the sale of the good. Determine the amount of output that yields maximum revenue if this profit constraint is to be fulfilled.

9

Theory of Costs

A firm's set of feasible production processes may be quite large. From it the firm chooses the input-output mix that maximizes the firm's profits. As a profit maximizing entity, the firm chooses a particular level of output. Whatever output level is chosen, profit maximization implies that the output should be produced at the minimum possible cost. This can be demonstrated as follows. Since profit is equal to revenue minus total costs, if the firm is not a cost minimizer, then there exists a way to increase profit. But this contradicts the assumption that the firm is a profit maximizer. Therefore, a profit maximizing firm minimizes cost of production.

We subdivide the problem of profit-maximization into two components. First, we will look at the problem of cost minimization for any arbitrarily given level of output, and then we will choose the most profitable output level. In this chapter we will consider the first step, that is, the problem of choosing inputs for producing any given level of output as cheaply as possible. The second step, that is, the choice of output by the profit maximization criterion, will be discussed in Chapters 10–12.

OPPORTUNITY COST

Before we analyze the theory of costs, we will discuss economists' approaches to the idea of cost from the decision making viewpoint. Since resources are limited, any decision in our economy to produce some good necessitates giving up some other good. For instance, when an automobile is produced, an implicit decision has been made to do without, say 10 bicycles that could have been produced using the labor, metal, glass, etc., that went into the production of the automobiles. In this case we say that the opportunity cost of one automobile is 10 bicycles. The opportunity cost of an action is the value of the best available opportunity foregone because of the action. Since it is not possible always to express opportunity costs in terms of physical goods, it is convenient to use monetary units. Indeed the price of an automobile may adequately reflect the goods that were given up due to its production. If this were true, we would say that the opportunity cost of an automobile is Rs. 60,000 worth of other goods. The opportunity cost doctrine is extremely important in economic analysis. Many problems of social choice are made conceptually clearer by recognizing the alternatives that are inherent in the economic process.

Unless specified explicitly, in this chapter we will assume that all factor markets are perfectly competitive in the sense that buyers and sellers take the market price as given, independent of the quantities of different factor services bought or sold. Under this assumption, with free entry and exit from markets, the market price for any input service equals the opportunity cost of input. To understand this, suppose that there is some industry that values the services of an input at Rs. 50/hour, this being the maximum value in the economy. If this input is not paid Rs. 50/hour in all industries in a competitive economy, then the owners of the input will move from the low-paying to high-paying jobs, bringing about equality of the rate of payment throughout the economy. Clearly, competition and free entry and exit are crucial in arriving at this conclusion.

SHORT RUN AND LONG RUN COST FUNCTIONS

We have already made a distinction between short run and long run situations in the context of analysis of a firm's production plan. In the short run a firm operates within a given industry with a fixed plant capacity. In the long run a firm can freely choose all the inputs and move from industry to industry.

The short run can be regarded as a situation in which the firm contracts for the use of a specific plant and chooses the best input-output mix for the plant. In the long run the firm looks around in the rental markets for plants and markets for other inputs also.

The short run is a situation in which the law of variable proportions applies to the technology of a firm and the number of firms in an industry is fixed. On the other hand, in the long run the returns to scale characterize the technology of the firm and the number of firms in an industry is a variable.

As in Chapter 8, we will assume that in the short run there are two variable inputs, labor services L and capital services K (labor hours and machine hours), combined with a fixed plant capacity P . Let w be the wage rate per unit time (say, per week) of labor services, and let r be the rental rate per unit of time (say, per week) of capital services. In the short run we will take the fixed cost, the cost of the fixed factors, as the rental payment for the use of plant facilities. In the long run K , L and the plant capacity P are all variable. Recall that in this case the firm's production function is $F(K, L, P)$. We will denote the rental rate per unit of time of the plant services by α . A firm's short run total cost function (or briefly, short run cost function) $C_s(r, w, \alpha, x)$ is defined as the minimum cost of producing x units of output when factor prices are (r, w, α) , given that the firm's fixed plant capacity is P . That is,

$$C_s(r, w, \alpha, x) = \min_{K, L} rK + wL + \alpha P \quad (9.1)$$

such that $F(K, L, P) = x$. In general, the minimum cost to produce x units of output in the short run will depend on the amount and the cost of the fixed plant.

The long run total cost function (or briefly, long run cost function) $C(r, w, \alpha, x)$ is defined by

$$C(r, w, \alpha, x) = \min_{K, L, P} rK + wL + \alpha P \quad (9.2)$$

such that $F(K, L, P) = x$. Here all factors are variable. The long run total cost function depends on the output level and factor prices only.

Several interesting properties follow directly from the definition of a cost function (short run or long run). For instance, the cost function is increasing in factor prices. That is, if one or more factor prices increase, then total cost increases. The cost function is linearly homogeneous in factor prices—an equiproportional variation in all factor prices changes the cost proportionally.

When factor prices are assumed to be constant, we can write cost as a function of output x alone. Therefore, under constancy assumption of input prices we can denote the short run cost function by $C_s(x)$ and the long run cost function by $C(x)$.

SHORT RUN ANALYSIS

Classification of resources in the short run as fixed and variable allows us to classify the costs also as fixed and variable. Variable costs are the costs of variable factors. Fixed costs are those of fixed resources. We will see in our discussion of total, average and marginal costs that the distinction between fixed and variable costs is a very important issue.

Short Run Cost Minimization Problem

The variable costs for the firm are, by definition,

$$rK + wL. \quad (9.3)$$

The problem for the firm is to choose K and L so as to minimize Eq. (9.3). As stated earlier, mathematically this is a constrained optimization problem and we write the problem formally as

$$C_v(x) = \min_{K,L} wL + rK \quad (9.4)$$

such that $F(K, L, P) = x$. Therefore, we can write the firm's total cost function $C_s(x)$ defined in Eq. (9.1) as

$$C_s(x) = C_v(x) + FC \quad (9.5)$$

where $FC = \alpha P$ is the firm's fixed cost.

In order to understand the solution to the optimization problem of Eq. (9.4), let us present the firm's cost and technological constraints graphically in the same diagram (see Fig. 9.1). The isoquant shown in the diagram gives us the technological constraints—it shows all the combinations of K and L that can produce the output level x .

Let us plot all capital-labor combinations that have a constant level of cost c . We can write this as

$$rK + wL = c. \quad (9.6)$$

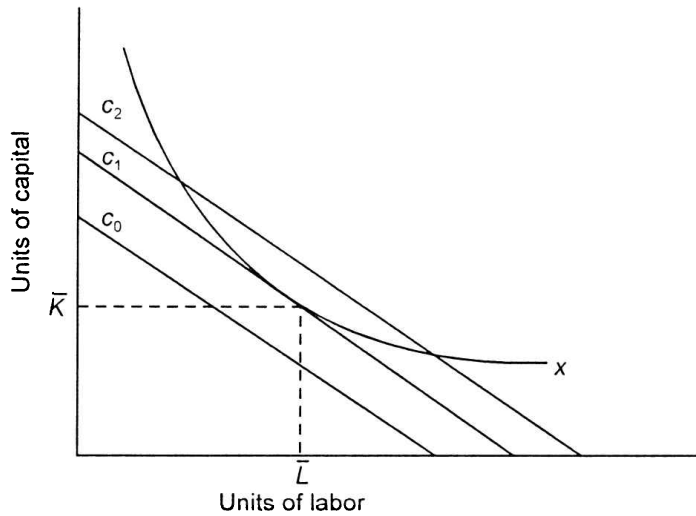


Fig. 9.1 Cost Minimization

Such a line is called an isocost line—a line along which the variable cost of the firm is a constant. Given that the firm's variable cost is c , it can rent a maximum of c/r units of capital. Similarly, c/w gives the maximum of labor that could be hired. If we allow the constant c to vary, we get a family of isocost lines. Every point on an isocost line has the same cost and higher isocost lines are associated with higher costs. Note also that the slope of an isocost line equals $-w/r$.

The cost minimization problem can now be restated as: determine the point on the isoquant that has the lowest possible isocost line associated with it. Such a point is illustrated in Fig. 9.1. Three isocost lines are shown in Fig. 9.1: $c_0 < c_1 < c_2$. It is clear from the figure that the minimum total cost for producing x is given by c_1 where the isocost line is tangent to the isoquant. The cost minimizing input combination occurs at $\bar{K}, \bar{L}$. At a point of tangency, the slope of the isoquant equals the slope of the isocost line. Thus, $MRTS = w/r$, equivalently, $MP_L/MP_K = w/r$.

Therefore, given any output level, to minimize costs the firm employs capital and labor services such that the ratio of their marginal products equals the ratio of their per unit costs.

We can also demonstrate this result algebraically. Let us consider a change $(\Delta K, \Delta L)$ in the pattern of the production that keeps the output constant, that is, the change $(\Delta K, \Delta L)$ takes place along the isoquant $f(K, L) = x$. Then

$$MP_K \Delta K + MP_L \Delta L = 0. \quad (9.7)$$

Clearly, ΔK and ΔL are of opposite signs. That is, if ΔK is positive (negative), then ΔL is negative (positive).

If costs are minimized, then this change cannot lower costs for producing x , so

$$r\Delta K + w\Delta L \geq 0. \quad (9.8)$$

Now look at the change $(-\Delta K, -\Delta L)$. This change also keeps output constant and cannot reduce cost. Hence

$$-r\Delta K - w\Delta L \geq 0. \quad (9.9)$$

From Eqs. (9.9) and (9.8) it follows that

$$r\Delta K + w\Delta L = 0. \quad (9.10)$$

Solving Eqs. (9.7) and (9.10) we get

$$-\frac{\Delta K}{\Delta L} = \frac{w}{r} = \frac{MP_L}{MP_K} \quad (9.11)$$

Fig. 9.1 exhibits the formal similarity between the firm's cost minimization problem and a consumer's utility maximization problem. In the consumer's problem the budget line replaces the isocost line. The consumer moves along the budget line to find the most preferred position. In the firm's problem the firm moves along the isoquant to find the optimal position. In both cases we take prices as fixed parameters and derive the tangency conditions.

In the consumer's problem we constructed the familiar downward sloping demand curve of a good by varying its price. An interesting question is whether an analogous development can be made for the firm's demand for an input. The analogy of the utility maximization process can be misleading here. The choice of inputs that yield minimal costs for the firm are conditional factor demands or derived factor demands, based on the firm's output. We cannot answer questions about an input demand, as its price changes, without looking at the interaction of the supply and demand in the goods market. We have to know what happens to the output level. The derivation of a firm's demand for an input involves issues about the firm's desired output level that do not arise in the consumer's problem. A complete discussion on the firm's demand for inputs is presented in Chapters 14 and 15.

Revealed Cost Minimization

Suppose that we observe two sets of input prices (r^1, w^1) and (r^2, w^2) . The corresponding input choices of the firm are (K^1, L^1) and (K^2, L^2) . Each of these

choices is assumed to produce the same amount of output. Given that each choice of inputs is a cost minimizing choice at the prevailing prices, we have

$$r^1 K^1 + w^1 L^1 < r^1 K^2 + w^1 L^2, \quad (9.12)$$

$$r^2 K^2 + w^2 L^2 < r^2 K^1 + w^2 L^1. \quad (9.13)$$

These inequalities can be referred to as the Weak Axiom of Cost Minimization (WACM).

We rewrite Eq. (9.13) as

$$-r^2 K^1 - w^2 L^1 < -r^2 K^2 - w^2 L^2. \quad (9.14)$$

Adding Eqs. (9.12) and (9.14) we get

$$(r^1 - r^2)K^1 + (w^1 - w^2)L^1 < (r^1 - r^2)K^2 + (w^1 - w^2)L^2, \quad (9.15)$$

which can be rearranged as

$$(r^1 - r^2)(K^1 - K^2) + (w^1 - w^2)(L^1 - L^2) < 0. \quad (9.16)$$

If we adopt the Δ notation to denote changes in factor prices and factor demands, then inequality (9.16) becomes

$$\Delta r \Delta K + \Delta w \Delta L < 0. \quad (9.17)$$

Inequality (9.17) is a consequence of the cost minimizing assumption of a firm.

Suppose $\Delta r < 0$ and $\Delta w = 0$, that is, the price of capital decreases and wage rate remains fixed. Then for (9.17) to hold we must have $\Delta K > 0$, which means that the demand for capital services goes up. Hence the conditional factor demand functions must be downward sloping.

Dual to the Cost Minimization Problem

A result identical to that derived in Eq. (9.11) can be obtained by considering the firm's dual output maximization problem. For a given level of variable input costs (say c_1) maximize the level of output. Mathematically, the problem is

$$\max_{K, L} f(K, L) \quad (9.18)$$

such that $rK + wL = c_1$. The solution to this problem is identical to that derived in Eq. (9.11). A graphical demonstration is given in Fig. 9.2. There the maximum output attainable with variable cost is x and this results when the capital-labor combination $(\bar{K}, \bar{L})$ is employed. All other combinations of input that lie below c_1 are below x -isoquant. Therefore, the solution derived in Fig. 9.2 will coincide with that in Fig. 9.1

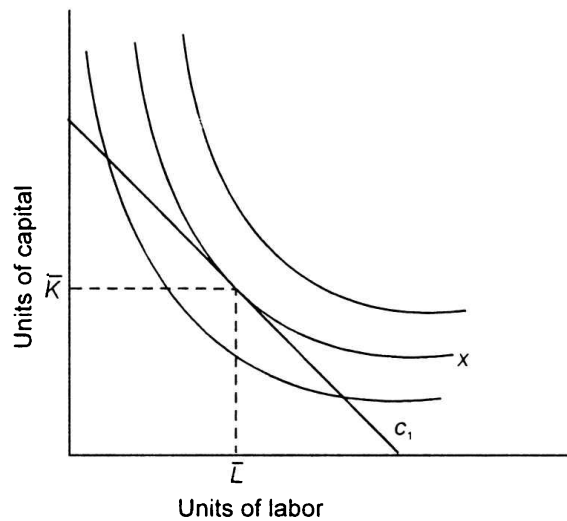


Fig. 9.2 Dual Output Maximization Problem

The Firm's Expansion Path

For each output level x , a firm can find the input choice that minimizes the cost of producing x . If input prices remain constant for all amounts of the inputs the firm may demand, we can trace out the locus of cost-minimizing choices. This procedure is shown in Fig. 9.3. The line OE connects the cost-minimizing tangency points for different levels of output. For instance, the minimum cost for producing output level x_1 is given by c_1 and (K_1, L_1) is the chosen capital-labor combination. Other tangency points in the figure can be interpreted similarly. For given plant capacity and input prices, the combinations of input hirings that represent cost minimization for the firm, as a function output level, is called the firm's short-run expansion path. In Fig. 9.3, the line OE is the short run expansion path because it records how input usage expands as output expands while holding input prices and plant capacity fixed.

The expansion path of a firm need not be a straight line. The use of some inputs may increase faster than others as output expands. Which inputs increase more rapidly will depend on the shape of the isoquant. The shape of the expansion path reflects the substitution that occurs as output increases. Since cost minimization requires equality between MRTS and the factor price ratio, and since the factor price ratio is a constant, the shape of the expansion path will be determined by where a particular MRTS occurs on successively higher isoquants. If the production function (involving only two factors) exhibits constant returns to scale, then the expansion path will be linear since the MRTS depends on the ratio K/L .

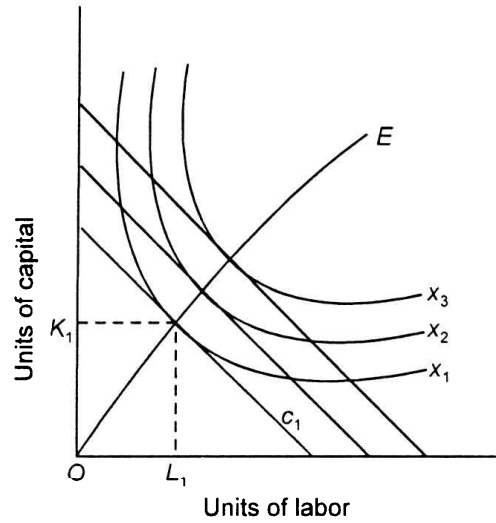


Fig. 9.3 Short Run Expansion Path

Average and Marginal Cost Functions

Although the total cost function gives complete information about the output-cost relationship, sometimes it is convenient to analyze costs on a per unit of output basis. The per unit approach corresponds quite closely to our analysis of demand which is based on the per unit price of a commodity. Two different notions of unit costs are widely used in economic analysis: (i) average cost, that is, the cost per unit of output, and (ii) the marginal cost, that is, the cost of one more unit of output.

The short run average total cost (ATC) function is found by computing total costs per unit of output:

$$ATC(x) = \frac{C_s(x)}{x}. \quad (9.19)$$

The average variable cost (AVC) function measures the variable cost per unit of output:

$$AVC(x) = \frac{C_v(x)}{x}. \quad (9.20)$$

The average fixed cost (AFC) function is obtained by dividing fixed cost by output:

$$AFC(x) = \frac{FC}{x}. \quad (9.21)$$

Since $C_s(x) = C_v(x) + FC$, it must be the case that

$$ATC(x) = AVC(x) + AFC(x). \quad (9.22)$$

Suppose that a firm is producing output level x and that we wish to see how costs will change if we change output by some small amount Δx . The marginal cost $MC(x)$ measures the change in costs relative to change in output:

$$\begin{aligned} MC(x) &= \frac{C(x + \Delta x) - C(x)}{\Delta x} \\ &= \frac{C_s(x + \Delta x) + FC - C_s(x) - FC}{\Delta x} \\ &= \frac{C_s(x + \Delta x) - C_s(x)}{\Delta x} \end{aligned} \quad (9.23)$$

since $C(x) = C_s(x) + FC$ and the fixed costs do not change if x changes. The marginal cost is the increase (decrease) in total costs brought about by one unit increase (decrease) in output. It indicates a rate of change: change in costs divided by change in output.

It is important to note that average and marginal costs depend both on the level of output being produced and the prices of inputs. In the next subsection we graph relationships between different costs and output. All such graphs are drawn on the assumption that the prices of inputs remain constant. If input prices change, cost curves generally shift to new positions. In Chapter 10 we look at the impact of input price changes on a firm's output.

Cost Curves

Fig. 9.4 shows the firm's short run cost curves. In panel (a) of the figure total costs and total variable costs are plotted against the output levels. The summation of total fixed costs and total variable costs for various output levels comprise the total costs of the firm for those output levels.

Suppose we have determined the least cost combination of labor and capital for producing one unit of output. Denote this combination by (K_1, L_1) . Thus, $C_s(1) = rK_1 + wL_1$. It is also true that $F(K_1, L_1, P) = 1$, where P is the firm's fixed plant capacity.

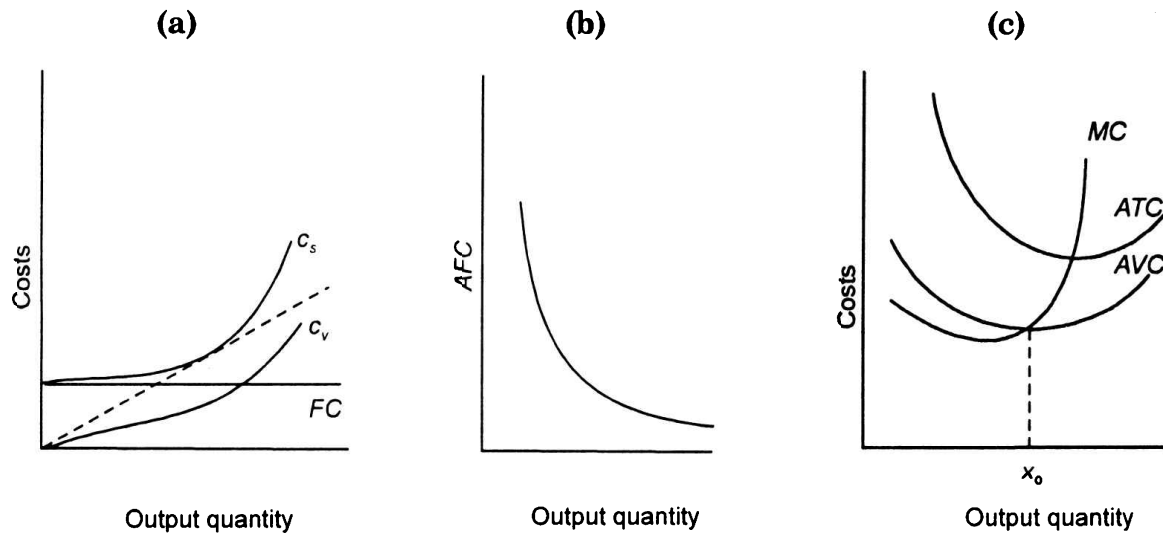


Fig. 9.4 Short Run Cost Curves

We know that in the short run the law of variable proportions applies. Suppose that x_0 is the output level beyond which the law of variable proportions works. Let $x_0 = F(K_0, L_0, P)$, where $K_0 = t_0 K_1$ and $L_0 = t_0 L_1$ for some $t_0 > 1$. Since beyond x_0 the law of variable proportions begins to operate, increasing the variable inputs proportionally leads to a less than proportionate increase in output. Therefore, for any $t > t_0$, $F(tK_1, tL_1, P) < tF(K_1, L_1, P)$. Since factor prices are constant, a proportional variation in inputs (capital and labor) changes the variable cost proportionally. Thus, for any output greater than x_0 , less than proportionate increases in output, which result from proportionate increases in capital and labor, are accompanied by proportionate increases in variable costs. Equivalently, we can say that when output $x(> x_0)$ increases proportionally, variable costs increase more than proportionally.

For capital-labor combinations below (K_0, L_0) equal proportional additions in capital and labor are assumed to increase output more than proportionally. At low levels of variable inputs, the number of units of the fixed input per unit of each variable input is quite large. This largeness increases the productivity of the variable inputs and output increases at an increasing rate. Using arguments similar to those employed in the preceding paragraph we can now establish that when output $x(< x_0)$ increases proportionally, variable costs increase less than proportionally.

Mathematically we can say that $C_v(tx)$ is smaller or larger than $tC_v(x)$ for all $t > 1$ accordingly x is less than or greater than x_0 . Therefore, the variable cost curve possesses the shape shown in panel (a) of the figure.

The total cost curve and the variable cost curve must have the same shape, since for each increase in output we get total costs by simply adding the fixed costs with variable costs. That is, the total cost curve is parallel to the variable cost curve, and the constant difference between the two curves at each output level is the fixed cost. Thus, the total cost curve also possesses the shape shown in the figure.

The average fixed cost measure is shown in panel (b) of Fig. 9.4. The greater the output of the firm, the smaller AFC will be. Since fixed costs remain the same regardless of output, as output expands, each unit of output bears a smaller share. Therefore the AFC curve is downward sloping to the right throughout its entire length. Since $AFC = FC/x$, $x AFC = FC$. Therefore the AFC curve is a rectangular hyperbola. As output increases it approaches—but never reaches—the quantity axis. Thus it is clear that firms with large fixed costs—for instance, the railways with their tremendous charges on roadbeds and rolling stock—can substantially reduce their per unit costs by increasing outputs.

The AVC , ATC and MC curves are shown in panel (c) of Fig. 9.4. The AVC curve has a U shape and it achieves its minimum at x_0 . Since for output x less than x_0 , proportional increase in x is accompanied by less than proportionate increase in variable costs, $AVC(x)$ is decreasing for $x < x_0$. (See Appendix 9.1 for a rigorous proof.) Similarly, we can explain why $AVC(x)$ is increasing for $x > x_0$. It achieves minimum at $x = x_0$.

To illustrate the U-shape further let us consider the following example. Suppose that a factory is designed to employ approximately 100 workers. The size of plant is fixed and labor is the only variable factor. The amount of output if only one worker is employed will be very small, but if an additional person is employed, the two can divide the job and more than double the output. In other words, the average product of labor increases with the employment of an additional worker. If doubling of labor costs will more than double output, average labor cost will decrease. When enough workers are employed, the average production of labor decreases, or equivalently, average variable costs increase. Thus, the average variable cost curve is a sort of monetized mirror image of the average product curve of labor. The ATC curve is the sum of the AFC and AVC curves. It will have the U-shape indicated in panel (c) of Fig. 9.4. The initial decline in ATC is due to decline in AFC ; the eventual increase in ATC is a consequence of increasing AVC . The combination of these two effects generate the U-shape shown in the figure.

MC is the slope of either—total cost curve or the variable cost curve—they are parallel and hence have the same slope at each x . The slope of the vari-

able cost curve is decreasing between zero and output x_0 and increasing beyond x_0 . Thus, marginal cost decreases up to x_0 and then increases as output rises.

We now determine the position of the MC curve relative to that of the AVC curve. For the first unit of output, AVC and MC are equal. To see this, note that when zero unit of output is produced, the variable cost is zero. Therefore, marginal cost for the first small unit amount of output is

$$\frac{C_v(1) - C_v(0)}{1} = \frac{C_v(1)}{1} = AVC(1) \quad (9.24)$$

which is the average variable cost for the single unit of output.

Now, suppose that AVC decreases. Thus, it must be the case that the cost of each additional unit of output is less than the average cost up to that point. This means that marginal costs are less than the average variable costs, when average variable costs are decreasing. Similarly, when average variable costs are rising, marginal costs are greater than average variable costs. The higher marginal costs push the average up in this range.

Thus, the MC curve must lie below the AVC curve to the left of its minimum point and above it to the right. This implies that the MC curve must intersect the AVC curve at its minimum point.

We can demonstrate analogously that if ATC is falling, MC must be less than ATC . On the other hand, if ATC is rising, MC must be greater than ATC . The MC curve passes through the minimum point of the ATC curve.

We summarize the above results in the following theorem.

Theorem 9.1

- a) *When average variable costs are decreasing, the marginal costs are less than the average variable costs.*
- b) *When average variable costs are increasing, the marginal costs are greater than the average variable costs.*
- c) *The marginal cost curve must pass through the minimum points of the average variable and the average total cost curves.*

The downward sloping portions of the U-shaped ATC , AVC and MC are a consequence of the assumption that the given plant capacity promotes increasing returns to variable inputs initially. However, eventually all these curves turn out to be increasing, since beyond a certain output level the law of variable proportions applies.

If the law of variable proportions applies at the outset, AVC rises to the right over all output levels and MC lies above AVC for all outputs. MC and AVC coincide at the first unit of output. Since AFC decreases continuously, the ATC curve will take a U-shape in this case also. Further, the MC curve which increases continuously passes through the minimum point of the ATC curve. We show these curves in Fig. 9.5.

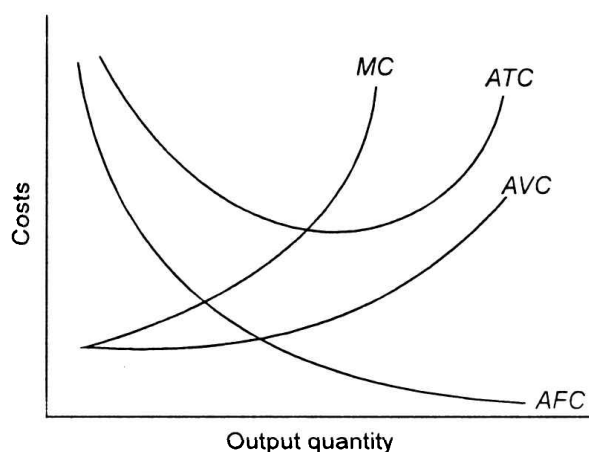


Fig. 9.5 Short Run Cost Curves with Diminishing Returns to Variable Inputs

LONG RUN ANALYSIS

In the long run planning period, the firm is free to choose the plant capacity it wishes. The firm can change the quantities per unit of time of land, buildings, heavy machinery, management and all other resources. There will be no fixed costs. We need to determine the shapes of the long run total cost curve, the long run average cost curve and the long run marginal cost curve. As stated earlier, long run total cost for a firm for a given level of output is the total cost associated with the least-cost combination of inputs (including plant capacity) required for producing that level of output. Long run average cost is the long run total cost per unit of output level. Long run marginal cost is the long run total cost when output increases by one unit.

The specific characteristics of the long run cost curves depend largely on the returns to scale properties of the production function. Below we discuss shapes of the cost curves under different notions of the returns to scale.

Returns to Scale and Cost Curves

Suppose first that constant returns to scale characterize the technology. Let K_1, L_1, P_1 be the least-cost combination of capital, labor and plant for producing

one unit of output. Thus $C(1) = rK_1 + wL_1 + \alpha P_1$. Now, what is the cheapest way to produce x units of output? We know that $F(K_1, L_1, P_1) = 1$. Therefore, by constant returns to scale $F(xK_1, xL_1, xP_1) = xF(K_1, L_1, P_1) = x$. It thus follows that the minimal cost $C(x)$ to produce x units of output is $x C(1) = \beta x$, say, where $\beta = C(1) > 0$. That is, the long run total cost curve is a straight line passing through the origin.

In the constant returns to scale case, the average cost (AC) function is given by

$$AC(x) = \frac{C(x)}{x} = \beta. \quad (9.25)$$

Thus, the long run average cost is a constant [equal to $C(1)$] for every level of output. Similarly the long run marginal cost (MC) equals β , since increasing output by one unit increases total cost by β . In other words, the slope of the long run cost curve is β for every x . We show long run total cost, average cost and marginal cost curves in panels (a) and (b) of Fig. 9.6.



Fig. 9.6 Long Run Total Costs (a); and Average and Marginal Costs (b) (under constant returns to scale)

Consider now the case of increasing returns to scale. Suppose the firm doubles the inputs K_1 , L_1 and P_1 . Then in view of increasing returns to scale $F(2K_1, 2L_1, 2P_1) > 2F(K_1, L_1, P_1)$. That is, the firm will more than double its output. Since factor prices are constant, by using twice as much of every unit, we simply double costs. Therefore, if the firm decides to double its inputs, it doubles its cost but more than doubles its output. That, is, the firm is

producing *more* than double the output, at double the costs. This is just equivalent to saying that the cost function will increase at a decreasing rate or less than linearly with respect to output. We show this function graphically in panel (a) of Fig. 9.7.

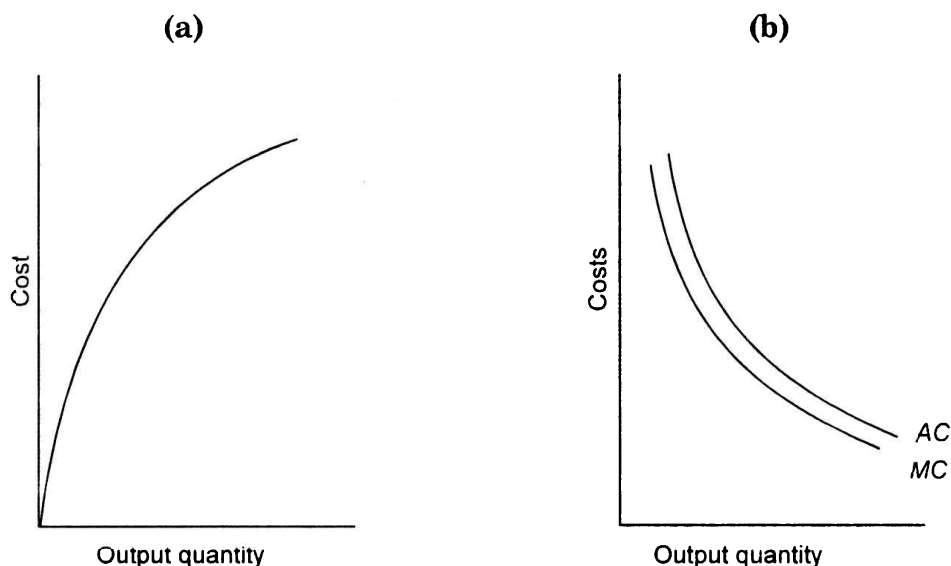


Fig. 9.7 Long Run Total Costs (a), and Average and Marginal Costs (b) (under increasing returns to scale)

The long run average cost decreases as output increases under increasing returns to scale, because doubling all inputs results in more than a doubling of output. More precisely, in the long run since all inputs are variable, $C(0) = 0$. Consequently, strict concavity of the cost function $C(x)$ will imply decreasingness of $C(x)/x$. (See Appendix 9.1 for a rigorous demonstration.) We say that a cost function has economies of scale at production level x , if the average cost is decreasing around x .

Since the marginal cost function is the slope of the total cost function and since under increasing returns to scale total cost increases at a decreasing rate, the marginal cost must be declining in this case. Further, decreasingness of average cost ensures that marginal cost always lies below the average. We show both average and marginal cost functions graphically for the increasing returns to scale case in panel (b) of Fig. 9.7.

In panel (a) of Fig. 9.8 we show the long run total cost function under decreasing returns to scale. The total cost function increases at an increasing rate or more than linearly with respect to output. Thus, if output doubles, costs will more than double. The average and marginal cost functions in this case will rise as output rises and the marginal cost curve will always lie

above the average cost curve. These two functions are shown in panel (b) of Fig. 9.8. When average cost is increasing around output level x , we say that we have diseconomies of scale.

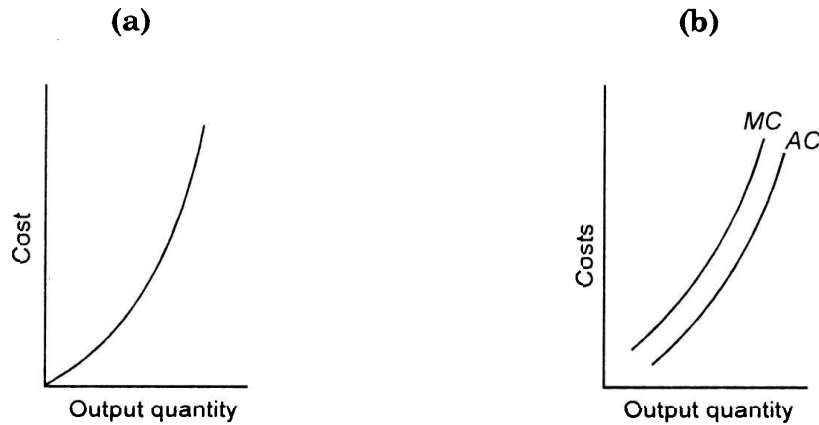


Fig. 9.8 Long Run Total Costs (a), and Average and Marginal Costs (b) (under decreasing returns to scale)

It may happen that the returns to scale associated with a firm's production function change with the firm's output. For outputs less than x_1 , increasing returns to scale operate; greater than x_1 , decreasing returns to scale occur and at x_1 constant returns occur. Then we obtain a U-shaped long run average cost curve, as shown in Fig. 9.9. At x_1 long run marginal cost equals long run average cost. The long run marginal cost lies below the long run average cost to the left of x_1 , and above it lies to the right of x_1 .

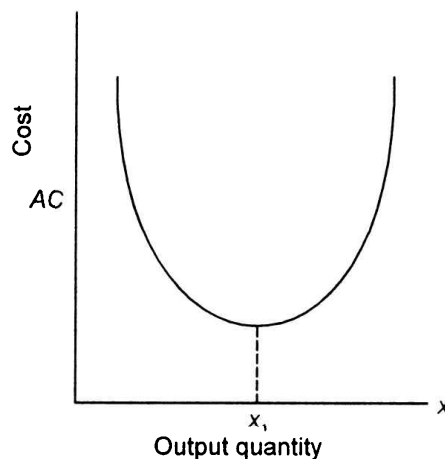


Fig. 9.9 U-Shaped Long Run Average Cost

Long Run and Short Run Costs

By considering all possible variations in plant size, we can establish the relationship between the short run cost curves and the fully flexible long run cost curves. The short run cost function, given that the firm has a plant of size P , will be denoted by $C_s(x, P)$.

For any given output level x , there will be some plant size which is optimal (that is, long run cost minimizing) to produce that output. We denote this plant size by $P(x)$. Clearly, this is the firm's conditional factor demand for plant as a function of output. Then $C_s[x, P(x)]$ is the minimum cost when plant size is fixed at the level that minimizes long run costs. This is the total cost of producing the output level x , given that the firm is free to adjust its plant size optimally. The long run cost function is simply the short run cost function evaluated at the optimal choice of the plant size:

$$C(x) = C_s[x, P(x)]. \quad (9.26)$$

Since in the short run the firm has a fixed plant size and in the long run the firm is free to adjust its plant size, the long run total cost can never exceed the short run total cost at any level of output. Let us consider an arbitrary output level x_0 . Suppose that $P_0 = P(x_0)$ is the optimal plant size for producing x_0 . The short run cost function for a plant of size P_0 is $C_s(x, P_0)$. Since one of the firm's long run choices can be P_0 , its optimal choice for producing x units of output, must have total costs at least as small as $C_s(x, P_0)$. Thus,

$$C(x) \leq C_s(x, P_0) \quad (9.27)$$

for all x . The above inequality becomes an equality if $x = x_0$, that is, at the output level x_0 , $C(x_0) = C_s[x_0, P(x_0)]$. In words, we say that the short run cost is always greater than the long run cost and they are equal at one level of output. This establishes the graphical relation between the long run and short run (total) cost curves.

By dividing both sides of inequality. Eq. (9.27) by x , we note that $AC(x) \leq ATC(x, P_0)$, where $ATC(x, P_0)$ is the short run average cost for producing x units of output with plant size P_0 . For $x = x_0$, $AC(x_0) = ATC(x_0, P_0)$. That is, the short run average cost curve always lies above the long run average cost curve and they touch at one output level only. This shows tangency between the two average cost curves.

If we do this exercise for several output levels, say, $x_1, x_2, \dots, x_n$, then we get a picture of the type shown in Fig. 9.10 for a U-shaped long run average cost curve. We say that the long run average cost curve is the envelope of the short run average cost curves—it touches each and intersects none.

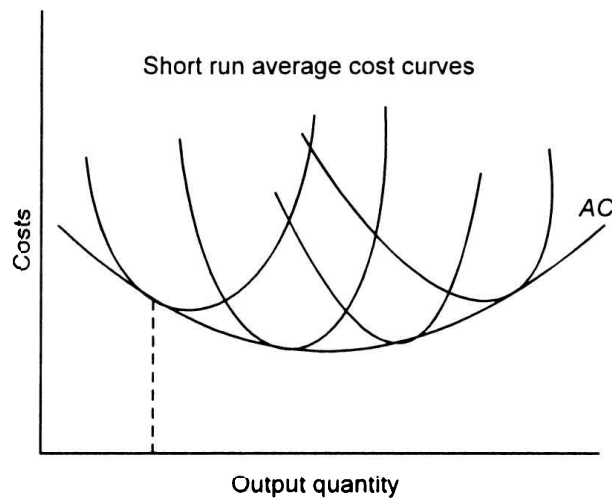


Fig. 9.10 Long Run and Short Run Average Cost Curves

We have shown how to construct the U-shaped long run average cost curve as the envelope of the short run average cost curves. We can also show that the long run total and average curves under alternative assumptions about returns to scale can be derived as envelopes of short run total and average cost curves respectively.

Long Run Marginal Cost

When the firm has constructed the proper size of plant for producing a given output, the short run marginal cost will be equal to the long run marginal cost at that output. Suppose, for instance, that the given output level is x_0 in panel (b) of Fig. 9.11. The firm would use the plant represented by the

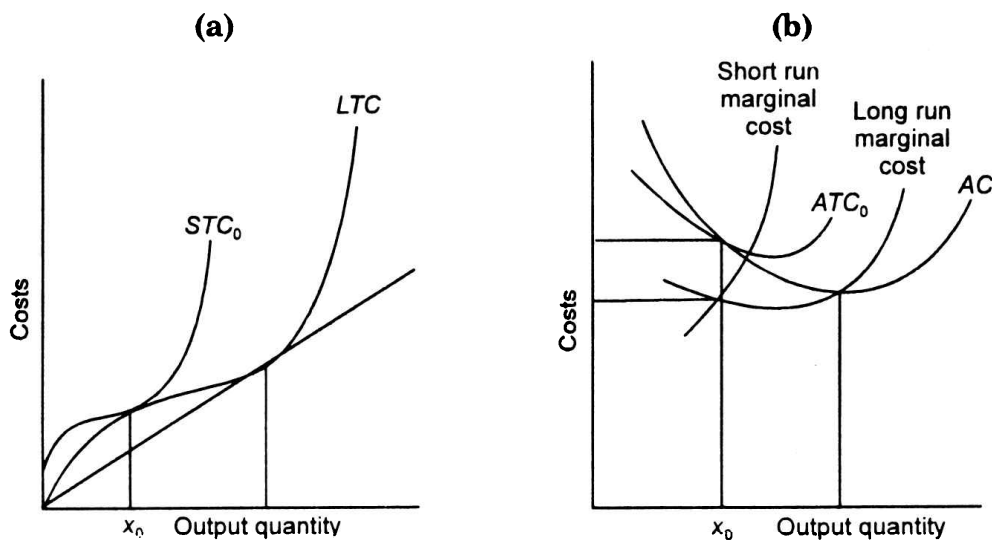


Fig. 9.11 Short Run and Long Run Costs: Total Costs (a), and Marginal Costs (b).

short run average cost curve ATC_0 , which is tangent to the long run average cost curve AC at that output. The corresponding short run and long run total cost curves are STC_0 and LTC in panel (a) of Fig. 9.11. STC_0 lies above LTC for $x < x_0$ because ATC_0 is greater than AC at these output levels. Equality of ATC_0 and AC at x_0 implies that STC_0 and LTC are equal at x_0 . At outputs greater than x_0 , STC_0 will exceed LTC again because ATC_0 exceeds AC for these outputs also. At x_0 , where ATC_0 is tangent to AC , STC_0 must also be tangent to LTC . At outputs around but below x_0 , STC_0 must have a smaller slope than LTC ; at outputs greater than x_0 , STC_0 must have a greater slope than LTC . At x_0 , where STC_0 is tangent to the LTC , both curves have the same slope.

Since the slope of the STC_0 curve is the short run marginal cost and the slope of the LTC curve is the long run marginal cost, it follows that short run marginal cost is greater or less than long run marginal cost accordingly output is greater or less than x_0 . At x_0 , the two marginal costs are equal.

ELASTICITY OF SUBSTITUTION RECONSIDERED

We know that if all input prices increase by the same proportion t , the total cost for producing any given output level will also increase by t . If only one input price changes, the situation becomes more complicated. Such a price change alters the ratios of the input's prices, the firm's cost minimizing choice of inputs will be affected and a new expansion path must be derived.

One way to measure the change in input choices in response to a change in price of an input is to examine how the ratio of input usage (K/L) changes as w/r changes, while keeping output constant. That is, we wish to examine $\Delta(K/L)/\Delta(w/r)$ along an isoquant. Putting this in proportional terms as

$$h = \frac{\Delta(K/L)w/r}{\Delta(w/r)K/L} \quad (9.28)$$

gives an alternative and more intuitive definition of the elasticity of substitution. The measure h is non-negative. An increase in w/r must be accompanied by an increase in K/L . (In the fixed proportion case K/L is constant.) Large values of h mean that the firm changes its input proportions significantly in response to changes in input prices, while small values indicate that input price changes will have little impact.

If there are only two inputs, the elasticity of substitution considered in Chapter 8 is identical to h . This follows from the fact that a cost minimizing

firm will equate its MRTS (of K for L) to the input ratio (w/r). In Appendix 9.1 we employ the definition Eq. (9.28) to calculate the elasticity of the CES production function considered in Chapter 8.

An advantage of the measure h is that it can be easily generalized to the many-input case. We define the partial elasticity of substitution between inputs i and j with prices w_i and w_j as

$$h_{ij} = \frac{\Delta(y_i/y_j)w_j/w_i}{\Delta(w_j/w_i)y_i/y_j} \quad (9.29)$$

under the assumption that output and all other input prices are held constant, where $y_i(y_j)$ is the quantity of input $i(j)$. The measure h_{ij} is quite flexible since it allows the firm to alter usage of inputs i and j when input prices change, whereas other inputs were held constant in the earlier definition.

Example 9.1

A firm's production function is given by $f(K, L) = K^{\frac{1}{2}} + L^{\frac{1}{2}}$. We want to find the expansion path of the firm and its total cost as a function of output x .

To determine the expansion path we maximize output x subject to the expenditure constant $rK + wL = c$. For this let us consider the Lagrangian expression

$$\mathcal{L} = K^{\frac{1}{2}} + L^{\frac{1}{2}} - \lambda(rK + wL - c)$$

where λ is the Lagrange multiplier. The first order conditions for the constrained maximization problem are

$$\frac{\partial \mathcal{L}}{\partial K} = \frac{1}{2} K^{-\frac{1}{2}} - \lambda r = 0,$$

$$\frac{\partial \mathcal{L}}{\partial L} = \frac{1}{2} L^{-\frac{1}{2}} - \lambda w = 0,$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = -rK - wL + c = 0.$$

Using the first two conditions we get

$$\frac{L^{\frac{1}{2}}}{K^{\frac{1}{2}}} = \frac{r}{w}. \quad (9.30)$$

Thus, with r and w given, the right hand side is a constant. If we denote r/w by b , then the above equation becomes $K = b^{-2}L$, which is a straight line expansion path through the origin.

To determine total cost as a function of output we need to consider the problem of minimization of cost for each level of output. We have noted that the dual to this primal cost minimization is: For a given total cost of inputs, maximize the level of output. It has also been observed that the solution to the primal problem is identical to that in the dual case. We therefore consider here the dual problem.

Now, from the first order conditions for output maximization subject to a given cost constraint we have $K = L(r/w)^{-2}$ (see Eq. [9.30]). Substituting this into the production function we have

$$\begin{aligned} x &= (r/w)^{-1} L^{\frac{1}{2}} + L^{\frac{1}{2}} \\ &= [r/w)^{-1} + 1] L^{\frac{1}{2}} \end{aligned}$$

Letting $a = (r/w)^{-1} + 1$, we have $L = (x/a)^2$ and $K = b^{-2} (x/a)^2$.

Thus, the cost function is

$$\begin{aligned} C(x) &= rK + wL \\ &= r^2 b^{-2} \left(\frac{x}{a}\right)^2 + w \left(\frac{x}{a}\right)^2 \\ &= r \left(\frac{wx}{r+w}\right)^2 + w \left(\frac{rx}{r+w}\right)^2 \end{aligned} \tag{9.31}$$

Clearly, $C(x)$ is increasing in x , w and r , and linearly homogeneous in w and r .

Applied Example 9.1 Cost Indices

Two economic measures that are used quite frequently in the literature on production theory are the total and marginal cost functions. The total cost function can be employed to compare the total cost of producing an arbitrarily given output level under two factor price regimes. Similarly, the marginal cost function can be used to compare the cost of producing an additional unit of output under two different factor price vectors.

To see how the total and marginal cost functions can be used for comparing costs in two different factor price set-ups, let us suppose that there are

k factors of production. Assume that we wish to compare production costs between periods 0 and 1. We refer to these periods as the base and current periods respectively. Let $w^0 = (w_1^0, w_2^0, \dots, w_k^0)$ be the vector of input prices in period 0. Similarly, we write $\underline{w}^1$ for the input price vector $(w_1^1, w_2^1, \dots, w_k^1)$ in period 1. The total costs of producing the output level x in periods 0 and 1 are $C(\underline{w}^0, x)$ and $C(\underline{w}^1, x)$ respectively. The corresponding marginal cost functions are $MC(\underline{w}^0, x)$ and $MC(\underline{w}^1, x)$.

Baye, Deily and Jansen (1991) suggested the ratio of the total cost of producing a given amount of output in the current period to the corresponding cost in the base period as a total cost index. More precisely, their total cost index is given by

$$T_R = \frac{C(\underline{w}^1, x)}{C(\underline{w}^0, x)}. \quad (9.32)$$

The marginal cost index suggested by these authors is

$$M_R = \frac{MC(\underline{w}^1, x)}{MC(\underline{w}^0, x)} \quad (9.33)$$

As an index of total cost Chakravarty and Chattopadhyay (1994) suggested the use of

$$T_A = C(\underline{w}^1, x) - C(\underline{w}^0, x). \quad (9.34)$$

Clearly, T_A is a measure of the absolute change in the total cost for producing the output level x in the current period over what it was in the base period. Analogously, the absolute change M_A in the cost of producing one more unit of output in period 1 over the corresponding cost in period 0 can be taken as a marginal cost index. More precisely,

$$M_A = MC(\underline{w}^1, x) - MC(\underline{w}^0, x). \quad (9.35)$$

Chakravarty and Chattopadhyay (1994) illustrated these indices using Indian cotton textile industry data for the years 1974–85, a volatile period as far as changes in the governments and policies are concerned. They estimated the following simplified form of the CES production function

$$F(\underline{y}) = \left(\sum_{i=1}^k y_i^{-\theta} \right)^{-1/\theta} \quad (9.36)$$

where $y = (y_1, \dots, y_k)$ is the vector of inputs, with y_i being the quantity of the i th input and then derived the associated cost function. The cost function associated with the CES production function is

$$C(\underline{w}, x) = x \left(\sum_{i=1}^k w_i^{\frac{\theta}{\theta+1}} \right)^{\frac{\theta+1}{\theta}} \quad (9.37)$$

where $\underline{w} = (w_1, \dots, w_k)$ is the input price vector. (This is left as an exercise.)

They used aggregated data on the cotton textile industry. The factors of production considered are materials, energy and labor. The corresponding factor prices were also available. The estimated value of the only parameter θ turned out to be 0.58255. This value of the parameter along with the values of the factor prices were used to calculate the different cost indices.

Both the cost of producing the base year output (8,283 million meters of textiles) and the cost of producing an additional million meters of textiles were calculated for the base year and for all other years. The total cost index T_A gives the change in cost of producing 8,283 million meters of textiles and the marginal cost index M_A shows changes in the cost of producing an additional million meters of textiles at that level of output. Similarly, T_R makes a relative comparison between production costs of 8,283 million meters of textiles and M_R compares costs of producing an additional million meters of textiles in proportional terms. The indices T_R and M_R showed a non-decreasing trend except for the year 1981. T_A reaffirmed this except for the year 1985. However, M_A showed the non-decreasing trend (except for 1981). The relative price of the textiles exceeded T_R from 1979 onwards. The relative change in price turned out to be much higher than the corresponding change in cost during the period. This shows that under the assumption of constant output, profitability in the Indian cotton textile industry has increased after 1979.

Appendix 9.1

COST MINIMIZATION

In this appendix we will first study the cost minimization problem considered in the text by the Lagrange criterion. Since P is fixed, we write $F(K, L, P)$ as $f(K, L)$. Then the problem can be restated as

$$\min_{K,L} rK + wL$$

subject to $f(K, L) = x$.

The Lagrangian expression for this problem is

$$\mathcal{L} = rK + wL - \lambda[f(K, L) - x],$$

where λ is the Lagrange multiplier. The first order conditions for a constrained minimum are

$$\frac{\delta \mathcal{L}}{\delta K} = r - \lambda \frac{\delta f(K, L)}{\delta K} = 0,$$

$$\frac{\delta \mathcal{L}}{\delta L} = w - \lambda \frac{\delta f(K, L)}{\delta L} = 0,$$

$$\frac{\delta \mathcal{L}}{\delta \lambda} = x - f(K, L) = 0.$$

The last condition is simply the constraint. From the first two conditions we have

$$\frac{w}{r} = \frac{\delta f(K, L)/\delta L}{\delta f(K, L)/\delta K}, \quad (9.38)$$

which says that the factor price ratio must equal the marginal rate of technical substitution.

From the first two conditions we also have

$$\frac{1}{\lambda} = \frac{\delta f(K, L)/\delta K}{r} = \frac{\delta f(K, L)/\delta L}{w}. \quad (9.39)$$

This means that the ratio of marginal benefit (that is, increased output) to the marginal cost should be the same for all inputs.

AVERAGE VARIABLE COST

We have argued in the text that if the variable cost function $C_v(x)$ increases at a decreasing rate, that is, if $C_v(x)$ is an increasing and strictly concave function of output x , then the average variable cost $C_v(x)/x$ is decreasing. Here we establish this claim rigorously. We know that variable cost is zero

for zero unit of output. That is, $C_v(0) = 0$. Therefore, for $0 < x_1 < x_2$, by strict concavity of C_v ,

$$C_v\left[\frac{x_1}{x_2}x_2 + \left(1 - \frac{x_1}{x_2}\right)0\right] > \frac{x_1}{x_2}C_v(x_2) + \left(1 - \frac{x_1}{x_2}\right)C_v(0).$$

Since $C_v(0) = 0$, the above inequality becomes

$$\frac{C_v(x_1)}{x_1} > \frac{C_v(x_2)}{x_2}$$

Since the output levels x_1 and x_2 are arbitrary, the above inequality establishes decreasingness of the average variable cost function $C_v(x)/x$.

Using an analogous argument we can show that for a strictly convex increasing variable cost function, the average is increasing. In order to show that in panel (c) of Fig. 9.4, average variable cost is increasing beyond x_0 , let us shift the origin of production to x_0 . We suppose that it is possible to designate different output levels by a subscript. Then for any $x_i > x_0$ let us define $z_i = x_i - x_0$. Clearly, $C_v(z_0) = C_v(0) = 0$. Assume that $x_i > x_j$ if $i > j$. We can now show using strict convexity of C_v that $C_v(z_i)/z_i > C_v(z_j)/z_j$ if $z_i > z_j$.

We have thus shown that for any real-valued function f defined on $[0, \infty]$ if $f(0) = 0$, then $f(x)/x$ is increasing or decreasing accordingly f is strictly convex or concave. This result also becomes useful in rigorously demonstrating that the average productivity of a variable factor initially increases but eventually decreases due to the assumption that the variable factor initially exhibits an increasing marginal productivity, but after some point marginal productivity diminishes (see panel [b] of Fig. 8.2).

PROOF OF THEOREM 9.1

Next, we give a rigorous proof of Theorem 9.1. When average variable cost is decreasing,

$$\frac{d[C_v(x)/x]}{dx} = \frac{xdC_v(x)/dx - C_v(x)}{x^2} < 0,$$

or,

$$\frac{dC_v(x)}{dx} - \frac{C_v(x)}{x} < 0,$$

or,
$$\frac{dC_v(x)}{dx} < \frac{C_v(x)}{x},$$

which means that $MC < AVC$. If AVC is increasing,

$$\frac{d[C_v(x)/x]}{dx} = \frac{xdC_v(x)/dx - C_v(x)}{x^2} > 0,$$

or,
$$\frac{dC_v(x)}{dx} > \frac{C_v(x)}{x},$$

that is, $MC > AVC$.

When AVC is minimum

$$\frac{d[C_v(x)/x]}{dx} = 0,$$

which will imply that $MC = AVC$. It thus also follows that the MC curve passes through the minimum point of the ATC curve.

ELASTICITY OF SUBSTITUTION

Finally, we calculate the elasticity of substitution h associated with the CES production function $f(K, L) = a[bK^{-\theta} + (1 - b)L^{-\theta}]^{-1/\theta}$. We know that

$$\text{MRTS (of } K \text{ for } L) = \frac{MP_L}{MP_K}.$$

Now, $MP_K = \frac{\delta f(K, L)}{\delta K} = ab[bK^{-\theta} + (1 - b)L^{-\theta}]^{-\left(\frac{1+\theta}{\theta}\right)} K^{-\theta-1}$ and $MP_L = \frac{\delta f(K, L)}{\delta L} = a(1 - b)[bK^{-\theta} + (1 - b)L^{-\theta}]^{-\left(\frac{1+\theta}{\theta}\right)} L^{-\theta-1}$. Therefore, $\text{MRTS} = \frac{(1 - b)K^{\theta+1}}{bL^{\theta+1}}$. By the first order conditions

$$\frac{(1 - b)K^{\theta+1}}{bL^{\theta+1}} = \frac{w}{r}.$$

Therefore,

$$\frac{K}{L} = \left(\frac{b}{1 - b}\right)^{\frac{1}{\theta+1}} \left(\frac{w}{r}\right)^{\frac{1}{\theta+1}},$$

from which it follows that

$$\frac{d(K/L)}{d(w/r)} = \frac{1}{\theta + 1} \left(\frac{b}{1-b} \right)^{\frac{1}{\theta+1}} \left(\frac{w}{r} \right)^{\frac{-\theta}{\theta+1}}.$$

So
$$h = \frac{d(K/L)}{d(w/r)} \frac{w/r}{K/L} = \frac{1}{\theta + 1}.$$

EXERCISES

Multiple Choice

- The marginal costs must equal average costs at the point of minimum average costs. This statement is true:
 - Always.
 - Never.
 - Sometimes.
 - Not possible to say.
- The area under the marginal cost curve is the
 - Variable cost.
 - Fixed cost.
 - Variable cost + fixed cost.
 - Variable cost $\times$ fixed cost.
- If L is the only variable factor used in the production of a good, and if the marginal product of L is 4 and the price of L is 5, then the marginal cost of producing the good is
 - 4/5.
 - 5/4.
 - 1.
 - None of the above.
- If a firm has the long run production function $x = K^3 L^7$ and input prices are constant, then long run marginal cost
 - Equals long run average cost.
 - Is less than long run average cost.

- c) Is greater than long run average cost.
 - d) Any of the above is possible.
5. If a firm is producing its current output at the minimum necessary level of long run average cost for that output, and its long run average cost curve is upward sloping at that output, it follows that the short run marginal cost is
- a) Greater than short run average cost.
 - b) Less than the short run average cost.
 - c) Equal to the short run average cost.
 - d) Any of the above is possible.
6. If we plot labor on the vertical axis and capital on the horizontal axis, and denote the constant prices of labor and capital by p_L and p_K , then the slope of an isocost line will be
- a) $-p_L/p_K$.
 - b) p_L/p_K .
 - c) $-p_K/p_L$.
 - d) p_K/p_L .

Quantitative Problems and Questions

1. Prove or disprove the following.
 - a) If the marginal product of labor, multiplied by the quantum of labor used, is always equal to the marginal product of capital, multiplied by the quantity of capital used then if the price of labor is twice the price of capital, a cost minimizing firm will use half as much labor as it does capital.
 - b) If the marginal cost curve is below the corresponding average cost curve, the average cost curve must be falling.
 - c) A profit maximizing firm always minimizes cost.
 - d) Since fixed costs are constant, and equal to the difference between total costs and variable costs, it follows that the difference between average cost and average variable cost must be constant.
 - e) The assumption of diminishing marginal product implies that the long run average cost curve is upward sloping.
 - f) Suppose that a firm is producing at a point where the ratio between the marginal product of factor 1 and the factor price exceeds the

corresponding ratio for factor 2. Then the firm should increase the amount of factor 1 and decrease that of factor 2.

- g) A firm produces identical product using two production functions. If the marginal cost for the first function is smaller than that for the second, then the firm should reduce production by the second function and increase production by the first.
2. A firm produces a good using plant (P) and labor (L) as inputs. The firm's production function is $x = P^a L^{1-a}$, where $0 < a < 1$. The factor prices are constants.
 - a) Derive the firm's long run total cost function.
 - b) Let $a = 0.5$. Derive the long run average and long run marginal cost functions.
 - c) Let $a = 0.5$ and P be fixed at $\bar{P}$. Derive the short run total cost function.
 - d) Let $a = 0.5$ and $P = 20$. Derive the short run average cost, short run average variable cost, and the short run marginal cost.
 - e) Let $a = 0.5$, $P = 20$ and the prices of P and L be 9 and 4. At what output x will the long run average cost be equal to the short run average cost?
 3. What does the rectangle under the fixed cost curve represent at any level of output? Is the area of the rectangle independent of output?
 4. Explain why the profit maximizing firm will operate where the isocost line is a tangent to an isoquant, but will never operate where an isocost line cuts an isoquant.
 5. Suppose that the cost of capital (K) is Rs. 5 per unit and the cost of labor (L) is Rs. 9 per unit, while the firm's budget is Rs. 100. Draw the isocost line. Suppose that the isoquant is of the form $KL = \bar{a}$, where $a > 0$ is a constant. Calculate the level of output and the levels of K and L that will be hired.
 6. What is meant by the expansion path of a firm? Suggest reasons why the expansion path of a textile producing firm may be labor-intensive at low levels of output and capital-intensive at high output levels. Draw such an expansion path.
 7. A firm has the following long run production function $x = aK^{\frac{1}{5}} L^{\frac{1}{5}} P^{\frac{2}{5}}$, where $a > 0$ is a constant, K , L and P are the inputs of three factors. The prices of the factors are Re. 1, Rs. 9 and Rs. 8.
 - (a) Derive the firm's long run total cost function, long run average cost function and the long run marginal cost function.

- (b) In the short run, P is fixed, and K and L are variable. Derive the firm's short run
- Total cost function.
 - Variable cost function.
 - Average variable cost function.
 - Marginal cost function.
- (c) Obtain an equation of the form $P = f(x)$ showing the optimum quantity of the fixed factor P for the firm to acquire as a function of the intended output x .
8. A firm producing a good from factors K and L can use either of the two following production processes. The production functions are: Process I, $x = aK^{.25} L^{.75}$, process II, $x = aK^{.75} L^{.25}$. The price of the factor K is Re. 1.
- At what price of the factor L will the firm be indifferent between the processes?
 - If the price of L is higher than that calculated in (i), which process will the firm use?
9. A firm uses several factors to produce a single product. In the short run the plant is fixed, while all other factors are variable. We are concerned with two of the possible plants. In the long run all factors are variable. The cost functions are:
- Long run total cost (LC) = $0.005x^3 - 1.4x^2 + 280x$.
- Plant 1: Short run total cost (SC_1) = $0.006x^3 - 1.33x^2 + 201.6x + 6860$
- Plant 2: Short run total cost (SC_2) = $0.0057x^3 - 1.424x^2 + 205.6x + 10,240$.
- Derive equations for
 - Long run average cost.
 - Long run marginal cost.
 - Short run average costs, short run average variable costs, and short run marginal costs under plants 1 and 2.
 - At what output does the firm achieve the minimum of long run average cost?
 - Does either plant permit achievement of the minimum of long run average cost?
 - At what output is the short run average cost under plant 1 minimized?
 - At what output is the short run average cost under plant 2 minimized?

- f) At the output level 160, determine the short run average cost and short run marginal cost under plant 2.
 - g) which of the two plants will the firm use if it wishes to produce the output level given by (e)?
 - h) For what output level will plant 2 be the best choice?
 - i) Would plants 1 and 2 operate if the product price is 120?
 - j) At what product price would the firm produce the same output in the short run whichever of the two plants it used?
 - k) Which plant would be more profitable if the product price is 120?
10. Demonstrate with the help of isoquants and isocost curves that the combination of inputs that maximizes output for a given cost to the firm is the same combination that minimizes cost for the same output level. Explain the logic behind your answer.
11. Given the production function $x = \sqrt{KL}$ and prices per unit of K and L Rs. 2 and 4 respectively, suppose that the total production cost is Rs. 80. Determine the maximum output subject to cost constraint.
12. Given the production function $x = \sqrt{K^{11}L^{11}}$, and prices per unit of K and L are Rs. 8 and 4 respectively, determine the maximum amount of output that can be produced at a given cost of Rs. 400.
13. Suppose that a firm's production function is given by $x = \min(5P, 10L)$ and the prices of the inputs P and L are 1 and 3 respectively.
- a) Calculate the firm's short run marginal, average and total cost functions if $P = 10$.
 - b) Calculate the firm's long run total, average and marginal, cost functions.
14. A person owns two firms. Each firm produces identical product and each has a production function given by $x_i = \sqrt{P_i L_i}$. The firms differ in the amount of equipment each has. In particular $P_1 = 25$ and $P_2 = 100$. The prices of both P and L are 1.
- a) If the person wishes to minimize short run total cost of production, how should output be allocated between the firms? Calculate the short run total average and marginal cost of production, assuming that output has been allocated optimally.
 - b) How should output be allocated in the long run? Calculate the long run total, average and marginal cost of production.

15. Suppose that a firm's production function is of the type $x = K^{a_1} L^{a_2}$ where $a_1, a_2 > 0$. The prices of the inputs K and L are w_1 and w_2 respectively.
 - a) Show that if $a_1 + a_2 = 1$, the firm's total cost function is proportional to x .
 - b) Calculate the firm's marginal cost. Calculate the elasticities of the marginal cost with respect to input prices and show that they are less than one.
 - c) What is the shape of the expansion path for this firm?
16. Consider a two-factor production function whose elasticity of substitution is infinity. Show that the cost minimizing input combination, if it is unique, will generally require the use of only one factor. In such a case, identify the firm's expansion path. On what do the shapes of the marginal and average cost curves depend? How will these curves shift as the price of the input that is used increases?
17. A Firm uses two factors, K and L , to produce a good. The production function is $x = \sqrt{KL}$, where x is output. The price of K is Rs. 3 and the amount of K employed is 900 units. The price of L Rs. 12.
 - a) Determine the amounts of L that the firm needs to employ if $x = 150, 300, 450$.
 - b) What are the marginal costs of the 150th, 300th and 450th units of output?
18. Consider the production function $f(K, L) = \min\{K, L\}$. Find the cost function associated with this production function. What will be the cost function if the production function is $f(K, L) = K + L$?
19. Illustrate the law of variable proportions graphically.
20. Why does the area under the marginal cost curve represent total variable costs?
21. Why does the rectangle under the average total cost curve represent total costs at any level of output?
22. Determine the cost function associated with the production function $f(K, L) = K^\alpha L^{1-\alpha}$, $0 < \alpha < 1$.
23. Show that the partial derivative of the cost function $C(r, w, \alpha, x)$ with respect to a factor price (r, w or α) is the conditional demand function for the factor. (This result is known as Shepherd's Lemma. For proving the result, use the envelope theorem.)

10

Perfect Competition

The purpose of this chapter is to construct a perfectly competitive model of price and output. Perfect competition refers to a market structure in which there is complete absence of rivalry among the constituent firms. We explain how production is organized by prices and profits in a perfectly competitive industry. Perfect competition provides a benchmark against which the behavior of markets can be judged. Even though perfect competition is rarely observed in the real world, it can be valuable to investigate in what respects actual markets deviate from perfectly competitive ones, and to find the circumstances under which the deviations are likely to be the most. However, if a market deviates from the perfectly competitive structure it does not necessarily mean that the performance of the market can be improved. In order to understand these points, it is necessary to understand what is meant by perfect competition and how firms behave in a perfectly competitive framework.

ASSUMPTIONS

The desirable properties of a perfectly competitive market explain why economists generally favor competition. The main assumptions of perfect competition are:

- a) Product homogeneity: Firms produce a homogeneous product.
- b) Many buyers and sellers: There are many buyers and sellers of the product and each buyer and each seller of the product under consideration is too small in relation to the entire market.
- c) Perfect information: All buyers and sellers possess perfect information about the price and quality of the product.
- d) Freedom of entry and exit: Sellers and buyers are free to move in and out of the market in the long run.

Condition (a) ensures anonymity of firms. It means that the product of one firm is indistinguishable from the products of the other firms: trademarks, brand labels, etc., do not exist. There is no way in which consumers can differentiate among the products of different firms.

According to assumption (b) many sellers face many buyers. On the selling side, an individual seller supplies such a small fraction of the total industry output so that if he drops out of the market, total output will not be decreased enough to cause any price rise. Or, if an individual seller sells the maximum amount he can sell, total output will not increase enough to generate any price reduction. On the buying side, an individual buyer takes such a small fraction of the total industry supply so that he is unable to influence the price. Individually consumers have no impact on the prices of most of the items we buy, e.g., meat, milk, etc. Therefore, assumptions (a) and (b) imply that buyers and sellers act as price takers: they take the price at which the product can be sold or purchased as given. Price is determined by the market and no buyer or seller can influence it.

Assumption (c) means that there exists perfect information on both sides of the market. Buyers and sellers possess complete knowledge with respect to the nature and quality of the product and the prevailing price. Given homogeneity of the product and perfect information, a single price must prevail in a competitive market. To see this, suppose on the contrary that there exist two prices for the product. Given perfect information, consumers are aware that the product can be purchased at two different prices and that each unit of the product is identical in all respects. As utility maximizers, consumers will buy the product at the lower price only. Therefore, a single price must prevail in the market. The existence of a single price also ensures that the consumers are identical from the sellers viewpoint in the sense that there are no advantages or disadvantages associated with selling to a particular buyer. Condition (d) means that there is no barrier to entry or exit from the industry. Entry or exit may take time, but there is freedom to decide

whether to stay in or to leave the market. Firms move into markets in which they can earn profits and leave the ones in which they incur loss. Along with movement of firms there will be movement of resources also. The condition ensures free flow of resources between occupations. Resources move into occupations from which they derive maximum benefit.

In our applied example discussed later in this chapter we show how stock transactions may approximate assumptions of a perfectly competitive market.

In some texts price-taking behavior, instead of being derived as an implication of assumptions (a) and (b) stated above, is taken as an assumption of a competitive market. In such a case even if there are only a few buyers and a few sellers in the market, they will treat the market price as being outside their control. An example may be the market for fresh fish. Even if there are a small number of sellers, each seller may have to take the prices of others as given. Given product homogeneity, consumers will buy from the seller charging the lowest price. If any of the other sellers wants to sell his product, he must sell at the lowest price. Thus the lowest price is the market price. In Chapter 12 we show how price-taking behavior can be derived as a particular model of a general conjectural variation model. The analysis carried out in this chapter remains the same whether the price-taking behavior is assumed at the outset or it is drawn as an implication of assumptions (a) and (b).

SHORT RUN ANALYSIS

In short run analysis the number of firms in an industry is fixed. Firms do not have enough flexibility either to enter or to quit a given industry. However, firms in an industry are able to adjust the quantities they are producing in response to changing conditions. They will do this by changing the levels of employment of those inputs that are variable in the short run, and we will investigate this supply decision first for a typical firm and then for the industry as a whole.

Revenue, Average Revenue and Marginal Revenue

A perfectly competitive firm treats the price per unit of its output as a constant. Let p denote the price per unit of output: p is independent of the number of units produced and sold by the firm. We write x to denote the amount of output produced by the perfectly competitive firm under consideration.

The firm's total revenue or simply revenue (R) is its total receipt from the sale of its product. Since the perfectly competitive firm can sell any amount of output at the same price per unit, its revenue or total receipts curve will be a linear, upward sloping curve starting at zero. If the sale of the firm is zero, so is revenue; if the sale is one unit of output, the total receipt is the price of the product. Each one unit addition to the firm's sales will increase revenue by a constant amount—the price per unit of output. Hence the revenue or total receipts curve is upward sloping and linear.

Average Revenue (AR) is the revenue per unit of output and Marginal Revenue (MR) is the addition to the firm's revenue when its sales increase by one unit. Under conditions of competition since the product price is fixed, average revenue is simply the price per unit of output. The increase in total receipts brought about by one unit increase in sales is also the product price—average revenue, marginal revenue and the product price are the same thing for the competitive seller.

Symbolically, the revenue $R(x)$ from the sale of output level x is

$$R(x) = px. \quad (10.1)$$

Average Revenue AR is

$$AR = \frac{px}{x} = p. \quad (10.2)$$

Marginal Revenue MR is

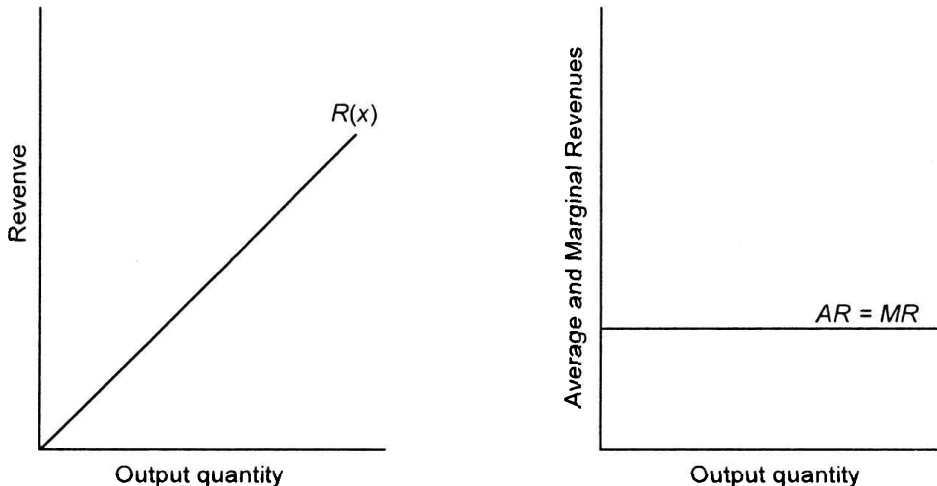


Fig. 10.1 Revenue (a), Average and Marginal Revenues (b), for a competitive firm

$$MR = \frac{\Delta R(x)}{\Delta x} = \frac{p\Delta x}{\Delta x} = p, \quad (10.3)$$

where Δx is the change in output of the firm.

Firm's Short run Profit Maximization

As a starting point we assume that a firm's objective is to maximize its profits or minimize its losses if it cannot make profits. Economic profits or simply profits are defined as total receipts over all costs of production incurred by the firm. In the expression for cost we need to include all the factors of production used by the firm, valued at their market price. For instance, if a person works in his own firm, then his labor should be regarded as an input and it should be counted as part of costs. His wage rate is the market price of labor—the amount he would be earning if he sold his labor in the open market. This means that we should value all inputs at their opportunity costs. When a firm is earning zero economic profit it receives only a normal profit (equal to the implicit opportunity cost) on its owned factors. Thus, in this case the firms receive just enough to pay for their costs of production and no more.

While for labor wages are expressed in terms of money per hour, the analogue for machines is the rental price—the price at which a machine can be hired for the given time period. In many cases firms buy capital equipment since there is not a very well-developed market for the rental of capital. In such a case the implicit rental rate should be calculated using the original price of the machine and its selling price at the end of the period.

The concept of economic profits differs from the accountant's concept of a corporation's net income or profits. The accountant's concept of profits considers historical cost—the original purchase price of a factor rather than economic costs, what a factor would cost if purchased now. For our purpose we will stick to the economic definition of profits.

Let $C(x)$ stand for the total cost or simply the cost of production of the output level x by the competitive firm. Then profits $\pi(x)$ are given by

$$\pi(x) = R(x) - C(x) \quad (10.4)$$

$$= px - C(x). \quad (10.5)$$

To show this graphically we superimpose the cost curve of a firm on the revenue graph as in Fig. 10.2. At any value of x , the vertical distance between the revenue curve and the cost curve denotes profits for the level of output.

At outputs between x_m and x_M , profits are positive. The point at which profits are maximized is the level of output x_0 where the slope of the cost curve equals that of the revenue curve, that is MC (marginal cost) = MR . When profits are maximized, marginal revenue is equal to the marginal cost, that is, the extra revenue gained by one more unit of output equals the extra cost of producing the additional unit.

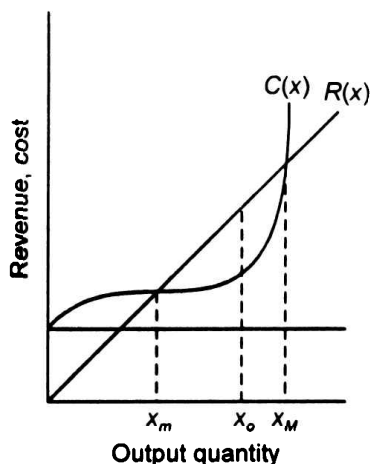


Fig. 10.2 Revenue, Cost and Profits

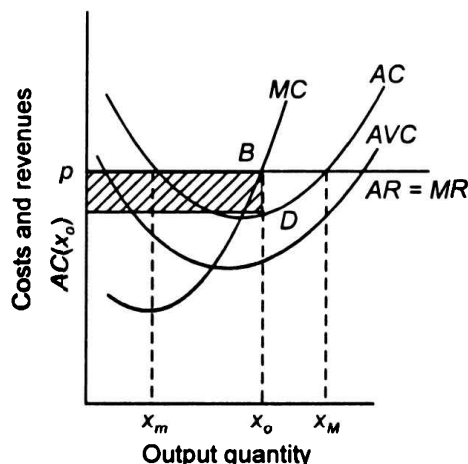


Fig. 10.3 Average Revenue, Average Cost and Profits

The necessary condition for profit maximization can be illustrated using average cost (AC), average variable cost (AVC), MC , AR and MR graphs, as in Fig. 10.3. Profits are zero at x_m and x_M , because $AC = AR = p$ at these points. The profit maximizing output level x_0 is that output which the MC curve crosses the $MR (= AR = p)$ curve. That is, in the short run a profit maximizing, competitive firm will choose a level of output x_0 , where the marginal cost it faces at x_0 is just equal to the market price. Symbolically,

$$p = MC(x_0). \quad (10.6)$$

This profit maximizing condition is also referred to as the equilibrium condition of the firm.

To show that profits are indeed a maximum at x_0 , note that for output levels less than x_0 , price exceeds marginal cost. Consequently, reducing output below x_0 would cut back on revenue more than on costs. Profits would fall. Hence when output is less than x_0 , profits increase by increasing output. Similarly, for output levels greater than x_0 , marginal costs exceed price. Producing more than x_0 would cause costs to rise more rapidly than revenues. Again profits would fall. Therefore, when output is greater than x_0 we

can increase profits by reducing output. Consequently, if a firm produces either more or less than x_0 (where $MR = MC$), it will lower its profits. Only at x_0 , are profits at a maximum.

The profits of the firm appear in Fig. 10.3 as the area of the rectangle $pBDAC(x_0)$. Profit per unit is price p minus average cost at output x_0 . Total profit equals profit per unit multiplied by sales, that is, the total profit equals the product of the difference between price and $AC(x_0)$, the average cost of producing output x_0 , and the number of units of output produced.

We need to consider two points related to the profit maximization problem. First, suppose there are two levels of output where price equals marginal cost, such as the case shown in Fig. 10.4. Which one will the firm choose?

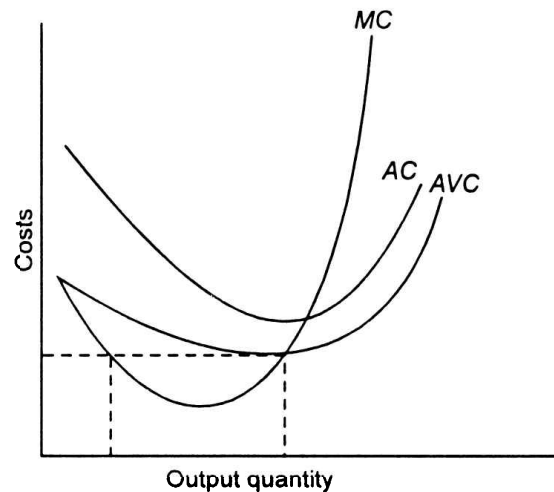


Fig. 10.4 Marginal Cost and Profit Maximizing Output

Look at the first intersection, where the marginal cost is sloping down. Now, if we increase output, the marginal cost, that is, the cost of each additional unit of output, decreases. But since the market price is the same, profits go up. In the case when marginal cost is falling, profits are always larger, the more output is produced. There is no profit maximizing output here.

That price equal marginal cost is a necessary condition for profit maximization. It is not in general a sufficient condition. That is, given price equal to marginal cost at some output level, we cannot conclude that it is the profit maximizing output level. But if profits are maximized at some output, then price must equal marginal cost.

The profit maximizing output level of a competitive firm must lie along the upward sloping portion of the marginal cost curve. If the MC curve crosses

the MR curve twice, then the only point where the MC curve rises to the right is the point of maximum profit.

The second point that we wish to discuss is loss minimization.

Loss Minimization

If it happens that the market price of a product is less than the short run average costs at all output levels, the firm will incur losses instead of making profits. The choices open to the firm are whether to (i) produce at a loss; or (ii) discontinue production. The firm will have to pay fixed costs even if the second alternative is chosen.

The firm's decision rests on whether the price covers average variable costs. If the market price is less than the minimum average variable costs, as in Fig. 10.5, then the firm's revenue is less than the total variable cost. In Fig. 10.5 at x_1 , $MC = MR$, but $p = AR = MR$ lies below the AVC curve. At x_1 , price p is less than AVC $C_v(x_1)/x_1$. The firm loses more than its fixed costs, since at x_1 the firm not only pays fixed costs but also $C_v(x_1) - px_1$, the difference between variable costs and revenue. Its revenue does not even cover the variable costs. However, if the firm produces nothing, it will lose fixed costs only. It loses more if it continues to produce. Therefore, if price is less than the minimum of AVC , then the profit maximizing decision in the short run is to shut down. We refer to the condition

$$AVC(x) = \frac{C_v(x)}{x} > p \quad (10.7)$$

as the shutdown condition.

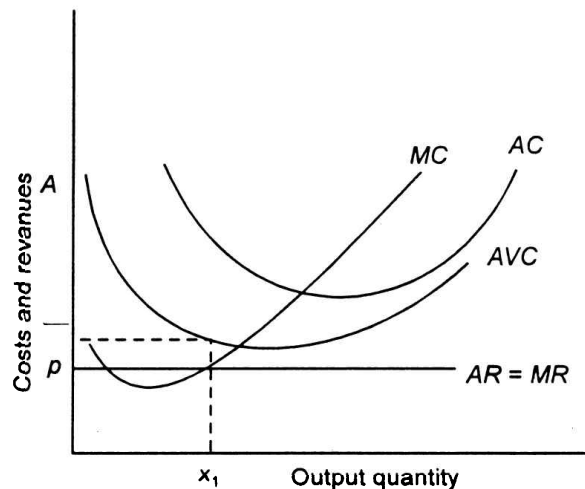


Fig. 10.5 The Shutdown Decision

If variable costs are just covered by revenues, then the firm's loss will be equal to the total fixed costs. It makes no difference whether the firm produces or not—the losses will be equal to fixed costs in either case.

At a price greater than the minimum average variable costs but lower than the average cost, the firm should continue production. At such a price, total receipts exceed the variable costs of production. The excess of total receipts over variable costs can be applied against fixed costs, thus reducing losses to an amount less than fixed costs. Hence it makes it worthwhile for the firm to operate with a positive output.

Profit Maximization, Output Supply and Input Demand

Given any set of prices for output and inputs, the rule of profit maximization enables us to determine the amount of output that competitive firms will supply or produce at that set of prices. Once we know the profit maximizing output level for a firm at given prices of inputs and output, we can determine the number of units of different inputs that will minimize the cost of producing the profit maximizing level of output. If a firm is maximizing profits, and if it chooses to supply some output x , then it must be minimizing the cost of producing x . We show after a while in this section that by varying output price, and keeping input prices constant, we can generate the firm's supply curve of output. Similarly, by varying the price of one input while holding the prices of other inputs and the output price fixed, we can find the firm's demand curve for that input. This is shown in Chapter 14.

Revealed Profitability

Suppose that given the competitive price p^1 , a firm's profit maximizing output level is x^1 . For a different competitive price p^2 , the profit maximizing output of the firm is x^2 . Then

$$p^1 x^1 - C(x^1) > p^1 x^2 - C(x^2) \quad (10.8)$$

and
$$p^2 x^2 - C(x^2) > p^2 x^1 - C(x^1) \quad (10.9)$$

The first inequality says that the profits that the firm achieves for the price-output combination (p^1, x^1) are larger than those for the price-output combination (p^1, x^2) . We can interpret the second inequality analogously. If either of these inequalities is violated, the firm cannot be a profit maximizing firm. The fulfillment of these inequalities is a profit maximizing behavior and therefore it can be referred to as the Weak Axiom of Profit Maximization (WAPM).

Adding the inequalities in Eqs. (10.8) and (10.9) we get

$$p^1x^1 + p^2x^2 - p^1x^2 - p^2x^1 > 0, \quad (10.10)$$

which can be rearranged as

$$(p^1 - p^2)(x^1 - x^2) > 0. \quad (10.11)$$

Let $\Delta p = p^1 - p^2$ and $\Delta x = x^1 - x^2$. Then the above inequality can be expressed as

$$\Delta p \Delta x > 0. \quad (10.12)$$

This inequality says that change in the price of output (Δp) times change in the output (Δx) is positive. This inequality comes solely from the definition of profit maximization.

Suppose that the price of output has increased, so that $\Delta p > 0$. Then the change in output must be positive, $\Delta x > 0$. That is, as the output price goes up, the competitive firm's output supply also goes up. Equivalently, we say that the profit maximizing supply function of a competitive firm must have a positive slope.

Homogeneity of Degree Zero of a Firm's Output Supply Function

We now consider the impact of equiproportionate changes in the prices of all inputs and output. Suppose that in the short run set up, given the output price p_0 , rental price of capital r_0 and the wage rate w_0 , the profit maximizing competitive firm chooses to produce x_0 units of output, employing L_0 units of labor and K_0 units of capital with a fixed plant capacity. Then the variable cost associated with output x_0 is

$$C_v^0 = w_0L_0 + r_0K_0. \quad (10.13)$$

Now, suppose we double all prices, hence (p_0, r_0, w_0) will be replaced by $(2p_0, 2r_0, 2w_0)$. Under this new set of prices, the slope of the firm's isocost line will be $(-2w_0/2r_0)$, which reduces to the original ratio— w_0/r_0 . That is, the slope of the isocost line does not change when all input prices are doubled. Under the new set of prices, the firm will choose (K_0, L_0) as before for producing x_0 . The variable cost associated with producing x_0 now becomes $2C_v^0$:

$$2C_v^0 = 2w_0L_0 + 2r_0K_0.$$

Thus, the variable cost of production gets doubled at every output level. This in turn implies that the marginal cost also doubles at every output level.

Since the output price also increases from p_0 to $2p_0$, the new profit maximizing output will still be x_0 . That is, doubling all prices leads to the same choice of inputs and output for the firm. This is shown in Fig. 10.6.

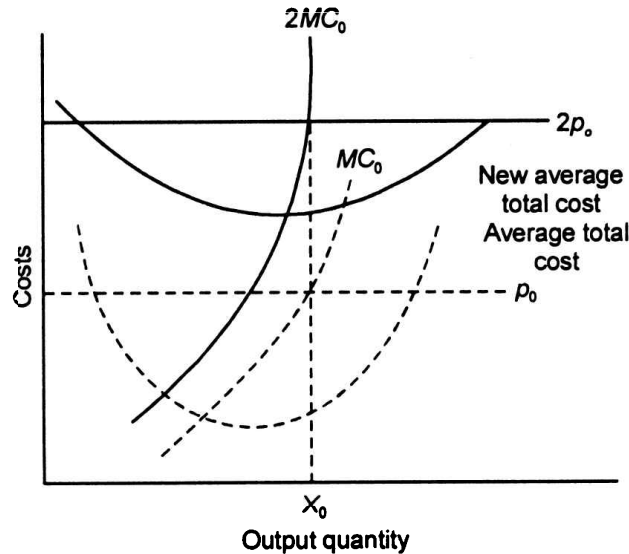


Fig. 10.6 Effect of Doubling All Prices on Output Choice

The new marginal cost curve $2MC_0$ crosses the new price line $2p_0$ at x_0 , the same level of output at which $MC_0 = p_0$, where MC_0 is the original marginal cost. Thus, the profit maximizing output choice remains unaltered under doubling of all prices.

To see the effect of doubling all prices on profits, note that fixed costs remain unaffected by changes in variable input prices. The new average total cost will be $AFC + 2C_v^0/x_0$, where AFC is the original average fixed cost. The new per unit profits are $2p_0 - AFC - 2C_v^0/x_0 > 2(p_0 - AFC - C_v^0/x_0)$. Thus, total profits more than double in the short run when all prices are doubled. More generally, we can say that in the short run an equiproportionate increase in all prices leaves the quantities demanded of all inputs and output produced unaltered. This means that a competitive firm's short run input demand functions and output supply function are homogeneous of degree zero in input and output prices. However, the profit function does not fulfill this property. Profits increase more than proportionately with price increase in the short run.

Short Run Output Supply Curve

Looking at the short run decision making of the firm we note that only that portion of the marginal cost curve that lies above the minimum average

variable cost curve is the short run supply curve for the firm. For prices below the minimum average variable cost, the profit maximizing output is zero.

Given that all input prices are constant, the short run supply curve of the firm specifies the number of units of output that the firm will place on the market at different possible prices. Consider panel (a) of Fig. 10.7. The graph indicates various prices for output of the firm and the associated levels of output that the firm will choose. At price p_5 , the firm chooses x_5 , since at x_5 , $MC = p_5$. At p_4 , the firm chooses x_4 , and so on. As long as the MC curve is above the minimum of AVC and rising to the right, positive output levels are determined by equating MC with price. Consequently, the upward rising portion of the MC curve that lies above the minimum of AVC is the supply curve of the firm. We show this supply curve in panel (b) of Fig. 10.7. There is a discontinuity at p_1 , the price equal to the minimum of AVC . At p_1 , the firm is indifferent between the output level x_1 , and zero output. For any price less than p_1 , the firm's supply is zero.

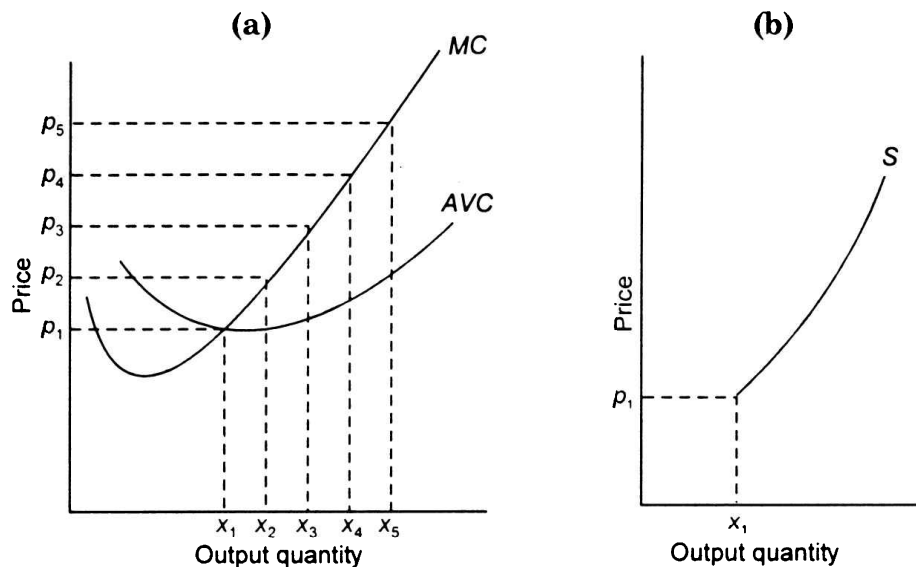


Fig. 10.7 Short Run Profit Maximization and Supply Curve

The curve we have drawn in panel (b) of Fig. 10.7 is in fact the inverse supply curve: price as a function of output. This function measures what prices would have to be to get the producer to supply different units of the good. We will see later that this way of expressing a supply curve is quite useful. It shows that an increase in the price of output always increases the quantity supplied by the competitive firm. Thus, the short run supply curve of output for a competitive firm slopes upward to the right.

Producer's Surplus

Since a competitive firm's marginal cost, or supply curve, of a product usually slopes upward to the right, it appears that we should be able to develop a measure of the producer's surplus analogous to the measure of the consumer's surplus. This is indeed the case.

Suppose that we have the supply curve of a good. Let BS , as shown in Fig. 10.8, be this supply curve. The producer is able to sell x_0 units of his product in a market at a price p_0 . What is the surplus enjoyed by the producer? The producer's surplus is represented diagrammatically by the area above the supply (marginal cost) curve and below the price at which that quantity is sold, namely, the area p_0AB .

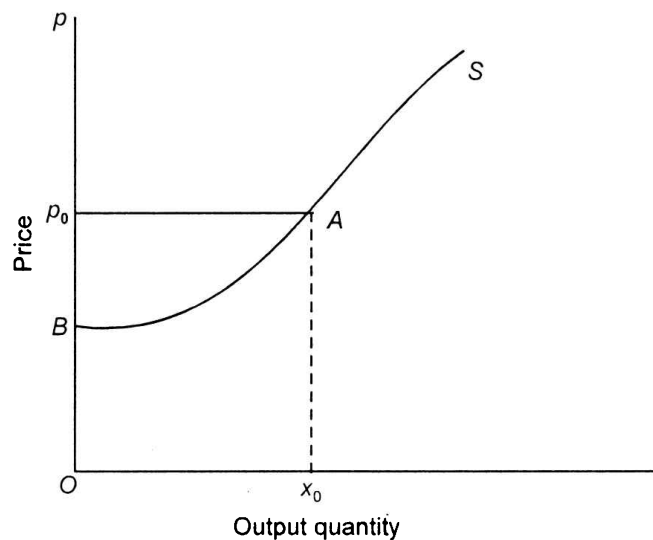


Fig. 10.8 Producer's Surplus

To understand this, let $p_s(x)$ stand for the producer's supply as a function of price. Think about supply for a discrete good. In this case the producer is willing to sell the first unit of the good at price $p_s(1)$, but actually he gets p_0 for this unit. Therefore he earns a surplus of $p_0 - p_s(1)$ by selling the first unit. Similarly, he is willing to sell the second unit of the good at $p_s(2)$, but he actually gets p_0 for it. Therefore, his surplus from the sale of the second unit is $p_0 - p_s(2)$. Continuing in this fashion we see that the producer will be willing to sell the last unit of the good at $p_s(x_0) = p_0$. The surplus earned for the last unit is zero. The sum of all such surpluses is the producer's surplus. That is, the difference between the minimum amount the producer will be willing to sell x_0 units for, and the amount he actually sells x_0 units for, is the producer's surplus. It is given by the area p_0AB .

Now, p_0x_0 is the producer's revenue from the sale of x_0 units of output. The area under the marginal cost or supply curve is the total variable cost. This follows because the area under the marginal cost curve is the cost of producing the first unit plus the cost of producing the second unit, and so on. Therefore the area $OBAx_0 = C_v(x_0)$, the variable costs of producing x_0 . It then follows that

$$\begin{aligned}
 \text{Producer's surplus} &= \text{area } p_0AB \\
 &= \text{area } p_0Ox_0A - \text{area } OBAx_0 \\
 &= \text{revenue from the sale of } x_0 \\
 &\quad \text{units of output} - \text{variable costs of producing } x_0 \\
 &\quad \text{units of output} \\
 &= p_0 x_0 - C_v(x_0).
 \end{aligned} \tag{10.14}$$

Thus, producer's surplus is equal to revenue minus variable cost or equivalently, profits plus fixed costs.

Short Run Market Supply Curve

The quantity of output supplied to the market as a whole is the sum of the quantities supplied by all the firms in the industry. Since each firm uses the market price to determine how much to produce, the total amount supplied to the entire market by all firms will obviously depend on the market price. The relationship showing the quantities of the commodity that all firms together will place on the market at various possible prices is called the short run market supply curve.

Fig. 10.9 illustrates the construction of the market supply curve. For simplicity assume that there are only two firms, 1 and 2. The short run supply (that is, marginal cost) curves for firms 1 and 2 are shown in panels (a) and (b) of Fig. 10.9.

The market supply curve is the horizontal sum of these two curves. For instance, at a price of p_1 , firm 1 is willing to supply x_{11} and firm 2 is willing to supply x_{12} . Therefore, at this price the total supply in the market is given by X_1 , which is equal to $x_{11} + x_{12}$. The other points on the curve are determined in an identical way. We show the market supply curve in panel (c) of the figure. Since each firm's supply curve has a positive slope, the market supply curve will also be positively sloped. The positive slope reflects the fact that short run marginal costs increase as firms increase their outputs.

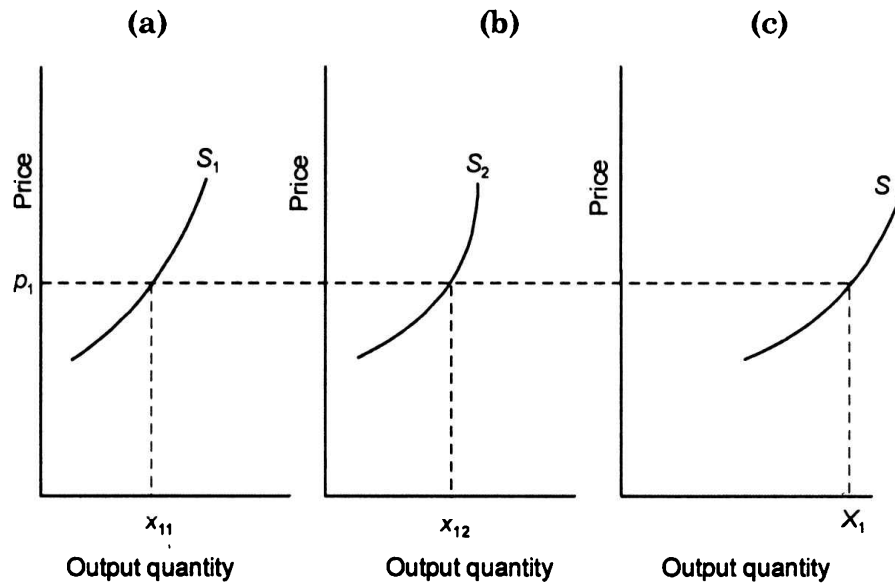


Fig. 10.9 Short Run Market Supply Curve

Elasticity of Short Run Supply

We can summarize the responsiveness of firms in an industry to price changes by the short run supply elasticity. This measure is defined as

$$e_s = \frac{\text{Percentage change in quantity supplied}}{\text{Percentage change in price}} \quad (10.15)$$

$$= \frac{\Delta S(p)/S(p)}{\Delta p/p} \quad (10.16)$$

where $S(p)$ indicates the quantity supplied as function of price p , Δp is the change in the price of the good and $\Delta S(p)$ is the change in the quantity supplied associated with this price change.

Since the slope of the supply curve equals $\Delta(p)/\Delta S(p)$, the formula in (10.16) can be written as

$$e_s = \frac{p}{S(p)} (\text{slope of the supply curve})^{-1}. \quad (10.17)$$

Because the quantity supplied is an increasing function of price, $\Delta S(p)/\Delta p > 0$. Hence the supply elasticity is positive. If e_s is between 0 and 1, supply is inelastic; if $e_s = 1$, the supply is of unitary elasticity; if e_s is greater than 1, supply is elastic. The supply curve is elastic when small changes in

market price lead to a relatively large supply response by firms, since marginal costs do not rise steeply. On the other hand, an inelastic supply curve implies that it takes relatively large changes in price to induce firms to change their output levels, since marginal costs rise rapidly.

Many supply curves become less elastic the higher the price of the commodity is. One reason behind this characteristic is that firms face increasing difficulty in expanding output in response to price increase because of fixed plant capacity or limited availability of skilled laborers or other specialized inputs. As with demand curves, supply becomes more elastic in the long run than in the short run since more time becomes available for adjustment to price changes.

Two polar cases are of interest here. Supply curves that are perfectly elastic ($e_s = \text{plus infinity}$) and perfectly inelastic ($e_s = 0$) are shown in panels (a) and (b) of Fig. 10.10. As we will see later in this chapter, a perfectly elastic supply curve fits a long run situation with free entry and exit of firms. The perfectly inelastic supply curve shows rigid constraints on output that are due to limited supply of factors, for example, the supply of tickets for a football match in a stadium with a limited number of seats.

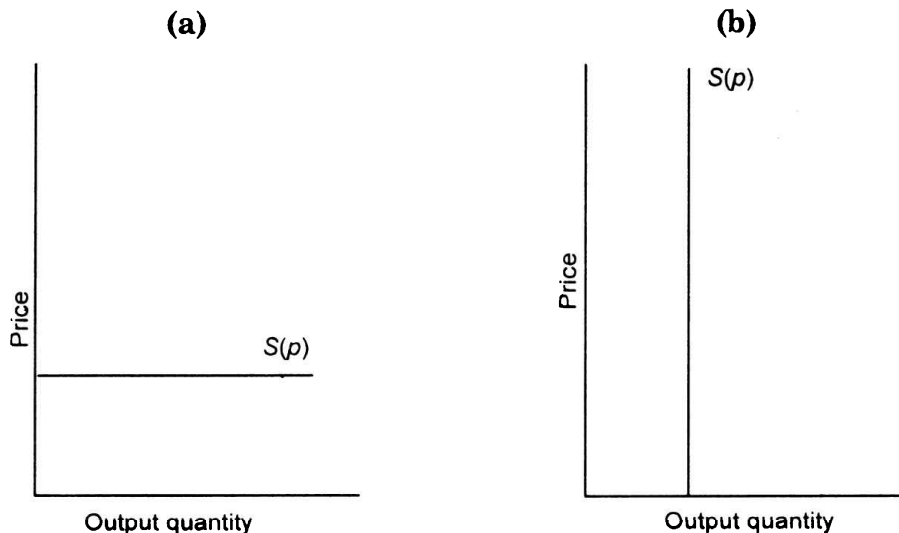


Fig. 10.10 Perfectly Elastic and Perfectly Inelastic Supply

Short Run Equilibrium

We are now ready to combine the market demand and supply curves to determine the equilibrium price in the market. Fig. 10.11 shows the market demand curve $D(p)$ and the short run market supply curve $S(p)$. The two

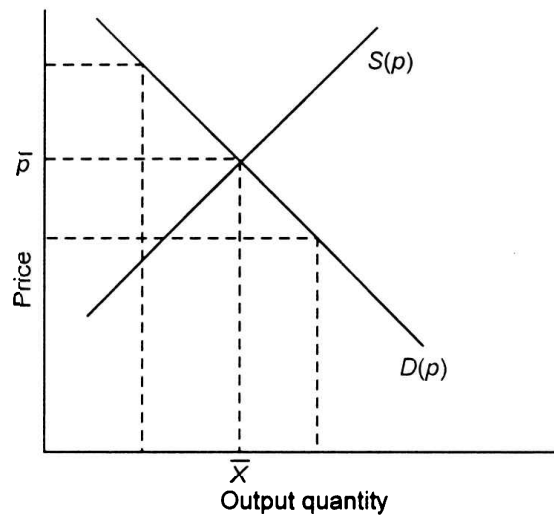


Fig. 10.11 Short Run Equilibrium

curves intersect at a price of $\bar{p}$ and quantity of $\bar{X}$. This price-quantity combination represents an equilibrium between the demands of individuals and the supplies of the firms. Formally, the short run equilibrium price for a competitive industry is that price where the market for the industry's product is cleared—that is, supply equals demand.

The equilibrium price $\bar{p}$ solves the equation

$$D(p) = S(p). \quad (10.18)$$

More formally, if we let $x_i(p)$ be the demand function of person i , $i = 1, \dots, n$ and $MC_j(p)$ be the short run supply function of firm j , $j = 1, \dots, m$; then a short run equilibrium price is simply a solution to the equation

$$\sum_{i=1}^n x_i(p) = \sum_{j=1}^m MC_j(p) \quad (10.19)$$

At the equilibrium point there is a precise balancing of the forces of supply and demand.

An economic equilibrium is a situation where each person's behavior is consistent with that of others. At any price which is not an equilibrium price the behavior of the demanders and suppliers are not compatible. At a disequilibrium price there would be a reason for behavior to change. A price that is not an equilibrium price cannot be expected to persist since at least some persons would have an incentive to change their behavior. Equilibrium will tend to persist unless one of the factors underlying the supply and demand curves changes.

For illustration, suppose that we consider some price $p > \bar{p}$, so that demand is less than supply. Then some suppliers will not be able to sell the amount they expected to sell. They will, however, be able to sell additional output at a lower price. But since all suppliers are selling an identical good and since some suppliers are willing to sell at a lower price, the other suppliers must match that price. This shows that excess supply creates a downward pressure on the market price, and the price will be pushed down to the point where supply equals demand.

Similarly, if $p < \bar{p}$, demand is greater than supply. Then some suppliers will observe that they can sell their product at a higher price than p to the disappointed customers. As more and more suppliers will observe this, the market price will be pushed up to the point where demand equals supply.

Thus, when $X < \bar{X}$ ($p > \bar{p}$) there is someone who is willing to supply an extra unit of the good at a price that is less than the price someone is willing to pay for an extra unit of the good. Consequently, if the good were produced and exchanged between two people at any price between demand price and supply price, they would be made better off.

Similarly, when $X > \bar{X}$ ($p < \bar{p}$), the amount someone would be willing to pay for an additional unit of the good is less than the price that someone must be paid to supply an extra unit of the good. Only at market equilibrium $\bar{X}$ is the willingness to pay for an extra unit equal to the willingness to be paid to supply an extra unit. In this situation there is no way to make any person better off without making someone else worse off. An allocation that fulfils this property is called Pareto efficient.

The equilibrium price acts as a signal to producers by providing them with information with which to decide how much should be produced. In order to maximize profits, firms will produce those output levels for which marginal costs are equal to $\bar{p}$. In the aggregate then, production will be $\bar{X}$.

Given the market price $\bar{p}$, the utility maximizing individuals will decide how much of their limited income should be devoted to buying the particular good. For any consumer the marginal rate of substitution between the good and 'all other goods' is equal to their price ratio.

We summarize all the properties of the market equilibrium discussed above as follows:

- a) Each firm producing the good operates at the same short run marginal cost, which equals the market price of the good.

- b) The marginal rate of substitution between the good and 'all other goods' is the same for all consumers.
- c) The demand reservation price (the most that the consumers will pay to obtain an extra unit of the good) equals the supply reservation price (the least the firms will accept to put an extra unit on the market), and both are equal to the price of the market. That is, the competitive market determines a Pareto efficient amount of output.
- d) Each firm producing the good minimizes the cost of producing the quantity of the good it produces, given the firm's fixed plant capacity.

Given the equilibrium price $\bar{p}$, we can go back to individual firms to look at their output levels and profits. Fig. 10.12 shows three firms, 1, 2 and 3. In this example firm 1 is operating at an output level x_1 such that $\bar{p} = C_1(x_1)/x_1$, that is equilibrium price equals average cost for firm 1.

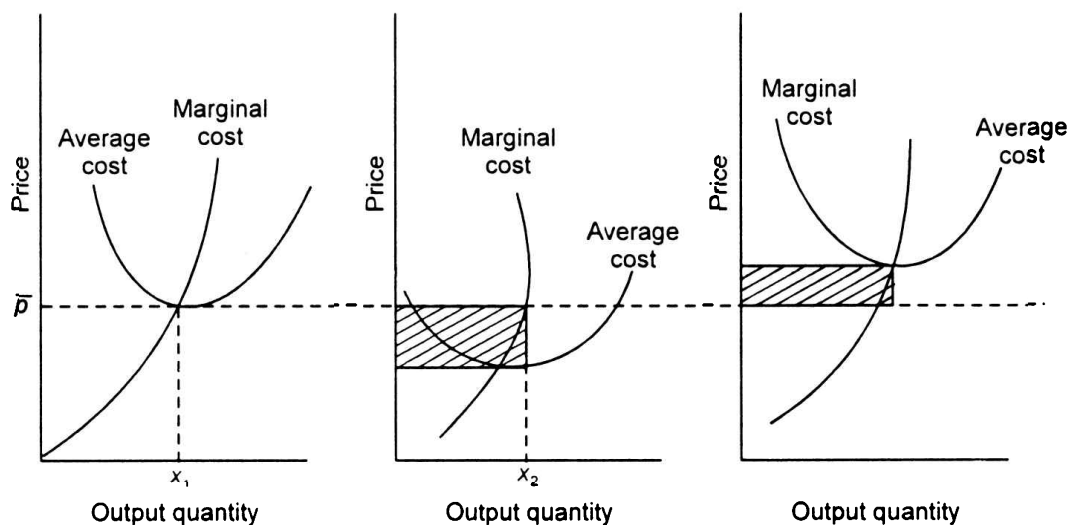


Fig. 10.12 Short Run Equilibrium Price and Individual Firms

Thus, firm 1 is making zero profits. Firm 2 is operating at a point where its average cost is less than the equilibrium price: $C_2(x_2)/x_2 < \bar{p}$, which means that firm 2 is making a profit in the short run equilibrium. Firm 3 is operating where its average cost exceeds the equilibrium price $\bar{p}$, so it is making a loss.

Stability of Equilibrium

If we introduce a disturbance into the industry which is in equilibrium, then the actual price will differ from the equilibrium price. An equilibrium is said to be stable if a disturbance results in a return to equilibrium, and unstable

if it does not. In our discussion in the previous section it was implicitly assumed that the market equilibrium is stable. In this section we make a more formal discussion on stability.

A disturbance usually gives rise to an adjustment process in the market. For instance, as stated earlier, if the actual price is higher than the equilibrium price, some suppliers will tend to sell their product at a lower price. The stability analysis with which we are concerned here does not deal with the time path of the adjustment process and considers only the nature of change, that is, whether it is towards, or away from the equilibrium.

We define the excess demand $E(p)$ as a function of price p as

$$E(p) = D(p) - S(p). \quad (10.20)$$

That is, at any price p excess demand $E(p)$ is the quantity demanded less the quantity supplied at that price. Given downward sloping demand and upward sloping supply curves (see Fig. 10.11), the excess demand curve is negatively sloped. Excess demand $E(p)$ is negative, zero or positive according to whether price p is greater than, equal to, or less than the equilibrium price.

A suggestion that explains the movement of price to its equilibrium level was proposed by Leon Walras. The Walrasian stability condition is based on the assumption that buyers tend to raise the price if excess demand is positive, and sellers tend to lower the price if it is negative. That is, for a market to be stable in the Walrasian sense we require excess demand to be a decreasing function of price. Such a mechanism is called a *tâtonnement* (groping) process. This mechanism is pictured graphically in Fig. 10.13. In this figure, the equilibrium price $\bar{p}$ is stable; there are forces that move p to $\bar{p}$. The rule of increasing price when quantity demanded exceeds quantity supplied, and decreasing price when quantity supplied exceeds the quantity demanded, always produces stability.

The Walrasian adjustment process regards price as the motivating force in the adjustment of markets to equilibrium. Consumers and suppliers respond to price changes by moving along their respective demand and supply curve until an equilibrium price-quantity combination is achieved. A somewhat different adjustment process was suggested by Alfred Marshall. He argued that consumers and suppliers adjust quantity in response to imbalances in quantities demanded and supplied, and that the price changes follow from these quantity changes.

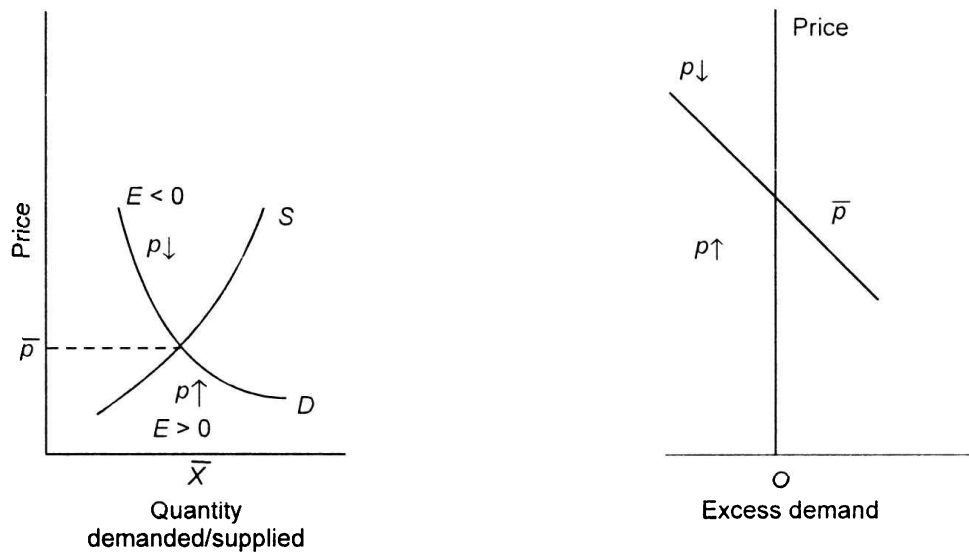


Fig. 10.13 Walrasian Stability

Let D^{-1} and S^{-1} be the inverse demand and supply functions. That is, $D^{-1}(X)$ represents the per unit price that the demanders are willing to pay for output level X . Similarly, $S^{-1}(X)$ stands for the per unit price that the suppliers require for the output level X . Then excess demand price is defined by

$$E^{-1}(X) = D^{-1}(X) - S^{-1}(X). \quad (10.21)$$

Movements in quantity toward equilibrium are motivated by discrepancies between the price consumers are willing to pay and what suppliers wish to receive. When these two prices coincide, quantity adjustment ceases.

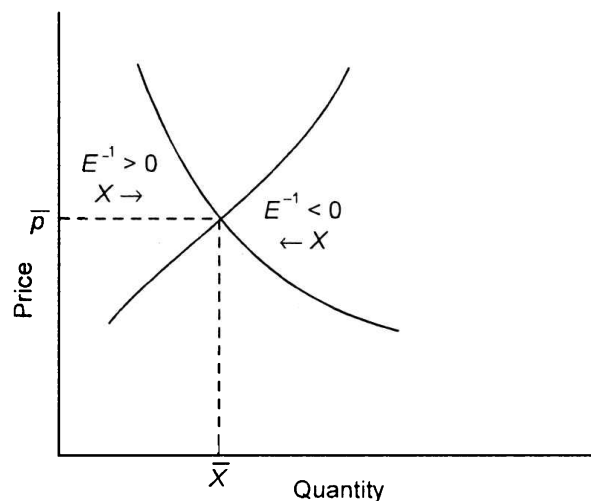


Fig. 10.14 Marshallian Stability

We illustrate this idea in Fig. 10.14. For quantities less than equilibrium quantity $\bar{X}$, demand price is greater than supply price. In this case suppliers will tend to raise their output, because they realize that the consumers are willing to offer a higher price than they are charging. For quantities above equilibrium, supply price exceeds demand price, providing incentives for quantity reduction. The Marshallian mechanism pictured in Fig. 10.14 implies that the equilibrium is a stable one. Starting from any initial position, forces move economic agents towards equilibrium.

Equilibrium is stable in the Walrasian sense if the supply curve is steeper than the demand curve and unstable in the opposite case. Equilibrium is stable in the Marshallian sense if the supply curve is less steep than the demand curve and unstable in the opposite case. In Fig. 10.15(a) an equilibrium exists, but it is not stable in the Walrasian sense. When price is above equilibrium, the price adjustment rule pushes price further away from the equilibrium. If price is below equilibrium, the adjustment rule specifies that price should be reduced, so again price moves further away from the equilibrium.

In panel (b) of Fig. 10.15, there are three points d , e and f at which equilibrium occurs. Points d and f are stable in the Walrasian sense, while e is unstable. In general, if multiple equilibria occur, points of stable equilibrium are separated from one another by unstable equilibrium points. Because of the very nature of an unstable equilibrium, it is rarely observed in an actual market. Economists often assume that price quantity pairs observed in markets occur not only at equilibrium, but at a stable equilibrium.

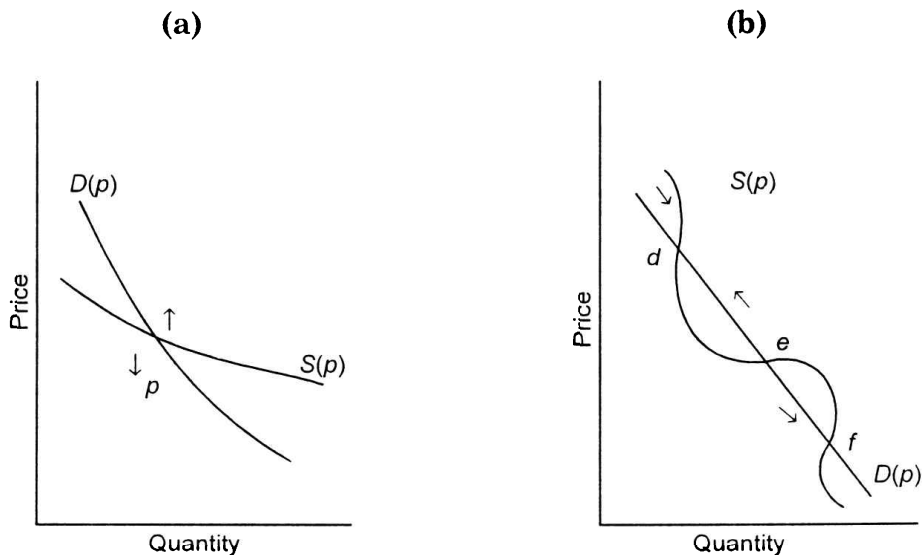


Fig. 10.15 Unstable Equilibrium (a), Multiple Equilibria (b)

One cannot say a priori whether Walrasian or Marshallian stability is more realistic. The stability of an equilibrium can be assessed only after empirical verification.

Comparative Statics

A demand curve or a supply curve may shift for many reasons. Such shifts may be caused by changes in price in other markets, in productive technologies, in the tastes of the consumers, and so on. These types of changes occur constantly in real world markets. When either a supply curve or a demand curve shifts, the equilibrium of a price-quantity combination will change. In this subsection we investigate relative magnitudes of such changes graphically. This is an example of comparative statics analysis: we will compare equilibria at two different time points. In Appendix 10.1 we present a brief discussion on how prices move from one equilibrium level to another.

Let us consider first an upward shift in the short run supply curve for a good. Such a shift may result from increased prices of inputs used by firms. The effect of the shift on equilibrium output-price combination will depend on the shape of the demand curve. In panel (a) of Fig. 10.16, a shift in the supply curve from S to S' will cause equilibrium price to rise only moderately from p to p' whereas quantity declines sharply from X to X' . The increase in the firm's input costs is met primarily by a quantity decrease (a movement down the firm's marginal cost curves) rather than by a sharp price increase. This results from the elastic shape of the demand curve.

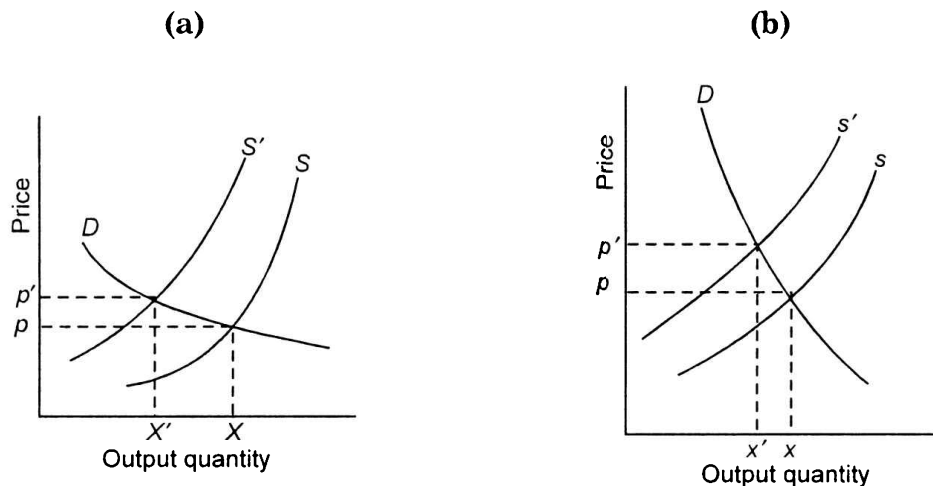


Fig. 10.16 Effect of Shift in Short Run Supply Curve

A reverse situation will arise if the market demand curve is inelastic. In panel (b) of Fig. 10.16, a shift in the supply curve causes equilibrium price to rise substantially, where the quantity change is rather minor. The reason for this is that individuals do not reduce their demands substantially if prices rise. Therefore, almost the entire effect of the upward shift of the supply curve is passed on to the demanders in the form of prices.

We can also show that a given shift in the market demand curve will have different implications for an equilibrium price-quantity combination, depending on the shape of the short run supply curve. In panel (a) of Fig. 10.17, the short run supply curve is inelastic. In this situation a rightward shift of the market demand curve will cause prices to increase sharply and quantity to increase only slightly. The increase in demand has caused firms to move up their steeply sloped marginal cost curves. The large price increase serves to ration demand.

In panel (b) of Fig. 10.17 we have a relatively elastic short run supply curve. In this case an increase in demand produces a substantial increase in quantity. However, because of the nature of the supply curve, this increase is not met by high cost increases. Consequently, the price rise is moderate only.

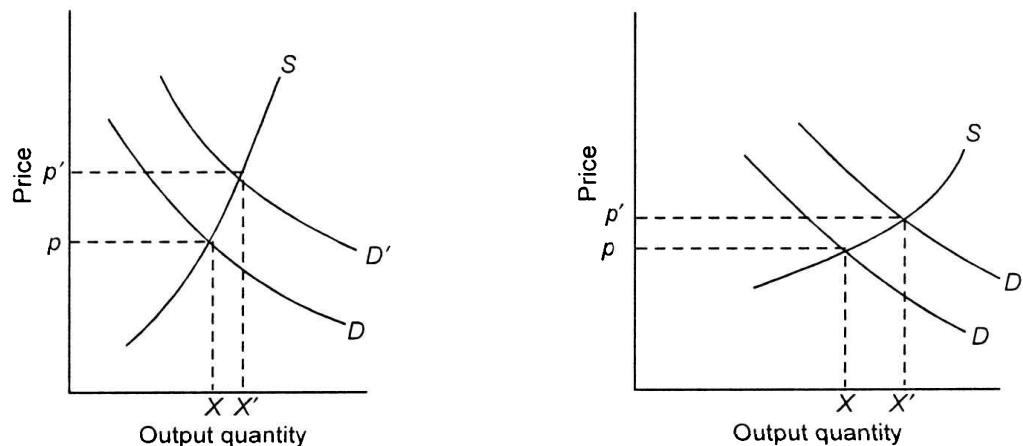


Fig. 10.17 Effect of Shift in Demand Curve

Tax Incidence Analysis

The supply-demand analysis can be employed to illustrate some basic principles of tax incidence theory. The objective of tax incidence theory is to determine the extents of burden of a tax borne by different segments of the society.

The main thing to understand about taxes is that when a tax is present in a market, the demand price and the supply price differ by the amount of the tax. Of the different kinds of taxes that can be imposed, two examples are quantity tax and ad valorem tax (also called value tax).

To illustrate quantity tax and value tax symbolically, let p_D be the price paid by demanders and p_s be the price received by suppliers. If t is the amount of quantity tax per unit sold, then

$$p_D = p_s + t. \quad (10.22)$$

On the other hand, a value tax is expressed in percentage terms. If the tax rate is given by r then

$$p_D = (1 + r)p_s. \quad (10.23)$$

We wish to examine here the impact of a quantity tax. Suppose that the suppliers are required to pay the tax. Then the amount that the suppliers get will be the amount the demanders pay minus the amount of the tax. Also the amount supplied will depend on the supply price and the amount demanded will depend on the demand price. We thus have two equations:

$$\begin{aligned} D(p_D) &= S(p_s), \\ p_s &= p_D - t. \end{aligned} \quad (10.24)$$

Substituting the second equation into the first we get

$$D(p_D) = S(p_D - t). \quad (10.25)$$

We can also rearrange the second equation to get $p_D = p_s + t$; substitution of this into the first gives

$$D(p_s + t) = S(p_s). \quad (10.26)$$

We can use either Eq. (10.25) or (10.26) for our purpose.

If instead of suppliers paying the tax, the demanders pay it, then we have

$$p_D - t = p_s. \quad (10.27)$$

Substituting this into the demand-equals-supply condition we get

$$D(p_D) = S(p_D - t), \quad (10.28)$$

which is same as the condition that arises when the suppliers pay the tax. Therefore, as far as the equilibrium price is concerned, it does not matter who pays the tax—what matters is that the tax must be paid by someone.

Fig. 10.18 illustrates the tax incidence analysis. Imposition of a per unit quantity tax of t creates a divergence between p_D and p_s . In panel (a) of the figure it appears that the burden of the tax is approximately equally shared

by demanders and suppliers, but this result stems from the similarity in demand and supply elasticities of the case pictured. Different distributions of burden would result, depending on the values of elasticities. For instance, if the supply curve is perfectly elastic, the entire amount of tax is paid by the consumers (panel [b] of Fig. 10.18). On the other hand, if the supply curve is perfectly inelastic, the suppliers pay the full tax (panel [c] of Fig. 10.18). We examine this incidence question more rigorously using calculus in Appendix 10.1.

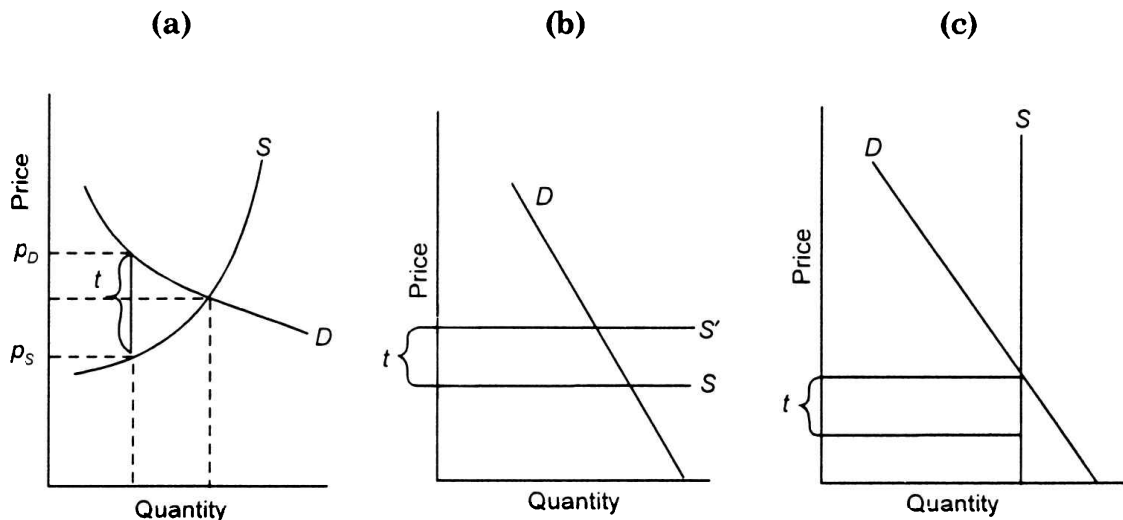


Fig. 10.18 Tax Incidence Analysis

LONG RUN ANALYSIS

In the long run, all the incentives available to the firm are fully exploited. A firm is free to choose the levels of all inputs (including plant capacity). Firms have time to alter their plant sizes. New firms can enter and existing ones can leave the market in the long run. In mathematical terms we allow the number of firms, n , to vary in response to economic incentives. New firms will be lured into any market in which economic profits are positive. Similarly, firms will leave any industry in which profits are negative. For long run analysis we use firms' long run cost curves, since these curves reflect flexibility in all inputs.

Long Run Supply Curve of a Firm

In the long run a profit maximizing firm will produce that output level for which price (p) is equal to the long run marginal cost (MC). Therefore the

argument presented in the short run case can be applied again to the long run case to derive the supply curve of a firm. That is, so long as MC is above the minimum of the long run average cost (AC) with the MC curve sloping upward to the right, the MC curve is the firm's long run supply curve. (Since in the long run all costs are variable, long run average costs = long run average variable costs.) The argument for deriving the long run supply curve is correct, but it does not cover one important case, the case of constant average costs, which deserves a separate treatment.

When the long run technology of firm exhibits constant returns to scale, the long run average costs of the firm becomes a constant. In this case the firm will be willing to supply any amount of output as long as the output price p equals long run average cost (AC) which in turn equals the long run marginal cost (MC). If the price is above AC , the firm produces an unlimited amount since profits increase without bound as output increases. If price is below AC , the firm will produce nothing. Hence under constant average costs, the output price p equals AC for any firm that actually produces the good and the long run supply curve will be the MC , which coincides with AC . This is shown in panel (a), Fig. 10.19

Under decreasing returns to scale and competitive input markets, the firm's supply curve is its long run marginal cost curve: the MC curve. This follows because MC is greater than AC , the long run average costs for all output levels and MC slopes upward to the right. This is shown in panel (b) of Fig. 10.19.

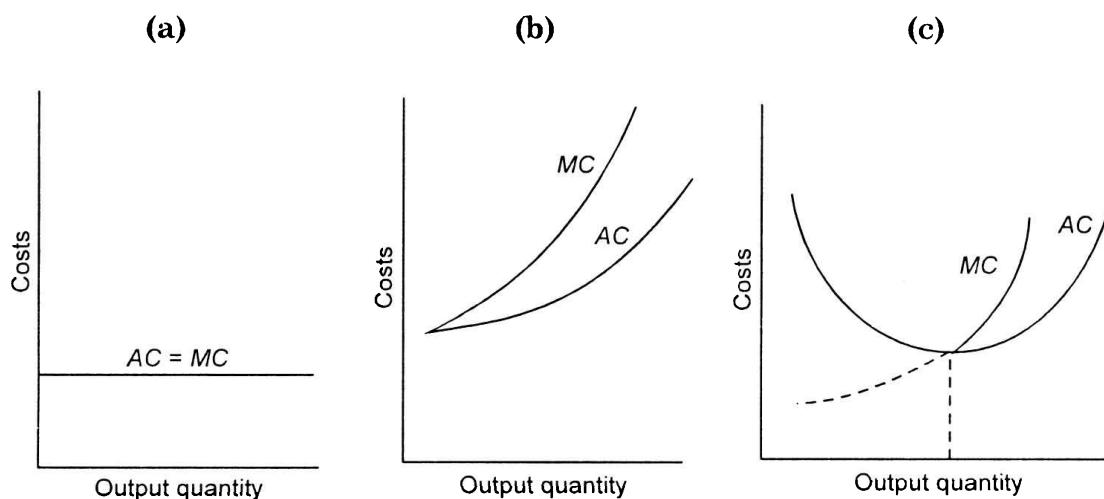


Fig. 10.19 Long Run Supply Curve of a Competitive Firm

In panel (c) of the figure we consider the U-shaped AC curve. Here the MC curve above the minimum of the AC and sloping upward to the right is the long run supply curve.

If increasing returns to scale characterize the technology of the firm with competitive input markets, then perfect competition breaks down. Hence a long run supply curve for the competitive case does not exist in this case.

In the long run, one choice of the firm is to leave the industry. Since by going out of business, the firm can make zero profit always, in the long run the profits that a firm can earn have to be at least zero. Thus, in the long run

$$px \geq C(x)$$

$$\text{or,} \quad p \geq \frac{C(x)}{x}. \quad (10.29)$$

This says that in the long run the price has to be at least as high as average cost.

Short Run and Long Run Elasticities of Supply

The short run supply curve of a firm involves the marginal cost of output holding plant capacity fixed, while the long run supply curve involves the marginal cost of output when plant size is adjusted optimally.

At price p_1 , in Fig. 10.20, associated with the optimal plant size is the short run marginal cost curve SM . Long run marginal cost $MC = SM$ at output X_1 , which is both long run and short run profit maximizing output level. That is, the short run and long run supply curves of the firm coincide at X_1 .

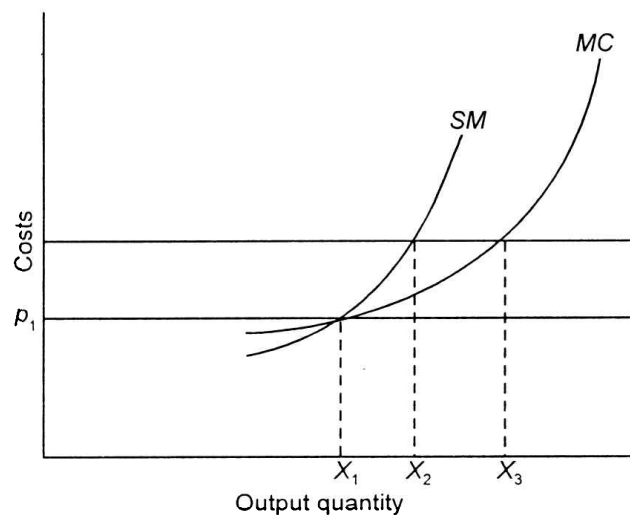


Fig. 10.20 Short Run and Long Run Supply Curves

When output price increases the firm has choices to adjust its plant size in the long run which it cannot do in the short run. Therefore the long run supply increases more than does the short run supply. This means that the long run supply curve will be more responsive to price—more elastic—than the short run supply curve. The short run supply curve is steeper than the long run supply curve.

Long Run Equilibrium

In the long run firms are free to expand or contract plant capacity and enter or leave the industry. The commitments that restrict choices in the short run are not binding in the long run. In the short run firms maximize profits within the constrained set up generated by fixed plant capacities. But in the long run all the incentives available to the firm are fully exploited.

In the short run equilibrium occurs when markets are clear, given the fixed plant capacities of the firms. The concept of long run equilibrium of a competitive industry captures the characteristics of a situation in which there is no incentive for any firm to change its plant capacity or move from industry to industry. As long as the technology, consumer preferences and the structure of prices remain constant, under long run equilibrium there is neither entry nor exit of firms, and each firm's production remains fixed. On the other hand, in the short run equilibrium incentives exist for firms to make changes. In our discussion of long run equilibrium, it will be assumed that entry and exit barriers are absent. However, there are organizational expenses of creating a firm.

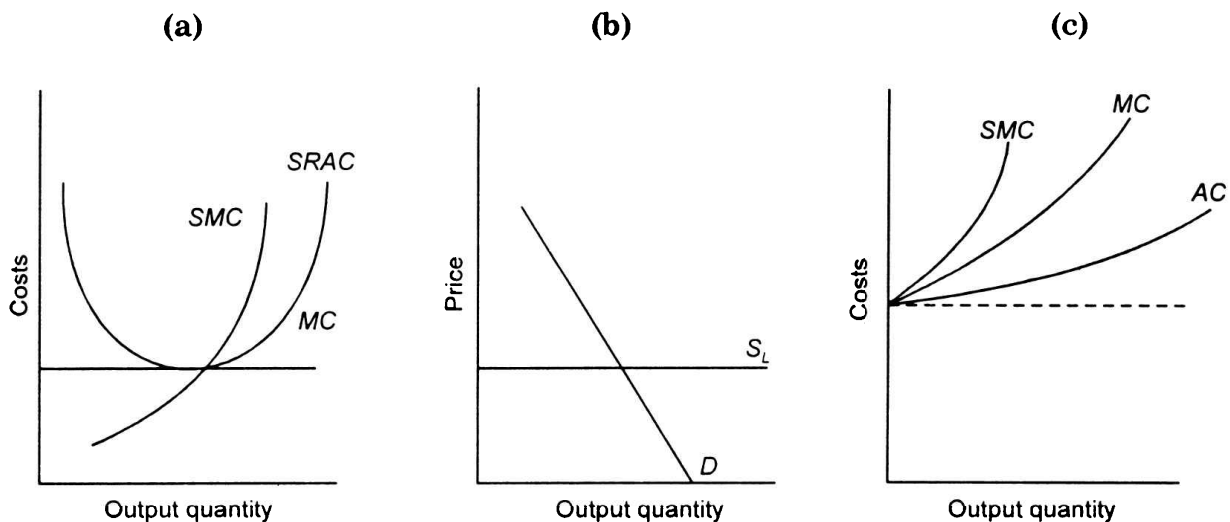
Long run equilibrium for a competitive industry occurs when the market for the industry's output is cleared given that the number of firms in the industry and the industry output are such that there are no incentives for firms to enter or leave the industry or for changes in the plant capacities of the firms.

We assume that each firm has full knowledge of the technology. The factor prices are assumed to be given. Therefore, for any given output level each firm will choose a plant capacity, the technology, and the combination of inputs that minimize the cost of producing that output level. In other words, we say that each firm in the industry has the same long run cost curves. Now, if price exceeds the minimum of AC (long run average cost), then new firms have an incentive to enter the industry since they can make positive profits by operating at the output where AC is a minimum. On the other hand, if AC

is greater than price, every firm will incur loss. Therefore, there would be incentives for leaving the industry. (By leaving the industry a firm can reduce its loss to zero.) But in the long run there can be no incentive either to enter or leave. Therefore, price must be equal to minimum of AC .

These ideas are illustrated in Fig. 10.21 for constant returns to scale and decreasing returns to scale technologies, and also for a U-shaped long run average cost curve. Note that for constant returns to scale the average cost is constant, and for decreasing returns to scale the average cost is increasing. The input markets are competitive. The firm's short run supply curve is the short run marginal cost (SMC) curve. For decreasing returns to scale and U-shaped average cost curve cases, the firm's long run supply curve is its long run marginal cost (MC) curve, above the minimum of long run average cost (AC) and rising to the right. These are shown in panels (c) and (e) of Fig. 10.21. In panel (a) we show that the firm's MC curve, which is its long run supply curve, is a horizontal line. This is for constant returns to scale case. Panel (a) also shows the short run average cost ($SRAC$) curve of the firm.

The industry long run supply curve S_L is interpreted as the quantity supplied by the entire industry as a function of the price per unit of output, given that firms make use of all incentives for entry and exit and for changing plant capacities. Thus, the long run supply curve of the competitive industry is not obtained by horizontally adding up the supply curves of a fixed number of firms in the industry, as in the case of short run supply. The long run industry supply curve also reflects entry and exit of firms.



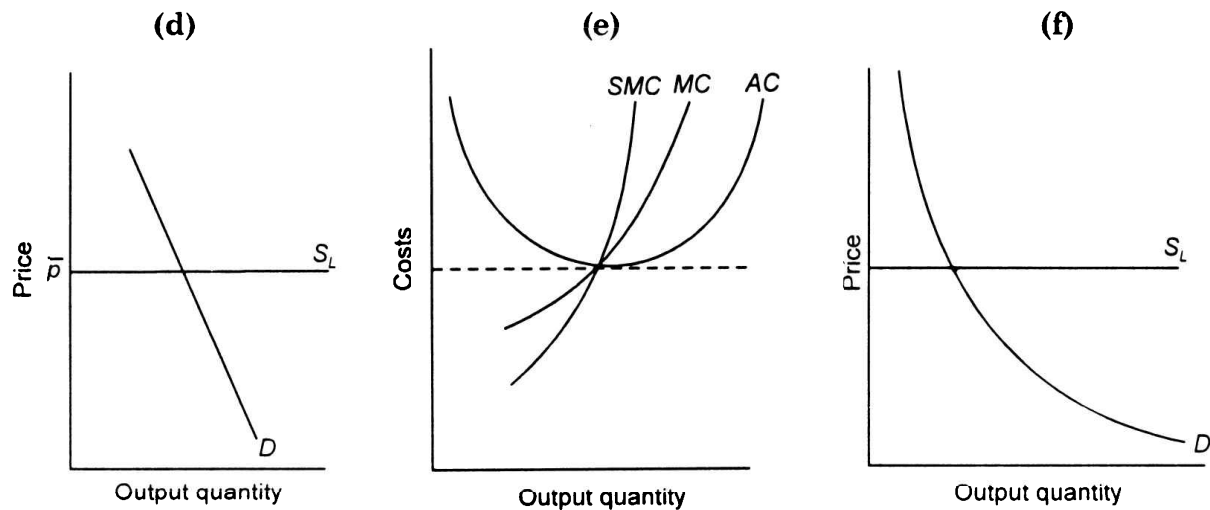


Fig. 10.21 Long Run Equilibrium: Constant>Returns to Scale (d) Decreasing Returns to Scale (e) U-Shaped Average Costs (f)

The long run industry supply curve S_L is perfectly elastic at the minimum of the firm's AC curve, irrespective of the shape of the firm's supply curves. This property reflects the view that entry will take place if price is above the minimum of AC , and exit will occur whenever price is below the minimum of AC . When price equals the minimum of AC , we have perfectly elastic industry output supply, since any amount of output can be supplied by varying the number of firms in the industry, where each firm chooses that output, plant capacity and technology which involve an average cost equal to the minimum of AC , and earns just normal profits.

We conclude that, whatever be the returns to scale (except for increasing returns, which are not compatible with competitive industry), in long run equilibrium price equals the minimum of long run average cost for any firm in the industry. Since the long run supply curve of the industry is perfectly elastic, market demand has no influence on the long run equilibrium price. Market demand determines the level of output which in turn determines the number of firms in the industry. For constant average cost, that is, for constant returns to scale with free entry, the number of firms is indeterminate. We can have one hundred firms each producing one unit of the output, or ten firms each producing ten units of output. For decreasing returns to scale, that is for an increasing average cost curve, with free entry, the smaller the production scale the lower will be the average cost. Hence in the long run equilibrium there will be an infinite number of firms each producing an infinitesimally small amount of output. For a U-shaped average curve with

free entry, the exact outcome will be determined by the relative size of a firm and the market size.

We summarize the properties of long run equilibrium in a competitive industry as follows:

1. Market price = long run marginal cost = short run marginal cost = minimum of long run average cost, all of which are the same for every firm in the industry.
2. Except for constant returns to scale, in which case firm size is indeterminate, each firm in the industry chooses the same plant, namely that size with the lowest short run average cost and long run average cost. Under decreasing returns to scale, the size of a firm is infinitesimally small.
3. In long run equilibrium price is the same, irrespective of the level of demand. Therefore whatever the returns to scale are, the industry works 'as if' there are constant returns to scale.
4. In long run equilibrium each firm earns only normal profits; excess profits (in excess of normal profits) are zero in any competitive industry.

In short run equilibrium, excess profits may arise in various industries. In the short run a firm may operate either below or above the minimum of long run (and short run) average costs.

The Adjustment Process

We discuss here the process of adjustment from one long run equilibrium to another. Suppose that the output-price combination when a competitive industry is at long run equilibrium is given by $(\bar{p}, \bar{X})$. Suppose also that the demand curve for the product of the industry shifts to the right from D to D' . This may result from an increase in the price of a good which is a gross substitute for the competitive industry's product. The short run industry supply curve is S_R , the sum of short run supply curves of the given number of firms in the industry. The new short run equilibrium occurs at $p' > \bar{p}$. (See panel [a] of Fig. 10.22.)

At p' , the firms in the industry are making excess profits, since p' exceeds the short run average costs. This situation generates incentives for entry into the industry. Entry will continue until the number of firms in the industry is such that the new short run industry supply curve S'_R cuts the industry's new demand curve at $\bar{p}$ (see panel [b] of Fig. 10.22.) The long run industry

supply curve is horizontal at $\bar{p}$, hence the adjustment process establishes the original market clearing price. There is a large number of firms in the industry, but each firm has the same plant capacity that was chosen originally by existing firms. This follows because we started with a long run equilibrium and hence each firm had chosen the plant capacity that minimizes the long run average cost. Consequently, the same plant capacity exists in the new long run equilibrium as in the old one. Each firm will adopt the same production process as the original one.

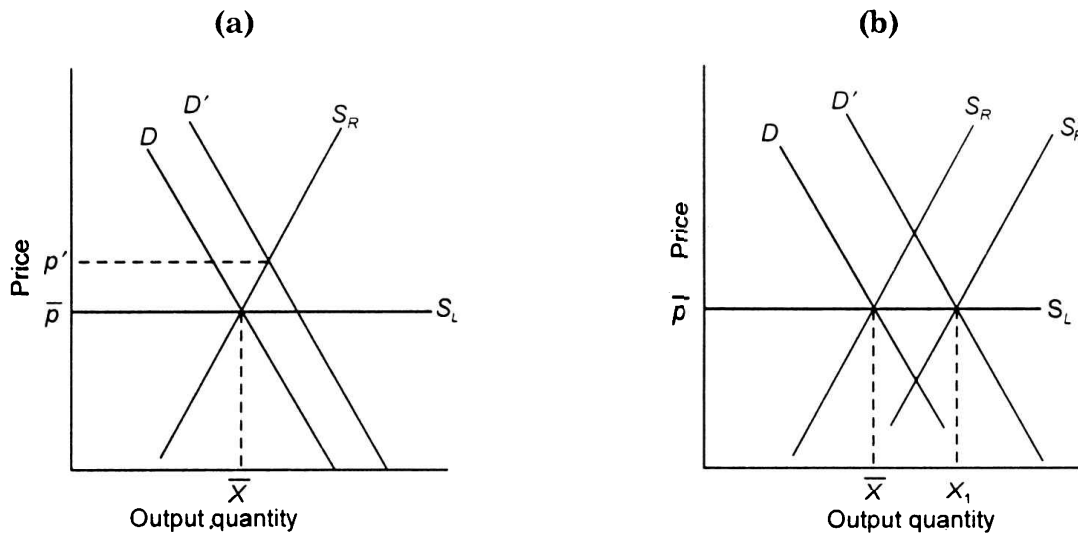


Fig. 10.22 Long Run effect of demand shift

We have thus shown that starting from a position of long run equilibrium for the perfectly competitive industry, if the market demand for the commodity increases, more firms will enter the industry, and under constancy of factor prices the market supply of the commodity will increase until the original market equilibrium price is established. In this case entry has no effect on costs of a firm's input. No matter how many firms enter (or leave) an industry, every firm will retain the same plant capacity that was chosen initially. Thus, we are making here a constant input cost assumption and that is why this type of industry is referred to as a constant cost industry.

Long Run Market Supply with External Economies and Diseconomies

There are several reasons why the entry of new firms may cause the average cost of all firms to rise or fall. New and existing firms may have to compete for scarce inputs, thus driving up their prices. New firms may impose external costs on the existing firms (and on themselves) in the form of air pollution.

Since firms may have to adopt some pollution control measure, this will increase costs for all firms.

On the other hand, new firms may reduce the costs of firms in an industry. For instance, the entry of new firms may provide a larger pool of trained labor from which to hire than was previously available. This reduces the costs associated with the hiring of new workers. Similarly, the mass industrialization caused by entry of new firms may permit development of more efficient transportation and communication networks.

An industry is referred to as an increasing or decreasing cost industry accordingly the average costs of all firms rise or fall as new firms enter. In microeconomic theory these effects are called diseconomies and economies external to the firm. These effects are external to the firm because they are beyond the firm's control. They are internal to the industry because they are controlled by changes in industry output and the gains and losses are captured within the industry by the constituent firms.

In their decision making, firms in a competitive industry ignore potential economies or diseconomies that are due to the expansion of the industry output. At any given industry output, each firm operates in the long run equilibrium at the minimum of long run average cost (AC). But with external economies or diseconomies, AC for the firm depends on the industry output.

Fig. 10.23 demonstrates market equilibrium in an increasing cost industry. The initial equilibrium price is p_1 . At this price the typical firm produces x_1 , the industry output is X_1 . Suppose now that the industry demand curve shifts to D' . Price will rise to p_2 in the short run, since the short run supply curve of the industry S_R intersects D' at p_2 . At this price the typical firm will produce x_2 and earn an excess profit. This profit attracts entrants and as a result the short run supply curve shifts to S'_R .

This entry of new firms causes the cost curves of all firms to rise. The new firms may compete for scarce inputs, thereby driving up prices of inputs. A typical firm's new (higher) set of cost curves are shown in panel (b) of Fig. 10.23. The new long run equilibrium price is p_3 ($p_3 = AC =$ long run marginal cost), and at this price X_3 is demanded. We therefore have two points (p_1, x_1) and (p_3, x_3) on the long run supply curve. Similar points can be found in an analogous way considering shifts of the demand curve. Thus the locus of all such points is the long run supply curve LS . The curve LS is positively sloped because of the increasing cost nature of the industry. Observe that LS is flatter than the short run curves. This indicates greater flexibility in supply response in the long run.

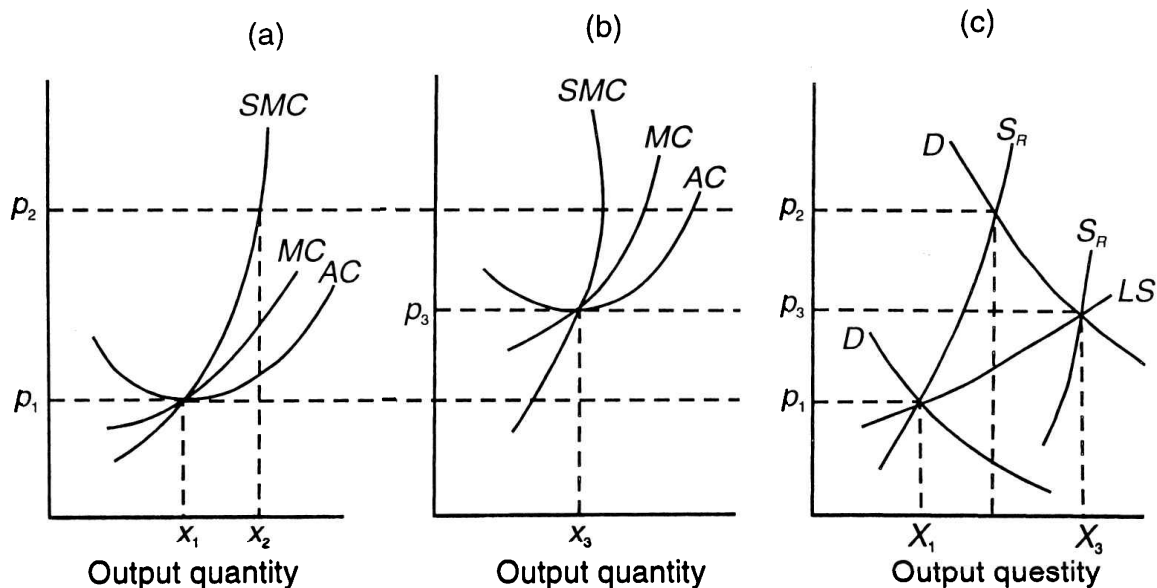


Fig. 10.23 Supply Curve for Increasing Cost Industry

In an analogous manner it can be shown that for a decreasing cost industry, the industry long run supply curve is negatively sloped, indicating that in the long run higher outputs will be available at lower prices.

THEORY OF CONTESTABILITY

Several economists argued that by focusing only on the behavior of actual entrants' long run equilibrium under perfect competition ignores the effect of potential entrants on market equilibrium. They made a distinction between competition *in* the market and competition *for* the markets by showing that the latter concept provides a more appropriate perspective for analyzing the entry assumption. Within this broad perspective competitive equilibrium is more likely to emerge.

The expanded examination of entry begins by defining a perfectly contestable market. Baumol, Panzer and Willig (1982) defined a contestable market as one in which entry and exit are absolutely free. Equivalently, a perfectly contestable market is one in which no outside potential competitor can enter by price cutting and still making profits (since if such opportunities existed, potential entrants would take advantage of that). Thus, a perfectly contestable market drops the perfectly competitive assumption of price-taking behavior, but expands the notion of free entry by allowing the potential entrants to operate in hit-and-run manner, whatever profit opportunities become available.

To define equilibrium in a contestable market formally, let us first introduce the notion of feasibility. An industrial configuration with price p and firm outputs $x_1, x_2, \dots, x_n$ is feasible if (i) the market equates supply and demand: $\sum_{i=1}^n x_i = D(p)$, and (ii) each firm at least covers its production cost, $px_i \geq C(x_i)$, $i = 1, \dots, n$, with C being the common cost function. A feasible configuration $(p, x_1, \dots, x_n)$ is sustainable if it does not offer profitable entry, that is, $p_e x_e \leq C(x_e)$ for all $p_e \leq p$ and $x_e \leq D(p_e)$. In a perfectly contestable market equilibrium can be formed by a sustainable industry configuration only.

Under perfect contestability, in equilibrium firms earn zero profit and price equals marginal costs. As we saw, this requires that firms produce at the minimum points of their long run average cost curves where price = marginal costs = average costs. Hence even in the absence of price taking behavior in markets with a relatively small number of firms, perfect contestability possesses characteristics that guide market equilibrium to a competitive type result. It thus can be said that contestability replaces price-taking behavior by rapid entry and exit.

Note that the potential entrants in a perfectly contestable market possess two properties. First, they will be able to serve before an incumbent can change its price so that they are not subject to any cost disadvantage. (They employ the same production techniques that are accessible to incumbent firms.) Second, entrants must assess the profitability of entry on the basis of existing price obtained by incumbents.

One way to think about how a contestable market works is to imagine that buyers are solicited by firms that wish to enter the market. Each firm offers a long term contract that allows a buyer to purchase from the seller at a fixed price. If buyers collectively agree to choose only one seller, they will **never** pay a price greater than average cost.

Example 10.1

A firm in a perfectly competitive industry has the following long run cost function $C(x) = x^3 - 60x^2 + 1500x$. We wish to answer the following questions:

1. What are the long run average and marginal costs?
2. If the firm can sell its output at a price of $p = 975$, how much will it produce to maximize profit?
3. Is the output of the firm in (2) compatible with industry equilibrium?
4. If the industry is one of constant average cost, derive the equation for the long run supply curve of the industry.

5. If the market demand curve is $p = 9,600 - 2X$, how many firms will there be in the long run equilibrium?

Answer:

1. Long run average cost $C(x)/x = x^2 - 60x + 1,500$. Long run marginal cost $dC(x)/dx = 3x^2 - 120x + 1,500$.
2. Setting $p =$ marginal cost, we have $3x^2 - 120x + 525 = 0$, that is $(x - 5)(x - 35) = 0$. Therefore $x = 35$ (at $x = 35$, average cost = 625 < marginal cost).
3. At $p = 975$ and average cost = 625, the firm makes excess profit, which is incompatible with industry equilibrium.
4. Find the minimum of average cost by setting the average cost = marginal cost.

$$x^2 - 60x + 1,500 = 3x^2 - 120x + 1,500,$$

$$\text{or,} \quad 2x^2 - 60x = 0.$$

Therefore $x = 30$. At $x = 30$, marginal cost = 600. Therefore, the industry long run supply curve is $p = 600$.

5. Equate supply and demand:

$$p = 9,600 - 2X = 600.$$

$$\text{Therefore } X = 4,500.$$

Each firm produces $x = 30$. Thus there are $4,500 \div 30 = 150$ firms.

Applied Example 10.1 Stock Market Transactions

On the Mumbai Stock Exchange, price adjustments in each stock are made by the specialists in that stock. A specialist is a broker on the floor of the exchange. He accepts purchase and sale orders from other brokers in a particular stock and carries out these orders according to the instructions he receives from other brokers. Some of these orders ask the broker to buy or sell a particular number of shares at the best price he can. We refer to these orders as market orders. Some of the orders specify a price at which the purchase or sale of a specified number of shares is to be made. Such orders are called limit orders. At any given time, a broker has an order book that lists the limit buy and sale order he has received.

When a market buy order comes in, if there is positive excess demand at the old price, the buy order will be carried out at some price higher than the old price. Similarly, when a market sale order comes in, if there is positive excess supply (negative excess demand) the sale order will be executed at some price less than the old price. Thus, if there is positive excess demand, price will rise, and if there is negative excess demand, price will fall. The specialist's order book possesses some of the characteristics as the market demand and supply schedules and the adjustment process in the stock market satisfies the intuitive description of adjustments in a perfectly competitive market.

The terms of stock transactions are flashed on the Exchange Board when they are concluded. The information is distributed immediately to interested parties. This shows that information is perfect. There are many buyers and sellers in the market. They are operating with a homogeneous product, and entry and exit are free. Price adjustments in the stock market are almost instantaneous and the supply and demand curves remain unchanged during the process of adjustment.

Appendix 10.1

PROFIT MAXIMIZATION

The profit maximization problem of competitive firm is:

$$\max_x \pi(x) = px - C(x).$$

The first order condition for maximum profit is

$$p - C'(x) = 0$$

where $C'(x)$ is the derivative of the cost function $C(x)$ with respect to x . Clearly $C'(x)$ is the marginal cost of production. The second order condition for profit maximization is

$$-C''(x) < 0$$

which means that the marginal cost must be increasing.

TAX INCIDENCE ANALYSIS

We will analyze the incidence of a tax that is charged per unit of output of the commodity. The imposition of such a tax creates a “wedge” between the price paid by demanders and the price paid by suppliers, and it is the effect of that wedge that we wish to analyze here. Let t be the per unit amount of the tax; p_D be the price paid by demanders, and p_S be the price received by suppliers.

We represent the demand curve as

$$X_D = D(p_D)$$

and the supply relationship as

$$X_S = S(p_S).$$

In equilibrium

$$X_D = X_S.$$

Assuming that before tax $p_D = p_S$, now

$$p_D - p_S = t$$

and $dp_D - dp_S = dt$.

The supply-demand system gives

$$dX_D = D_p dp_D$$

$$dX_S = S_p dp_S$$

$$dX_D = dX_S$$

where $D_p = dD(p_D)/dp_D$ and $S_p = dS(p_S)/dp_S$.

From $dp_D = dp_S + dt$ we have

$$\frac{dp_D}{dt} = 1 + \frac{dp_S}{dt}$$

Now,

$$\begin{aligned} S_p dp_S &= dX_S \\ &= dX_D \\ &= D_p dp_D \\ &= D_p (dp_S + dt), \end{aligned}$$

from which it follows that

$$S_p dt = S_p dt + S_p dp_S - D_p dt - D_p dp_S.$$

We can rewrite the expression as

$$1 + \frac{dp_S}{dt} = \frac{S_p}{S_p - D_p}$$

Hence
$$\frac{dp_D}{dt} = \frac{S_p}{S_p - D_p}$$

Multiplying the numerator and denominator of the right hand side of the above expression by p/X where $p = p_S = p_D$ and $X = X_S = X_D$, we get

$$\frac{dp_D}{dt} = \frac{e_s}{e_s - e}$$

where e_s (e) is the elasticity of supply (demand).

A similar series of algebraic steps would lead to the result

$$\frac{dp_S}{dt} = \frac{e}{e_s - e}.$$

Because of signs of the elasticities, these calculations immediately provide the obvious results that

$$\frac{dp_D}{dt} \geq 0 \text{ and } \frac{dp_S}{dt} \leq 0.$$

In particular, if $e = 0$ (demand is perfectly inelastic), $dp_D/dt = 1$ and $dp_S/dt = 0$, so that the tax is fully paid by consumers. In contrast, if demand is infinitely elastic ($e = -\infty$), $dp_S/dt = -1$ and $dp_D/dt = 0$, so that the suppliers bear the full tax. Similar polar cases can be calculated for extreme values of the supply elasticity $e_s = 0$ and $e_s = \infty$. In the intermediate cases the tax burden is shared by both demanders and suppliers.

COBWEB MODEL

The traditional models dealing with adjustment process in a market picture supply and demand decisions as being made simultaneously. This is not very

realistic. In this framework there are two equations (supply and demand curves) to be solved for two unknowns (price and quantity). The theory offers no guidance on how demanders or suppliers behave in disequilibrium situations. If the simultaneity assumption can be relaxed (by assuming, for instance, that supply decisions are made first), the problem becomes simplified. For example, if firms' current output decisions are based on the market price occurring in the previous period, then current period output can be regarded as fixed and there will be no current period supply response. The analysis of pricing in the very short run where quantity is absolutely fixed would then be the relevant model. Here price acts only as a device to ration demand. The primary examples are auctions. Price has no influence on production in the current period, since current period output is fixed. In the next period, the market price established in the current period will be the one relevant to output decisions and thus there is some lagged response of output to price changes. This kind of phenomenon is also relevant to the cycles of prices in corn and hogs. Hogs are fed on corn and there is a time lag (say, 10 to 12 months) between the time that a hog is first born and the time that it is sent to the market.

These ideas can be illustrated by a simple cobweb model of price determination. The term cobweb arises from the diagram that one can construct to describe the cyclical pattern of price and output. Suppose the firm's supply decisions in period t depend on the price which prevailed in period $(t - 1)$:

supply in period $t = X_t^S = a + bp_{t-1}$.

Market demand depends on current price:

demand in period $t = X_t^D = c - dp_t$.

Equilibrium in the market at time t requires that

$$X_t^D = X_t^S.$$

Firms decide on X_t^S depending on p_{t-1} , demanders bid for X_t^S and in doing so establish p_t . In period $(t + 1)$ firms' output decisions are based on p_t . An equilibrium price $\bar{p}$ in this setup is one that will be repeated from period to period: $\bar{p} = p_t = p_{t+1} \dots$. This equilibrium price can be obtained by setting $X_t^D = X_t^S$;

$$c - dp_t = a + bp_{t-1}.$$

At an equilibrium price $\bar{p}$,

$$c - d\bar{p} = a + b\bar{p},$$

which gives

$$\bar{p} = \frac{c - a}{b + d}.$$

Also, any price p_t can be predicted from the market clearing equation

$$p_t = -\frac{b}{d}p_{t-1} + \frac{c - a}{d}.$$

Hence, for any arbitrary initial price p_0 ,

$$p_1 = -\frac{b}{d}p_0 + \frac{c - a}{d}$$

$$p_2 = -\frac{b}{d}p_1 + \frac{c - a}{d}.$$

$$= \left(-\frac{b}{d}\right)^2 p_0 + \frac{c - a}{d} \left[1 + \left(-\frac{b}{d}\right)\right].$$

By repeated substitution,

$$\begin{aligned} p_t &= \left(-\frac{b}{d}\right)^t p_0 + \frac{c - a}{d} \left[1 + \frac{-b}{d} + \left(\frac{-b}{d}\right)^2 + \dots + \left(-\frac{b}{d}\right)^{t-1}\right] \\ &= \left(-\frac{b}{d}\right)^t p_0 + \frac{c - a}{d} \left(\frac{d}{b + d} \left[1 - \left(-\frac{b}{d}\right)\right]\right) \\ &= \left(p_0 - \frac{c - a}{b + d}\right) \left(\frac{-b}{d}\right)^t + \frac{c - a}{b + d} \\ &= (p_0 - \bar{p}) \left(\frac{-b}{d}\right)^t + \bar{p}. \end{aligned}$$

This equation shows the time path of p_t starting from an arbitrary initial position p_0 . It says that market price will oscillate around $\bar{p}$ (since the term

$[-b/d]^t$ is positive and negative alternatively). p_t converges to $\bar{p}$ if $(-b/d)^t \rightarrow 0$ as t becomes sufficiently large. This will be the case if $b < d$. That is, the demand curve is flat relative to the supply curve. This means that the initial price ultimately coincides with $\bar{p}$. We show this possibility in the cobweb supply-demand graph in Fig. 10.24. The sequence of events shown in this figure is as follows: p_0 dictates the amount X_1^S to be supplied in period 1; this determines p_1 by reference to the demand curve; p_1 is used to set X_2^S ; and the process continues. From the figure it follows that the market price will proceed toward $\bar{p}$.

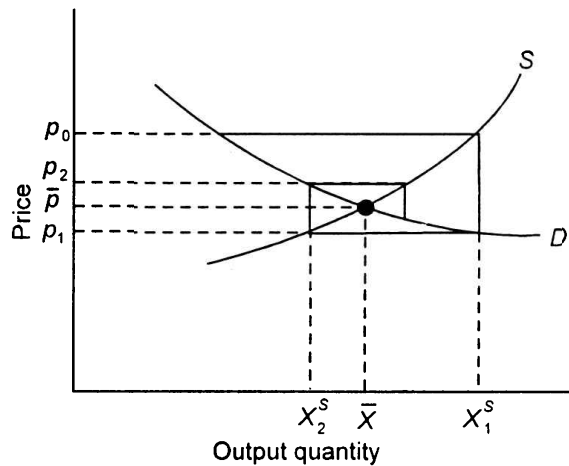


Fig. 10.24 Cobweb Model

If $d > b$, then the amplitude of price and production over time increases and the system is unstable. The reason behind this is that demand is relatively inelastic. Price changes are rather higher in response to quantity changes. But these price changes interact with an elastic supply curve so that quantity changes are even higher, and, consequently the process explodes.

RATIONAL PRICE EXPECTATIONS

The difficulty with the cobweb model is the following. Suppose that the cobweb model actually predicts how prices evolve over time. Then we have a situation of economists knowing how to predict a time pattern of prices. Then the sellers, who have stronger incentive to predict price accurately, instead use the naive rule that this period's supply decisions depend only on the market price which prevailed in the previous period. This means that the suppliers are refusing to learn from experience. Muth (1961) argued that this pattern of behavior is not consistent with the basic postulate of microeconomics, namely that economic agents follow their self-interest and behave rationally.

The alternative model that Muth proposed is that market participants would correctly deduce the consequences of any predictive theory of the market on the basis of their own observations of the market, and would incorporate all available information into the formation of their behavior patterns. This is Muth's concept of a rational expectation model of market. Rational expectations anticipate future price changes. People do not passively wait for a particular outcome to manifest itself and then revise their expectations; instead, they recognize the future consequences of present events and form their expectations accordingly. Specifically, a supplier who knew the precise forms of supply and demand functions could calculate the equilibrium price:

$$\bar{p} = \frac{c-d}{b+a},$$

and then use this to form the price expectation:

$$E(p_t) = \bar{p} = \frac{c-d}{b+a},$$

where $E(p_t)$ is what suppliers expect market price to be at time t . Using this expected price, supply will be at its equilibrium level and market will be free from the fluctuations observed in the cobweb model.

The information requirements of the rational expectations solution are quite severe. In order to guarantee a rational expectations equilibrium, we need to assume that not only can each supplier infer the equilibrium price, that is, each supplier knows the precise values of the structural parameters a , b , c and d ; but that each supplier acts on the basis of this information. This makes sense only when each supplier believes that other suppliers will act in this way as well, because if there are suppliers who follow some other behavioral rule, then price will not end up at the equilibrium level and all suppliers should alter their behavior as a consequence. Thus, we have a common knowledge condition: everyone knows what everyone else is going to do and everyone acts on the basis of this assumption. Rational expectations and common knowledge go together. As might be expected, the results of the rational expectations do not turn out to be so simple once more realistic assumptions are employed, but the approach has certainly revolutionized economists' thinking.

EXERCISES**Multiple Choice**

1. The following information applies to a perfectly competitive firm: $MR_K = 6$, $MP_L = 15$, r (rental price of capital) = 3, w (wage rate) = 5 and output price = 1. For maximizing profit the firm should use
 - a) More of capital and less of labor.
 - b) More of labor and less of capital.
 - c) More of both capital and labor.
 - d) Less of both capital and labor.
2. A perfectly competitive firm might be maximizing profits in the short run when
 - a) Average variable cost is falling.
 - b) Average cost is falling.
 - c) Marginal cost is falling.
 - d) Average product of a variable factor is rising.
3. At the shut down point
 - a) Price > average variable costs.
 - b) Revenue > variable costs.
 - c) Total losses of the firm > fixed costs of production.
 - d) None of the above.
4. When the perfectly competitive firm but not the industry is in long run equilibrium,
 - a) Price = marginal revenue = short run marginal cost = short run average cost.
 - b) Price = marginal revenue = short run marginal cost = long run marginal cost $\neq$ short run average cost = long run average cost.
 - c) Price = marginal revenue = long run marginal cost = long run average cost.
 - d) None of the above.
5. An increase in output in a perfectly competitive and constant cost industry which is in the long run, equilibrium will result from
 - a) New firms only.
 - b) Existing firms only.

- c) Both new and existing firms.
 - d) Any one of the above.
6. When the perfectly competitive firm and the industry are in long run equilibrium,
- a) Price = short run marginal cost = long run marginal cost.
 - b) Price = minimum of long run average cost.
 - c) Price = short run average cost = long run average cost.
 - d) All of the above.
7. If factor prices go up as factor usage increases, we have
- a) A constant cost industry.
 - b) A decreasing cost industry.
 - c) An increasing cost industry.
 - d) None of the above.
8. At the optimum output level a firm
- a) Maximizes its profit.
 - b) Minimizes its loss.
 - c) **Maximizes revenue and profit.**
 - d) Either (a) or (b).

Quantitative Problems and Questions

1. A perfectly competitive firm in the short run, with labor L as the only variable input, has the following production function: $x = -L^3 + 8L^2 + 140L$.
 - a) Determine the minimum employment level at which the firm will continue its production.
 - b) What is the maximum labor requirement?
 - c) Determine the level of output at which marginal cost equals average variable cost.
2. A perfectly competitive firm operates in the short run with labor L as its only variable input. Its production function is given by $x = -L^3 + 10L^2 + 88L$.
 - a) The wage rate is 324 units of money and the product price is 3.24. Determine the firm's profit maximizing employment of L .

- b) Calculate the elasticity of labor demand at the profit maximizing output.
 - c) The price of the product now rises to 4 and wage rate does not change. By how much does the firm's profit increase?
 - d) Suppose the output price is 4, and the wage rate increases to 480. How much should the firm produce?
3. A competitive firm uses one variable factor and one fixed factor to produce a single good. The price of the fixed factor rises by 7.5%, while the prices of the variable factor and output rise by 4.5%. Will there be any change in the quantity of output produced by the firm? Justify your answer.
4. Do you think that studying perfectly competitive markets is useful? Give careful analytical reasons for your answer.
5. Why does a firm close down in the short run if prices fall below the minimum average variable costs and continue to operate if the price lies in the range between the minimum average variable cost and the average cost?
6. What requirements have to be fulfilled for a market to be perfectly competitive? Discuss certain implications of these requirements.
7. Why is it necessary for the short run supply curve of a firm to be given only by the rising section of the marginal cost curve? Suppose product price and factor prices rise equiproportionally: what happens to the curve?
8. The market demand and supply schedules for a commodity are given by $D(p) = 1,000 - 100p$ and $S(p) = 500 + 100p$.
 - a) Graph the supply and demand schedules and determine the equilibrium price and quantity.
 - b) Calculate the elasticity of demand and the elasticity of supply at the equilibrium price.
 - c) Suppose that a tax of Re. 1 per unit of good is imposed. Graph the new supply schedule and determine the new equilibrium price and quantity.
9. The demand function for oil per day in India is given by $D(p) = 70,000,000 - 1,400,000p$. The domestic supply of oil in India is $S(p) = 10,000,000 + 100,000p$, where p is the per unit price. Suppose that the Organization of Petroleum Exporting Countries (OPEC) is the only foreign supplier, offering unlimited amounts of oil at Rs. 35 per unit.

- a) Plot the domestic supply and demand curves for oil together with the OPEC supply curve. Determine the price of oil in India, the amounts of domestic production and domestic consumption.
 - b) Suppose that imports are limited to 4 million units per day. What will be the new domestic price, domestic production and domestic consumption of oil?
 - c) Suppose that the government imposes an import duty on oil of Rs. 8 per unit. Determine the impact of this duty on imports, on domestic production and consumption, and on the price of oil.
 - e) Suppose that OPEC is interested in maximizing its revenues from the Indian market. What price would OPEC set?
10. Let the demand curve for a competitive industry's output be $D(p) = 100 - 5p$. Assume that the production function for each firm in the industry is $x = \sqrt{LP}$, where x is output, L is labor and P is plant capacity. The wage rate is 1, and the rental per unit of plant is 2. In the short run there are 10 firms in the industry, each with $P = 1$.
- a) Plot the short run total cost, average total cost, average variable cost and marginal cost for a firm.
 - b) Graph the short run industry demand and supply curves and find the equilibrium price-quantity combination.
 - c) Find short run profits for each firm.
 - d) Determine the long run supply curve of a firm and show that it is more elastic than the short run supply curve of the firm.
 - e) Graph the long run supply curve for the industry and determine the long run equilibrium price-quantity combination. Can you determine the number of firms?
11. Examine graphically the effect of per unit tax on the production of a good for the typical case. Show graphically the part of the tax borne by the producer and the part borne by the consumer. Do the same for the following special cases:
- a) Downward sloping demand, perfectly inelastic supply.
 - b) Downward sloping demand, perfectly elastic supply.
 - c) Perfectly inelastic demand, upward sloping supply.
 - d) Perfectly elastic demand, upward sloping supply.
- Give examples of goods for which cases (a), (b), (c) and (d) hold.

12. Using the demand curve and production function of problem 10, determine the adjustment that occurs in moving from one long run equilibrium to another, when wage rate increases from 1 to 2. Suppose that $P = 1$ for all firms, before and after the change in the wage rate. Graphically determine the following.
- The short run effect on labor demand.
 - The short run effect on firm's supply curve, the industry supply curve, the equilibrium price-quantity combination.
 - The long run effect on the choice of plant-labor combinations to produce different outputs.
 - The effect on the long run industry supply curve, and the long run equilibrium price-quantity combination.
13. Suppose that the demand and supply schedules for labor are given by $D(w) = 100 - 20w$, $S(w) = -60 + 80w$, where w is the wage rate.
- Graph the supply and demand schedules for labor. Determine the equilibrium wage rate and level of employment.
 - Suppose that minimum wage is set at $w = 2$. Determine the level of unemployment. What happens to the total earnings of laborers under fixation of minimum wage?
14. Examine the effects on the short run output supply curve of a competitive firm of:
- A property tax of plant capacity;
 - A 50% tax on profits;
 - A per unit tax of Rs. 10 on each unit of output produced;
 - A fee of Rs. 500 to operate the business, to be paid when the business starts;
 - A 1% tax on wages, half of which is borne by the employer and the remaining half by the workers.
15. The long run cost function of a firm is $C(x) = x^2/2$. What is its long run supply curve?
16. The supply curves of two firms are given by $S_1(p) = p - 8$ and $S_2(p) = p - 12$. Determine the industry supply curve.
17. In the short run an increase in price has different effects on the incomes of different segments of persons in the society. Do you think that the short run effects of price increase will be more favorable to firms' owners than to workers?

18. Determine the effect of a subsidy in a market with a
 - a) perfectly elastic supply curve,
 - b) perfectly inelastic supply curve.
19. Consumers view goods 1 and 2 as perfect substitutes. The supply curve of good 1 is upward sloping. What will be consequences of a per unit tax on good 1?
20. Suppose that India needs to import a substantial portion of oil it requires. India can import as much oil as it wants at a constant price of Rs. 50 per unit. What will be the impact on the price of domestic oil if a quantity tax of Rs. 5 is imposed on foreign oil?
21. The supply function of a firm is given by $S(p) = 4p$. Suppose that price changes from 10 to 15. What is the change in the producer's surplus?
22. The cost function of a firm is given by $C(x) = x^2/2 + 50$. At what output is average cost minimized?
23. The market demand for a good is $X_D = -2p + 13$ and the industry supply curve is $X_S = 2p^2 - 12p + 21$. Determine the equilibrium prices for this market. Which of these prices is stable by the Walrasian rule?
24. The demand curve for corn at time t is $X_t = 100 - 2p_t$ and the supply in period t is given by $X_t = 70 + E(p_t)$, where $E(p_t)$ is the supplier's expectation about price in period t .
 - a) Suppose that in equilibrium $E(p_t) = p_t$. Determine the price-quantity combination.
 - b) Suppose that suppliers use the last period's price as their expectation of this year's price (that is, $E[p_t] = p_{t-1}$). If the initial price of corn is Rs. 8, how long will it take for price to get within 0.25 of the equilibrium price?
 - c) Suppose that farmers have rational expectations. How do they choose $E(p_t)$?
25. A perfectly competitive market has 1,000 firms. In the very short run each firm has fixed supply of 100 units. The market demand is given $X = 1,60,000 - 10,000p$.
 - a) Calculate the equilibrium price in the very short run.
 - b) Calculate the demand schedule that any one firm faces.
 - c) Calculate the equilibrium if one of the sellers decided to sell nothing or if one seller decided to sell 200 units.

- d) At the original equilibrium point, calculate the elasticity of any one seller and of the industry demand.
26. Explain how the supply curve of a competitive industry is derived from the individual firms' supply curves. How could an improvement in technology affect the market supply curve?
27. Long run equilibrium requires minimum of long run average cost to equal minimum of short run average cost to equal price. Explain this condition. What would happen if price was greater than minimum of short run average cost, which in turn was greater than minimum of long run average cost?
28. Explain why, under conditions of perfect competition, all firms produce with minimum price and efficient use of plant capacity.
29. Consider two perfectly competitive industries operating side by side in a city. Assume that good 1 is used as an input in good 2's production. Suppose that the demand for good 2 increases as the fine quality of the good becomes known. What would happen to price and output in each industry?
30. Why does a perfect competitor have no effect on price? What happens if a competitor tries to sell below the price?
31. Suppose that the quantity demanded of good 1 depends on p_1 (price of good 1), p_2 (price of good 2) and m (income). Price elasticity of demand for good 1 is -1.0 , the cross price elasticity of demand for good 1 is 1.0 , the income elasticity of demand for good 1 is 0.5 , and the short run elasticity of supply for good 1 is 2.0 . Initially, the price of good 1 is Re .5 per unit, the price of good 2 is Rs. 1.5 per unit and income is Rs. 800 billion. At these prices and income, 2 billion units of good 1 are bought.
- Now suppose that the price of good 2 rises to Rs. 1.80 per unit and income increases to Rs. 960 billion:
- How will the demand for good 1 shift?
 - By how much will the price of good 1 rise in the short run? What will be the new equilibrium price of good 1?
 - How much of good 1 will now be bought? Draw graphs to illustrate your results.
32. A perfectly competitive industry has a large number of potential entrants. Each firm has an identical cost structure such that long run average cost

is minimized at an output of 20 units. The minimum average cost is Rs. 10 per unit. The total market demand is $X = 1500 - 50p$.

- a) What is the industry's long run supply schedule?
- b) What are the long run equilibrium price, the total industry output and the number of firms?

33. There are 100 identical firms in a competitive industry. Each firm has a short run cost function of the form

$$C(x) = \frac{1}{300}x^3 - 0.2x^2 + 4x + 10.$$

- a) Determine the short run supply curve of a firm.
- b) Calculate the short run industry supply curve.
- c) The market demand curve is $X = -200p + 12,000$. Determine the equilibrium price-quantity combination.

34. There are 100 identical firms producing a good and the short run total cost function for each firm is of the type $C(x) = x^2 + \bar{w}x$, where x is the firm's output and $\bar{w}$ is the wage rate.

- a) If $\bar{w} = 10$, what will be the firm's short run supply curve? Determine the amount of the good that will be produced if the price p of the good is 20. Determine the additional units of the good that will be supplied if p increases to 21.
- b) Suppose that the wage rate and total industry output X are related by $\bar{w} = 0.002X$, where $1,000x = X$. What is the short run supply curve of the firm in this situation? What is the industry supply curve? How much will be produced if $p = 20, 21$?

35. Demand for a good is given by $X = 100 - 2p$ and the supply by $X = 20 + 6p$.

- a) Find the equilibrium price-quantity combination.
- b) The government imposes a tax of Rs. 4 per unit of the good. Determine the equilibrium price-quantity combination. How is the burden of tax shared by buyers and sellers?
- c) How would your answers to (a) and (b) change if the supply curve is of the form $X = 70 + p$?

36. Individual producers of a good have U-shaped, long run average cost curves that reach a minimum average cost of Rs. 3.00 per unit when 1,000 units are produced.

- a) If the demand curve for the good is $X = 2,600,000 - 200,000p$, determine the long run price output combination. How many firms will there be?
- b) The demand curve shifts to $X = 3,200,000 - 200,000p$. The producers cannot adjust their output in the short run. What will be the market price?
- c) What will be the new long run equilibrium?

37. The cost functions of firms 1 and 2 are given by

$$C_1(x_1) = 0.04x_1^3 - 0.8x_1^2 + 10x_1;$$

$$C_2(x_2) = 0.04x_2^3 - 0.8x_2^2 + 20x_2.$$

Determine the two firms' supply curves and the total supply curve.

38. The long run cost function for each firm that supplies some good is $C(x) = x^3 - 4x^2 + 8x$. Describe the industry's long run supply function. The demand function for the good is $X = 2,000 - 100p$. Determine the number of firms in long run equilibrium.
39. Each firm in a perfectly competitive constant cost industry has the following long run cost function $C(x) = x^3 - 50x^2 + 750x$. The market demand curve for the product is $X = 2,000 - 4p$.
- a) What is the long run supply curve of the industry?
 - b) How many firms are there in the long run equilibrium?
 - c) A sales tax of 20% of the market price is imposed on the product. How many firms will there be in the new long run equilibrium?
 - d) Suppose instead of sales tax an excise tax of Rs. 50 per unit is imposed. How many firms will there be in the new long run equilibrium?
 - e) Suppose the government, instead of imposing any tax, pays some subsidy so that the number of firms in the industry increases by three from the situation in (b). How much is the per unit subsidy?
40. The production function of a perfectly competitive firm is given by $x = a\sqrt{(y_1y_2)}$ where $a > 0$ is a constant and y_1 and y_2 denote the quantities of two inputs used by the firm. The input markets are also competitive. Derive the shapes of the short run average variable cost curve and the long run average cost curve.
41. A firm's long run average cost curve is U-shaped and possesses its minimum at an output of 25,000 units. In the short run the firm has a U-

shaped average cost curve that coincides with the long run average cost at an output of 26,000 units. Is the short run average cost rising, falling, or neither at an output of 26,000 units? Is the short run marginal cost greater than, or equal to the long run marginal cost at an output of 25,999.9 units?

42. Show that in the long run, proportionate increase in the prices of output and factors lead to proportionate increase in profits.
43. The market demand and supply curves for an agricultural product are given by $X_D = 4,500 - 250p$, $X_S = 200p$. The government wishes to achieve a higher point on the supply curve than the initial equilibrium. The desired point would require both price and quantity to be 10% higher than the original equilibrium levels. The government is considering either a subsidy or a support price.
 - a) If the subsidy were used, how much would the subsidy be? What would be the government's costs?
 - b) If the support price is used, how much would the government buy? What would be the government's total costs?
 - c) Do you think (a) is a better policy than (b)?
 - d) Would your conclusion in (c) be different if the demand curve had been $X_D = 3,600 - 160p$?
44. A competitive firm's production function is given by $X = -0.01L^3 + L^2 + 36L$, where L is input in man-hours. The output price is Rs. 10, and the wage rate is Rs. 4.80. If the total fixed cost is Rs. 50, how much is the profit?
45. A good is produced by a perfectly competitive industry that is in the long run equilibrium with 100 firms, each employing 25 workers. Each firm's production function is of the type: $x = -0.1(PL)^{3/2} + 10(PL) + 50(PL)^{1/2}$, where L is the number of workers employed, and P is the plant size. The market demand is $X = 2.25 - 0.5p$. Determine the wage rate and the rental price of the plant.
46. A perfectly competitive constant cost industry contains a number of firms, each of which has the following long run cost function $C(x) = 0.01x^3 - 1.2x^2 + 111x$. The market demand is $X = 6,000 - 20p$.
 - a) Determine the long run equilibrium output.
 - b) How many firms are there in the industry in the long run equilibrium?
 - c) The government decides to reduce the number of firms to 60 and sells only 60 licences. The licensed firms operate competitively. What is

the new equilibrium price? What is the competitive equilibrium price of the licence?

47. A firm uses plant and labor to produce a good. In the short run the plant size is fixed, while labor is variable. The rental price of plant is Rs. 14 and the wage rate is Rs. 3.40. On the basis of these factor prices the firm builds its plant so as to maximize profits at the expected price of the good over which it has no control. The firm's short run and long run cost functions are given by

$$\text{Short run cost} = 2x^3 - 24x^2 + 157x + 98,$$

$$\text{Long run cost} = x^3 - 12x^2 + 136x.$$

- a) What will be the expected price of the good?
 - b) Determine the amount of the good that will be produced by the firm.
 - c) What is the lowest price at which the firm will continue to produce in the long run?
 - d) If the price of the good is Rs. 211, how much output will the firm produce?
 - e) If the rental price of plant rises to Rs. 15 and the wage rate rises to Rs. 4, how much output will be produced in the short run?
48. A commodity is produced domestically by a perfectly competitive industry under conditions of increasing costs. Each firm has the following long run cost function:

$$C(x) = 0.01x^3 - 1.6x^2 + (70 + 0.005X)x$$

where x is the output of the firm and X is the industry output. The product can be imported or exported at a constant price of Rs. 10.00. The domestic market demand curve is given by $X = 27,900 - 1,500p$. What quantity is imported or exported? How many domestic firms are there?

The government imposes a 20% tariff on imports of the good. The market adjusts to a new long run equilibrium. What quantity is imported or exported. How much is the domestic price now? How many firms are there now? By how much have domestic profits changed as a result of the tariff? Are the domestic consumers better off or worse off as a result of the tariff? How much is the government revenue from the tariff?

49. In the long run, a specific tax imposed on a competitive decreasing cost industry will raise market price by more than the amount of the tax. Show that this is true.

50. The long run total cost function for a competitive firm is given by $C(x) = wx^2 - 10x + 100$, where w is the wage rate of workers. The market demand for the output of the competitive industry is $X = -1,000p + 40,000$.
- If $w = 1$, what will be the long run output for the typical firm?
 - Assume that the industry exhibits constant costs and that all firms are identical. What will be the long run equilibrium price of output and how many firms are there?
 - Suppose w increases to Rs. 4, say, because the government imposes a tax of Rs. 3 on each worker hired. How will your answers to (a) and (b) change?
 - How will your answers to (a) and (b) change if the market demand becomes $X = -1,000p + 60,000$?
51. As output in the steel industry goes up, the cost of iron ore is bid up. This causes the industry to exhibit increasing costs. Suppose that one steel firm owns its own iron ore. Do that firm's costs increase as output expands?
52. Each firm in a competitive industry has the cost function: $C(x) = 200 + 10x + 2x^2$. The market demand curve for the industry's product is $p = 100 - X/8$. There is free entry. What is the equilibrium output-price combination?
53. The market demand curve for a competitive industry's output is $p = 100 - X/8$. Each firm has a minimum average cost of $50 - 0.025X$. There is free entry. Examine the consequences of : (a) a maximum price of 40, (b) a minimum price of 40, (c) a maximum price of 30.
54. A competitive firm's production function is given by $x = P^{1/4} L^{1/4}$, where P = plant size, and L = labor input. Suppose that the rental price of capital = wage rate = 1.
- Determine the long run supply curve.
 - Suppose $P = 1$. Find the short run supply curve. Which is more elastic, the long run or the short run supply curve? For what output is $P = 1$ the optimal plant size?
 - Suppose $P = 2$. Compare the short run cost curve with the one you get in (b).
 - Suppose that the output price = 4. A levy of 2.5 is imposed on all firms. Determine the (i) long run output of the firm, (ii) short run output of the firm if $P = 1$.

55. The production function of a competitive firm is $L = x - x^2 + x^3$, where L is the only input of the firm and x is the firm's output.
- Derive the firm's output supply curve if the input price is Rs. 2.
 - How much will the firm produce if the output price $p = \text{Rs. } 2$, and if $p = \text{Re. } 1$? What will the firm's profit be if $p = \text{Rs. } 2$?
 - For what values of x are returns to scale increasing, decreasing or constant?
 - Suppose that the output price is Re. 1. What will the firm do if a levy of Re. 1/2 is imposed on the firm?
 - Suppose a profit tax of 25% is imposed on the firm. How much tax will the firm pay if $p = \text{Re. } 1$ and the input price is Re. 1?
56. Prove or disprove the following:
- If a competitive firm finds that its minimum possible level of average cost is higher than the price of its product, then the firm's profit maximizing decision is to shut down production.
 - A competitive firm cannot maximize its profits if the last worker hired adds to its output more than the average productivity of the already existing workers.
 - A perfectly competitive firm operating in the short run has a falling average variable cost curve. Then to maximize profits the firm should increase the input of the variable factor.
 - In the short run the demand for a good is perfectly inelastic and in the long run it is perfectly elastic. Then in the short run the suppliers pay the entire tax and in the long run the consumers pay the entire tax.
 - In a perfectly competitive market, the market price must be equal to the marginal cost of production for all firms in the industry.
 - In the long run equilibrium no firm will earn negative profits.
 - In the short run a tax on profits is born entirely by owners of firms, but in the long run the tax is entirely shifted in the form of price increases.
 - Although fixed costs may be the cause of losses, they can never be the reason for closing down.
57. Explain how free market forces can lead to efficient allocation of resources.
58. Graphically derive the long run supply curve of a decreasing cost industry.

11

Monopoly

The market for a particular good is called a monopoly if there is only one producer of the good. This single firm faces the entire market demand. Using its knowledge about this demand, the monopoly chooses that level of output that maximizes its profits. Unlike a perfectly competitive firm's output decision (which does not have any effect on the market price), the monopolist's output decision will determine the market price. That is, in contrast to a firm in a perfectly competitive market, a monopolist is a price setter, not a price taker. In this sense monopoly markets and markets characterized by perfect competition are opposite polar cases.

In some texts a monopoly is treated as having the power to set price, and in a sense a monopolist possesses this power. Technically, a monopoly firm can choose that point on the market demand curve at which it prefers to operate. It may choose either price or quantity, but not both independently. It may choose the price and let the consumers decide how much they wish to buy at that price, or it may select the quantity. Whenever a firm can influence the price it receives for its product, the firm is said to have market power.

In this chapter we will assume that the monopoly firm under consideration chooses that quantity of output which maximizes its profit and then

settle for the market price that output level yields. It would be a simple matter to reformulate the discussion in terms of price setting.

COSTS AND REVENUES UNDER MONOPOLY

The cost concepts we have developed in Chapter 9 are applicable to monopoly as they are to competition. The differences between monopoly and competition are brought about by the conditions under which goods and services are sold, not by the conditions under which resources are purchased. We will assume throughout this chapter that the monopolist seller of a product is a competitive buyer of resources and does not have any influence on resource prices. The monopoly firm can purchase as much of any resource as it desires without affecting the resource price. (Discussions on cost functions that take into account the influence of firms on resource prices are made in Chapter 15. They would make no major difference in the development of this chapter.)

As we have stated, there are differences on the selling side between a competitive firm and the monopolist. A competitive firm faces a horizontal demand curve and a marginal revenue curve that will coincide with it, since the firm can sell any desired amount at the prevailing price. However, the monopolist faces the market demand curve which, if the law of demand holds, is downward sloping to the right; charging a higher price for a lower number of units sold. This downward sloping demand curve has important implications for the monopolist's marginal revenue curve in relation to price.

We write X for the output of the monopolistic firm. Let us use $p(X)$ to denote the market (inverse) demand function and $C(X)$ to denote the cost function. The price per unit of output is written as $p(X)$, to show that the price p , the monopolist can charge is related to the output X that can be sold through the market demand function.

At any level of output the monopolist's revenue R equals the price multiplied by the number of units of output sold. That is, at the output level X , revenue $R(X)$ is

$$R(X) = p(X)X. \quad (11.1)$$

To identify the monopolist's revenue diagrammatically, let us consider the downward sloping market demand curve that the monopolist faces. The firm's demand function is illustrated in Fig. 11.1. Revenue at output X_1 , is $R(X_1)$, which is the area of the rectangle labeled R_1 .

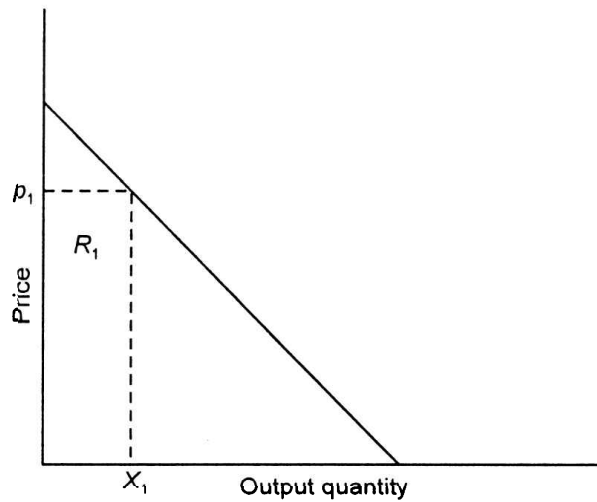


Fig. 11.1 Monopolist's Demand and Revenue

Now, what is the relationship between output and revenue as output increases? The exact relationship is revealed by the elasticity of the demand curve. We have seen in Chapter 5 that if at a given level of output the demand is elastic, then an increase in output (decrease in price) leads to an increase in revenue. On the other hand, if demand is inelastic, then an increase in output (a decrease in price) generates a reduction in revenue. If demand is of unitary elasticity, then revenue remains unchanged as output changes.

A monopolist's average revenue is defined as the revenue per unit of output. Average revenue $AR(X)$ at the output level X is given by

$$\begin{aligned}
 AR(X) &= \frac{R(X)}{X} \\
 &= \frac{p(X) \cdot X}{X} \\
 &= p(X).
 \end{aligned} \tag{11.2}$$

Thus, the monopolist's average revenue curve is the market demand curve itself.

Marginal revenue at X , $MR(X)$ is the change in revenue when the firm increases its output by one more unit. If the monopolist decides to increase his output by ΔX , there are two effects. First, the monopolist's revenue increases by the amount $p\Delta X$ from that. Second, the monopolist pushes the

price down by Δp and gets lower price from all the units it was selling previously (that is, X units) to sell the added ΔX units. Therefore the total change in revenue ΔR due to ΔX change in output is $\Delta R = p\Delta X + X\Delta p$. Consequently, marginal revenue MR , change in revenue divided by the change in output, is

$$\begin{aligned} MR(X) &= \frac{\Delta R(X)}{\Delta X} \\ &= \frac{p\Delta X + X\Delta p}{\Delta X} \\ &= p + X \frac{\Delta p}{\Delta X} \end{aligned} \quad (11.3)$$

where $\Delta p/\Delta X$ is the slope of the demand curve. The expression of Eq. (11.3) can be contrasted with the marginal revenue of a competitive seller. In the competitive case if a firm lowers its price below the existing price charged by other firms, then it will capture the entire market from its competitors. But in the case of the monopoly, since the monopoly controls the whole market, it has to take care of the effect of a price reduction (quantity increase) on all units it was selling previously.

In view of our discussion in Chapter 5, we can also express the marginal revenue in terms of the price elasticity of demand using the formula

$$MR(X) = p(X) \left(1 + \frac{1}{e} \right) \quad (11.4)$$

where e is the price elasticity of demand (see Eq. 5.9). When demand is inelastic (that is, $e > -1$), MR is negative, when demand is unit elastic ($e = -1$), $MR = 0$; and when demand is elastic ($e < -1$), MR is positive.

Given that the demand curve is downward sloping, $e < 0$. From the expression derived for MR in Eq. (11.4) it then follows that at any positive output level, MR is smaller than price, that is, the MR curve lies below the demand curve.

Note that in the competitive case the firm faces an infinitely elastic demand curve, that is, $e \rightarrow -\infty$ in this case. Therefore, from Eq. (11.4) it follows that for a competitive firm marginal revenue coincides with price, an observation which we have made earlier.

LINEAR DEMAND CURVE AND MONOPOLY REVENUES

To illustrate monopoly average revenue, marginal revenue and the relationship between change in revenue and change in quantity of output diagrammatically, let us consider the linear demand curve of Fig. 11.1. The equation for a linear demand curve is

$$p(X) = a - bX, \quad (11.5)$$

where a and b are constants. The slope of this demand curve is $-b$. The constant a is the maximum price that the monopolist can charge. The revenue function is

$$R(X) = p(X)X = aX - bX^2. \quad (11.6)$$

The change in revenue when output increases by ΔX is $\Delta R(X) = p\Delta X + X\Delta p = (a - bX)\Delta X - X(b\Delta X) = (a - 2bX)\Delta X$. Therefore the marginal revenue is

$$MR(X) = \frac{\Delta R(X)}{\Delta X} = a - 2bX. \quad (11.7)$$

The marginal revenue curve has the same vertical intercept, a , as the demand curve, but it has a horizontal intercept which is half of that of the demand curve. We show this in panel (a) of Fig. 11.2.

The elasticity of the linear demand curve of Eq. (11.5) is $-(a-bX)/bX$. The point of unitary elasticity occurs at the midpoint of the demand curve, that is, at the output level $a/2b$ and marginal revenue is zero at that output. Demand is elastic or inelastic accordingly the output level is less than or greater than $a/2b$.

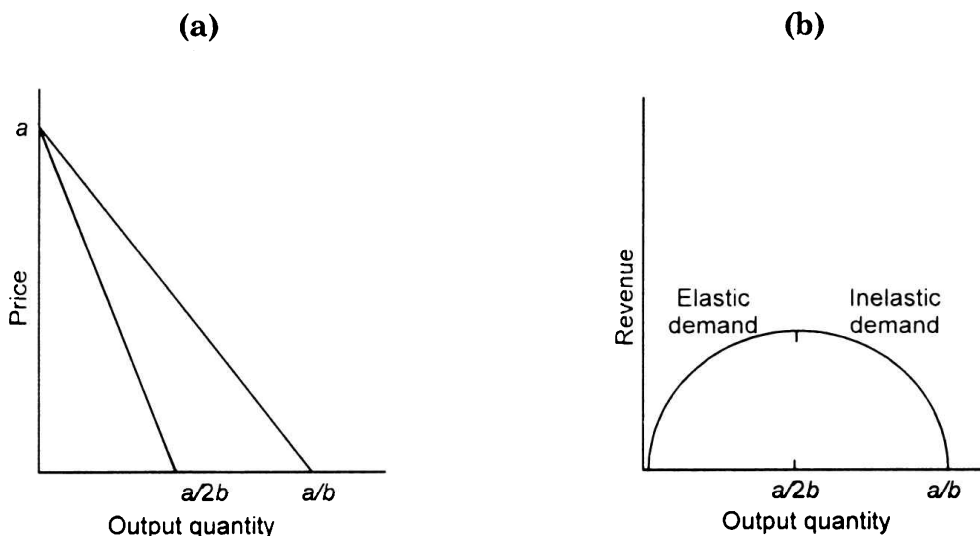


Fig. 11.2 Monopoly Revenue, Marginal Revenue: Linear Demand

We know that if the demand for a monopolist's output is elastic (inelastic), then an increase in output leads to an increase (a decrease) in the monopolist's revenue. Therefore as output increases from 0 to $a/2b$, where marginal revenue is zero, revenue is at a maximum. Finally, as output increases from $a/2b$ to a/b , revenue starts declining. This is shown in panel (b) of Fig. 11.2.

In terms of the revenue graph at any output level, average revenue equals the slope of the line drawn from the origin to the revenue graph (see panel [a] of Fig. 11.3). Average revenue at X_1 , $AR(X_1) = R(X_1)/X_1 =$ slope of the line drawn from the origin to the $R(X)$ curve at X_1 . On the other hand, marginal revenue can be interpreted as the slope of the revenue function. In panel (b) of Fig. 11.3, $MR(X_0)$, marginal revenue at X_0 , is the slope of the line AB , which is tangent to $R(X)$ at $X = X_0$.

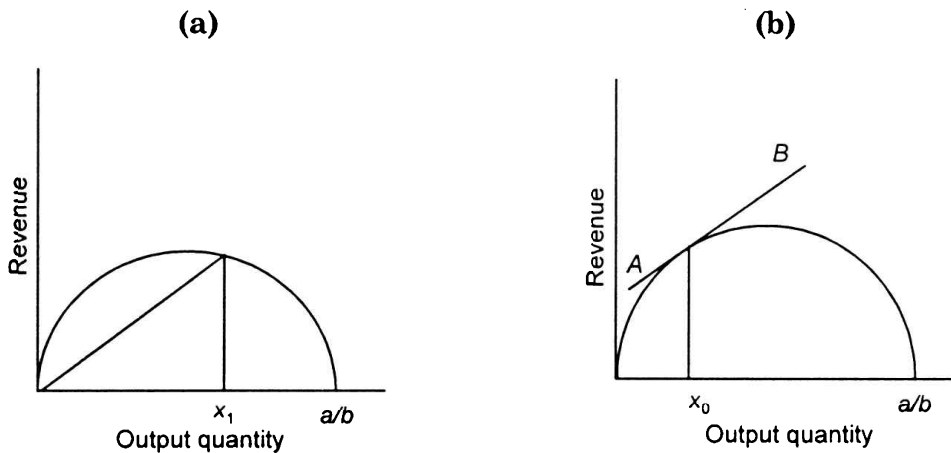


Fig. 11.3 Monopoly Average Revenue and Marginal Revenue: Linear Demand

MONOPOLY PROFIT MAXIMIZATION

A monopolist faces a downward sloping demand curve and must decide how much of his product to sell for maximizing profits: the more he sells, the lower the price at which he has to sell. The monopolist's profit maximization problem takes the form

$$\max_x \pi(X) = \max_x R(X) - C(X), \quad (11.8)$$

where $\pi(X) = R(X) - C(X)$ is the monopolist's profit. The optimality condition for the problem of Eq. (11.8) is that marginal revenue equals marginal cost. That is, the profit maximizing monopolist produces that quantity for which

marginal revenue equals marginal cost, the slope of the total cost curve. Algebraically, the profit maximizing condition is

$$MR(X) = MC(X). \quad (11.9)$$

Using Eq. (11.4) we can rewrite Eq. (11.9) as

$$p(X) \left(1 + \frac{1}{e} \right) = MC(X). \quad (11.10)$$

If output is such that $MR > MC$, then profits can be increased by increasing output, if output is such that $MR < MC$, then profits can be increased by decreasing output. It is only when $MR = MC$ that profits are maximized.

We have noted that if the demand for the monopolist's output is inelastic ($e > -1$), then marginal revenue is negative. In such a case it is not possible to equate marginal revenue with marginal cost, which is positive. That is, the profit maximizing condition of Eq. (11.10) is not applicable. Thus, a profit-maximizing monopoly firm will never operate at a price and level of output such that the demand curve is inelastic at that output. If demand is inelastic, then a reduction of output increases the revenue and decreases the cost, hence profits increase. Consequently, a monopoly firm cannot maximize profits at an output where demand is inelastic. In terms of Fig. 11.2 (panel [b]), profit maximization implies that $X < a/2b$.

Since the monopoly firm buys inputs in competitive input markets, the monopolist's cost curves are derived in exactly the same manner in which the competitive firm's curves are derived. In panel (a) of Fig. 11.4 we draw the (total) cost curve and the revenue curve. At any output X , profits $\pi(X) = R(X) - C(X)$ are measured by the vertical distance between the $R(X)$ and $C(X)$ curves. Profits are maximized at output $\bar{X}$, where the difference between the $R(X)$ and $C(X)$ curves is the maximum. The output at which the difference between the revenue and the total cost curves is the greatest is that at which their slopes are equal (i.e. the tangents to the curves at this output are parallel). Since the slope of the total cost curve is marginal cost and that of the revenue curve is marginal revenue, profits are maximum at the output at which marginal revenue equals marginal cost. We show this in panel (b) Fig. 11.4

At $\bar{X}$, marginal revenue equals marginal cost. The price charged by the monopolist is given by $\bar{p}$, the price associated with $\bar{X}$ on the demand curve. Profits are represented by the area of the rectangle $\bar{p}FEG$, that is, by the product of the per unit profits ($AR[\bar{X}] - AC[\bar{X}]$) and $\bar{X}$, the number of units of output produced, where AC denotes average cost.

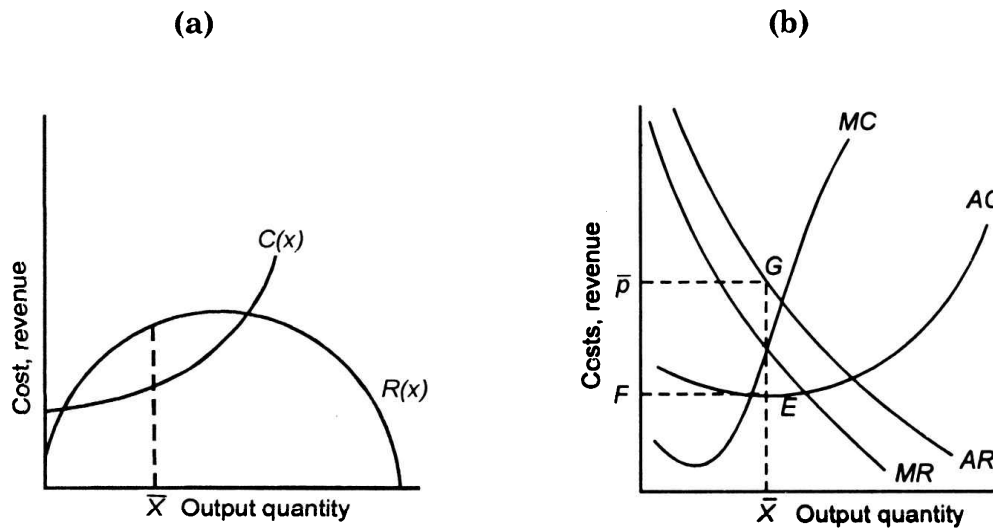


Fig. 11.4 Monopoly Profit Maximization

Monopoly profits will be positive if market price exceeds average total cost. Since a monopoly is a single-firm industry, the monopolist's profits can exist even in the long run. That is why some authors refer to the profits that a monopoly earns in the long run as monopoly rents. These profits can be regarded as a return to the factors that form the basis of monopoly (a patent, or a favorable location, for example).

The monopolist will produce a positive output only if he can cover his cost. If price is so low that it is less than average costs (AC), then the profit maximizing output level is zero; consequently, the monopolist will shut down his plant. This situation can arise only if the demand curve is below the AC curve at all output levels (Fig. 11.5). Note that at $\bar{x}$ in Fig. 11.5, where MR

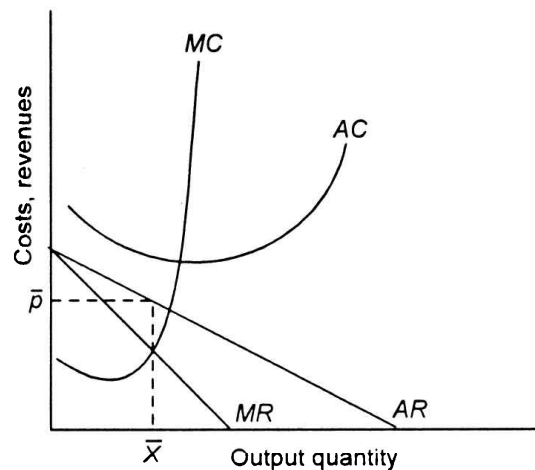


Fig. 11.5 Monopoly Loss Minimization

$= MC$, $\bar{p}$ is less than AC , because AR lies below AC at all output levels. If the firm decides to operate, its losses are part of its costs. Therefore if $AR < AC$ for all output levels the profit maximizing (or loss minimizing) decision is to close the plant. (However, as in the competitive case, the monopolist may incur losses in the short run and continue to produce if price covers more than average variable costs. For this situation to arise we require some output levels at which average revenue exceeds average variable cost. In this case losses are less than fixed costs.)

MONOPOLY SUPPLY OF OUTPUT

In the case of the competitive firm the $MR = MC$ rule can be employed to derive the supply curve of output of the firm, where the supply curve becomes the MC curve above the minimum of AC . (Under constant returns to scale a firm's supply curve coincides with its marginal cost curve.) Is there anything that corresponds to this in the monopoly case? The answer is no. A monopoly firm has no supply curve. That part of the MC curve lying above the average cost curve does not show the quantities that the monopolist would place on the market at alternative prices above minimum AC ; instead it depicts the quantities that would be placed on the market at alternative levels of marginal revenue for that firm. Given the market demand curve the monopolist simply picks out one profit maximizing output-price combination.

Any given level of marginal revenue at the profit maximizing output level is consistent with several demand possibilities. To illustrate this suppose that a monopolist faces the demand curve D_1 in Fig. 11.6. Profits are maximum at X_1 , the product price is p_1 and the level of marginal revenue is r_1 .

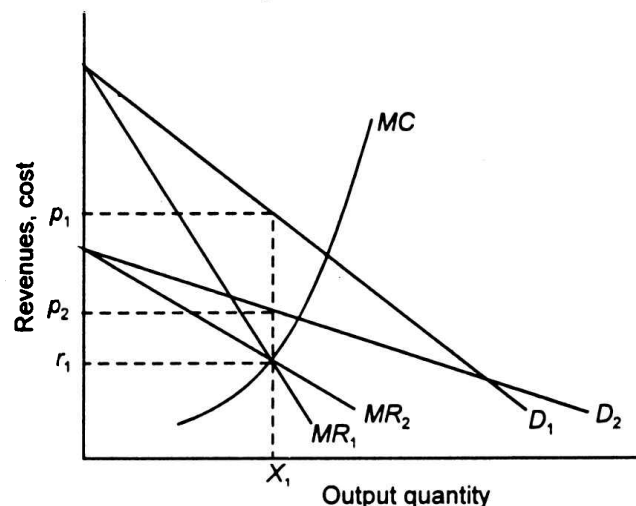


Fig. 11.6 Absence of Supply Curve in Monopoly

Now suppose that a new demand curve D_2 is such that the corresponding marginal curve MR_2 intersects the marginal cost curve at X_1 where marginal revenue is r_1 . The price at which output X_1 will be sold under D_2 is p_2 instead of p_1 . Thus, the monopolist's output is not determined by marginal cost and price but by marginal cost and marginal revenue.

MONOPOLY PRICING AND MEASURE OF MARKET POWER

Lerner (1934) defined the market power or monopoly power of a firm as the proportional excess of price over marginal cost. Rearranging the monopolist's profit maximization condition of Eq. (11.10) we can express the Lerner market power index L for the monopolist as

$$L = \frac{p(X) - MC(X)}{p(X)} = -\frac{1}{e} \quad (11.11)$$

The measure L is an indicator of the ability of the firm to set price above the marginal cost. Since the monopolist operates in the region where the demand curve is elastic, $0 < L < 1$. Thus, market price is a mark-up over marginal cost for the monopolist. We can use L or an increasing transformation of L as a measure of mark-up for the monopolist. One such increasing transformation of L is $L' = 1/(1 - L) = 1/(1 + 1/e)$. Clearly, $L' > 1$. As L increases, for instance, due to a fall in the marginal cost of the firm, L' increases.

A zero value of the Lerner market power index for a firm expresses the inability of the firm to influence the price and hence set it above the marginal cost. Clearly, for a competitive firm, market power index will take on the value of zero. A positive value of the Lerner index means that the firm has some market power—it has some ability to influence the price level.

The measure L defined in Eq. (11.11) depends on the monopolist's optimal output level. If the monopolist faces a constant elasticity demand curve, then L , the proportionate excess of price over marginal cost, becomes a constant.

COMPARATIVE STATICS ANALYSIS

A very general property of monopoly pricing is that monopoly price will increase as the marginal cost increases. To see this, suppose that the monopolist's average cost of production is constant. Hence the monopolist's marginal cost is a constant. Let C_1 and C_2 be two marginal costs, where $C_1 < C_2$. Suppose that (X_1, p_1) is the monopolist's output-price combination if

the marginal cost is C_1 . Similarly, for the marginal cost C_2 , (X_2, p_2) is the output-price combination.

Since the monopolist is a profit maximizer,

$$p_1X_1 - C_1X_1 > p_2X_2 - C_1X_2,$$

and
$$p_2X_2 - C_2X_2 > p_1X_1 - C_2X_1.$$

The first inequality means that the monopolist has a higher profit if he chooses (X_1, p_1) instead of (X_2, p_2) if his marginal cost is C_1 . Interpret the second inequality similarly. Adding these two inequalities we get

$$C_2(X_1 - X_2) > C_1(X_1 - X_2),$$

or
$$(C_2 - C_1)(X_1 - X_2) > 0. \quad (11.12)$$

Because $C_2 > C_1$, inequality Eq. (11.12) implies that $X_1 > X_2$. Given that the monopolist faces a downward sloping demand curve, it must be the case that $p_1 < p_2$. Therefore monopoly price is an increasing function of marginal cost.

This result seems quite reasonable. Since the monopolist controls the entire output market, an increase in marginal cost will force him to increase price for maintaining maximality of his profits.

PRICE DISCRIMINATION

In some cases a monopolist finds it possible and profitable to sell different units of output at different prices. Selling different units of output at different prices is referred to as price discrimination. When price discrimination exists we say that prices are non-uniform. Everyday experience confirms that pricing is more complicated than simple uniform price per unit. It should be mentioned that three conditions are essential for successful price discrimination. Their presence does not always ensure price discrimination, but it can never happen without them. First, a firm must have some market power, otherwise it can never succeed in charging non-uniform prices. (A monopoly firm always possesses some market power). Second, the firm must know the consumer's willingness to pay. Third, a firm must be able to prevent or limit resale from consumers who pay the lower price to those who pay the higher price. In such a case consumers who buy directly from the firm at a higher price will never do that: they will buy at a lower resale price from the group which is charged a lower price by the firm. Economists generally consider these three types of price discrimination: first-degree, second-degree and third-degree. We will discuss each of these three kinds of price discrimination separately.

First-Degree Price Discrimination

First-degree price discrimination, also called perfect price discrimination, occurs when the monopolist is able to sell different units of output at different prices, extracting the highest possible price for each unit sold. Hence under first-degree price discrimination each unit is sold to the individual who values it most highly, that is, who is willing to pay the maximum price for it. Thus, in this case prices may differ from person to person. It is clear that in such a market consumers have no consumer's surplus remaining. (Recall that consumer's surplus is the maximum amount that could be extracted from a consumer above and beyond what he is paying.)

Suppose that the monopolist can identify each consumer and knows the maximum amount that each one is willing to pay. Then the firm will charge each consumer a different price and extract the full consumer's surplus from each one. As long as the price to the last buyer exceeds the marginal cost, the monopolist will sell more of his product to additional consumers, because he makes an extra profit from additional sales. Therefore under first-degree price discrimination the monopolist charges the last buyer the marginal cost. Perfect price discrimination entirely avoids the effect of decreased revenues that result from passing along a lower price to all who were already buying the good, because a low price to one consumer can no longer be passed on to another. Elimination of this effect ensures that the demand curve becomes the marginal revenue curve.

An interesting feature of perfect price discrimination, is that it generates the same output as would be produced under perfect competition. The reason is that under both cases consumers purchase the good as long as they value it above marginal cost. Therefore, under perfect price discrimination the sum of the producer's and consumer's surplus is maximized. However, the producer captures all the surplus generated in the market. Consequently, perfect price discrimination entails no efficiency loss, but alters the distribution of income (more of the consumer's money is transferred to the monopolist).

Real world examples of first-degree price discrimination are hard to find. But it generates a resource allocation mechanism other than a perfectly competitive market that achieves Pareto efficiency. Since perfect price discrimination requires detailed knowledge about buyers, it is more likely to occur when one-by-one bargaining occurs. A doctor may be able to successfully price discriminate if he can identify the wealthy people in the area.

Second-Degree Price Discrimination

According to second-degree price discrimination, the monopolist sells different units of output at different prices, but every consumer who purchases the same quantity of the good pays the same price. Thus price differs across units of the good, but not across buyers. In second-degree price discrimination, total expenditure on the good is not equal to a constant price multiplied by the quantity of the good bought. That is, a consumer's total expenditure on the good does not increase linearly or proportionately with the quantity bought. Therefore second-degree price discrimination can also be called non-linear pricing.

A simple type of non-linear pricing schedule is a two-part tariff. A two-part tariff consists of a lump sum fee plus a charge per unit. For example, in an amusement park, visitors pay an admission fee and then separate fees for each ride. In a club, all members pay a common annual membership fee together with usage fees that vary with the intensity of the use of the club's facilities. Telephone companies charge a subscription fee and then a usage fee that depends on the number of calls made. Many firms that rent out copy machines charge a rental fee plus a fee that depends on the extent of use of the machine. Electricity bills are often computed according to block schedules, in which the first block of units of usage incurs one charge, and subsequent blocks of units incur a different charge. A buyer of a polaroid instant-picture camera buys the camera for one price and its film for another. The price of the camera is the lump sum fee, and film price is a usage-fee. This is also an example of tie-in sales. A tie-in sale is the sale of one product under the implicit or explicit condition that the buyer also buys another. A second example of tie-in sales is the purchase of a durable machine under the condition that the consumer also purchases from the seller all repair parts and repair services. Further examples can be given. In general, analysis of non-linear pricing is quite complicated and is beyond the scope of this book. The interested reader is referred to Carlton and Perloff (1990).

Here we will analyze a simple case of determination of an optimal two-part tariff. Let us suppose that the consumers are identical and demand more units of the monopolist's good as price falls. That is, each consumer is identical to all other and has the same downward sloping demand curve for the monopolist's good as price falls. Marginal cost is assumed to be constant, say C .

The monopolist could charge a price of C per unit consumed and also charge each (identical) consumer S for the right to buy the product, where S is the common consumer's surplus. A consumer pays the amount S irrespective of

the number of units of the product he buys. The monopolist will extract all of the surplus (consumer's surplus plus producer's surplus) generated in this market if he sets a price which is equal to marginal cost. When price equals marginal cost, total surplus is maximized. Therefore, since the monopolist is maximizing the total surplus in this case, the optimal (profit maximizing) two-part tariff is to charge each consumer S for the right to buy and a per unit price C . As stated earlier this pricing method yields competitive output.

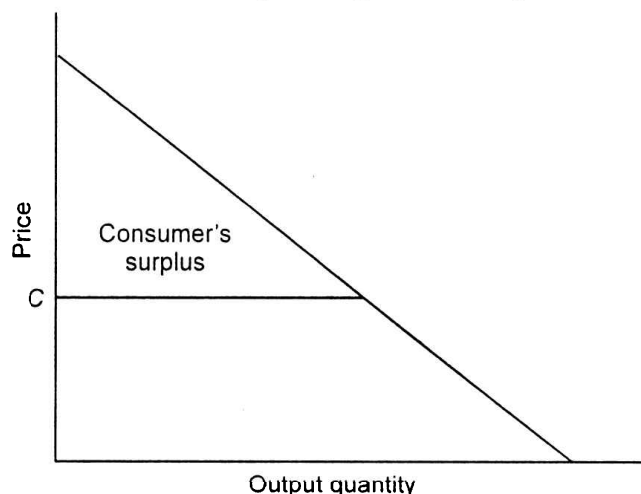


Fig. 11.7 Optimal Two-Part Tariff

If the consumers have non-identical downward sloping demand curves, then the monopolist will still charge each consumer C per unit consumed but a different lump sum fee to extract all the consumer's surplus. Therefore, the monopolist must have knowledge about each consumer's demand curve for designing a pricing policy that captures the consumer's surplus of each consumer. In some cases it may be possible to determine the value the consumers place on the product. For instance, a firm renting out copy machines may use a meter in the machine to determine the number of copies made by a consumer and then set rent depending upon the number of copies made. This pricing principle will maximize profits if those who make the most copies value the machine most.

Third-Degree Price Discrimination

Third-degree price discrimination is the most common form of price discrimination. It occurs when the monopolist sells output to different persons at different prices, but the price of output sold to a given person is the same for all units. It may arise in the following way. Suppose a firm does not have

sufficient information to identify each consumer and is unable to practice first-degree price discrimination. But suppose the firm can determine whether a particular customer belongs to one group or another. If it is possible to prevent resale between the two groups, and if the firm knows the aggregate demand curve of each group, then it becomes profitable to charge different prices to different groups. This is third-degree price discrimination. For instance, magazines charge a higher price to people who buy from newsstands than to those who subscribe. Many cinema halls offer discounts to students. Products are often packed with discount coupons that allow the bearer to purchase the product at a low price the next time.

How does the monopolist determine optimal prices for the two groups of customers? Let X_1 and X_2 be the quantities of the product sold to two groups (markets). Suppose that $p_1(X_1)$ and $p_2(X_2)$ are the prices of the good in the two markets, and that $C(X_1 + X_2)$ is the cost of producing the total output ($X_1 + X_2$). Then the monopolist's profits are given by

$$\pi = p_1(X_1)X_1 + p_2(X_2)X_2 - C(X_1 + X_2).$$

The first order conditions for profit maximization are

$$MR_1(X_1) = MC(X_1 + X_2), \quad (11.13)$$

$$MR_2(X_2) = MC(X_1 + X_2) \quad (11.14)$$

where $MR_i(X_i)$ is the marginal revenue from market i ($i = 1, 2$) and MC is the marginal cost. Note that since the monopolist produces in a single plant, the marginal cost of production is the same no matter to which market the output is sold. Eqs. (11.13) and (11.14) imply that the monopolist equates marginal revenue across the two markets.

$$\text{Now,} \quad MR_1(X_1) = p_1(X_1) \left(1 + \frac{1}{e_1} \right) \quad (11.15)$$

$$\text{and} \quad MR_2(X_2) = p_2(X_2) \left(1 + \frac{1}{e_2} \right) \quad (11.16)$$

where e_1 and e_2 are the price elasticities of demand in markets 1 and 2 respectively. Therefore, from Eqs. (11.13), (11.14), (11.15) and (11.16) we have

$$p_1(X_1) \left(1 + \frac{1}{e_1} \right) = p_2(X_2) \left(1 + \frac{1}{e_2} \right)$$

or
$$\frac{p_1(X_1)}{p_2(X_2)} = \frac{1 + \frac{1}{e_2}}{1 + \frac{1}{e_1}}. \quad (11.17)$$

So if $e_1 > e_2$ (market 2 has a more elastic demand), p_1 will exceed p_2 . That is, the market with lower elasticity of demand will have a higher price. This seems quite reasonable. An elastic demand is a price sensitive demand. A price discriminating firm will set a low price for the price sensitive group and a higher price for the price insensitive group.

Certain conditions are necessary for a monopolist to practise third-degree price discrimination. First, the monopolist must be able to keep the markets apart so that resale is ruled out. Market separation may be assured by high transportation costs. Second, the elasticities of demand at each price level must differ among markets. Third, the sources of price differentiation among the markets must be something other than variation in costs of producing and selling the product in different markets.

Two sources of inefficiency are present in third-degree price discrimination. The first one is the usual one associated with standard monopoly: price exceeds marginal cost. The second is a consumption inefficiency. Since different consumers pay different prices for a product, each consumer's marginal willingness to pay is not the same. Note that third-degree price discrimination can benefit the society compared to simple monopoly if it produces higher amount of output.

In Fig. 11.8 we present third-degree price discrimination diagrammatically. The monopolist's marginal cost of production is a constant C . D_1 and D_2 are the demand curves in markets 1 and 2. D_2 is more elastic than D_1 . Therefore a price discriminator charges a higher price in market 1.

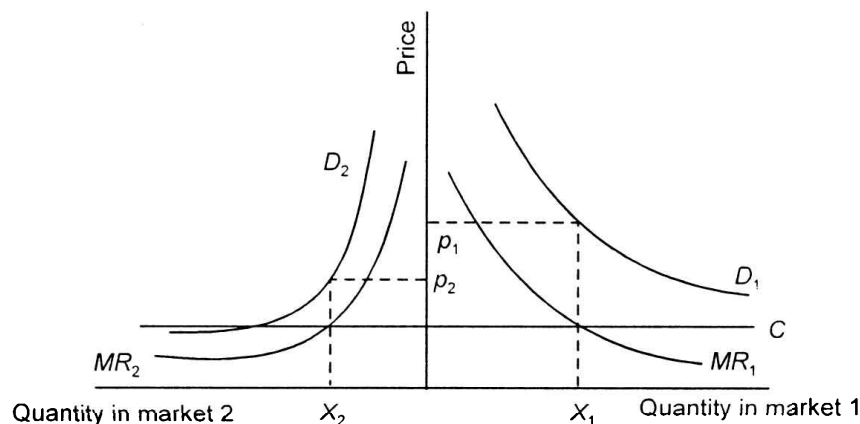


Fig. 11.8 Third-Degree Price Discrimination

INEFFICIENCY OF MONOPOLY

What impact would the introduction of monopoly into the purely competitive world have on consumer welfare? A competitive industry produces an output level where price equals marginal cost. On the other hand, a monopolist operates at a point where price exceeds the marginal cost. Assume that a product could be produced either by a monopolist or a competitive industry using the same cost function. There are reasons to believe that this assumption is unrealistic because monopolies often arise due to cost advantages resulting from the economies of scale. We make this crude assumption to look at the implications of monopoly power.

In Fig. 11.9 the monopoly maximizes profits by choosing the output price combination (X_m, p_m) , where marginal revenue equals marginal cost. Under competition, the market price and output are set at the point where the industry supply curve (the marginal cost curve) cuts the demand curve D , which means the competitive industry will operate at the point (X_c, p_c) . Monopoly produces less output ($X_m < X_c$) and increases price ($p_m > p_c$). Therefore, consumers are worse off in a monopolized industry than in a competitive industry.

But the monopoly firm enjoys higher profit than competition, since given the demand curve D , profits are maximized at (X_m, p_m) . Therefore, a monopoly firm's shareholders receive more income. Since monopoly restricts output, input hiring will be restricted as well. Therefore, there will be smaller earnings within the industry for the different inputs used in the production of the good. This means that there is a change in the distribution of income and consumers as owners of inputs receive smaller income now.

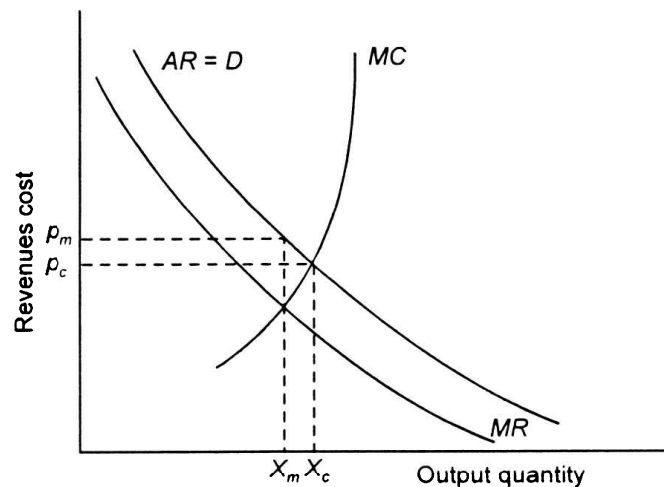


Fig. 11.9 Inefficiency of Monopoly

If we consider firms and consumers together, it is not quite clear at this stage whether competition is better than monopoly. We now show that it is possible to argue against monopoly on efficiency grounds. When we consider the inverse demand curve, at each level of output X , $p(X)$ is a measure of the consumer's willingness to pay for an additional unit of the good. For any output X lying between X_m and X_c , $p(X)$ is greater than the marginal cost $MC(X)$. Therefore, in the output interval (X_m, X_c) , people are willing to pay more for a unit of output than it costs to produce that unit.

Now, since $p(X_m) > MC(X_m)$, there must be someone who is ready to pay more for an additional unit of output than it costs to produce that additional unit. Suppose the firm produces an additional unit of output at the marginal cost $MC(X_m)$ and sells it to this person at a price p where $p(X_m) > p > MC(X_m)$. Then the consumer becomes better off because he was willing to pay $p(X_m)$ and he actually pays $p < p(X_m)$. Similarly, the monopolist spent $MC(X_m)$ to produce that additional unit but sold it at a price $p > MC(X_m)$. We have assumed that all the other units of output are being sold at the same price as before. So the monopolist moves to a better off situation. The consumers as a whole are also better off. Therefore, there is a Pareto improvement here. Hence monopoly control of output results in inefficient production.

MONOPOLY DEADWEIGHT LOSS

The analysis in the previous section establishes that monopoly is inefficient. We wish to measure here the welfare loss arising from the existence of monopoly. The concepts of the consumer and producer surpluses are used to measure the welfare loss of monopolization. When consumers pay the monopoly price instead of the competitive price, we look at the change in the consumer surplus. This change determines how much the consumers would have to be compensated for accepting the higher price. Similarly, when the firm charges a monopoly price instead of a competitive price, we consider change in the producer's surplus. This change—the change in the firm's profits—determines how much the firm would be willing to pay to get the monopoly price effective. The sum of these two changes gives us a measure of the cost of monopolization.

To determine this measure explicitly let us refer to Fig. 11.10. To keep things simple we consider a one-consumer world. The consumer demand curve is given by D . The industry's marginal cost curve is MC . Under competition, the product price is p_c , the amount of output produced is X_c , the

To interpret the deadweight loss measure in terms of the value of the lost output, we note that the demand curve represents the willingness of the consumer to pay for a good and the marginal cost curve shows the cost of producing an additional unit of the good. Since the demand curve lies above the marginal cost curve between output X_c and X_m , expansion of output from X_m to X_c produces a net benefit to the consuming person. At all outputs between X_m and X_c , the good is valued at more than the cost its production would impose on society. The deadweight loss represents the benefits from producing additional units of the good beyond X_m until X_c . The deadweight loss is a measure of the inefficiency that the monopoly creates.

WHY MONOPOLIES EXIST

The reason for the existence of monopolies is that other firms find it unprofitable or impossible to enter the market. Therefore, entry barriers are the sources of all monopoly power. If other firms are able to enter the market, then monopoly would be a short run phenomenon only. After all, when the monopolist sets price above the competitive level and restricts output to maximize profits, the resulting excess profits would create incentive for the entry of new firms into the industry, until price is forced down to the long-run competitive level, that is, until price equals minimum of long run average cost. If other firms can enter the market, there would, by definition, no longer be a monopoly. Monopoly power can persist only if there are natural or artificial entry barriers to the industry.

By a natural barrier into an industry we mean a barrier that arises because of the firm's technological characteristics. That is, a natural barrier is inherent in the nature of a firm's production function. That is why, natural barriers are also called technical barriers to entry. A primary natural barrier to entry is the presence of increasing returns to scale. The technology of production is such that relatively large scale firms are efficient. In such a situation one firm may find it profitable to drive others out of the industry by price cutting. Similarly, once a monopoly has been established, entry becomes difficult because any new firm must produce at relatively low output levels and therefore at relatively high costs. Increasing returns to scale will be treated further in the section which deals with natural monopoly.

The most important artificial barriers to entry are those supported or enforced by the state. That is, in such cases monopoly is created by law rather than by economic considerations. Because of this, these barriers are also called legal barriers. One important example of a government granted

monopoly position is the legal protection of a productive technique by a patent. Polaroid camera is a notable example of this. Since the basic technology for a product is uniquely assigned to one firm, a monopoly position is established. The argument for creating such a governmentally granted monopoly position is that the patent system makes innovation more profitable and therefore acts as an incentive to technical progress. Exclusive franchises to carry on particular lines of business are frequently granted by governments to protect the public interest. Examples are railroads, airlines, television and radio stations and so on.

Monopoly firms themselves can use artificial or self-help means of blocking entry. A common way is through ownership of control of the sources of raw materials from which a product is manufactured. This type of barrier becomes very effective when raw material sources are concentrated geographically; magnesium and nickel provide examples. Firms may develop unique products or technologies and take steps to keep these from being copied by competitors. Monopoly may also occur if firms in an industry may be able to collude and restrict output in order to raise prices and increase profits. Finally, a would-be monopolist may request government help in devising barriers to entry. He may lobby for legislation that will restrict entrants so as to maintain an "organized" market. Since the monopolist has special knowledge of his business and significant incentives to pursue these goals, he may be successful considerably in creating such entry barriers.

NATURAL MONOPOLY

In some industries it becomes most efficient for only one firm to produce all the output. When total production costs would rise if two or more firms produced instead of one, the situation is referred to as a natural monopoly. That is, a firm is called a natural monopoly if it can produce the entire industry output more efficiently (at lower cost) than two or more firms. Formally, let $x_1, x_2, \dots, x_n$ be the output of n firms such that $x_1 + x_2 + \dots + x_n = X$, the industry output. Let the cost of producing x_i be $C(x_i)$. Then we say we have a natural monopoly if

$$C(X) < C(x_1) + C(x_2) + \dots + C(x_n). \quad (11.18)$$

Thus, the least expensive (most efficient) way to produce is to have one firm produce all X units. A cost function is said to be subadditive at X if inequality Eq. (11.18) holds. So subadditivity is a necessary condition for the existence of a natural monopoly.

If production is characterized by economies of scale everywhere, then average cost decreases as output increases, and it is always less costly for one firm to produce any given output than to divide the output among several firms. That is, everywhere decreasing average costs imply subadditivity. Thus, when average cost falls with output, there is a natural monopoly.

A natural monopoly can occur even if average cost is not decreasing everywhere with output. For instance, a U-shaped average cost curve with the minimum average cost occurring at an output of 50 means that it may be most efficient (least expensive) for only one firm to produce if output is 51, even though average cost is rising at that output. Therefore, economies of scale is a sufficient but not a necessary condition for natural monopoly.

A natural barrier to entry exists when an industry is characterized by natural monopoly. A natural monopolist facing competitive factor markets can undercut any firm producing a smaller output level because per unit costs decrease as output increases. Thus, a natural monopoly firm can charge prices that generate excess profits without entry occurring. The monopolist will squeeze out any prospective entrant by lowering price in the short run to a level where the entrant loses, only to raise the price again after the entrant has become bankrupt.

The public utility and communications industries can be regarded as representative of natural monopolies. Concrete examples are electric power companies, natural gas pipelines and telephone companies. A local telephone company involves very large fixed costs for providing wires and switching networks, but the marginal costs of an extra unit of telephone service is quite low.

We can get a situation of the type shown in Fig. 11.11 when there are large fixed costs and marginal costs are low. The marginal revenue curve lies below the average cost curve or cuts it to the left of its minimum point, as illustrated in Fig. 11.11. The monopolist's market is limited. The monopolist will build a less-than-most-efficient size of plant and operate at a less-than-most-efficient rate of output. The market is not large enough to expand the plant sufficiently to take advantage of all economies of scale. The size of the plant will have some excess capacity.

The unregulated monopolist will produce X_1 units and charge price p_1 . Total sales volumes exceeding X_1 can be sold at prices below p_1 . At such prices, the horizontal distances from the vertical axis to the AC curve show the minimum quantities a newcomer must produce to break even. These

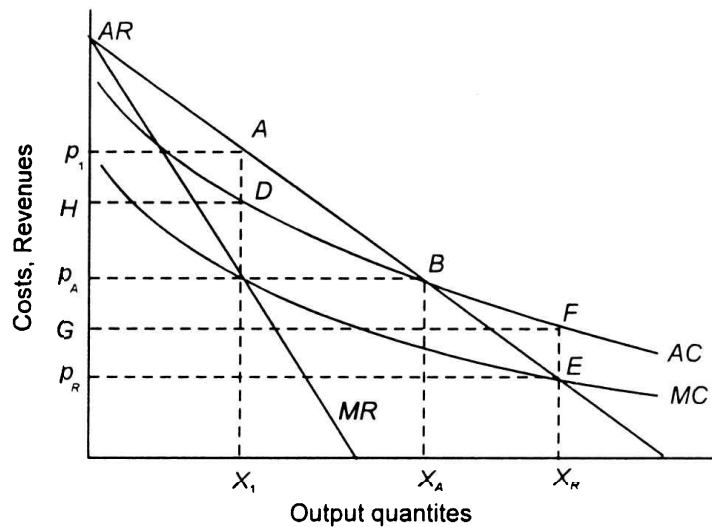


Fig. 11.11 Natural Monopoly

quantities, added with X_1 , may cause price to be less than AC for both firms, and entry by the new firm is forestalled. The original firm has the additional option of increasing output, forcing the price down, to preclude profit possibilities for the entrant.

REGULATION OF MONOPOLY

The regulation of monopolies is an important issue in applied economic analysis. Government regulation of monopolies may increase welfare. Public utilities, such as gas and electric power companies, communications and transportation industries are highly regulated in many countries. The economic problem involved is the determination of the rate that will induce the monopolist to furnish the maximum amount of product consistent with its costs and with consumer demand. Here we will examine a few aspects of monopoly regulation that relate to pricing policies.

Marginal Cost Pricing and Natural Monopoly

Many economists believe that price charged by regulated monopolies should reflect the marginal costs of production. This parallels the perfectly competitive case. In this way the welfare loss from monopolies is minimized. When price is greater than marginal cost, as in an unregulated monopoly firm, the economy ends up at a non-Pareto efficient state.

The major problem raised by an enforced policy of marginal cost pricing is that it may require natural monopolies to operate at a loss. Let us refer

to the natural monopoly case considered in Fig. 11.11. With increasing returns to scale over all outputs, average cost decreases as output increases, with marginal cost being lower than average cost at all outputs. As stated earlier, in the absence of regulation the monopolist will produce output level X_1 , and receive a price p_1 for his product. Profits in this situation are given by the rectangle p_1ADH . Pareto efficiency is violated at X_1 because $p_1 >$ marginal cost at X_1 .

In fact, Pareto efficiency requires that price be set at p_R . At this price X_R is demanded. Note that $p_R =$ marginal cost at X_R . Thus, marginal cost pricing has been achieved. Unfortunately, because of the declining nature of the firm's cost curves, the price p_R falls below average cost. With this regulated price the monopoly must operate at a loss of Gp_REF . Since a firm cannot operate indefinitely at a loss, this poses a dilemma for the regulatory agency: Either the goal of marginal cost pricing must be abandoned, or the government must subsidize the monopoly.

A basic problem with implementing the marginal cost pricing policy for a regulated firm is that marginal cost is difficult to measure, and of course the correct estimation of marginal cost can be made by the concerned firm only. But if the policy of regulation involves providing a subsidy equal to the firm's loss, then the firm has no incentive to keep cost (including marginal cost) under control. An alternative to the marginal cost pricing principle is the average cost pricing regulation, where price is set equal to average cost. Average cost pricing leads to the choice of the output-price combination (X_A, p_A) . Under average cost pricing the government does not have to subsidize the firm out of government revenue. But if the regulatory authority increases price as costs increase, then the firm does not have any incentive to keep costs under control. This situation forces the regulatory authority to determine allowable costs for the regulated firm. Allowable costs are ideally the minimal costs that a firm requires for producing its output.

Tow-Tier Pricing System

An alternative regulatory approach is the implementation of a two-tier pricing system. Under such a system the monopoly firm is allowed to charge some users a higher price while maintaining a lower price for marginal users. Therefore, in a sense the consumers paying the high price subsidize the losses of low-price consumers. We show this in Fig. 11.12. The regulatory commission has derived that some users will pay a relatively high price p_1 . At this price X_1 is demanded. Other users (presumably those who would not buy the

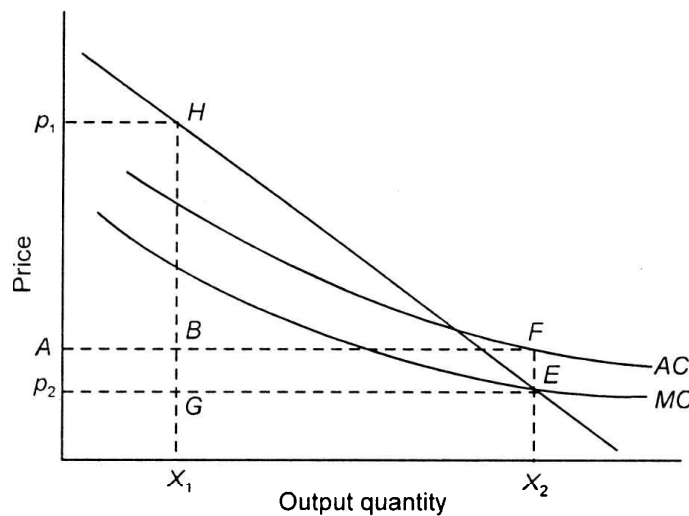


Fig. 11.12 Two-Tier Pricing

good at p_1) are offered a lower price p_2 . This lower price gives rise to an additional demand of $X_2 - X_1$. Consequently, a total of X_2 units of output are produced at an average cost of A . Because of this two-tier pricing system the profits from the sales to high-priced customers (given by the rectangle p_1ABH) will balance the losses incurred on the sales to the low-priced customers ($BGEF$). For the low-priced customers, who are marginal users, the marginal cost pricing principle is followed. The intramarginal users subsidize the firm so that it does not operate at a loss.

It is often argued that under the two-tier pricing system the intramarginal users pay both the fixed and the variable components of average costs, whereas marginal users pay only the variable marginal costs. In this approach also there are problems of determining marginal and average costs and of keeping costs under control. But it has the advantage of moving the setup to Pareto efficient one by fixing price equal to marginal cost. Still, most regulatory agencies use average cost pricing as a regulatory principle.

Rate of Return Regulation

Sometimes regulatory agencies implement average cost pricing by using the rate of return regulation for such natural monopolies as power and light companies so that the regulated firm earns a “fair” rate of return on capital. Much analytical effort has been devoted to defining the fair rate of return and to developing ways of measuring it. Although this regulation may help consumers, it does not encourage firms to behave efficiently. For instance, if the rate of return allowed to firm exceeds what owners might obtain under

competitive conditions, there would be a tendency to over-invest in capital. In general such a bias is called the Averch-Johnson effect. This effect results in inefficient production methods plus a further movement away from the Pareto optimal allocation at which price equals marginal costs.

BENEFITS OF MONOPOLY

There are two major arguments against monopolies. First, too few resources are devoted to the production of a good for which the supply is monopolized. Second, monopolization of a market creates substantial profits and these profits accrue to rich owners at the expense of relatively poor consumers. These arguments, particularly the first one, provide the basis for regarding monopolies as undesirable. This conclusion is not shared by all economists. Some contributors, most notably J.A. Schumpeter (1950) pointed out that monopoly profits can play a beneficial role in the process of economic development. Below, we mention some of the possible beneficial aspects of monopolies.

Monopoly profits provide funds that can be invested in research and development. Monopolies have surplus funds with which they can undertake the risky process of research. The prospect of receiving monopoly profits may motivate a firm to develop new products, improve products, or find cost saving production techniques. Without such a monopolistic position, firms might innovate less and not try hard to improve their efficiency.

Monopolization of a market may make it less costly for a firm to plan its activities. Since a monopoly firm is the only source of supply of a product, it does not have to spend much on selling expenses (such as advertising). A monopoly may be able to know more about the market demand and may more readily adapt to changing demand conditions.

MONOPOLISTIC COMPETITION

The demand curve for the product of a monopolist depends only on the price set by the monopolist. But there may exist close substitutes of the monopolist's product, and prices of these substitutes will certainly affect the demand for the monopolist's output. Each of these substitutes may be differentiated in some way from every other and also from the monopolist's product. Product differentiation can be brought about through trade names, advertising, packaging, and other marketing devices. Product differentiations lead some consumers to prefer the product of one seller over those of others. The theory of monopolistic competition, which has been advanced by Chamberlin (1956),

provides descriptions of those competitive industries in which product differentiation occurs—men's clothing and cotton textiles, for instance—in the sense that it recognizes small monopoly elements, and consequently different prices are charged by different sellers of a particular type of product. Another example of a monopolistically competitive market is provided by convenience stores. There are many small firms in this industry and the product of each is slightly differentiated from that of others, mainly because of locational advantages. Obviously there are limits on the ability of one store to take advantage of any beneficial location it may have, because others provide closely substitutable products.

We consider a variant of the theory of monopolistic competition under three basic assumptions:

1. The industry consists of a large number of small firms, each facing a downward sloping demand curve for its own differentiated product.
2. Each firm ignores the effect of its actions on the decisions taken by other firms.
3. New firms producing close substitutes for the products of the existing firms can enter the industry.

The structure of the industry is monopolistic because each firm faces a downward sloping demand curve for its own product. Consequently, it has some market power in the sense that it can set its own price, rather than accepting the market price like a competitive firm. On the other hand, the last two assumptions give a competitive flavor to the industry. This follows from the fact that the firms compete for customers in terms of price and the kinds of products, which are close substitutes, they sell. Also new firms may enter the industry. Thus, the industry shares elements of both competition and monopoly, and that is why it is referred to as monopolistic competition.

We assume that the industry consists of n producers who sell similar but not identical products. Therefore each of these n sellers is a monopolist so far as the demand for his own product is concerned. The price p_i that the consumers are willing to pay for the output of firm i will depend on the output of firm i and also on the levels of outputs of the remaining $(n - 1)$ firms. Let x_i denote the output of firm i . The industry output vector is $\underline{x} = (x_1, \dots, x_i, \dots, x_n)$. Since for each i , price p_i will depend on $\underline{x}$, we write it as $p_i(\underline{x})$ to indicate the dependence of p_i on outputs of all the firms in the industry.

The elasticity of a typical firm's demand curve depends on packaging, advertising and other devices that can increase the distinctiveness of the firm's product. The more distinctive the firm's product is (that is, the fewer the number of close substitutes), the more inelastic the demand curve is. The firm's demand curve will shift to the left or right depending upon the prices of other firms, their production characteristics, advertising expenditures, etc.

Suppose $C_i(x_i)$ stands for firm i 's cost of producing output x_i . Then each firm i wants to choose its output level x_i that will maximize its profits

$$\pi_i = p_i(\underline{x})x_i - C_i(x_i). \quad (11.19)$$

Since demand for firm i 's output depends on what other firms produce, firm i needs to forecast other firms' behavior.

In view of assumption (2), firm i assumes that other firms' behavior will be constant. Therefore, the optimizing (profit maximizing) condition for firm i is

$$MR_i = MC_i, \quad (11.20)$$

where MC_i is firm i 's marginal cost and MR_i is firm i 's marginal revenue. Evidently we can write MR_i as $p_i(1 + \frac{1}{e_i})$, where e_i is the elasticity of demand for firm i 's product.

In Fig. 11.13, we show that for each firm its marginal revenue equals marginal cost, given the actions of other firms.

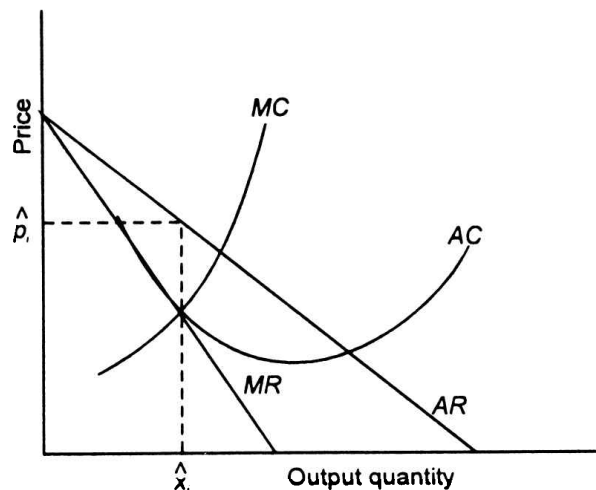


Fig. 11.13 Profit Maximization by a Firm in a Monopolistic Competitive Industry (short run)

The industry will be in equilibrium if each firm's forecast about the behavior of other firms is compatible with what other firms actually do. The vector $(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n)$ is an equilibrium output vector for the industry if $\hat{x}_1$ is the profit maximizing output level of firm 1 when it assumes that the remaining firms are going to produce $\hat{x}_2, \dots, \hat{x}_n$. Similar statements hold for other firms' output levels. Short run equilibrium in the industry does not imply that all firms charge the same price. Identical price is not expected, since the firms do not produce the homogeneous products. Each firm looks for its own profit maximizing position and equates its marginal cost with its own marginal revenue. Yet the prices charged by different firms will not be far apart. We expect prices to be clustered but not necessarily equal. Although each firm has the liberty of setting its own price, it faces competition from many close substitutes for its product.

In the short run, a monopolistically competitive firm enjoys some market power (since marginal revenue exceeds marginal cost). The typical firm in the industry earns short run excess profits. This encourages entry into the industry by firms producing close substitutes for the products of the existing firms. The increase in the number of firms—expansion of the industry output—acts to shift the demand curve of a typical firm to the left and to make it more elastic since a wider range of close substitutes for the product of the firm becomes available. It is natural to assume that entry (hence left ward shifting of the demand curves of the firms) will continue as long as there are excess profits in the industry. Long run equilibrium in a monopolistically competitive industry can be defined as it is under perfect competition viz. a situation in which there is no incentive for entering or exiting the industry with the typical firm earning a zero profit.

For earning zero profit firm i must charge a price $\hat{p}_i$ and produce an output $\hat{x}_i$, such that

$$\hat{p}_i \hat{x}_i - C_i(\hat{x}_i) = 0,$$

$$\text{or} \quad \hat{p}_i = \frac{C_i(\hat{x}_i)}{\hat{x}_i} \quad (11.21)$$

Eq. (11.21) means that price must be equal to the average cost for each firm in the long run equilibrium. Further, each firm operates at the point of maximal profit on its demand curve, that is, there can be no point on its demand curve that will generate positive profit. (Positive profit generating

points are those which are above the average cost curve, that is, where price becomes greater than average cost.) Consequently, a zero profit condition means that the typical firm's demand curve must be tangent to its average cost curve. Since the demand curve facing the firm has a negative slope, each firm i will produce at an output level $\hat{x}_i$ lying to the left of the output level $\bar{x}_i$ where the minimum of its average costs occurs. The monopolistic competitive equilibrium therefore exhibits excess capacity.

The monopolistic competition model is an intermediate case between monopoly and competition. It does not differ much from pure competition with respect to its welfare effects. It does, however, have some minor restrictive impacts. If a purely competitive market in long run equilibrium were to become monopolistically competitive welfare would be reduced by a slight reduction of output and a slight increase of price. The demand curve faced by the monopolistic competitor, though very elastic, is less than perfectly elastic. Its marginal revenue is less than price, and the output falls short of that at which marginal cost equals product price. The more elastic the demand curve faced by the firm, the less will be the deviation from purely competitive price and output.

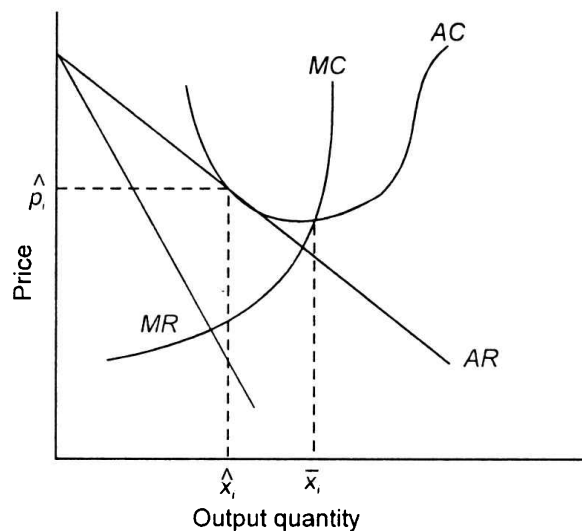


Fig. 11.14 Monopolistic Competitive Equilibrium (Long run)

As we have seen a necessary condition for Pareto optimality is that products be priced at marginal cost; hence monopolistic competition leads to inefficiency in the economy. The cause of welfare loss is that market power is being exercised by each firm, though the market power is limited in the long run. Just as in the pure monopoly, consumers can be made better off by

increasing the industry output (which reduces per unit costs), because firms in the industry operate with excess capacity. The fact that equilibrium is at a non-Pareto efficient state implies that the consumers can more than compensate the losers when industry output is increased so that a Pareto superior allocation is obtained.

Another source of inefficiency should also be mentioned. A firm has incentives for shifting its demand curve to the right: its revenue increases at any output. One way to achieve this is to make its product distinctive through advertisement. However, a portion of advertising expenditures represents defence advertising that is required to keep a firm's demand curve in the present position. We can think of an equilibrium situation in which substantial amounts are spent by firms on advertisement, with no net change in the distribution of sales among firms. Therefore all advertising then would be defensive. Clearly, this is inefficient. In fact empirical works suggest that a substantial portion of advertising is defensive.

HOTELLING'S MODEL OF LOCATION AND PRODUCT DIFFERENTIATION

H. Hotelling (1929) developed a model to explain the location and pricing behavior of firms. Although he concentrated on geographic space, the model can be employed to study monopolistic competition. In the Hotelling model products differ with respect to the location of the stores that sell them.

Consider a long, narrow city that has only one street which is L kilometres long. Consumers are uniformly distributed along this street, so that in any block there is an equal number of consumers. All consumers are identical except for location, and each consumer buys one unit of the product in each time period.

Two stores sell identical products in this city. Store 1 is located k_1 kilometres from one end of the city, and store 2 is located k_2 kilometres from the other end of the city. Each consumer faces a transportation cost of c per kilometre. Each consumer buys from the least expensive shop, taking transportation cost into account.

Consider consumer j who lives at the location shown in Fig. 11.15. Let $d_1(d_2)$ be the distance between his residence and store 1(2). Since $d_1 > d_2$, he will go to store 2. (His transportation costs for going to store 2 are lower.) If someone lives exactly halfway between the shops, he will be indifferent as to which shop to patronize.

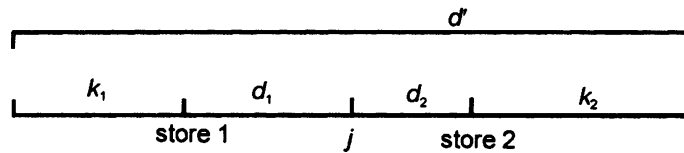


Fig. 11.15 Hotelling's Location Model

Suppose that the government sets the price of the product, so that the shops are required to choose their locations. How will shop 1 choose its location given that shop 2 is k_2 kilometres from the right end of the city? Since customers only care about how much they travel, shop 1 would like to be the nearest shop for the maximum possible number of customers. Shop 1 will maximize its profits by locating itself just to the left of shop 2, d' kilometres from the left end of the city. There it gets all the customers to its left, which is the majority.

Now, If firm 2 can costlessly relocate its position after firm 1 has located its position, then it would move slightly to the left of firm 1's new location. This process will continue until both firms are in the middle of the city, where each firm has half the customers. This seems quite reasonable. Quite often we see the propensity of firms to locate near each other in a variety of markets.

Example 11.1

Suppose that a monopolist sells his output in two distinct markets between which price discrimination is possible. His cost function and two demand curves are given by $C = 8X + 100$, $X_1 = 10 - 0.5p_1$, $X_2 = 40 - p_2$, where $X_1 + X_2 = X$. We wish to determine the profit maximizing values for p_1 , p_2 , X_1 and X_2 .

The monopolist's profit maximizing condition in market 1 is

$$MR_1(X_1) = MC(X_1 + X_2).$$

Now, revenue from market 1 is $R_1(X_1) = p_1(X_1)X_1 = [(10 - X_1)/0.5]X_1 = 20X_1 - 2X_1^2$. Therefore $MR_1(X_1) = d[R_1(X_1)]/dX_1 = 20 - 4X_1$. Next, note that $C(X) = C(X_1 + X_2)$. Therefore $[\delta C(X_1 + X_2)]/\delta X_1 = 8$. Thus, $MR_1(X_1) = MC(X_1 + X_2)$ gives $20 - 4X_1 = 8$, from which we have $X_1 = 3$ and hence $p_1 = 14$.

Proceeding in a similar manner one can see that the profit maximizing condition in market 2 becomes $40 - 2X_2 = 8$ from which it follows that $X_2 = 16$ and hence $p_2 = 24$.

The elasticities of the demand curves at the profit maximizing output price combinations are $e_1 = -0.5(14/3) = -7/3$ and $e_2 = -1(24/16) = -3/2$. The lower elasticity is in market 2 where price is higher.

Applied Example 11.1 Natural Monopoly in India

The Indian postal system provides services as a regulated government monopoly. The rationale behind this is that the system is so complicated and so vast that more than one supplier would be uneconomic. Duplication of the system would be enormously expensive; a monopoly seems to be the best solution. Thus, this is an example of a natural monopoly in India. Other examples of natural monopoly in India are electricity distribution systems, the Indian Railways and the Oil and Natural Gas Commission.

The costing and pricing policy carried out by a Government monopoly does not follow that of a private enterprise monopoly. For instance, consumers purchase cooking gas at a highly subsidized price. One argument for justifying this kind of pricing is that the value of the service includes not just the market value, but also consumers' surplus and other benefits that flow to users and non-users. Communications, electricity and natural gas are public utilities and therefore require non-commercial pricings.

Appendix. 11.1

MONOPOLY PROFIT MAXIMIZATION

The monopolist's revenue is $Xp(X)$. Then his profit maximization problem is

$$\max_X [Xp(X) - C(X)].$$

The first order condition for this problem is

$$\frac{d}{dX} [Xp(X)] - \frac{d}{dX} [C(X)] = 0.$$

Now, $\frac{d}{dX} [Xp(X)] = p(X) + X \frac{dp(X)}{dX}$, which is the monopolist's marginal

revenue. We rewrite $p(X) + Xdp(X)/dX$ as $p(X) \left(1 + \frac{1}{e}\right)$, where e is the elasticity of demand. Therefore, the first order condition for the monopolist's profit maximization is

$$MR(X) = MC(X),$$

where $MC(X) = \frac{d}{dX}[C(X)]$ is the marginal cost function.

The second order condition for a maximum profit is

$$\frac{d^2}{dX^2}[Xp(X)] - \frac{d^2}{dX^2}[C(X)] < 0,$$

or,
$$\frac{d}{dX}[MR(X)] - \frac{d}{dX}[MC(X)] < 0.$$

This implies that the slope of the marginal cost curve must exceed the slope of the marginal revenue curve.

MONOPOLISTICALLY COMPETITIVE EQUILIBRIUM

Next, we define a monopolistically competitive equilibrium. There are n firms. Firm i 's profits are

$$\pi_i(\underline{x}) = p_i(\underline{x})x_i - C_i(x_i),$$

where x_i is the output of firm i and $\underline{x} = (x_1, \dots, x_n)$. Now each firm maximizes its profits assuming that outputs of other firms remain constant. Thus, firm i will choose its output level $\hat{x}_i$ so as to satisfy

$$\frac{\delta p_i(x_1, \dots, x_{i-1}, \hat{x}_i, \hat{x}_{i+1}, \dots, x_n)}{\delta x_i} + p_i(x_1, \dots, x_{i-1}, \hat{x}_i, x_{i+1}, \dots, x_n) - \frac{dC_i(\hat{x}_i)}{dx_i} = 0.$$

For an output vector to be a monopolistically competitive equilibrium output vector, each firm's forecast about the behavior of other firms must be compatible with what other firms actually do. Thus, a monopolistically competitive equilibrium output vector $(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n)$ must satisfy

$$\frac{\delta p_i(\hat{x}_{i-1}, \dots, \hat{x}_{i-1}, \hat{x}_i, \dots, \hat{x}_n)}{\delta x_i} \hat{x}_i + p_i(\hat{x}_1, \dots, \hat{x}_{i-1}, \hat{x}_i, \dots, \hat{x}_n) - \frac{dC_i(\hat{x}_i)}{dx_i} = 0, \quad i = 1, 2, \dots, n.$$

EXERCISES**Multiple Choice**

1. When a monopolist maximizes profit in the
 - a) Short run, he will do so in the long run also.
 - b) Long run, he will do so in the short run also.
 - c) In the long run, he may not be able to do so in the short run.
 - d) Any of the above is possible.
2. If a monopolist incurs loss in the short run, then in the long run he will
 - a) Go out of business.
 - b) Stay in business.
 - c) Produce at the point where price = average cost.
 - d) Any of the above is possible.
3. The profit maximizing output level for a monopolist occurs at the point where
 - a) Total cost is minimized.
 - b) Total revenue is maximized.
 - c) Total revenue = total cost.
 - d) The total cost and total revenue curves are parallel.
4. Suppose that a monopolist has to follow the marginal cost regulation. Then he will
 - a) Incur loss.
 - b) Break even.
 - c) Make a profit.
 - d) Any of the above is possible.
5. If a per unit tax is imposed on the monopoly output, then
 - a) Only the marginal cost curve will shift upward.
 - b) Only the average cost curve will shift upward.
 - c) The average cost and the marginal cost curves will shift upward.
 - d) None of the above.
6. Which kind of monopoly regulation is most advantageous for consumers?
 - a) Price regulation.
 - b) Lump sum tax.

- c) Per unit tax.
 - d) All of the above are equally effective.
7. If the market demand curve is given by $X = 50 - 2p(X)$, then
- a) Marginal revenue will be negative for quantities at which demand is elastic.
 - b) Total revenue is maximized when price = 25.
 - c) Marginal revenue is positive for any price above 15.
 - d) The market price is less than 12.
8. A monopolist is able to discriminate prices between two markets. Then he will
- a) Charge prices such that the price elasticities of demand in two markets are the same.
 - b) Charge a higher price in the market with more elastic demand.
 - c) Sell a higher quantity in the market where the demand curve is more elastic.
 - d) None of the above is true.
9. If the elasticity of demand is -3 at a price of 30, then the marginal revenue at the corresponding output is
- a) 20.
 - b) 10.
 - c) 5.
 - d) 30.
10. Under conditions of monopolistic competition, in long run equilibrium it is necessary that
- a) Price be greater than long run average cost.
 - b) Some barrier must exist to prevent the entry of more firms.
 - c) Resources get wasted on advertising.
 - d) None of the above.

Quantitative Problems and Questions

1. A monopolist can produce at constant average and marginal cost of 5. The firm faces a demand curve given by $X = 53 - p(X)$.
- a) Calculate the profit maximizing output-price combination for the monopolist. Also calculate the monopolist's profits.
 - b) What output level would be produced by this industry under perfect competition?
 - c) Calculate the producer's surplus under competition and monopoly.

- d) Calculate the consumer's surplus in case (b). Show that this exceeds the sum of the monopolist's profits and the consumer's surplus received in case (a). What is the area of the welfare triangle, that is, the value of deadweight loss from monopolization?
 - e) Determine the fraction that monopoly welfare loss is of the value of output of the competitive industry.
2. The total cost function for a monopolist's output X is $C = 6X + 0.05X^2$. The demand function is $X = 360 - 20p(X)$.
 - a) Calculate the price, quantity produced and the monopolist's profits.
 - b) The government wishes to impose a maximum price at a level that will induce the monopolist to produce the maximum possible output. What is the price, how much output will be produced, and what is the profit of the monopolist?
 - c) What are the break-even output and price levels, that is, the output and price levels if the monopoly is regulated under the average cost regulation? What are the consequences of imposing a maximum price at that level?
 3. A monopolist's cost function for his output X is given by $C = 0.1X^3 - 6X^2 + 140X + 3,000$. The monopolist produces 40 units of output and his profit is 1,000 units.
 - a) Calculate the elasticity of the demand curve at the profit maximizing output level.
 - b) Derive the equation for the demand curve assuming that it is linear.
 - c) Suppose a tax is imposed on the good; as a consequence the monopolist's output reduces to 39 units. How much is the tax per unit?
 4. Why are barriers to entry necessary for a monopoly to be maintained? Give some examples of different kinds of barriers to entry.
 5. Show why the monopolist will not produce on the inelastic region of the demand curve.
 6. What is natural monopoly? Show how natural monopoly stems from entry barriers.
 7. Suppose that a monopolist produces output at a constant marginal cost of 5. The monopolist sells his output in two different markets separated by some distance. The demand curves in the two markets are given by $X_1 = 55 - p_1$ and $X_2 = 70 - 2p_2$. Calculate the (a) total output, (b) quantity sold in each market, (c) price in each market, (d) monopolist's total profits.

8. Suppose that the demand function for a monopolist's product is $X = a - bp$. Show how the slope of the marginal revenue curve is related to that of the average revenue curve.
9. A monopolist's cost function for its output and demand curve are given by $C = 6.9X + 0.003X^2$, $p = 12.65 - 0.022X$.
 - a) Calculate the output, price and the monopolist's profits.
 - b) Suppose an ad valorem sales tax of 10% of the price is imposed. Alternatively, imposition of per unit excise duty or a licence fee is possible, so that the same tax revenue is yielded as the ad valorem tax.
 - i) Determine the output, price, and profit (net of tax) for each of these options.
 - ii) Compare the impact of these taxes on the monopolist and the consumers. Which tax would you recommend and why?
10. Use the concept of consumer's surplus to demonstrate that the imposition of a specific tax in a perfectly competitive economy generates a deadweight loss similar to that shown for the case of a monopoly.
11. A monopolist sells his output in two separate markets. His total cost function is $C = 0.01X^3 - 9X^2 + 330X + 3,700$. The demand functions that the monopolist faces in the two markets are $p_1 = 300 - 1.5X_1$, $X_2 = ap_2^{-2}$. The monopolist earns the maximum profit by charging a price of 240 in market 1. What is his total profit.
12. Suppose that a perfectly competitive industry can produce a good at a constant marginal cost of 10. Under monopolization marginal cost increases to 12 because a fee of 2 is to be paid for rent seeking. Suppose that the market demand curve for the good is $X = 1,000 - 50p$.
 - a) Determine the monopoly and competitive outputs and prices.
 - b) Determine the loss of consumer's surplus under monopolization
13. Consider a monopolist who produces his output at a constant per unit cost. The monopolist has to follow a two-tier pricing system under regulation. Is this efficient?
14. A monopolist produces his product from a single plant using the cost function $C = X^3/3 - 30X^2 + 1,000X$. He sells the product in two markets between which price discrimination is possible. The demand function in the first market is $p_1 = 1,100 - 13X_1$. His total output is 48, and the elasticity of the second market demand curve at the profit maximizing price is -3 . How much is his profit?

15. A monopolist's cost function is $C = 0.05X^3 - 6X^2 + 335X$. He faces the linear demand function $p = 330 - 2.2X$. The monopolist maximizes his profits. The government now imposes a sales tax on the good. The tax is imposed on the customer and is 10% of the price. The government also imposes a maximum price (net of tax) of 180. The monopolist adjusts his profit maximizing position.
 - a) What is the government revenue from the tax?
 - b) By how much are the consumers better off in terms of change in the consumer's surplus?
 - c) What is the change in the monopolist's profit?
16. A monopolist produces his product using the cost function $C = X^2 + 10X$. He sells his output in two markets and the demand functions in these markets are $X_1 = 32 - 0.4p_1$, $X_2 = 18 - 0.1p_2$.
 - a) The monopolist is able to price-discriminate between **the two markets**. Determine his price-quantity combination in each **market**. What is his total profit?
 - b) Suppose price discrimination is prohibited and the monopolist charges the same price in both markets. Determine his price, output and profit.
17. Suppose that a monopolist produces output at a constant marginal cost c and faces the linear demand curve $X = a - bp$ for his output. Also suppose that as a first option the monopolist can earn the consumers' surplus associated with the price-output combination when he goes into business as well as his profit. What do you think of this option relative to the one in which he can keep only the producer's surplus?
18. A monopolist produces his output using the cost function $C = 0.000005X^3 - 0.0075X^2 + 10X + 60,000$. He sells the product in three markets among which price discrimination is possible. The demand curve in the first market is $p_1 = 100 - 0.025X_1$. In market 1 the monopolist sells four times as much as he sells in market 2, in which his price is 80. In market 3 he sells 500 units, at which quantity the elasticity of the market demand curve is -5 . What is his profit?
19. Under what circumstances might consumers actually be better off with monopoly than with a competitive structure?
20. Show diagrammatically the total cost and total revenue curves of a monopolist. Draw the total profit curve on the same diagram, showing zero profit on the cost and revenue curves.

Mark the point of maximum profit. What can you say about the slopes of the total revenue and total cost curves at this point.

21. A monopolist can be thought of as choosing the price to determine how much quantity to sell. Show that you get the same profit maximizing condition as in the case when the monopolist chooses quantity to determine price.
22. Determine in each of the following cases whether or not third-degree price discrimination exists. Explain your answer.
- The same cinema hall charges all persons less for matinee shows than for evening shows.
 - Electric power companies charge different rates to the household sector and to the commercial sector.
 - Some medicines, such as antibiotics, are sold by veterinarians at lower prices for animal consumption than they are by pharmacies for human use.
 - A cinema hall with an exclusive run on a recent release charges less to old persons and to students with identity cards than to general adults.
23. A monopolist produces a product from two plants (A and B) and sells the product in two markets (1 and 2). The average cost levels are such that it is always profitable to produce. The two cost functions and the market demand functions are as follows:

$$C_A = \frac{2X_A^2}{3} - 12X_A, \quad C_B = 2X_B^2 + 36X_B,$$

$$X_1 = 300 - p_1, \quad X_2 = 180 - p_2.$$

Consider each of the following cases and answer the questions.

- The monopolist can charge only one price, that is, the same price is charged in both markets.
 - How much does the monopolist produce from the plants A and B?
 - How much does he sell in the two markets?
- The monopolist follows a third-degree price discrimination.
 - How much does he produce from plants A and B?
 - How much does he sell in the two markets?

- iii) What are his prices in markets 1 and 2?
 - iv) Does he earn a higher profit now than under case (a)? Demonstrate your claim.
- c) The monopolist follows first-degree price discrimination.
- i) How much does he produce from plants A and B?
 - ii) How much does he sell in the two markets?
 - iii) Compare his profit with that under case (a).
24. Consider the following production function $X = Y_1^{1/2} Y_2^{1/2} Y_3^{1/3}/9$, where Y_i is the quantity of input i , $i = 1, 2, 3$. The price of the three inputs are 1, 9 and 18. The market demand curve for the product is $p = 216X^{-1/2}$.
- a) Is there a perfectly competitive equilibrium for this industry? If not, why?
 - b) Suppose that the industry is taken over by a monopolist. What is the monopolist's profit?
 - c) Determine the price and output if the industry is regulated under the (i) average cost regulation, (ii) marginal cost regulation.
25. A monopolist produces his output in a single plant using the cost function $C = \frac{X^3}{3} - 40X^2 + 1,800X + 5,000$. He sells his product in two markets between which price discrimination is possible. The demand functions in the two markets are $X_1 = 320 - 0.4p_1$, $p_2 = a - bX_2$. The monopolist's total output and profit are 60 and 5,000. Determine the values of a and b .
26. A single firm is the only domestic producer of a good. Its monopoly power is protected by entry barriers and a tariff that prevents competition from foreign firms. Tariffs imposed by other nations do not make exports profitable. The monopolist's cost function and the domestic market demand function are given by $C = 0.5X^2 + 10X$, $X = 100 - p$.
- a) Calculate the firm's output, price and profit.
 - b) The government now revokes the tariff. A foreign product, identical in all respects, can be imported for selling in the domestic market at a price of 55. How much will the domestic producer produce? What is his profit? What quantity will be imported?
 - c) The government imposes a maximum price of 50 on all domestic sales of the product. Determine the firm's output and profit. Is there a shortage/surplus in the market?

27. A single firm produces a good in country A . A different firm produces the good in country B . Trade between the two countries is not allowed. The cost and demand function associated with the good are: $C_A = 0.1X_A^2 + 40X_A$, $C_B = 0.05X_B^2 + 50X_B$, $X_A = 3,200 - 20p_A$, $X_B = 900 - 5p_B$, where the subscripts identify the countries.

- a) Determine the price and output in each country.
- b) Suppose that the two firms are allowed to merge. The combined firm practices price discrimination. Calculate the price and output in each country.
- c) Free trade develops between the two countries and transport costs are negligible. Competitive shippers can buy in the low-price country and sell in the high-price country. The merged firm adjusts to this so as to maximize its profit. Determine its profit.

28. A monopoly firm in country A produces a good using the cost function $C = 0.025X^2 + 45X + 400,000$. The good is sold in countries A and B , where the latter is the only export market. The demand curves in the two countries are $p_A = 744 - 0.3X_A$, $p_B = 400 - 0.3X_B$. The exchange ratio between the currencies in the two countries A and B is 1:2.

Country A imposes a sales tax of 20% of the retail price of the product. There is no sales tax in country B , but there is flat import duty of 15 units of money per unit of output. Transport costs to country B are 50 units of money per unit of output.

The producer is a profit maximizer. Determine the (i) price in country A , before sales tax (ii) volume of sales in country A , (iii) price in country B , (iv) volume of sales in country B , (v) tax receipts from sales tax in country A , (vi) tax receipts from import duty in country B , (vii) revenue of the transport company, (viii) producer's profits.

29. A profit maximizing monopolist faces a linear demand curve that cuts the price axis at 150 and the quantity axis at 75. The cost function is $C = X^2$. Suppose that a maximum price of 80 is imposed on the monopolist. Determine the amount by which his profit falls.

30. Prove or disprove the following:

- a) A monopolist will not advertise his product if it makes his demand curve more elastic.
- b) The market demand curve for a good is -3 . Suppose that the market is one of monopoly and the monopolist's average cost curve is a horizontal straight line. Then the monopolist's profit will be one-third of his total cost.

- c) Suppose that a monopolist produces from two separate plants the cost functions of which are:

$$C_1 = a_1X_1^2 + b_1X_1 + c_1;$$

$$C_2 = a_2X_2^2 + b_2X_2 + c_2$$

where $a_i, b_i, c_i, i = 1, 2$, are all positive. Then if his profit maximization requires some output from each plant, the plant with greater quantity will have lower average cost.

- d) If a monopolist faces a linear demand curve, then the imposition of an ad valorem tax collected from the consumer will raise the price by the amount of the tax.
- e) Suppose that a monopolist's cost function is given by $C = 0.125X^2 + 30X$. The monopolist faces the demand function $p = -0.1X + 48$. Then the monopolist's profit maximizing output, if he is regulated by a maximum price of 42, will be greater than that if he is regulated by a maximum price of 43.
- f) When a monopolist charges different prices in different markets, the price difference must reflect a difference in the costs of supplying the commodity to the markets.
- g) A monopoly will never provide a Pareto efficient level of output.
- h) A monopolist faces a market demand curve that has constant elasticity of -2 . The government imposes a quantity tax of 4 per unit of output. Then his output price rises by 6.
- i) Decreasing average costs everywhere imply decreasing marginal costs everywhere.
- j) Under conditions of monopolistic competition, equilibrium requires that for each firm in the short run, marginal revenue = marginal cost, whether or not average revenue = average cost, whereas in the long run average revenue = average cost whether or not marginal revenue = marginal cost.

31. A monopolist's demand function is given by

$$X = 0 \text{ if } p > 10$$

$$= 50/p \text{ if } p \leq 10$$

- a) What is the profit maximizing output if his marginal cost is 1?
- b) What output will the monopolist produce under a marginal cost price regulation, when marginal cost is $c \neq 1$?

32. A monopolist produces an output using a cost function of the firm $C = ax$, $a > 0$. He sells outputs in two markets where the demand functions are $X_1 = a_1 - b_1p_1$, $X_2 = a_2 - b_2p_2$. Determine the conditions on (a_1, b_1, a_2, b_2) when the monopolist chooses not to act as a price discriminator.
33. A monopolist produces a good using the cost function $C = ax$, $a > 0$. There is an ad valorem tax on the price of the good so that when the consumer pays p , the monopolist receives $p_1 = (1 - \alpha)p$, $0 < \alpha < 1$. As an alternative to the ad valorem tax, consider a quantity tax t on output such that $p' = p + t$. The elasticity of the market demand curve is constant. Determine the value of t such that the final pricing will be the same under the two tax schemes.
34. Suppose that a monopolist faces a demand curve whose constant elasticity is -3 . What level of subsidy should the government provide for enabling him to produce the socially optimal output level?
35. Which, if any, of the following pricing practices constitute first-second or third-degree price discrimination?
- A country sells a commodity at a lower price abroad than in its home market.
 - A telephone company charges Rs. 2.50 per call for the first 100 calls and Rs. 2.00 per call for the next 100 calls and so on.
 - A bookseller offers 10% discount to a local college's students which he does not offer to others.
 - A hostel canteen charges students only 5% more for a weekly 21-meal plan than it charges for a weekly 15-meal plan.
 - An amusement park owner charges a different price for each ride.
36. Suppose that a quantity tax and a value tax affect the output produced by a monopolist in the same way. Show that the value tax will raise more revenue than the quantity tax.
37. Suppose that the market demand curve for a monopoly product is linear. The monopolist's marginal cost of production is a constant. A quantity tax is imposed on the output.
- How do the marginal cost, price and output change?
 - How does the Lerner index of market power change?
38. Which of the following cost functions are sub-additive?
- $C(x) = a + bx$, where a and b are positive constants.

- b) $C(x) = a$, where $a > 0$ is a constant.
 - c) $C(x) = ax^2$, where $a > 0$ is a constant.
 - d) $C(x) = ax$, where $a > 0$ is a constant.
 - e) $C(x) = ax^{1/2}$, where $a > 0$ is a constant.
39. Suppose that a monopolist's cost function is $C(X) = 2X$. The market demand function for the monopolist's product is $X = 10 - p$. Assume that all the four consumers are identical.
- a) Determine the optimal two-part tariff.
 - b) Determine the monopoly output when the monopolist is a perfect price discriminator.
40. The long run cost function of a monopolistically competitive firm is given by $C = 0.001x^3 - 0.425x^2 + 85x$, where x is the firm's output. The demand curve for the firm's product is given by $x = 300 - 2.5p$.
- a) Calculate the firm's long run price and output.
 - b) Calculate the elasticity of the firm's demand curve at the long run equilibrium position.
41. A number of firms produces a good. The long run cost function of each firm is given by $C(x) = 10^{-7} x^3 - 0.001x^2 + 6x$, where x is the output in kilograms. The output market is such that if all firms charge the same price, each sells the same quantity. Suppose that the demand curve that a firm faces is $x = (300,000 - 500,000p)/n$, where p = price of output and n = number of firms.
- a) Assume that a competitive equilibrium exists in the industry. Determine the equilibrium price and the number of firms.
 - b) The firms now discover that printing a producer's name on the packets leads consumers to acquire brand preferences. There is no cost for printing names. Each firm's linear demand curve is downward sloping where the slope is such that reduction of price by 0.01 unit of money increases sales by 50 kilograms. Long run equilibrium is achieved.
 - i) Determine the equilibrium price.
 - ii) How many firm's are there in equilibrium?
 - iii) Are the consumer's better or worse off in (b) than in (a)?
 - c) Consumers now realize that the product inside the packets is really homogeneous and they no longer demand brand preferences, but still find printing of names attractive. Now, there is also a printing cost of 0.4 unit of money per kilogram of output. Determine the equilibrium price and how many firms are there in equilibrium?

42. Monopolistic competition exists within an industry. The long-run cost function for a representative firm is given by $C(x) = 0.0025x^3 - 0.5x^2 + 384x$. The demand curve for the firm's product is given by $p = A(n) - 0.1x$, where $A(n)$ is a function of the number of firms n in the industry. The industry achieves long run equilibrium.

- a) What quantity does a typical firm produce? What is its price?
- b) What is the value of $A(n)$?

43. Firms in an industry produce a homogeneous good under conditions of perfect competition. Each firm has the following short run cost function:

$$SC = 0.002x^3 - 0.423x^2 + 67.32x + 820.8.$$

The market demand curve for the product is given by $p = 81 - 0.012X$, where X is the industry output.

The firms now discover that printing the producer's name on the packets leads consumers to acquire brand preferences. There is no printing cost. Total sales by the industry remain the same at any common price and all firms continue to enjoy equal share of the market if there is one price. Each firm also observes that by varying its own price from the average price charged by other firms it can change its volume of sales. The elasticity of that response for a typical firm is -5 .

Compare the short run competitive equilibrium with the short run monopolistically competitive equilibrium (in terms of price, output and profits).

44. A monopolist's cost function is given by $C(X) = \alpha + \beta X$, where α and β are positive. The monopolist faces the market demand function $p(X) = a - bX$, where a and b are positive. The government imposes a producer's tax of Rs. t per unit on the production and as a result the monopolist's price increases from p and p_1 . Calculate the difference $p_1 - p$ in terms of t and interpret your result.
45. The cost function of a monopolist is given by $C = X^3 + 3X^2$. The monopolist faces the following market demand function: $p = 12 - 5X$. Determine the change in price after imposition of a tax of Re. 1 per unit produced. Find the corresponding profit. What will be the profit if a sales tax of 10% is imposed?
46. Show graphically that a monopolist's price is an increasing function of his marginal cost.

12

Oligopoly

The purpose of this chapter is the discussion of pricing policy in oligopolistic industries. An oligopolistic industry is one in which the number of firms is small enough for the activities of any single firm to affect other firms and for the activities of other firms to affect that firm in turn. A single firm occupies a position of sufficient importance in the industry since changes in its own activities have repercussions on other firms. The changes in the output and the price charged by one firm will affect the amounts that the other firms can produce and the prices that they can charge. Other firms will react to price-output changes on the part of a single firm. An individual firm recognizes this interdependence, and when changing price, output or sales promotion activities must take into account the reactions of others. In such a market the activities of one seller are important for others.

Oligopolistic industries are frequently classified as : (i) homogeneous, and (ii) differentiated. A homogeneous oligopoly is one in which firms make an identical product. Consumers have no reason for preferring the output of one firm to that of another on any basis except price. Clearly, such an industry falls between the polar extremes of a perfectly competitive industry and a monopoly industry. Examples are steel, aluminum, cement and glass industries.

Many ordinary consumer products (toothpaste, soft drinks, automobiles) are produced by industries that can be described as differentiated oligopolies. A differentiated oligopolistic industry is one in which firms produce and sell differentiated products. The products of all firms in the industry are good substitutes for one another, but the product of each firm possesses its own distinguishing characteristics. The differences between products may be real or fancied. They may consist of differences in quality and design, as in the case of the automobile industry, or may be simply brand name differences e.g. toothpastes.

Since in many cases a firm may not be sure what rivals' reactions will be to its own activities and since oligopoly deals with a wide range of cases, there is no single model that can be used to explain all possible forms of oligopolistic industries. Consequently, in this chapter we will examine a few basic elements that are common to many of the models which are in current use. Several selected models are examined with this objective in mind. We begin our analysis with a discussion of pricing in homogeneous good oligopolies.

Next, we present a model of differentiated oligopoly that has been used to explain price rigidity. The final objective of this chapter is to take a brief look into some other types of research on industrial organizations which attempt to provide details on those aspects of market behavior that are not captured by formal oligopoly models.

PRICING IN HOMOGENEOUS GOOD OLIGOPOLY

In this section we will examine the theory of price determination in homogeneous good oligopoly markets. We assume that all the oligopolistic firms buy the factors competitively. The demanders of the oligopoly product are all price takers. Throughout the section we will assume that there are a fixed number of n firms (where n is taken to be relatively small). If there are only two firms in the industry, that is, if $n = 2$, then the situation is called one of duopoly.

Most of the issues to be discussed in this section center around how firms make respective profit maximizing choices. Several possibilities will be examined here. For instance, one firm may be able to set its price before the other firms. The price setting firm here is called the price leader and the other firms are price followers. Similarly, one firm may get to set its quantity first in which case it becomes a quantity leader and others are quantity followers. A third possible situation is that any firm i treats all other firms'

outputs as fixed in its decision. Analogously, any firm may treat all the remaining firms' prices as fixed in determining its own price. Another possibility is that firms may collude instead of competing against each other. In this case the firms jointly decide to set quantities that will maximize the sum of their profits.

The following general notation will be adopted throughout the section. We will denote firm i 's output by x_i ($i = 1, 2, \dots, n$). Total industry output is $X = x_1 + x_2 + \dots + x_n$. The cost of producing x_i by firm i is $C_i(x_i)$.

Quantity Leadership Model

Certain industries consist of a dominant firm or a leader—one with a large share of the market, and many fringe or small firms, each with a small share of the market. In such a case the dominant firm may set its output choice before the other firms, the followers. This is the Stackelberg model of oligopoly, after the German economist who developed the theory in the 1930s. Stackelberg's model is probably one of the first which have introduced asymmetry in oligopoly. (Here asymmetry arises in the following way. Although the dominant firm is not a monopolist in the sense used in Chapter 11, still it is assumed to dominate the industry in that it sets its quantity choice first. The followers' choices depend on the output choice of the leader. They simply view the leader's output as a constant. Since the followers do not set their quantities simultaneously with the leader, asymmetry arises.)

Industries in which one firm has a large share of the market are quite common. For instance, Kodak is often regarded as a dominant firm in the photographic film business. Another example can be the dominance of International Business Machines in the computer industry.

There are several reasons that are sufficient to guarantee that some firms in an industry may become dominant. The first reason is that the dominant firms may have lower costs than the fringe firms. Two major causes for lower costs are: (i) A dominant firm may be more efficient than the followers, it may have better technology or better management; (ii) a dominant firm as an early entrant to the industry may have lower costs from having learnt by experience; learning by doing. A second important factor is that a group of firms may jointly act as a dominant firm for increasing their profits. A third cause is that superiority of a firm's product may make the firm dominant.

To discuss the Stackelberg model analytically, let us suppose that there are only two firms, 1 and 2. The price of output is written as $p(x_1 + x_2)$ to

indicate that the price depends on the total industry output $X = (x_1 + x_2)$. Suppose firm 1 is the leader and chooses to produce a quantity x_1 . Firm 2 will choose a quantity x_2 assuming that x_1 is given. Since the dominant firm or the leader knows that, given the leader's output level, the follower will maximize its profit; in order to make a sensible decision about its own output, the leader has to consider the follower's maximization problem.

The follower wants to maximize its profits

$$\pi_2(x_1, x_2) = p(x_1 + x_2)x_2 - C_2(x_2) \quad (12.1)$$

with respect to x_2 assuming that x_1 is given. Evidently the first order condition for profit maximization is

$$MR_2(x_1, x_2) = p(x_1 + x_2) + \frac{\Delta p}{\Delta x_2}x_2 = MC_2(x_2), \quad (12.2)$$

where $MR_2(x_1, x_2)$ is firm 2's marginal revenue at x_2 assuming that firm 1's output x_1 is given and $MC_2(x_2)$ is firm 2's marginal cost at x_2 .

From Eq. (12.2) we can solve x_2 as a function of x_1 . This function will determine the profit maximizing output level of firm 2 as a function of the dominant firm's output. We write this relationship as

$$x_2 = r_2(x_1). \quad (12.3)$$

The function r_2 is firm 2's reaction function. The reaction function, which is assumed to have a negative slope, represents firm 2's reactions given alternative output choices of firm 1, the dominant firm.

Since the dominant firm knows that its output choice will influence the output choice of the follower, while making its output decision it should recognize this influence. Consequently, the dominant firm will maximize its profit

$$\pi(x_1, x_2) = p(x_1 + x_2)x_1 - C_1(x_1) \quad (12.4)$$

with respect to x_1 subject to the constraint $x_2 = r_2(x_1)$. If we substitute the constraint $x_2 = r_2(x_1)$ into Eq. (12.4), firm 1's profit becomes a function of x_1 only. Denoting $\pi[x_1, r_2(x_1)]$ by $\bar{\pi}(x_1)$, we can say that firm 1's objective is to maximize its profit

$$\bar{\pi}(x_1) = p[x_1 + r_2(x_1)]x_1 - C_1(x_1) \quad (12.5)$$

with respect to x_1 .

Suppose that the value of x_1 that solves Eq. (12.5), that is, the profit maximizing output of firm 1 is $\hat{x}_1$. The follower's profit maximizing output level

$\hat{x}_2$ is obtained by substituting $\hat{x}_1$ into the reaction function: $\hat{x}_2 = r_2(\hat{x}_1)$. Then we say that the output vector $(\hat{x}_1, \hat{x}_2)$ constitutes a Stackelberg equilibrium.

To see the Stackelberg equilibrium graphically, let us consider the isoprofit curve of firm 1. This curve is the locus of all output levels of the two firms that generate a constant profit for firm 1. Firm 1's profit increases as we move to lower isoprofit curves since firm 1's profit will increase as firm 2's output decreases. Given that the demand curve is negatively sloped, this becomes clear from Eq. (12.4). We know that firm 2 being a follower will select an output level along its reaction curve $r_2(x_1)$. Therefore, firm 1 will choose an output combination on the reaction curve that generates the highest possible profit for it. Since higher profits for firm 1 are associated with lower isoprofit curves, firm 1 will choose that output combination on the reaction curve where the reaction curve is tangent to the lowest possible isoprofit curve. That is, the Stackelberg equilibrium is determined by the point of tangency between firm 1's lowest possible isoprofit curve and firm 2's reaction function. This is shown in Fig. 12.1.

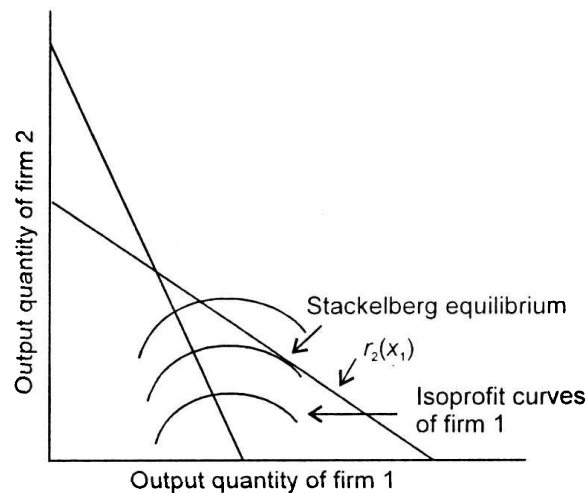


Fig. 12.1 Stackelberg Equilibrium

Price Leadership Model

As stated earlier, instead of setting quantity the dominant firm may set price. The fringe firms follow the price leadership of the dominant firm and take the price as given. For instance, in the world petroleum market, the major countries are leaders and the dependents are followers. As long as the leading firm lets the fringe firms sell as much as they want, they have no incentive to charge a lower price. If they charge a higher price than the dominant firm they will be unable to sell their products.

Several crucial assumptions underline this simple no-entry model. These are:

1. There is one firm which is much larger than any other firm.
2. All firms except the dominant firm are price takers.
3. The dominant firm knows the industry demand curve.
4. The dominant firm can forecast how much output the competitive fringe will produce at any given price, that is, it knows the competitive fringe's supply schedule.

The dominant firm's problem is much more complex than that of simple monopoly. It must consider how the competitive fringe responds to its actions. For instance, if the dominant firm lowers its output, price rises. The competitive fringe then increases its output because it is moving up its supply curve. Therefore, the industry output falls less than the dominant firm would like. Thus to maximize profits the dominant firm must take into account the actions of the fringe firms.

Suppose that the dominant firm has set a price p . For ease of exposition let us assume again that the number of fringe firms is one, that is, there are two firms in the industry: one is dominant and one is a follower. Assume that firm 1 is dominant and firm 2 is a follower. The follower takes the price p as given and maximizes its profits by choosing an appropriate output level.

Formally, the follower maximizes its profits

$$\pi_2(x_2) = px_2 - C_2(x_2) \quad (12.6)$$

with respect to x_2 . Since p is outside the control of the follower, its profit maximizing output will be the one where price equals marginal cost. That is, the profit maximizing output for firm 2 will fulfil the condition

$$p = MC_2(x_2) \quad (12.7)$$

where MC_2 is firm 2's marginal cost function. This determines the supply curve $S(p)$ of firm 2, the fringe firm.

The dominant firm now observes that if it sets a price p , the fringe firm's supply is $S(p)$. Consequently, the dominant firm's residual demand is $R(p) = D(p) - S(p)$, where $D(p)$ is the market demand as a function of price. Thus, the dominant firm's residual demand curve is the horizontal difference between the market demand and the fringe firm's supply.

The dominant firm behaves as a monopolist would with respect to the residual demand. It sets its price (or output) so that its marginal cost equals marginal revenue. (The marginal revenue of the dominant firm is the marginal revenue associated with the residual demand curve.) Formally, the dominant firm maximizes its profit.

$$\pi_1(x_1) = x_1 h_1(x_1) - C_1(x_1) \quad (12.8)$$

by choosing an output-price combination, where $h(x_1)$ is the inverse residual demand function. The inverse residual demand function, is obtained as follows. Given that $R(p) = D(p) - S(p) = x_1 + x_2 - x_2 = x_1$, we can solve p as a function of x_1 and write it as $p = h(x_1)$. Since we obtain p here by inverting the residual demand function $R(p)$ and since p depends only on the residual output x_1 , we call $h(x_1)$ the inverse residual demand function.

To illustrate the price leadership model graphically, let us suppose that the dominant firm has a constant marginal cost c . In Fig. 12.2 we depict the aggregate demand curve and the supply curve of firm 1. The residual demand curve, which is obtained by subtracting the supply curve of firm 1 from the aggregate demand curve is given by TT_1 . The dominant firm treats this as its demand curve and selects the profit maximizing output-price combination. Given that c is its marginal cost, p is the price, OT_3 is the quantity supplied by the dominant firm, and T_3T_2 is the supply of the fringe firm.

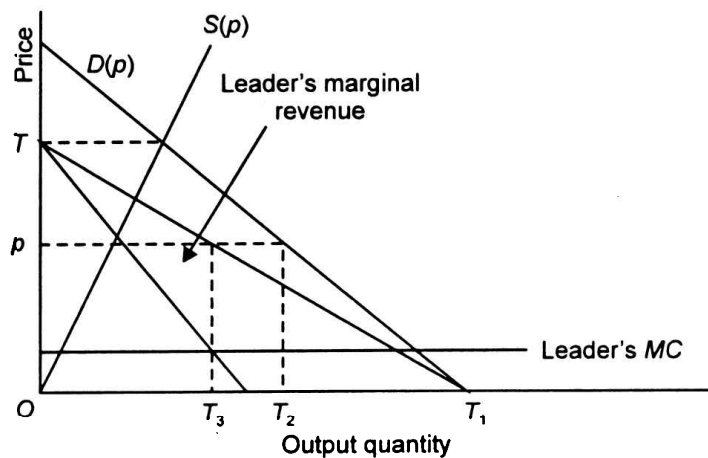


Fig. 12.2 Price Leadership Oligopoly

Cartel Model

Firms in an industry have an incentive to coordinate their production and pricing activities for increasing individual and joint profits by raising the

market price and restricting the output levels. If firms form an association that coordinates the activities of its members, then the association is called a cartel. A cartel that includes all firms in an industry is a monopoly and the members of the cartel share the monopoly profits.

Certain major factors are necessary to establish a cartel. First, firms in an industry will join a cartel only if the cartel is expected to raise the price and keep it high. The more inelastic the demand curve that the cartel faces, the higher the price the cartel can set and the greater are its profits. If the cartel's demand curve is inelastic, price rise will increase revenues and profits significantly. Next, in many countries cartels are illegal. Therefore, in such countries cartels form only if members do not expect the government to catch and severely punish them. Large expected penalties reduce the expected value of forming a cartel. The third factor is low organizational cost. Several factors can keep organizational cost low, facilitating the formation of a cartel. These are: (i) few firms are involved, (ii) a product of uniform quality is produced, and (iii) the industry is highly concentrated, that is, a few large firms make most of the sales in an industry. For instance, Spain and Italy, which control 80% of the world production of mercury, formed a successful cartel that did not involve the other producers.

We now consider a profit maximization problem for an n -firm cartel, that is, for the n firms as a group where the firms recognize that by collusion they can affect price and coordinate their decisions so as to achieve monopoly profits. In this case the cartel acts as a multiplant monopolist and chooses outputs $x_1, x_2, \dots, x_n$ so as to maximize total industry profits. That is, here

$$\begin{aligned}\pi(x_1, \dots, x_n) &= p(X) (x_1 + x_2 + \dots + x_n) - C_1(x_1) - \\ &\quad C_2(x_2) - \dots - C_n(x_n) \\ &= p(x_1 + x_2 + \dots + x_n) (x_1 + x_2 + \dots + x_n) - \\ &\quad C_1(x_1) - \dots - C_2(x_2) - C_n(x_n)\end{aligned}\quad (12.9)$$

is to be maximized with respect to $x_1, x_2, \dots, x_n$. The optimality conditions are,

$$p(x_1 + x_2 + \dots + x_n) + \frac{\Delta p}{\Delta x_i} [x_1 + \dots + x_n] = MC_i(x_i)$$

$$\text{or} \quad MR_i(X) = MC_i(x_i) \quad (12.10)$$

where $MR_i(X)$ is the industry marginal revenue when output of firm i increases by 1 unit, $i = 1, 2, \dots, n$.

Now, total revenue depends on the sum of all cartel members' output levels and the marginal revenue is the same no matter whose output level is changed. Thus, $MR_1(X) = MR_2(X) = \dots = MR_n(X) = MR(X)$, say. We therefore rewrite the optimality condition of Eq. (12.10) as

$$MR(X) = MC_i(x_i) \quad (12.11)$$

for $i = 1, 2, \dots, n$. That is, at the profit maximizing point the industrial marginal revenue will be equated with each firm's marginal cost. Clearly, at the optimal output vector, the marginal costs of production are equated across firms.

An important problem with the cartel or collusive solution is that there is always a temptation to cheat. Since each cartel member i produces an output level for which $p > MC_i(x_i)$, it has some market power. (The Lerner market power index for each firm is positive.) Therefore each firm has an incentive to expand output secretly. To see this more rigorously, let us rearrange the optimality condition for firm i .

$$p(x_1 + x_2 + \dots + x_n) + \frac{\Delta p}{\Delta x_i} x_i - MC_i(x_i) = - \frac{\Delta p}{\Delta x_i} (x_1 + x_2 + \dots + x_{i-1} + x_{i+1} + \dots + x_n) > 0. \quad (12.12)$$

The left hand side of Eq. (12.12) is the increase in firm i 's profit if it increases its output by one unit. Since the market demand curve has a negative slope $\Delta p/\Delta x_i$ is negative. Therefore the inequality on the right hand side of Eq. (12.12) follows. Thus, if firm i believes that other firms will keep their outputs constant, then it will believe that it can increase profits by increasing its production. At the optimal output levels it becomes profitable for each firm to unilaterally increase its output if it expects that the other firms will keep their outputs constant. It is this kind of instability that mainly leads to disruption of cartels. One of the difficulties of the Organization of Petroleum Exporting Countries (OPEC) cartel in dictating output levels to its members during the late 1970s attest to this problem.

In order to maintain an effective cartel, the firms need a way to detect cheating and enforce agreement. The factors that aid in the detection of cheating are: (i) there are few firms in the industry, (ii) prices do not fluctuate independently, (iii) prices are well-known, and (iv) all cartel members

sell identical products. To prevent cheating, successful cartels must agree on more than just a price. Cartels may succeed in preventing cheating by assigning each firm certain buyers or geographic locations, which will help easy detection of cheating. The two-country mercury cartel used a geographic division of markets: Spain supplied the United States and Italy supplied Europe. Another effective technique can be fixing market shares. As long as the shares are easily observable, no firm has an incentive to cut its price.

Cournot's Model

One of the first and probably still one of the most widely cited models of oligopoly was developed by the French mathematician Augustin Cournot in 1838. Cournot assumed that each firm acts independently and chooses output for maximizing its profits. While choosing its output level each firm assumes that other firms continue to produce the same levels of output no matter how it behaves. That is, each firm assumes that other firms sell their existing quantities of output. In other words, each firm believes that a change in its output by θ unit will change the industry output by θ unit only. Such a situation can arise when a firm has been convinced by its competitors that, no matter what it does, they will not change their outputs. An example of this case may be one in which sellers are bound to their buyers by long-term contracts for fixed amounts of supply.

Now, the profit function of firm i is

$$\pi_i(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n) = p(x_1 + x_2 + \dots + x_i + \dots + x_n)x_i - C_i(x_i) \quad (12.13)$$

where $i = 1, 2, \dots, n$. The Cournot assumption requires that each firm i maximizes π_i with respect to its output x_i assuming that outputs of all other firms do not change. The optimality conditions turn out to be

$$p(x_1 + x_2 + \dots + x_n) + x_i \frac{\Delta p}{\Delta x_i} = MC_i(x_i) \quad (12.14)$$

for $i = 1, 2, \dots, n$. Notice from this equation that firm i assumes that changes in x_i affect its total revenue through its direct effect on the market price of its own sales. Hence the equation differs from the cartel solution where the effect of a change in price on the industry revenue is taken into account.

We formally say that the firms in an industry are in a Cournot equilibrium, more precisely, an output vector $(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n)$ constitutes a Cournot

equilibrium if $(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n)$ is a solution of the n equations in Eq. (12.14). Thus, Cournot equilibrium must satisfy the n first order conditions in Eq. (12.14). The first order condition for firm i determines its optimal choice of output as a function of its beliefs about other firms' output choices. This functional relationship is known as firm i 's reaction function. It shows how firm i will react given its alternative beliefs about the choices of the other firms. A Cournot equilibrium is simply a point of intersection of the reaction functions.

To illustrate this graphically, let us assume that there are only two firms, that is, $n = 2$. Then by solving firm 1's first order condition for profit maximization for x_1 in terms of x_2 we get the reaction function of firm 1 and write it as $x_1 = r_1(x_2)$. Similarly, we can define $x_2 = r_2(x_1)$, firm 2's reaction function. The functions r_1 and r_2 are assumed to be negatively sloped. Then the output choices $(\bar{x}_1, \bar{x}_2)$ constitute a Cournot equilibrium if they satisfy

$$\begin{aligned}\bar{x}_1 &= r_1(\bar{x}_2), \\ \bar{x}_2 &= r_2(\bar{x}_1).\end{aligned}\tag{12.15}$$

In a Cournot equilibrium each firm maximizes its profits, given its beliefs about the other firm's choices, and these beliefs are confirmed in equilibrium. We show this in Fig. 12.3.

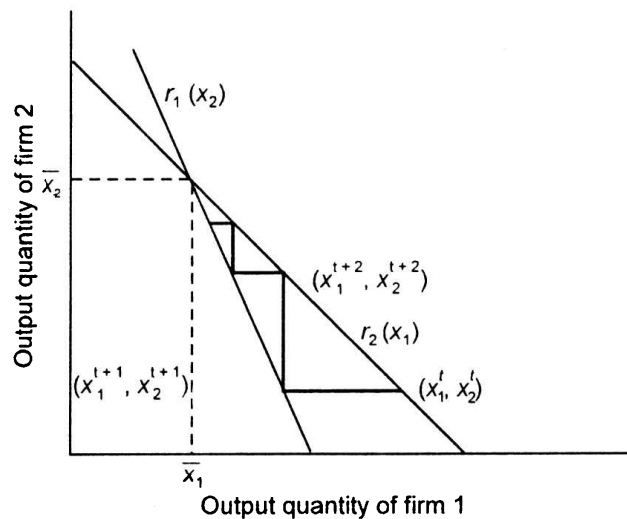


Fig. 12.3 Cournot Equilibrium

Stability of Cournot Equilibrium

The Cournot equilibrium is stable in the sense that no firm has an incentive to change its output if the industry is in equilibrium. But will the firms have a tendency to adjust their outputs if they are not in equilibrium initially?

Cournot's model has been used to see how firms adjust from non-equilibrium to equilibrium. When firms practising Cournot behavior adjust their outputs over time, their outputs gradually approach the Cournot equilibrium, and at the equilibrium no firm has an incentive to alter its output level.

To illustrate this, let us assume that $n = 2$. The two firms vary output at discrete time intervals (say, once a month). Each firm knows how much its rival has produced in the previous period. Since the firms follow Cournot behavior, each one assumes that its rival will choose the same output level in period $(t + 1)$ that it chose in period t .

Suppose at time t firms are producing (x_1^t, x_2^t) . Firm 1's choice in period $(t + 1)$ will be determined by

$$x_1^{t+1} = r_1(x_2^t). \quad (12.16)$$

Similarly, firm 2's choice will be

$$x_2^{t+1} = r_2(x_1^t). \quad (12.17)$$

These two equations describe how each firm will adjust its output when it faces the other firm's choice. The two firms alternately adjust their outputs until they reach the equilibrium point, where neither will have a tendency to adjust further. This adjustment path over time is shown as the thick line in Fig. 12.3. It forms a cobweb pattern.

The dynamic framework illustrated above helps in explaining Cournot's idea. But the Cournot model is not a dynamic model. It should be viewed only as representing a static equilibrium.

Market Power in Cournot Model

Recall that each firm i following Cournot behavior believes that if it changes its output by Δx_i unit then the industry output will change by Δx_i unit only: all other firms' outputs will remain constant. Now, each firm i maximizes its profit under this belief. Therefore we can replace Δx_i by ΔX , the change in industry output when firm i increases its output by Δx_i , in the profit

maximizing condition of firm i . That is, we can rewrite the optimality condition of Eq. (12.14) as

$$p(X) + \frac{\Delta p}{\Delta X} x_i = MC_i(x_i), \quad (12.18)$$

from which it follows that

$$p(X) \left[1 + \frac{X \Delta p}{\Delta X p(X)} \frac{x_i}{X} \right] = MC_i(x_i), \quad (12.19)$$

Let $s_i = x_i/X$ be firm i 's share of total market output. Then Eq. (12.19) becomes

$$p(X) \left[1 + \frac{s_i}{e} \right] = MC_i(x_i) \quad (12.20)$$

where e is the elasticity of demand. The Lerner index of monopoly power L_i for firm i now becomes

$$L_i = \frac{p(X) - MC_i(x_i)}{p(X)} = - \frac{s_i}{e}. \quad (12.21)$$

Thus, for a Cournot firm the Lerner measure of market power, the price mark-up over marginal cost as a fraction of price, depends on two factors: (i) The output share of the firm which gives us an idea about the size of the market, that is, the number of customers, and (ii) the price elasticity of demand, which shows how fast customers may leave the market because of price increase. The larger the measure L_i , the greater is the market power.

If $s_i = 1$, then firm i is a monopolist—it faces the entire market demand curve. As shown earlier, in this case the Lerner index becomes $-1/e$. If all the n firms in the industry have the same cost function (symmetric oligopoly), then the Lerner measure L_i in Eq. (12.21) becomes $-1/ne$. Notice that, in this symmetric case holding elasticity constant, as the number of firms increases, Lerner's measure must fall. As n approaches infinity, the Lerner measure approaches zero, and the market is competitive (price coincides with marginal cost). In this case the firm is a very small part of a large market, its market share is effectively zero. Thus, symmetric Cournot equilibrium with a large number of firms is effectively the same as pure competition.

Let us now consider a weighted average L_a of individual Lerner indices L_i in Eq. (12.21) to arrive at a measure of market power at the industry level,

where the weights are the output shares of the respective firms. The weighted average is given by

$$L_a = - \frac{\sum_{i=1}^n s_i^2}{e} = - \frac{H}{e} \quad (12.22)$$

where $H = \sum_{i=1}^n s_i^2$ is the Herfindahl–Hirschman measure of industrial concentration. The measure H indicates the extent to which output shares are controlled by large firms. A high value of H means that the industry is highly concentrated. The measure H increases if the sales of a large firm increase at the expense of the sales of a small firm. This may arise if some of the customers who were buying from the small firm now buy from the large firm.

Another property of H is that if two or more firms in the industry merge and produce their per-merger combined output, then H increases. These two properties show that whenever large firms are becoming larger, that is, the degree of dominance of large firms increases, H increases. Hence H is an indicator of concentration—the extent to which economic activity is dominated by giant firms. As expected, H takes on the maximal value of unity if the industry is monopolized, that is, $s_i = 1$ for some i , $1 \leq i \leq n$, and $s_j = 0$ for all $j \neq i$. In other words, H achieves its maximum attainable value when only one firm in the industry produces positive output and all other firms' output is zero.

If all the n firms in the industry enjoy equal output share $1/n$, $H = \frac{-1}{ne}$. This is the minimum value of H . In this case, H increases as the number of firms decreases.

Another example of a concentration measure is the k -firm concentration ratio, the output share of the k largest firms in the industry, where k can be 1, 2, 3, 4, etc. (See Jacquemin, 1987 and Chakravarty, 1996 for further discussions.)

Bertrand's Model

In 1883, i.e. 45 years after the publication of Cournot's book, Joseph Bertrand argued that in oligopolistic markets firms act as price setters rather than as quantity setters. In the Bertrand model each firm believes that it can charge whatever price it desires, while the other firms continue to charge the same price as before. Whereas Cournot firms believe that the rivals will keep their quantities constant, Bertrand firms believe that they will keep their prices

constant. Now, prices instead of outputs are the instruments controlled by firms.

To illustrate the Bertrand model let us consider a two firm homogeneous good industry where each firm produces the good at a constant average cost c . Given that consumers are perfectly informed of offered prices, they all will buy from firm 1 if its price p_1 is smaller of the two prices p_1 and p_2 . But if $p_1 > p_2$ then firm 1 sells nothing and all consumers buy from firm 2. Consumers are indifferent between the two firms if $p_1 = p_2$. In this case we can assume without loss of generality that firm 1 gets half of the consumers. Let $D(p)$ be the market demand function. Then firm 1's output demand function is

$$\begin{aligned} d_1(p_1, p_2) &= D(p_1) \text{ if } p_1 < p_2 \\ &= \frac{1}{2}D(p_1) \text{ if } p_1 = p_2 \\ &= 0 \text{ if } p_1 > p_2. \end{aligned} \quad (12.23)$$

The profit function of firm 1 is

$$\begin{aligned} \pi_1(p_1, p_2) &= D(p_1)(p_1 - c), \text{ if } p_1 < p_2 \\ &= \frac{1}{2}D(p_1)(p_1 - c), \text{ if } p_1 = p_2 \\ &= 0 \text{ if } p_1 > p_2. \end{aligned} \quad (12.24)$$

We can derive firm 2's output demand function and profit function analogously.

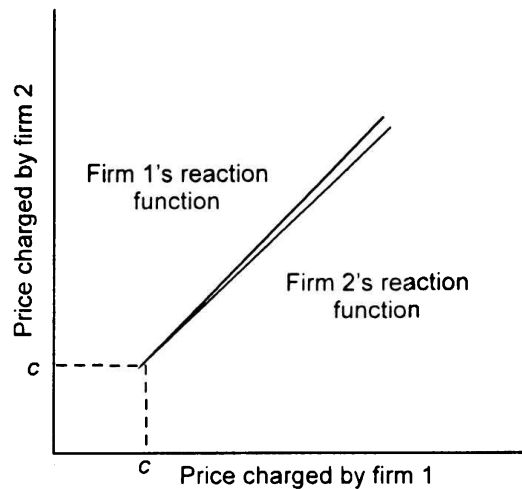


Fig. 12.4 Bertrand Equilibrium

When both firms charge c , neither firm profits by deviating. If a firm lowers its price, it earns negative profits (because price is below marginal and average cost). If either firm raises its price, it makes no sales at all. Thus $(p_1, p_2) = (c, c)$ is the only Bertrand equilibrium. In this equilibrium, both firms charge a price equal to marginal cost and earn zero profit, and thus, it is the same as the social optimum (competitive equilibrium). In the Bertrand equilibrium no firm enjoys market power.

We can explain the Bertrand equilibrium using reaction functions. Each reaction function shows that each firm will choose a price slightly below its rival's. The reaction functions are upward sloping. The only point of intersection of these two reaction functions is the point where price equals marginal cost, which is the only equilibrium.

COMPARISON OF SOLUTIONS

Since the cartel or collusive model is a multiplant monopoly model, it results in the smallest industry output and the highest price. On the other hand, Bertrand equilibrium, which is socially optimal, generates the highest output and lowest price. The other models give results that are in-between these two extremes depending on the specific assumptions made about firms' strategic interrelationships.

In general, a price leadership model will generate a price-quantity combination which is different from that generated by a quantity leadership model. Many economists find Bertrand's model more attractive than Cournot's model because it explains how prices are determined. In the absence of an auctioneer, it is difficult to imagine how prices are determined when firms set outputs instead of price, as in the Cournot framework. Since price is the decision variable, a Bertrand firm's residual demand curve differs substantially from that of a Cournot firm.

Alternatives and variations of the models discussed in the section are possible. For instance, in the price leadership oligopoly model we may assume that the number of dominant firms is more than one. Then the price policy may be determined either through a competition among the dominant firms or by a cartel. In the Cournot model it is assumed that a firm's conjectural variation is zero, equivalently, the conjectural industry output variation of the firm is unity. (See Appendix 12.1 to this chapter for a discussion on conjectural variation.) We may make an alternative assumption about the conjectural variation and determine the corresponding optimal output levels.

As stated in the introduction, another line of variation can be product differentiation where goods are not perfect substitutes. Intertemporal issues can also be examined. We, however, do not wish to examine these issues here. The interested reader may consult books on industrial organization, like Tirole (1988), Carlton and Perloff (1990) and Hay and Morris (1991)

KINKED DEMAND CURVE MODEL

An analytical device sometimes used to explain oligopolistic markets characterized by infrequent price changes is the kinked demand curve suggested by Paul M. Sweezy. It tries to explain price rigidity often observed in oligopolistic markets. This is assumed to occur under a special set of conditions in a market and for a given firm. First, the industry sells a differentiated product and has established a satisfactory set of prices. Second, other firms in the industry will not follow if one firm increases its price—they expect the firm to lose its market share as its price now becomes relatively higher than their own prices. Third, other firms in the market will follow any price reduction made by one firm for holding their market shares. Thus, there is a price rigidity and the oligopolists compete for a greater share of the market on the basis of quality, product design, packaging, advertisement, etc. This shows how product differentiation arises in the kinked demand curve model.

The demand curve of the oligopolist is DAB in Fig. 12.5. The firm sells x_1 at price p_1 initially. If it reduces its price, then other firms will follow, and it will be able to retain its share of the market shown by AB . If it increases

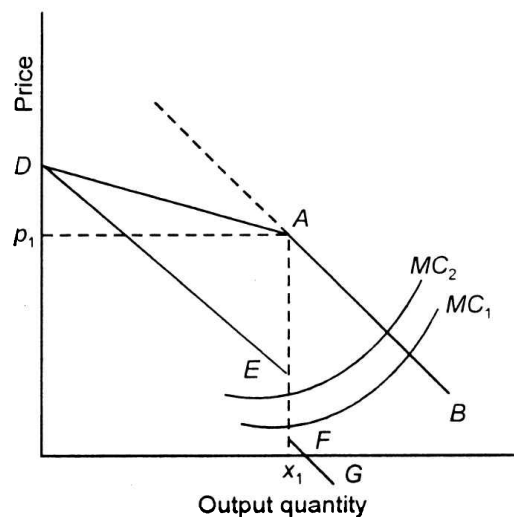


Fig. 12.5 Kinked Demand Curve

its price above p_1 , others will not follow and it will lose its market share. For price increases, it will face the demand curve DA . The entire demand curve faced by the firm has a kink at the original price p_1 .

The kink in the demand curve makes the marginal revenue curve discontinuous at output x_1 . The DA segment of the demand curve generates the DE segment of the marginal revenue curve, and the AB portion generates the FG part. These two parts of the marginal revenue curve do not meet at x_1 , so the marginal revenue curve is discontinuous at x_1 .

We can explain the discontinuity of the marginal revenue curve using elasticity of demand. There is a break in the demand curve at output level x_1 . This creates an abrupt drop in elasticity between output levels infinitesimally below and infinitesimally above x_1 . Since $MR = p(1 + 1/e)$, marginal revenue drops sharply at x_1 . The greater the difference of elasticities between the upper and lower parts of the demand curve, the wider the discontinuous segment of the marginal revenue curve.

The firm's profits are maximum at output x_1 and price p_1 , even though we do not have any information on how price was determined originally. The marginal cost curve MC_1 cuts the marginal revenue curve within its discontinuous segment. At any output less than x_1 , marginal revenue exceeds marginal cost, hence the firm's profits could be increased by increasing output upto x_1 . For any increase in output above x_1 , marginal cost exceeds marginal revenue, and profits fall.

Discontinuity of the marginal revenue curve can give rise to price rigidity. Suppose that the firm's costs increase because of increases in resource prices. The marginal cost curve shifts upward to a position such as MC_2 , but as long as the marginal cost curve continues to intersect the marginal revenue curve within the discontinuous segment, there is no incentive for the oligopolist to change the price. Similarly, resource price reductions shift the marginal cost curve downward, but as long as the marginal cost curve cuts the discontinuous part of the marginal revenue, no price change will occur.

It is easy to see that the demand curve may shift without a change in the price although quantity will change. In this case the demand curve shifts along the same price line. As the market expands (or contracts), the firm will not raise its price, because its given marginal cost curve continues to intersect the marginal revenue curve in the discontinuous region, and hence there is no incentive to change price although output will change. Thus price rigidity would also survive some demand changes.

The logic underlying the kinked demand curve is intriguing, and it does not explain the price-output decisions of firms. It can only explain price rigidity. The model does not tell us anything about the explicit determination of the price level at which rigidity will occur. The empirical existence of the model has not been supported by research efforts. It remains as an oligopolistic possibility only.

INDUSTRIAL ORGANIZATION THEORY

So far in this chapter we have been concerned with determination of prices in oligopolistic models. However, it is difficult to capture in a formal manner all of the factors that may influence market outcomes. In recent years, research in the field of industrial organization has attempted to provide those aspects of market behavior which cannot be captured through formal model formulation. In this section we briefly discuss this.

Industrial organization analysis is mainly divided into three topics: (i) market structure, (ii) firm conduct; and (ii) economic performance. The first of these describes the industry. The second topic deals with firm behavior. The performance aspect is concerned with the desirability of particular markets.

Market Structure

Market structure is concerned with industry description: How many firms are there? What is the number of giant firms? Is the industry highly concentrated? What types of products are produced? Are there increasing returns to scale? Are there significant entry barriers? Understanding such issues is a necessary precondition for analyzing firms' production.

Of particular interest to industrial organization economists is industrial concentration. Several measures of industrial concentration have been employed by researchers to investigate J.S. Bain's (1951) hypothesis which states that profit rates of firms in industries with high concentration should, on the average, be high. Another aspect of market structure that has been systematically examined is entry barriers, because it has been argued that firms in an industry can earn profits above the competitive level if they are protected by entry barriers. Concentration alone is not sufficient to guarantee high profits. There is a vast amount of empirical information on these topics. (A recent survey is available in Chakravarty, 1996.)

Another concern is integration or merger. Under merger a new firm is created by combining two or more firms. Integration can occur between firms producing the same types of goods (horizontal integration), between firms where one produces for another (vertical integration) and between firms in unrelated business (conglomerate integration). The merger between the Punjab National Bank and the New Bank is an example of horizontal integration. Merger of a steel firm with a firm producing pig iron is an example of vertical integration.

Firms may expand horizontally, vertically or into unrelated markets. There are several reasons why it may be beneficial economically for two or more firms to consolidate into one. We briefly discuss some of them.

1. The integrated entity derives benefits of combined services, manufacturing assets, manpower and cash flow. The combined technology and managerial reserves of the integrated firm may enhance the capability of the firm to implement more sophisticated projects to ensure rapid growth and enable it to compete globally. There will be reduction in advertising and other promotional expenditure because of integration. Integration may provide a stronger distribution network.
2. There may be cost saving efficiency due to integration. It may be cheaper to produce different output levels together rather than producing them separately. A firm that controls many firms may be able to close the less efficient ones.
3. Integration may give rise to synergies. For instance, if two firms own complementary patents then they may combine and be able to produce more efficiently than either could alone.
4. The search for market power may motivate integration. If a sufficient number of firms in a homogeneous good industry combine, then the integrated firm would face less competition and possess additional market power. This additional market power might lead to smaller output (higher price for consumers). Analogously, a supplier of an input may create market power in the following way. It may buy or vertically integrate with the manufacturing firms which use the input and monopolize the final output market.
5. Integration can be used for achieving diversification of activities in order to avoid risks. If firms in industries which are subject to oscillatory variations diversify into unrelated markets, they may be able to even out their instability in the income stream.

6. Integration may also pave the way for better and more efficient running of the firm through its ability to attract superior managerial talents. Some managers wish to control large firms since they can enjoy more power.
7. The ability of the merged firm to absorb new technology and spend on research and development may increase.
8. The transaction cost of buying from or selling to other firms may be avoided through vertical integration.
9. Steady supply of an essential input may motivate vertical integration. This will eliminate uncertainty regarding delivery of the input.
10. Firms may integrate vertically to increase rivals' input costs. A vertically integrated firm possessing market power on an essential input may exercise power over rivals which are non-integrated and in turn can place them at a cost disadvantage.
11. Integration may offer incidental benefits like reduction in tax liabilities. This will certainly be the case when the sum of taxes on individual profits will be greater than the tax on a merged entity's profit.
12. Apart from growing big, the fear of increasing competition from rivals may force firms to seek merger.

Mergers and acquisitions were not very uncommon in India as elsewhere in the world. In line with the changes in policy of the Union Government, several mergers and acquisitions had taken place. But it was on a low key, mainly because of several restrictions like the Monopolies and Restrictive Trade Practices (MRTP) Act, 1969. In fact, merger used to be regarded mainly as an instrument of revival of sick units and obtaining tax benefits rather than as a vehicle for achieving corporate growth and expansion. However, with the current liberalization process of the Indian economy the scenario is changing quite fast. The urge for vertical and horizontal expansion and technology upgradation with an eye on global markets to meet competition at international levels are some of the main purposes of the recent merger proposals. An important merger is the recent one of Reliance Petrochemicals with Reliance Industries, one of the largest producers of synthetic fiber textiles in India. This merger was aimed to increase the shareholders' values by realizing significant 'synergies' of both the companies. Another important merger was that of Renuagar Power Supply and Hindustan Aluminium Company. The merged entity expects to participate more vigorously in competitive markets, consolidate the position of the company and diversify into capital intensive projects. Similarly Tata Tea acquired a leading position

in the plantation sector after it took over Consolidated coffee, which has got the largest coffee growing activity in Asia. In addition to coffee and tea, these companies produce other agro products including spices, mushrooms, cardamom, etc. With the economy moving towards globalization, mergers and acquisitions are expected to take place on a much larger scale than in the past.

Firm Conduct

The concerns of this topic are firms' output and price decisions, product differentiation, advertising, investment decisions, research and development activities, and so on. Investigations have also been carried out to see whether oligopolistic coordination in pricing decisions is possible or likely. Economists have examined the roles of explicit agreements, tacit understandings, price leadership, dynamic changes in cost and demand conditions, etc., for effecting such coordination.

Industry Performance

Market performance determines the extent to which the market becomes beneficial to consumers. Researchers investigate how closely behavior in that industry approaches the perfectly competitive ideal. Industries with a fairly large number of firms, low entry barriers and pricing close to marginal cost can be called 'effectively' competitive. Important departures from these ideals will reduce social welfare. Several measures of industry performance such as industry growth and productivity improvements have been considered and attempts have been made to test specific hypotheses about the effect of market structures on various performance measures. The theoretical as well as empirical literatures developed along this line are quite voluminous. The interested reader may look at Schmalensee (1989) and Chakravarty (1996) for further discussions.

Example 12.1

We consider a two-firm homogeneous good industry. The market demand for the commodity is given by $X = x_1 + x_2 = 120 - p$. The cost functions of the firms are $C_1(x_1) = 10x_1$ and $C_2(x_2) = x_2^2/2$. We will determine the quantity leadership, price leadership, cartel and Cournot solutions to this duopoly problem.

Quantity Leadership Solution

Firm 1 is the leader and firm 2 is the follower. Firm 2 maximizes its profits assuming that x_1 is given. The profit function for firm 2 therefore becomes

$$\begin{aligned}\pi_2(x_1, x_2) &= [p(x_1 + x_2)x_2 - C_2(x_2)] \\ &= [(120 - x_1 - x_2)x_2 - x_2^2/2].\end{aligned}$$

The first-order condition for maximum profit is

$$\frac{\delta\pi_2}{\delta x_2} = 120 - x_1 - 3x_2 = 0.$$

Solving this equation in terms of x_1 , we get firm 2's reaction function:

$$r_2(x_1) = x_2 = \frac{120 - x_1}{3}.$$

Now, firm 1's profit maximization must take account of this reaction function. Firm 1 therefore maximizes

$$\begin{aligned}\pi_1(x_1, x_2) &= p(x_1 + x_2)x_1 - C_1(x_1) \\ &= (120 - x_1 - x_2)x_1 - 10x_1\end{aligned}$$

such that $x_2 = r_2(x_1)$.

Substituting $x_2 = r_2(x_1) = (120 - x_1)/3$ into π_1 , we get

$$\begin{aligned}\pi_1(x_1, x_2) &= (120 - x_1 - 40 + x_1/3)x_1 - 10x_1 \\ &= (70 - 2x_1/3)x_1.\end{aligned}$$

$\frac{\delta\pi_1}{\delta x_1} = 0$ implies that $x_1 = 52.5$. Thus, $x_1 = 52.5$ is firm 1's profit maximizing output. Firm 2's profit maximizing output x_2 is determined by substituting the value of x_1 into its reaction function: $x_2 = (120 - 52.5)/3 = 22.5$. The price of the commodity is: $p = 120 - x_1 - x_2 = 120 - 52.5 - 22.5 = 45$.

Price Leadership Solution

For any price p the follower, firm 2, wants to operate where price equals marginal cost. Firm 2's marginal cost function is $MC_2(x_2) = x_2$. Therefore price equal to marginal cost gives $p = x_2$.

This gives the follower's supply curve $x_2 = S(p) = p$.

The residual demand curve (the demand that the leader faces) is

$$\begin{aligned}x_1 &= R(p) = D(p) - S(p) \\&= 120 - p - p = 120 - 2p.\end{aligned}$$

Firm 1 now behaves like a monopolist with respect to the residual demand curve. The profit function of firm 1 is

$$\begin{aligned}\pi_1(x_1) &= px_1 - C_1(x_1) \\&= (60 - x_1/2)x_1 - 10x_1.\end{aligned}$$

$\frac{\delta\pi_1}{\delta x_1} = 0$ gives $x_1 = 50$, which is firm 1's profit maximizing output. The equilibrium price p is obtained by substituting $x_1 = 50$ into the residual demand function: $p = 60 - x_1/2 = 35$. At $p = 35$, firm 2's supply of output is $x_2 = 35$.

Cartel Solution

The cartel solution is found by maximizing total industry profits:

$$\begin{aligned}\pi(x_1, x_2) &= p(X)X - C_1(x_1) - C_2(x_2) \\&= (120 - x_1 - x_2)(x_1 + x_2) - 10x_1 - x_2^2/2.\end{aligned}$$

The first order conditions are

$$\begin{aligned}\frac{\delta\pi}{\delta x_1} &= 120 - 2x_2 - 2x_1 - 10 = 0 \\ \frac{\delta\pi}{\delta x_2} &= 120 - 2x_1 - 3x_2 = 0.\end{aligned}$$

Solving these two equations, we have $x_1 = 45$, $x_2 = 10$. The optimal price is $p = 65$.

Cournot Solution

The two firms' profits are given by

$$\pi_1(x_1, x_2) = (120 - x_1 - x_2)x_1 - 10x_1.$$

$$\pi_2(x_1, x_2) = (120 - x_1 - x_2)x_2 - x_2^2/2.$$

The first order conditions for maximum profits are:

$$\frac{\delta \pi_1}{\delta x_1} = 120 - 2x_1 - x_2 - 10 = 0$$

$$\frac{\delta \pi_2}{\delta x_2} = 120 - x_1 - 2x_2 - x_2 = 0.$$

These two equations can be solved simultaneously for the equilibrium values of x_1 and x_2 to yield $x_1 = 42$, $x_2 = 26$. The optimal price turns out to be $p = 120 - x_1 - x_2 = 52$.

Hence more will be supplied under Cournot's assumption than under the cartel case.

Applied Example 12.1 IATA as a Cartel

The International Air Transport Association (IATA) was formed after a meeting of airline representatives and Governments in 1944. IATA is a cartel which sets prices but no quotas. Airline representatives meet and decide on the price structure and the quality of service to be provided at each price. Air India is a member of this cartel.

This cartel allows free entry, assuming, of course, that the potential airline can overcome the enormous financial entry barriers in terms of aircraft costs. If the entrants follow IATA rules they add nothing to existing consumer choice but add to excess capacity.

The IATA price agreements mean that illegal fare discounts become profitable. Price cutting can take several forms like allowing excess baggage and simple price discounting. IATA uses a large fraction of its budget to investigate illegal price cuts and it attempts to enforce price agreements.

IATA is an example of the worst type of cartel. It is not a natural monopoly and therefore cannot take advantage of economies of scale. But it acts as a monopolist in raising prices for the consumer.

Another example of an unsuccessful cartel is the pepper cartel whose members are Brazil, India, Indonesia and Malaysia. Even though the cartel produces 95 percent of the world's pepper and could raise the price, it has never been able to do so. The reason behind this is that the cartel members keep undercutting the cartel's minimum price.

Appendix 12.1

CONJECTURAL VARIATIONS AND OLIGOPOLY

The profit function of firm i in a homogeneous good oligopoly industry consisting of n firms is given by

$$\pi_i(x_1, \dots, x_n) = x_i p(x_1 + \dots + x_n) - C_i(x_i). \quad (12.25)$$

Firm i will maximize its profits through a choice of its output x_i . The first order condition (marginal revenue = marginal costs) is:

$$p(x_1 + x_2 + \dots + x_n) + x_i \frac{dp(x_1 + \dots + x_n)}{dx_i} = \frac{dC_i(x_i)}{dx_i}.$$

Since $x_1 + x_2 + \dots + x_n = X$, we can rewrite this condition as

$$p(X) + x_i \frac{dp(X)}{dX} \frac{dX}{dx_i} = MC_i(x_i) \quad (12.26)$$

Now
$$\frac{dX}{dx_i} = \frac{dx_1}{dx_i} + \frac{dx_2}{dx_i} + \dots + \frac{dx_{i-1}}{dx_i} + 1 + \frac{dx_{i+1}}{dx_i} + \dots + \frac{dx_n}{dx_i}$$

The term $\frac{dx_j}{dx_i}$ is the rate of change in firm j 's output anticipated by firm i in response to its own change, $j \neq i$. The sum of all such rates $\left(\frac{dx_1}{dx_i} + \frac{dx_2}{dx_i} + \dots + \frac{dx_{i-1}}{dx_i} + 1 + \frac{dx_{i+1}}{dx_i} + \dots + \frac{dx_n}{dx_i} \right)$ measures firm i 's belief or perception about the responses of other firms to its own output change and is called firm i 's conjectural variation. Using a similar argument, we can call $\frac{dX}{dx_i}$ firm i 's conjectural industrial output variation.

We therefore observe that the firm must now not only be concerned with how its own output affects market price directly but also must consider how variations in its own output will affect market price through its effect on other firm's decisions. Because any number of plausible assumptions might be made about conjectural variation, there is no generally accepted type of equilibrium that is likely to emerge from the responses given by Eq. (12.26).

To see this, let us suppose first that the firms follow Cournot behavior. Under the Cournot assumption each firm believes that when it chooses its output all other firms keep their outputs constant. Therefore, $dx_j/dx_i = 0$ for all $j \neq i$ under Cournot behavior. That is, for such a firm i , the conjectural variation is zero. Equivalently, we say that the conjectural industrial output variation dX/dx_i for firm i following the Cournot assumption is 1. Hence if firm i follows Cournot behavior, the first order condition stated above becomes

$$p(X) + x_i \frac{dp(X)}{dX} = MC_i(X). \quad (12.27)$$

Observe that if we assume at the outset that the firms follow the Cournot assumption, then we can begin by stating that the first order condition for firm i 's profit maximization will be

$$\frac{\delta \pi_i(x_1, \dots, x_n)}{\delta x_i} = p(x_1 + \dots + x_n) + x_i \frac{\delta p(x_1 + \dots + x_n)}{\delta x_i} - MC_i(x_i) = 0,$$

or,
$$\frac{\delta \pi_i(x_1, \dots, x_n)}{\delta x_i} = p(X) + x_i \frac{dp(X)}{dX} \frac{\delta X}{\delta x_i} - MC_i(x_i) = 0,$$

which becomes

$$p(X) + x_i \frac{dp(X)}{dX} = MC_i(x_i),$$

the condition given by Eq. (12.27). An n -firm Cournot equilibrium must satisfy the n first order conditions.

We can also express the Cournot first order conditions in terms of output shares. To see this we rewrite condition Eq. (12.27) as

$$p(X) \left(1 + x_i \frac{1}{p(X)} \frac{dp(X)}{dX} \right) = MC_i(x_i),$$

or
$$p(X) \left(1 + \frac{x_i}{X} \frac{X}{p(X)} \frac{dp(X)}{dX} \right) = MC_i(x_i),$$

or
$$p(X) \left(1 + \frac{s_i}{e} \right) = MC_i(x_i),$$

where $s_i = x_i/X$ is the output share of firm i and e is the elasticity of market demand. Therefore, the Lerner measure of market power for Cournot firm i is $L_i = (p - MC_i)/p = -s_i/e$.

We have shown above that the Cournot oligopoly model drops out as a particular case of the general conjectural variation model for specific assumptions about conjectural variation. Other possibilities may also arise. For instance, if we suppose that firm i 's conjectural industry output variation dX/dx_i is zero, then we have price-taking behavior. That is, Eq. (12.26) becomes the first order condition for profit maximization when firms act as price takers. Note that $dX/dx_i = 0$ means that firm i believes that if it increases its output by 1 unit then the remaining firms as a whole will decrease their output by 1 unit, and hence price will not change. Next, suppose dX/dx_i is $1/s_i$, the inverse of the firm's output share. Then Eq. (12.26) becomes the first order condition for a monopolist's profit maximization. That is, the conjectural variation model becomes a monopoly if $dX/dx_i = 1/s_i$.

Thus, the conjectural variation provides a general framework for describing firm behavior in an oligopoly model. Specific assumptions about firm conjectures lead to specific models of market behavior.

CARTEL MODEL

In the collusive or cartel model, firms jointly maximize their profits:

$$\pi(x_1, x_2, \dots, x_n) = p(X)X - C_1(x_1) - C_2(x_2) - \dots - C_n(x_n)$$

where $x_1 + x_2 + \dots + x_n = X$. The first order conditions are

$$\frac{\delta \pi}{\delta x_i} = p(X) + X \frac{\delta p(X)}{\delta x_i} - \frac{dC_i(x_i)}{dx_i} = 0,$$

or

$$\frac{\delta \pi}{\delta x_i} = p(X) + X \frac{dp(X)}{dX} \frac{\delta X}{\delta x_i} - MC_i(x_i) = 0,$$

which become

$$p(X) + X \frac{dp(X)}{dX} = MC_i(x_i)$$

for $i = 1, 2, \dots, n$. Since $p(X) + X[dp(X)/dX]$ does not vary across firms, optimality requires that firms must equate their marginal costs of production with each other.

EXERCISES

Multiple Choice

1. Which of the following is not a characteristic of oligopolistic industries?
 - a) Mutual interdependence.
 - b) Substantial barriers to entry.
 - c) Relatively small sellers.
 - d) Fierce price competition.
2. A profit-maximizing cartel must
 - a) Set the marginal cost of each firm equal to the industry marginal revenue.
 - b) Have some system of sharing profit among firms.
 - c) Set a production quota for each firm.
 - d) All of the above are true.
3. All price leadership models of oligopoly
 - a) Assume that one firm has the power to force others to follow it.
 - b) Assume that the firm with the highest output sets the price.
 - c) Assume that the lowest-cost firm sets the price.
 - d) None of the above is true.
4. Consider a four-firm industry which faces the market demand curve $p = 100 - X$. The firms practise Cournot behavior and the marginal cost of production for each firm is zero. Then
 - a) Each firm produces 25 units of output.
 - b) The market price is 20.
 - c) The industry supply is 60 units of output.
 - d) The equilibrium does not exist.
5. In an oligopolistic industry producing a homogeneous good, industry profits are maximized when firms
 - a) Are Cournot maximizers.
 - b) Produce the same output, whatever their costs.
 - c) Charge the same price, but produce where marginal costs are equalized.
 - d) Charge the price dictated by the dominant firm.

6. Which of the following statements about oligopolistic models is not correct?
- a) The dominant firm price-leadership model assumes that the dominant firm allows the follower firms to sell all the output they wish to at the price dictated by the dominant firm.
 - b) All models of oligopoly assume that firms take account of the effects of their own action on the pricing and output decision of other firms.
 - c) Excess capacity may be an entry barrier.
 - d) In the Bertrand model each firm assumes that others keep their price constant.
7. In the case of the Cournot model it is necessary to assume that
- a) The firms have identical reaction functions.
 - b) Each firm assumes that other firm's prices are constant.
 - c) Marginal costs are zero.
 - d) None of the above.
8. The kinked demand curve model of oligopoly
- a) Assumes that if one firm increases its price, others will also increase their prices.
 - b) Explains why each firm will continue with its current price irrespective of what other firms do.
 - c) Explains why the price is rigid but does not explain how it is determined.
 - d) Leaves price indeterminate because there is a discontinuity in the marginal revenue curve.

Quantitative Problems and Questions

1. Each firm in a duopoly practices Cournot behavior. The cost functions of the firms are given by

$$C_1 = 0.1x_1^2 + 20x_1 + 100,000$$

$$C_2 = 0.4x_2^2 + 32x_2 + 20,000$$

The firms produce a homogeneous product, the market demand function for which is $X = 4,000 - 10p$. Calculate (i) the equilibrium price, and (ii) the equilibrium outputs of the firms.

2. Two firms produce a homogeneous product. The cost functions of the firms and the market demand for the product are given by $C_1 = 0.1x_1^2 + 2x_1$, $C_2 = 0.025x_2^2 + 2x_2$, $X = 500 - 10p$.
 - a) The firms practise Cournot behavior and achieve equilibrium. Derive the reaction functions of the firms and calculate
 - i) The outputs of the firms,
 - ii) The profits of the firms, and
 - ii) The price of the product.
 - b) The two firms form a cartel to maximize joint profit. They also agree that one firm will make a side payment to the other which is sufficient to ensure that forming the cartel will result in the same absolute increase in profit for each firm. Determine
 - i) The outputs of the firms,
 - ii) The price of the product, and
 - iii) The amount of the side payment and which firm receives it.
3. Consider a three-firm homogenous good industry. The market demand function is $X = 1,000 - 40p$. The cost functions of the firms are $C_1 = 20x_1$, $C_2 = 0.075x_2^2 + 13.5x_2$, $C_3 = 0.005x_3^2 + 16.3x_3$.
 - a) The firms practise Cournot behavior. Calculate
 - i) The price,
 - ii) The outputs of the firms and their profits.
 - b) The dominant firm buys the other two firms. The government responds to the merger by imposing a maximum price of Re. 0.50 unit less than the previous equilibrium price.
 - i) Is the monopolist better or worse off than he was before the merger?
 - ii) Are the consumers better or worse off than before the merger?
4. A monopolist can produce at a constant average cost of 5. The firm faces a market demand curve given by $X = 53 - p$.
 - a) Determine the price-quantity combination for this profit maximizing monopolist.
 - b) Suppose that a second firm enters the market. Let x_1, x_2 be the output of firms 1 and 2. Market demand is $x_1 + x_2 = X = 53 - p$. Assuming that firm 2 has the same costs as firm 1, calculate the profits of firms 1 and 2 as function of x_1 and x_2 .

- c) Assume that the firms practice Cournot behavior. Determine the reaction functions, the output levels and the profits of the firms. What is the equilibrium price?
 - d) Suppose now that there are n identical firms in the industry. The firms adopt the Cournot strategy. Determine the market price. Also find the profit maximizing output level and profit for each firm. Show that as n approaches infinity, the output levels, the market price, and the profits, will tend to those that would prevail in perfect competition.
5. Use the analysis developed in this chapter and elsewhere in the chapters on markets to explain the following price behavior:
- a) Doctors charge poor patients less for identical services than they charge rich patients.
 - b) It is less expensive to buy soft drinks in bottles than to buy in cans.
 - c) Known persons pay less than unknown ones for identical houses.
 - d) Banks announce a widely publicized interest rate and change it occasionally.
6. What is a cartel? Why does a cartel behave as a monopolist? Show diagrammatically how a cartel, by combining marginal cost curves, can set price and output to maximize profits. Now show how individual marginal cost can be used to allocate output among oligopolists.
7. What are the advantages to a country of forming a cartel to control price and output of an export good?
8. An industry consists of one big firm (designated b) and twelve small firms (designated i). The cost function of the big firm is $C_b = 0.0333x_b^3 - 2x_b^2 + 50x_b$. The market demand function is $X = 250 - p$. The big firm observes that each of the small firms has the cost function $C_i = 2x_i^2 + 10x_i$. The big firm is a price leader and sets the price so as to maximize its profit and the small firms follow this price. The big firm regulates its own output so that neither surplus nor shortage occurs in the market.
- i) Determine the price.
 - ii) What are the output and profits of the leader?
 - iii) What are the output and profits of each small firm?
9. There are a large number of potential firms all of which can produce a commodity. The i th firm's cost function is $C_i = 1 + x_i + x_i^2$, where x_i is the i th firm's output. The competitive demand for the good is $X = 60 - 12p$.

- a) Determine the competitive equilibrium for this industry. What is the equilibrium number of firms?
 - b) Suppose that the equilibrium number of firms determined in (a) forms a cartel. What will be the price? What will be the output level of a member of the cartel?
 - c) Suppose a firm that does not join the cartel accepts the cartel price as being given. How much output will it sell? What will be its profit? Compare this profit with that of a cartel member.
 - d) Suppose that there is free entry into the cartel. How many firms will join the cartel? (The cartel is assumed to possess the legal right to force any firm to join it.)
10. The market demand curve for a good is given by $X = 1 - p$. The industry that produces the good consists of two firms, each of which has zero marginal cost. (i) Each firm believes that, whatever the market demand, it will get one half the market demand. What is the equilibrium output of the industry? (ii) Suppose that firm 1 believes that it will get a share α ($0 < \alpha < 1$) of the industry output, and firm 2 believes that it will get $1 - \alpha$. Determine the industry output.
11. Along a 10-mile long highway, eleven persons are spaced one mile apart. Each has a demand function for a good given by $x = 20 - d - p$, where x is the quantity of the good demanded, d is the distance to be traveled in miles to obtain the good and p is the price of the good. Suppose that the constant marginal cost of production is 10.
- a) If only one firm is located in the middle of the highway, what is the level of output produced?
 - b) If two firms are to be located along the highway, what should be the optimal location? What are the price and the industry output if the firms agree to maximize industry profits?
 - c) Supposing that the two firms are competing with each other and charge the long run competitive price, what will be their location? Calculate the loss of consumer's surplus as compared with the case of optimal location.
12. The market demand curve for a good is given by $p = 10 - 0.1X$. The constant long run and short run marginal cost of producing the good by any firm in the industry is 5.
- a) Calculate the long run price and output under perfect competition.
 - b) Calculate the monopoly price, output and profits.

- c) Determine the Cournot solution if there are three firms in the industry. What are the industry profits? Why are they less than the monopoly profits? Do you think that the industry profits increase if the number of Cournot firms reduces from three to two?
13. The market demand curve for a good is $x = 200 - 3p$. Suppose that there are 10 small firms in the industry, where each firm has the identical short run marginal cost function $MC = 5 + 2x_i$, with x_i being the output of a typical firm i . The marginal cost of the dominant firm is 5.
- Calculate the short run equilibrium price and output of the industry. What are the outputs of the dominant firm and a small firm?
 - Suppose that the long run marginal cost of a small firm is 20 and that of the dominant firm is $0.1x_d$, where x_d is the output of the dominant firm. Determine the long run industry output, the output of a small firm and the output of the dominant firm.
14. Do you think a cartel of mercury producers will be more stable than that of consumers to fight price increase of some good? Justify your answer.
15. Consider a homogeneous good industry consisting of two firms. The constant marginal costs of these firms are C_1 and C_2 respectively. What are the Bertrand equilibrium, the competitive equilibrium here?
16. Consider an industry consisting of two firms, each of which has zero cost. The market demand function is $X = 120 - p$.
- Determine the competitive equilibrium level of industry output.
 - If each firm practises Cournot behavior, determine the reaction function of the firms. Determine their outputs, profits and the industry output.
 - Determine the cartel output of the industry.
 - Calculate the Stackelberg equilibrium output level if firm 1 is the leader.
17. Prove or disprove the following.
- The Cournot model of duopoly always results in a stable equilibrium.
 - In the price leadership model, the dominant firm sets a price that, in the long run will drive all other firms out of the industry.
 - A leader earns a lower profit in the Stackelberg equilibrium than he would get in the Cournot equilibrium.
 - The profit maximizing cartel obtains a stable equilibrium.

- e) In general, oligopolies do not produce an efficient output level.
 - f) Maximum total profits of the firms in a cartel are achieved when all firms produce equally.
 - g) An oligopoly firm facing a kinked demand curve will maximize its profit where $MC = MR = p\left(1 + \frac{1}{e}\right)$, where e is the elasticity of demand. Therefore, if we know the firm's marginal cost and price, we can determine the elasticity of demand.
 - h) There are n identical firms in a Cournot equilibrium. Then the elasticity of the market demand is less than $1/n$.
 - i) In a price leadership model, all firms other than the dominant firm behave as perfect competitors.
18. Why will the price leader set a price? Why will the price taker accept a lower price than he would otherwise prefer?
19. Consider an industry that consists of one large dominant firm and many small firms. Show graphically how market demand and supply determine output in the absence of the dominant firm. Now illustrate on your graph the effect of a lower price determined by the dominant firm. Show the joint output of the small firms and the output of the large firm.
20. Consider a two-firm homogeneous good industry where the firms follow Bertrand behavior and charge the same price. Show that a firm's residual demand curve is kinked.
21. An oligopolist's cost function is given by $C(x) = 0.00024x^3 - 0.0728x^2 + 64.32x$, where x is the oligopolist's output. The oligopoly firm is now producing 200 units of output and the selling price is 100 per unit. The firm estimates correctly that if it increases its price it would lose its customers, and elasticity of demand for its own product would be -5 . However, if it reduces its price, it would expect rival firms to retaliate with compatible price cuts, and elasticity of demand for its own product would be -2 .
- a) What is the maximum proportion by which factor prices can increase so that the firm will not change the selling price of the product?
 - b) Suppose demand for the product increases but factor prices remain unaltered. There is no change in the elasticity of demand. By what proportion do sales increase at the current price?

22. A homogeneous good is produced in a number of separate plants.

The long run cost function for each plant is given by $C = x^3 - 20x^2 + 120x$.
The market demand function is $X = 3,020 - p$.

- a) Suppose that the market is perfectly competitive. Each plant can be regarded as a separate firm. Find the long run equilibrium values for (i) price, (ii) firm output, (iii) number of firms, and (iv) industry output and industry profit.
- b) Now suppose that firms 1, 2, 3 and 4 each acquire ownership of one quarter of all the plants. These firms practice Cournot behavior. Find the long run equilibrium values for (i) price, (ii) firm output, (iii) the number of plants operated by each firm, and (iv) the total industry output and industry profit.
- c) Next, suppose that firms 1 and 2 merge, and that firms 3 and 4 merge. The resulting firms practise Cournot behavior. Determine the long run equilibrium values for (i) price, (ii) firm output, (iii) number of plants operated by each firm, and (iv) the total industry output and industry profit.
- d) Finally suppose that firms 1–2 and 3–4 merge to form a monopoly. Find the long run equilibrium values for (i) price, (ii) output for each plant operated, (iii) number of plants operated, and (iv) the total industry output and profit.
- e) Compare your results for cases (a), (b), (c) and (d). Do you think prohibition of mergers is justified?

23. In a homogenous good oligopoly industry consisting of n firms, let x_i denote the output of firm i . Denote the industry output $x_1 + x_2 + \dots + x_n$ by X . Let $e_i = (dX/dx_i) (x_i/X)$ be the elasticity conjectured by firm i of the total industry output with respect to changes in its own output.

- a) Interpret e_i .
- b) Calculate e_i when firm i is a (i) monopolist, (ii) price taker, and (iii) Cournot maximizer.

24. Derive a sufficient condition for the reaction function of a Cournot firm to be negatively sloped.

25. What is oligopoly? Why are prices likely to be rigid under oligopoly?

26. Mention and discuss any three types of non-price competition in oligopoly markets.

13

Game Theory

If the choices by a decision maker lead to reactions by other decision makers, with the resulting predictable feedback being taken into account when the original decision is made, we say that strategic elements are present in the framework. Strategic thinking and decision making are absent in a perfectly competitive market because of a firm's inability to influence the market price, and in a monopoly industry because the single firm controls the entire industry. But in a market with few firms, each firm must, to some extent, be concerned with what its rivals will do. As we saw in an oligopoly, making a profit maximizing decision requires that each firm make some conjectures about its competitors' behavior. Thus, in an oligopoly strategic behavior is an essential ingredient of the market structure problem.

One of the primary tools that economists employ for modelling competition among a group of decision makers who interact strategically is game theory. Game theory deals with multi-person decision problems. Applications of game theory arise in many areas of economics. As we have stated, one well-known area of application is oligopoly, that studies market interactions with a small number of firms. One can think of other applications as well. For instance, models of trading processes (such as auction and bargaining models) involve game theory. Labor and financial economics include game theoretic

models of the behavior of a firm in its inputs market (rather than its output market, as in an oligopoly). There may be multiperson decision problems within a firm. Many workers may vie for a promotion, several divisions may compete for a corporation's capital. In welfare economics, applications of game theory arise in the context of private provision of a public good and problems of externalities. International economics includes models in which countries compete (or collude) in choosing tariffs and trade policies. Macroeconomics considers models in which the monetary authority and wage or price setters interact strategies to determine effects of monetary policy.

BASIC ELEMENTS OF A GAME

A game can be defined as a situation in which decision making persons make strategic choices and in which the final outcome depends on the strategies adopted by each person. In the oligopoly context a decision maker or a player is a firm and its price, output, etc., will be determined by the strategies chosen by the firm. Game theory in the oligopoly set-up therefore describes how firms will frame their strategies and how these strategies will determine the values of the concerned variable. All games have three basic elements: (a) players, (b) strategies, and (c) payoffs.

Games may be cooperative or non-cooperative depending on whether decision makers are allowed to make some binding agreements to cooperate, or they act independently since they are not allowed to make binding agreements. The bulk of the developments of game theory during the 1950s and 1960s concerns cooperative game theory. However, recently interest has shifted mainly to non-cooperative game theory. Here we will deal primarily with non-cooperative games. The three basic elements we discuss below are included in such games.

Each decision maker in a game is called a player. A player may be a firm (as in oligopoly markets), a person (as in poker games), or the entire nation (as in military conflicts). All players have ability to choose from a set of possible actions. The number of players in a game is usually assumed to be fixed during the play of the game. In this chapter, for simplicity of exposition, we will be mainly concerned with two-player games i.e., games with two players, and denote these players as *A* and *B*. All of the concepts described in the chapter can be easily extended to multi-person games. An important assumption made in game theory is that the identity of the players is irrelevant: the only concern is the action chosen by a player that will yield the most favorable outcome.

An action open to a player in a game is called a strategy. Depending on the game being played a strategy may be very simple or complicated, but each strategy is assumed to be a well-defined action. The set of all strategies is called the strategy space.

The final returns to the players of a game at its conclusion are called pay-offs. Pay-offs are measured in utility levels obtained by the players, although quite often monetary pay-offs (such as profits, for firms) are used. Pay-offs incorporate all aspects associated with the outcomes of a game including explicit monetary pay-offs and implicit feelings by the players about outcomes such as whether they are embarrassed or satisfied. Players prefer pay-offs that offer more utility to those that offer less. Unless specified, we will deal with games of complete information. That is, each player's pay-offs and strategies are common knowledge between the players. In other words, each player knows his own as well as the other player's strategies and pay-offs. Furthermore, each player has the knowledge that the other player knows this, and so on.

Games may be divided into two groups according to their pay-offs: constant-sum and variable-sum games. In constant-sum games the pay-offs sum to a constant no matter what strategies are adopted. For instance, in a zero-sum game an increase in one player's pay-offs must be offset by an equal loss by another player. All possible pay-off distributions in a zero-sum game are Pareto efficient: one player cannot be made better off without making somebody else worse off. In variable-sum games the pay-off total may vary depending on the strategies chosen. Such games are more realistic. For instance, if all firms charge high prices, profits may increase rather than if all firms charge low prices.

An example of a constant-sum (but not zero-sum) game can be the division of an open space in a housing complex among members of the complex (perhaps in accordance with the strategies such as offering to help keep the space clean). The size of the space is fixed—only the relative shares are to be determined.

CONCEPTS OF EQUILIBRIUM

In our examination of the theory of perfectly competitive markets, we developed the concept of equilibrium, in which both demanders and suppliers remain content with the market outcome. Given the equilibrium price-quantity combination, no market participant has any incentive to change his behavior. Are there similar equilibrium concepts in game theory models?

Are these strategies which, once adopted, provide no incentive for the players to change their behavior further? Do resulting equilibria offer believable explanations of market outcomes?

Although there are several ways of formalizing equilibrium concepts in non-cooperative games, the best-known non-cooperative equilibrium is a generalization of Cournot's equilibrium concept, in which an equilibrium occurs if no player wants to change his strategy. This generalization is the work of John F. Nash published in 1951. A pair of strategies (a, b) is called a Nash equilibrium if a is player A 's best strategy when B plays b , and b is B 's best strategy when A plays a . That is, if a is A 's optimal choice, given that B chooses b , and b is B 's optimal choice given that A chooses a , then (a, b) is a Nash equilibrium. In other words, under Nash equilibrium each player maximizes his pay-off given the decision taken by the other player, and these decisions are mutually consistent with each other. Thus in Nash equilibrium, holding the strategies of the other player constant, no player can obtain a higher pay-off by choosing a different strategy. Even if one of the players reveals the (equilibrium) strategy he will use, the other player cannot benefit by knowing this. Thus, a Nash equilibrium can be interpreted as a pair of expectations about each player's choices such that when the other player's choices are revealed, neither player wants to alter his choice. We will see that for non-equilibrium strategies this is not the case. If one player knows what the other's strategy will be, he may be able to benefit from that knowledge, and may take actions that reduce the pay-off received by the player who revealed his strategy.

Nash equilibrium has some problems. Not every game has a Nash equilibrium set of strategies. In some cases a game may have multiple equilibria, some of which are more plausible than others. Sometimes other equilibrium concepts may be more reasonable than Nash equilibrium. Hence there is a complex relationship between game theory equilibria and traditional market equilibria. However, since Nash equilibrium is an initial working definition of equilibrium and since it is quite appealing intuitively, most of our discussion on equilibrium will take place in the neighborhood of Nash equilibrium.

REPRESENTATION OF GAMES

Two forms of model are used to represent a non-cooperative game: the extensive form and the normal or strategic form. Loosely speaking, these correspond respectively to presenting the game through a sequence of

movement of the players, that is, through a move by move description, and through specification of strategies available to each player and the pay-off received by each player for each combination of strategies that could be chosen simultaneously by the players. These two approaches are not, in general, equivalent, though each has its merits. We will first analyze the normal form, reserving the analysis of the extensive form for the section on sequential games.

To illustrate game theoretic approach to strategic modelling, let us consider a simple example in which two firms (*A* and *B*) must decide how much price to charge. Each firm has only two possible strategies: set either a high price (*H*) or a low price (*L*). In the normal form representation of the game, the firms choose their strategies simultaneously, and the combination of strategies chosen by the firms determines a pay-off for each firm. Each firm in the game knows the other firm's pay-offs and available strategies but does not know in advance the other firm's actual choice of strategies. Each chooses its action without knowledge of the other's choices. Their pay-offs, which depend on the strategies both choose, are shown in Table 13.1. This table presents the price setting game in normal form.

Table 13.1 Price Setting Game in Normal Form

		<i>Firm B</i>	
		<i>H</i>	<i>L</i>
<i>Firm A</i>	<i>H</i>	3,3	0,4
	<i>L</i>	4,0	1,1

In the table, firm *A*'s strategies (*H* or *L*) are shown at the left and *B*'s strategies are shown across the top. Pay-offs corresponding to the various strategic choices are shown in the body of the table. For instance, the pay-offs in the table show that if firm *A* chooses *H* and firm *B* chooses *L* the pay-off will be 0 for firm *A* and 4 for firm *B*. Other pay-offs are interpreted similarly. Table 13.1 therefore provides a summary of the game considered.

DOMINANT AND NASH EQUILIBRIA

From Table 13.1 it emerges that setting a low price is a dominant strategy for firm *B*. No matter what *A* does, the *L* strategy provides a higher pay-off to *B* than does the *H* strategy. Similarly, strategy *L* is a dominant strategy for firm *A*. Whichever choice *B* makes, *A* will get a higher pay-off if it adopts *L*. Hence these strategies dominate the alternatives.

A dominant strategy equilibrium is a choice of strategies by each player such that each strategy dominates every other strategy available to the player. Since a dominant strategy is the strategy that is the best no matter what the other player does, a dominant strategy equilibrium is a Nash equilibrium. But the converse is not true. That is, a Nash equilibrium need not be a dominant strategy equilibrium. A dominant strategy equilibrium, if it exists, is an especially attractive solution to the game, since there is a unique optimal choice of strategy for each player.

In our price setting example, considerations of strategy dominance suggest that we would expect an equilibrium outcome in which *A* adopts *L*, receiving an equilibrium pay-off of 1, and *B* chooses *L*, receiving an equilibrium pay-off of 1. To see why other strategy pairs in Table 13.1 do not meet the Nash criterion, let us suppose that *A* chooses *H* and *B* chooses *L*. This provides *A* with a change to improve its position—if *A* knows that *B* adopts *L*, it can get a higher pay-off by choosing *L*. The choice of *H* by *A* and *L* by *B* is therefore not a Nash equilibrium. Other cases can be analyzed similarly.

When no dominant strategy equilibrium exists, we have to resort to the notion of Nash equilibrium. But there may be more than one Nash equilibrium. We should then try to eliminate those Nash equilibria which are “unreasonable”. One sensible belief about players’ behavior is that it is unreasonable for a player to adopt a strategy that is dominated by other strategies. Therefore, given a game, we should first eliminate all strategies that are dominated and then determine the Nash equilibria. This procedure which may be referred to as elimination of dominated strategies may reduce the number of Nash equilibria significantly.

To see this, consider the two-player (*A* and *B*) game presented in Table 13.2. Each player has two strategies S_1 and S_2 . There are two Nash equilibria here: (S_1, S_2) and (S_2, S_2) . For player *B* the pay-off from S_2 is at least as large as that from S_1 . This means that for player *B* strategy S_1 is weakly dominated by strategy S_2 . Hence if *A* assumes that *B* will not adopt S_1 , then he can eliminate S_1 from *B*’s strategy space, leaving the second column of the game. But now S_2 dominates S_1 for *A*, leaving (S_2, S_2) as the outcome of the game and the corresponding pay-offs are (2, 2).

Table 13.2 Elimination of Dominated Strategies

		<i>Player B</i>	
		S_1	S_2
<i>Player A</i>	S_1	3,3	0,3
	S_2	3,0	2,2

MATCHING PENNIES GAME

There may be games in which no Nash equilibrium exists. This is shown by the two-player (A and B) matching pennies game presented in Table 13.3. Imagine that each player has a penny and must choose whether to display it with head (H) or tail (T) facing up. Thus, each player's strategy space is (H, T) . Once the strategies are chosen, there are pay-offs to each player which depend on the choices of the players.

Table 13.3 Game of Matching Pennies

		<i>Player B</i>	
		H	T
<i>Player A</i>	H	-1, 1	1, -1
	T	1, -1	-1, 1

The choices are made independently and neither player knows the other's choice when he chooses his strategy. If the two pennies match (that is, both are heads up or both are tails up), then B wins A 's penny; if the pennies do not match, then A wins B 's penny. Note that corresponding to each strategy pair, the sum of pay-offs of the players is zero. Thus, this is a zero-sum game. In this game the interest of the players are diametrically opposite.

There is no Nash equilibrium in this game. Any strategy pair is unstable because it offers at least one of the players an incentive to adopt another strategy. For instance, if players' strategies match viz. (H, H) or (T, T) , then player A prefers to switch strategies. On the other hand, if strategies do not match viz., (H, T) or (T, H) , then B prefers to switch. The irregular cycling behavior exhibited in the play of the game shows the absence of a Nash equilibrium.

Thus, the Nash approach may not always identify a well-defined solution to a game. However, if we enlarge our definition of strategies, we can find a new sort of Nash equilibrium for a game in which the Nash equilibrium defined earlier does not exist. This is argued below.

MIXED STRATEGIES

The distinguishing feature of the matching pennies game is that each player would like to outguess the other. In any game in which each player would be interested in outguessing the other(s), there is no Nash equilibrium be-

cause the solution to such a game involves uncertainty about what the players will do. But we have been thinking that each player is making one choice and sticking to it. That is, we have been thinking of each player as choosing a strategy with certainty. This is called a pure strategy.

We now introduce the concept of a mixed strategy, which we will interpret in terms of one player's uncertainty about what another player will do. (This interpretation was suggested by John C. Harsanyi in 1973.) Formally, a mixed strategy for a player is a probability distribution over (some or all of) his (pure) strategies. In the simultaneous-move game of complete information, a player's pure strategies are the different actions the player could take. For instance, in the matching pennies game, player i 's pure strategies are heads and tails, where $i = A, B$. Therefore, a mixed strategy for a player (say, A) is the probability distribution $(q_a, 1 - q_a)$, where q_a is the probability of playing head and $(1 - q_a)$ is the probability of playing tail by A , $0 \leq q_a \leq 1$. The mixed strategy $(0, 1)$ is simply the pure strategy tail; likewise, the mixed strategy $(1, 0)$ is the pure strategy head. We can define mixed strategy for player B similarly and denote it by $(q_b, 1 - q_b)$.

Now, each player has some belief about the other player's behavior. Suppose that A believes that B will play the mixed strategy $(r_b, 1 - r_b)$, that is, A believes that B will play head with probability r_b and tail with probability $1 - r_b$. Similarly, suppose that B believes A will play the strategy $(r_a, 1 - r_a)$. Here r_a is supposed to indicate the probability, as envisioned by B , that A will play head.

Now, each player makes his choice without knowing the other's choice. Player A 's probability that a particular outcome, say (H, H) , will occur is $q_a r_b$. This is simply the probability that A plays H times A 's assumption about the probability that B plays H . In this way we can find the two players' probabilities for different outcomes to occur. Then a player's expected pay-off is given by the sum of the probabilities of different outcomes multiplied by the corresponding pay-offs. Therefore, player A 's expected pay-off from the matching penny game is

$$\begin{aligned} E_A &= q_a r_b (-1) + q_a (1 - r_b) \cdot 1 + (1 - q_a) \cdot r_b \cdot 1 + (1 - q_a) (1 - r_b) (-1) \\ &= q_a (2 - 4r_b) + (2r_b - 1). \end{aligned} \quad (13.1)$$

Similarly, for player B the expected pay-off is

$$\begin{aligned} E_B &= q_b r_a \cdot 1 + q_b (1 - r_a) (-1) + (1 - q_b) r_a (-1) + (1 - q_b) (1 - r_a) \cdot 1 \\ &= q_b (4r_a - 2) + (1 - 2r_a) \end{aligned} \quad (13.2)$$

A mixed strategy Nash equilibrium consists of probability beliefs $(r_b, 1 - r_b)$ and $(r_a, 1 - r_a)$ over strategies, and the probability of choosing strategies $(q_a, 1 - q_a)$ and $(q_b, 1 - q_b)$ such that

- i) the beliefs are correct : $r_b = q_b$ and $r_a = q_a$ and
- ii) each player chooses probabilities (A chooses q_a and B chooses q_b) so as to maximize his expected pay-off given his beliefs.

Therefore, Nash equilibrium in this definition is an equilibrium in action and beliefs. In equilibrium each player's belief about the other player's choices coincides with the actual choice the other player wishes to make. That is, in equilibrium, each player correctly foresees the chances with which the other intends to make choices, and the beliefs of the two players are mutually consistent. Thus, a Nash equilibrium is a rational expectations equilibrium. Note that Nash equilibrium in pure strategies is simply a mixed strategy Nash equilibrium in which the probability of playing a particular strategy is one for each player.

For every game with finite strategies there exists at least one Nash equilibrium in mixed strategies. Since a mixed strategy Nash equilibrium always exists, and since the notion has certain plausibility, it has become a very popular concept of equilibrium. In the matching penny game considered in Table 13.3, it can be shown that $(q_a, 1 - q_a) = \left(\frac{1}{2}, \frac{1}{2}\right)$ and $(q_b, 1 - q_b) = \left(\frac{1}{2}, \frac{1}{2}\right)$ is a mixed strategy Nash equilibrium. (See Appendix 13.1 for a detailed derivation.)

BATTLE OF THE SEXES

Many games have unique Nash equilibria. But there may be games with multiple Nash equilibria and all these equilibria may stand out as equally compelling. This is shown by the classic example; the battle of the sexes. In this game a husband (A) and wife (B) are planning a vacation. B prefers mountain locations, A prefers the seaside. Both prefer a vacation spent together to one spent apart. The pay-offs of Table 13.4 reflect these preferences. Here both the joint vacations represent Nash equilibria. With A adopting the seaside and B adopting the seaside, neither player can gain by taking advantage of knowing the other's strategy. Similar remarks apply to A adopting the mountains and B adopting the mountains. Hence this is a game with two Nash equilibria, and both seem to be equally compelling. In such games Nash equilibrium loses much of its appeal as a prediction of play.

Table 13.4 Battle of the Sexes

		<i>Player B</i>	
		Seaside	Mountains
<i>Player A</i>	Seaside	2,1	0,0
	Mountains	0,0	1,2

PRISONER'S DILEMMA

Another problem with Nash equilibrium is that it may not lead to Pareto efficient outcomes. To see this consider the prisoner's dilemma game depicted in Table 13.5. This game was first discussed by A. W. Tucker in the 1940s. (See Luce and Raiffa [1957].) The title stems from the following game situation. Two persons (*A* and *B*) are arrested and charged with a crime. The police does not have sufficient evidence to convict the suspects, unless at least one confesses. The suspects are separated and explained the consequences that will follow from the actions they could take. If neither of them confesses then both will be convicted of a minor offence and sentenced to one year in jail. If both confess then both will be sentenced to jail for six years. Finally, if one confesses but the other does not, then the confessor will be released immediately but the other will receive a nine-year sentence. The pay-offs for this game are shown in Table 13.5.

Table 13.5 Prisoner's Dilemma

		<i>Prisoner B</i>	
		Confess	Not confess
<i>Prisoner A</i>	Confess	-6,-6	0, -9
	Not confess	-9,0	-1,-1

Now, in this game the 'confess' strategy dominates for both *A* and *B*. To see this, consider the position of *A*. If *B* decides not to confess, then *A* becomes better off by confessing, since he will be released then. Similarly, if *B* confesses, then *A* will be better off by confessing. Thus, whatever *B* does, 'confess' is a dominant strategy for *A*. In fact, the strategy 'confess' is the *maximin* strategy: a player maximizes the minima, or the worst possible outcome. An analogous argument shows that the maximin strategy 'confess' is a dominant strategy for *B*. Therefore, the strategies (confess, confess) constitute a Nash equilibrium.

But if both of them decide not to confess, then both become better off. Because an agreement by both not to confess would reduce the prison terms from six years to one year. Therefore, the strategy pair (not confess, not confess) Pareto dominates the pair (confess, confess). The strategy pair (not confess, not confess) is Pareto efficient—there is no other strategy choice that makes both players better off, while the pair (confess, confess) is Pareto inefficient.

However, the Pareto efficient solution is not stable—each prisoner has an incentive to squeal on his colleague. This is the dilemma: outcomes that appear to be optimal are not stable. The reason behind this is that there is no way for the two prisoners to coordinate their actions. If each can trust the other, then they both become better off.

The prisoner's dilemma-type problems may arise in many real-world market situations. The prisoner's dilemma game formalizes the proposition that in an oligopolistic industry, arrangements that benefit all firms in the industry, when they act as a cartel, create strong incentives for individual firms to deviate from any such arrangements. If all firms follow individual self-interest by deviating from the cartelized arrangements, then the arrangement falls apart. To see this more explicitly, consider a two-firm cartel. Now, interpret 'confess' as 'produce your quota of output' and 'not confess' as 'stick to your quota'. If one firm, say firm *A*, thinks that the other firm is going to stick to its quota, then it is profitable for firm *A* to produce more than its quota. And if firm *A* thinks that the other firm will produce more than its quota, then firm *A* will also adopt the same strategy. Hence we have a prisoner's dilemma-type situation.

In the prisoner's dilemma game, the strategy 'confess' seems to be reasonable if the game is played just once. This follows from the fact that the 'confess' strategy is dominant for both *A* and *B* if the game is played once. From the point of view of formal game theory, one way that the dilemma gets resolved is if the game is to be repeated over and over again, *ad infinitum*. We demonstrate this after a while. Before showing this we interpret the Cournot and Bertrand oligopoly models in terms of single period games and show the relevance of the prisoner's dilemma.

GAMES WITH COURNOT AND BERTRAND EQUILIBRIA

The conjectural variations oligopoly models have been criticized because firms make arbitrary assumptions about rivals' behavior. Game theory shows that

reasonable and rational strategies exist that can lead to Cournot, Bertrand and Stackelberg equilibria. Moreover, these equilibria may be viewed as special cases of Nash equilibria. We interpret the Cournot and Bertrand equilibria in this section. The interpretation of the Stackelberg equilibrium will be taken up later.

The reinterpretation of the Cournot model in game theory terms involves showing that the Cournot equilibrium is a special type of Nash equilibrium. Game theory can then replace the Cournot conjecture and the Cournot equilibrium can be obtained in a more satisfactory manner. For this, we assume, for simplicity, a two-firm industry.

In the Cournot model, the two firms (A and B) use output levels as strategies. The strategy of a firm's rival affects a firm's profit, since each firm's profit is a function of the other firm's output. A set of outputs (strategies) is said to be a Cournot (Nash) equilibrium if, holding the output levels (strategies) of the other firm's (player's) constant, no firm (player) can obtain higher profits (pay-offs) by choosing a different output level (strategy).

Since the Cournot equilibrium is a special case of the Nash equilibrium, it is often referred to as the Cournot-Nash equilibrium. Given that the firms in a Cournot model use outputs as strategies, the Cournot-Nash equilibrium can also be called the Nash equilibrium in quantities.

The Bertrand model can be similarly reinterpreted as a game in which the strategies of firms are the prices they choose. The Bertrand equilibrium can be referred to as Nash equilibrium in prices, since it is a Nash equilibrium in which firms choose prices as their strategies.

In the Bertrand case the profit function takes a quite different form. It is plausible to assume that consumers will buy from the firm with the lowest price, and they will split equally between the two (A and B) firms if they charge the same price. Letting $D(p)$ be the market demand function, the pay-off function for Bertrand firm A is

$$\begin{aligned}\pi_A(p_A, p_B) &= D(p_A) (p_A - c), \text{ if } p_A < p_B, \\ &= \frac{1}{2} D(p_A) (p_A - c), \text{ if } p_A = p_B, \\ &= 0 \text{ if } p_A > p_B,\end{aligned}\tag{13.3}$$

where $p_A(p_B)$ is the price charged by Bertrand firm $A(B)$, $D(p)$ is the market demand as a function of price p and c is the identical constant average cost of production for the two firms.

This game has a structure similar to that of the prisoner's dilemma. In this game the only Nash equilibrium is $(p_A = c, p_B = c)$. On the other hand, if the firms cooperate, that is, collude, then they can charge the monopoly price and each gets half of the total monopoly profit, which is greater than the profit enjoyed under the Nash solution. But the tendency is always there for one firm to cut its price slightly and thereby to capture the entire market. But if both firms get engaged in price cutting, then they both become worse off.

The contrast between the Bertrand and the Cournot games is striking: the former predicts competitive behavior whereas the latter predicts monopoly type inefficiencies. This shows that actual behavior in oligopoly markets may exhibit a variety of outcomes depending on the way in which competition takes place.

REPEATED GAMES AND ENFORCING COOPERATION

Communication and cooperation between participants of a game can be an important part of the game. In the prisoner's dilemma game, for example, the inability to reach a cooperative agreement not to confess leads to a Pareto dominated outcome. Since both players gain from cooperation, they have an incentive to communicate to avoid the dilemma, which stems from lack of trust.

If a single-period prisoner's dilemma game is repeated, cooperation is more likely. In the single-period game, each player took his rival's strategy as given and assumed that he could not influence it. If the game is repeated, however, each player can influence his rival's behavior by signalling and threatening to punish. In a repeated prisoner's dilemma game, an optimal strategy for each player can be announcement of a policy of punishing the other if he chooses the bad strategy, that of 'confessing'. One optimal policy can be the tit-for-tat strategy. According to the tit-for-tat strategy, a player should cooperate in the first round of the game, that is, should play 'not confess' in the first round. In every round thereafter if his opponent cooperated in the previous round, he should cooperate. If his opponent confessed in the previous round, he should confess in the current round. That is, whatever his opponent did in the previous round, he should do the same in the current round.

One interesting point about the repeated game is that if the game is going to be repeated only a finite number of times, the punishment strategy does not work. The reason is that both players know that in the last play of the game, they will both confess (because it is a simple one-period game). Since

they know that both will confess in the last play, the threat to punish does not have any clout with respect to the last but one play, so they will not cooperate in that round too. To see this more precisely, observe that we have just concluded that both will confess in the last round. So why cooperate in the previous round? If one cooperates, the other might confess and exploit his opponent's good nature. Since this argument applies to each player, both will confess. The same argument applies to the third-to-last play, and so forth.

The punishment strategy unravels unless there is going to be an infinite number of repetitions or a random number of repetitions. This is because if one confesses in the n th round, the other will confess in the $(n + 1)$ th round. As long as both players care about pay-offs, the threat strategy may be played repeatedly or randomly.

The tit-for-tat punishment strategy is quite attractive because it offers an immediate punishment and the punishment is imposed only once for each confession. It seems to be a good instrument for obtaining a Pareto efficient outcome in a prisoner's dilemma game which is repeated an infinite number of times or a random number of times.

The above query may be raised about the Bertrand game also: would repetition of the Bertrand game offer some mechanism for the players to offer some supra-competitive profits by pursuing a monopoly pricing policy? One possibility, as discussed in Chapter 12, would be for the firms to establish a cartel and set the price explicitly. We have seen there that such explicit agreements may be difficult to enforce. We therefore adopt an alternative approach by exploring models of tacit collusion. That is, we use game theory to see whether there exist equilibrium strategies which, though not explicitly coordinated, would allow firms to achieve monopoly profits.

In the single-period Bertrand model, the only Nash equilibrium is $(p_A = c, p_B = c)$. This situation would prevail if the game were repeated a finite number of times. Any strategy in which one firm, say A , chooses $p_A > c$ during the final period (T), offers firm B the possibility of capturing the entire market by setting p_B , where $p_A > p_B > c$. Thus, the threat of charging $p_A > c$ in T is not credible. Clearly, this argument applies to any period prior to T . Therefore the Bertrand model makes tacit collusion impossible over a finite time horizon. It is evident that in the finitely repeated Bertrand game, the only strategies which constitute Nash equilibrium in every period are the ones $(p_A = c, p_B = c)$ in every period, that is, in which firms charge a competitive price in every period.

If firms are viewed as having an infinite time horizon, matters change substantially. In this case collusive strategies exist. One such strategy is tit-for-tat. This strategy requires each firm (say, firm A) to charge the monopoly price p_M in the initial period but to charge in any period thereafter the price level that its opponent charged in the previous period. Using the argument similar to that employed in the context of the prisoner's dilemma game it should be clear that the threat implicit in the tit-for-tat game may allow firms to maintain the monopoly price p_M .

A second threat strategy can be the following. If one firm A does not cooperate (that is, charges a price lower than p_M) in one period, then B proceeds to charge the competitive price over and over again for, say, n periods. If the punishment is severe enough i.e. if n is large enough, then this can change the incentives to cheat.

A third threat strategy is the trigger strategy. In this strategy each firm (say, firm A) sets $p_A = p_M$ in every period for which firm B adopts the same price, but chooses $p_A = c$ if firm B has cheated in the previous period. A strategy of this type is called a trigger strategy since a single cheating by firm B triggers off non-cooperation for the rest of the period on the part of firm A .

Trigger strategies constitute a Nash equilibrium in every period. Suppose the firms have colluded for a time and firm B thinks of cheating in this period. Knowing that firm A will choose $p_A = p_M$, it can set p_B slightly below p_M and, in this period, obtain the entire market for itself. It will thereby earn (almost) the entire monopoly profits π_M in this period. But by doing this firm B will lose its share of monopoly profits $\pi_M/2$ forever thereafter because its treachery will trigger firm A 's retaliatory action. The present value of these lost profits is $\sum_{t=1}^{\infty} (1+r)^{-t} \pi_M/2 = \pi_M/2r$, where r is the per period interest rate. Therefore, cheating on firm B 's part is unprofitable if $\pi_M < \pi_M/2r$. This condition holds for $r < 1/2$. We can therefore conclude that the trigger strategies constitute a Nash equilibrium in every period if the rate of interest is very low. The collusion implicit in these strategies is non-cooperative. Firms never have to meet to adopt strategies that generate monopoly profits. These markets are characterized by tacit collusion among firms.

A repeated game is a multi-period game. The chief advantage of a multi-period model is that it allows for more realistic interactions between firms than in a single-period model. We have shown that a firm can signal to another firm that it wants to play cooperatively by lowering (increasing) its output

(price) level. If the other firm responds then they behave cooperatively and avoid competition. If either firm increases (lowers) its output (price) the other firm can retaliate by raising (lowering) its output (price) to punish the transgressor. Thus, in multi-period markets, firms that produce at the Cournot (Bertrand) level in a single-period model may further restrict (increase) output (price) and make larger profits.

SEQUENTIAL GAMES

The games we have discussed so far are either single-period games or repeated versions of single-period games. In these games the players have been assumed to act simultaneously. But in some games one player may know the other player's choice before he makes his own choice. That is, choices may be made sequentially. Chess and bridge are examples of games of this type. Many economic games have this type of structure. An example is the Stackelberg model where one firm is the leader and the other firm is the follower. The strategies are outputs and the equilibrium concept is the same as that in the Cournot model. However, by presenting the Stackelberg model as an explicit game with sequential moves, we see that the Stackelberg model is different from the static Cournot model. A second example can be the following: a duopolist may be able to observe his opponent's capital investment before he takes his own output decision. This type of interaction may exist between consumers and producers also—a monopolist needs to get an idea about the consumer demand before he takes his output decision.

In this section we try to describe a game of this type. In order to illustrate the formal aspects of such a game, we consider a game with two players *A* and *B*, where each player has two strategies, *l* or *h*.

Fig. 13.1 illustrates the details of the game. Player *A* makes the choice first and player *B* makes his choice after observing *A*'s behavior. Such a sequential game is presented in the form of a game tree. The game tree is a diagram that depicts the choices that each player can make at each point in time. More precisely, each node of the tree represents a decision point for the player indicated there. In the game tree presented in the figure, the action proceeds from left to right. Since by assumption the first move belongs to *A*, he must choose *h* or *l*. Because *B* takes his decision after *A*, the tree shows that *B*'s decisions occur to the right of *A*'s. The game tree is nothing but the extensive form presentation of the game.

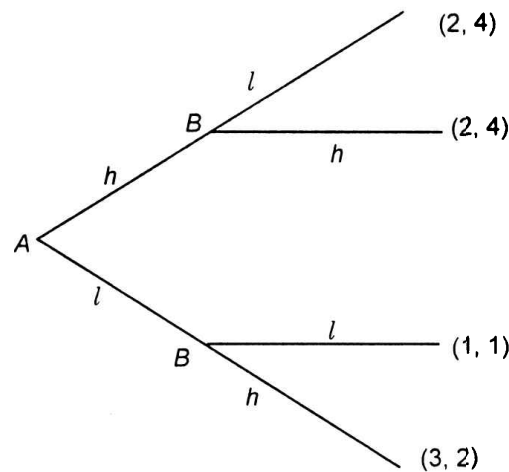


Fig. 13.1 Extensive Form Game

With reference to the game tree in Fig. 13.1, which is the extensive form presentation, it is possible to describe what is meant by a strategy. For example, for A a strategy consists of a specific initial move (that is, choosing l or h). Say l is chosen. Then B has two possible moves; l or h . The choice of l or h indicates B 's strategy. B 's choices are phrased in a way that takes the information he has into account.

The pay-offs to each player are shown as the leaves of the tree, as in Fig. 13.1. For instance, the pay-offs in Fig. 13.1 show that if A chooses l and B chooses h , the pay-offs will be 3 for A and 2 for B . The other pay-offs are interpreted similarly. The nice thing about the tree diagram shown in the figure is that it points out the dynamic structure of the game, that some choices are made after others. A choice in the game means a choice of a branch of the tree. Once a choice has been made the players are in a subgame. A subgame is a game that begins at one decision node and includes all future actions stemming from decisions at this node.

We can calculate the Nash equilibria in each of the subgames. If A chooses h , then it does not matter what B does, the pay-offs are (2,4). If A chooses l , then the sensible decision for B is to choose h and the pay-offs are (3,2).

If A chooses h initially, the outcome will be (2,4). But if he chooses l , he can get a pay-off of 3. Therefore it is reasonable for A to choose l . Hence the sensible equilibrium for this sequential game is (l, h) , so that the pay-offs to A and B are 3 and 2 respectively. That is, the strategies (l, h) are equilibrium strategies given the order in which players have to make their choices.

But the strategies (h, l) are not a reasonable equilibrium in this sequential game. If A believes that B will choose l , then his optimal choice is h . But why

should A believe that B will play l ? Once A has chosen l , the optimal choice for B in the resulting subgame is h . A choice of l by B at this point is not an equilibrium choice in the appropriate subgame.

Now, B can threaten to play l if A chooses l . If A thinks that B is going to act according to this threat he will adopt h . This follows since h gives him a pay-off of 2, which is greater than 1, the pay-off he obtains if he chooses l in the presence of the threat. But is this threat credible? A glance at Fig. 13.1 shows that this threat is not credible. If B were presented with the fact of A having chosen l , he will get a pay-off of 1 if he chooses l and 2 if he chooses h . The threat implicit in the (l, l) strategy is therefore not credible. Even though B 's strategy l is one component of (l, l) , A should be able to infer non-credibility of the threat implicit in it.

Process of elimination of strategies that involve non-credible threats is called backward induction. The term backward comes from the fact that we start at the last move of the game and take backward leaps to do the elimination. Thus, we see that by the elimination of strategies that involve non-credible threats, A can conclude that B will never play l . It turns out that always playing h is a dominant strategy for B . A can recognize this and will opt for l . The Nash equilibrium (l, h) has therefore been shown to be the only one that does not involve non-credible threats. Such an equilibrium is called subgame perfect or perfect equilibrium (Selten, [1975]). Formally, a perfect Nash equilibrium is one in which the strategy choices of each player do not involve non-credible threats. It is called subgame perfect equilibrium also because of its focus on subgames implicit in any extensive form game. For a Nash equilibrium choice of strategies to be perfect or subgame perfect, the strategies specified must constitute not only an overall Nash equilibrium, but also a Nash equilibrium in each subgame encountered during the play. In Fig. 13.1, the Nash equilibrium (l, h) is a perfect equilibrium because once the game reaches B 's decision node, the choice (l, h) is a Nash equilibrium. The Nash equilibrium (h, l) is not a perfect equilibrium because the choice l by B is not a component of Nash equilibrium for the subgame starting at B 's decision node after A plays l .

In view of our discussion in the previous section it is clear that for the finitely repeated Bertrand game, the only subgame perfect equilibrium is one in which firms charge the competitive price in each period. Similarly, for the finitely repeated prisoner's dilemma game the only subgame perfect equilibrium is (confess, confess).

Another interesting observation that can be made from our discussion in the previous section is that a wide range of collusive behavior can be explained in terms of trigger strategies. To see this, consider the infinitely repeated Bertrand game. Recall that in this game the strategy pair (p_M, p_M) constitutes a subgame perfect Nash equilibrium for a low discount rate. Now, suppose that instead of charging the monopoly price p_M , the firms agree on $p_M - \varepsilon$, where $\varepsilon > 0$ is such that $p_M - \varepsilon > c$. Then $(p_M - \varepsilon, p_M - \varepsilon)$ can be a subgame perfect Nash equilibrium for a low discount rate. Since $\varepsilon > 0$ is arbitrary, we can have an infinite number of subgame perfect equilibria in the infinite Bertrand game. This suggests that an infinite number of subgame perfect equilibria can be possible in games with an infinite number of periods. This is called the folk theorem. (This result is called the folk theorem because it was widely known among game theorists in the 1950s, even though it was not published.)

We conclude this section with a minor remark on extensive form presentation of a game. In the example of the extensive form game presented in the section we have considered sequential movements. The extensive form of the game is also capable of modelling cases where some of the choices are sequential and some are simultaneous. The required concept is that of an agent's information set, the set of all nodes of the agent that cannot be differentiated. For instance, suppose that in Fig. 13.1 player B cannot differentiate which of the two decisions (l or h) A has made at the time when B must take his own decision. Hence it is just as if the choices are made simultaneously. In fact, to every extensive form game there is a corresponding normal form game. But there may be many extensive form games with the same normal form game.

Therefore, the extensive form can be used to present everything contained in the normal form and information about choice sequence and information set. Consequently, the extensive form presentation is a more powerful concept than the strategic or normal form. (For further discussions, see Kreps, [1990].)

SUNK COSTS, FIRST MOVER ADVANTAGE AND ENTRY DETERRENCE

In our treatment of entry and exit in the context of competitive and non-competitive markets, strategic considerations were not taken into account. A potential entrant was regarded as being concerned with the relationship between prevailing market price and its own costs. Similarly, a firm was assumed to take quick exit from an unprofitable market. However, under strategic considerations, analysis of entry and exit becomes more complex.

The important point is that a firm wishing to enter or leave a market must make some conjectures about the impact of its action on market price in future periods.

Many game theory-based models of entry and exit stress the importance of a firm's commitment to a particular market. For instance, if the nature of production is such that a very high capital cost is required where the capital does not have a profitable alternative use on exit, then by making such an investment a firm has committed itself to being a market participant. Such kinds of expenditure are called sunk costs. More precisely, sunk costs of a firm are one-time investments that must be made to enter a market but such investments will have no residual value if the firm exits the market. Examples of sunk costs can be expenditure on job specific training for workers or expenditure on unique types of machines. Sunk costs have some similarities with fixed costs in the sense that both these costs are incurred even if no output is produced. However, sunk costs are incurred only once in connection with the entry process, but many fixed costs are incurred periodically. (For instance, executive salaries contracted for in the past may have to be paid on a monthly basis.)

By incurring expenditure as sunk costs a firm may be able to commit itself to a market and in the process can limit the kinds of actions its rivals find profitable. Some game theoretic formulations, therefore, stress the advantage of moving first.

An example can be the Stackelberg sequential game. Suppose output decisions of the firms reflect certain commitments. If one firm (say *A*) has the advantage of moving first, it will be the leader and thereby limit *B*'s options. Adoption of a very high initial production capacity by *A* gives it an advantage—not much room is left for *B* in the market. Given *A*'s advantage in moving first, *B*'s profitable decision is to become a follower.

Other situations in which a first mover may enjoy an advantage can be pursuance of product differentiation strategies or investment in research and development. Pursuit of brand proliferation strategies by an existing toothpaste company may make it difficult for those who come later to develop a sufficiently different product to guarantee a place in the market.

In some situations, the first mover can make the strategic choice to have a very large capacity which in turn will discourage all other firms from entering the market. For instance, in the Stackelberg game, the first mover has a successful entry deterrence strategy if it can satisfy the entire market

demand. If the first mover enjoys an economy of scale in production, the possibility of profitable entry deterrence increases. If the first mover adopts a large scale of operation, it will be able to discourage entry.

LIMIT PRICING AND PREDATORY PRICING

Our discussions on entry decisions in the previous section have focused on sunk costs and output commitments. Entry deterrence strategies can also be formulated in terms of pricing policies. For instance, there may be situations where a monopolist will purposely choose a low (limit) price policy with the objective of deterring entry into the market.

If an incumbent monopoly chooses a price $p_L < p_M$ (where p_M is the profit maximizing price), it has obviously given up some of its current-period profits. But this limit price will block entry in the future only if p_L is smaller than the average cost of the potential entrant. If the cost functions of the monopolist and the potential entrant are the same, then the only sustainable limit price is $p_L = \text{average cost}$. But in this case the monopoly profits are also zero. Therefore, the standard monopoly model offers little scope for limit price behavior—either there are entry barriers that allow monopoly to charge p_M , or there are no such barriers, in which case competitive price will prevail.

Models of limit pricing behavior will therefore differ from the standard set-up. Models incorporating incomplete information have played major roles along this line. If an incumbent monopoly firm possesses more information about a market situation, then it may be able to make use of this additional knowledge to block entry. To illustrate this, consider the game tree shown in Fig. 13.2.

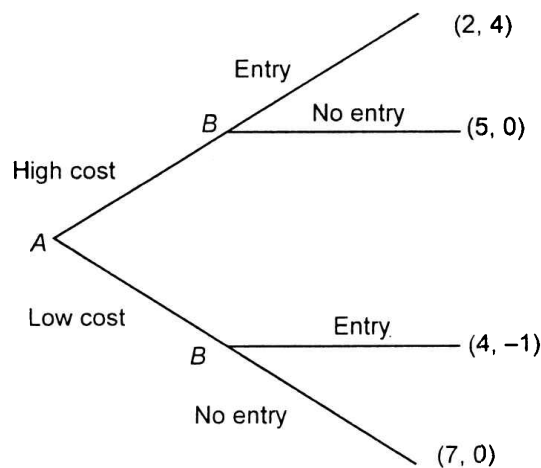


Fig. 13.2 An Entry Game

Firm *A*, the incumbent monopoly firm, has either a high or low cost structure as a result of past decisions. These costs are not known to *B*. Clearly, entry is attractive to *B* if *A* adopts high costs. Since *B* does not know *A*'s cost structure, *B* can develop a (subjective) probability estimate about it. (If a game requires one player to develop subjective probability estimates about other players' behavior, then it is called a 'Bayesian game of incomplete information' [see Harsanyi, (1967)].) This means that *B* must assign some probabilities to the states 'low cost' and 'high cost'. If *B* assigns a probability q to the state 'high cost' and hence the probability $(1 - q)$ to the state 'low cost' then entry will be profitable if *B*'s expected payoff is positive. That is, entry will occur if

$$4q + (-1)(1 - q) > 0; \quad (13.4)$$

that is, $q > 1/5$.

An interesting question that now arises is whether *A* can influence *B*'s probability estimates. Clearly, *A* is better off if *B* decides not to enter, and one way to assure that is to make *B* believe that $q < 1/5$. If *B* is convinced that *A* is a low-cost producer, then *B* will certainly not enter. For instance, if *A* charges a low price when it serves the market as a monopoly, this may act as a signal to *B* that *A*'s costs are low and as a result *B* will decide to adopt no entry strategy. Such a strategy may be helpful for *A* even though it has to give up some of its profits if its costs are high. This provides a possible justification for limit pricing as an entry-deterring strategy.

Sometimes an incumbent firm may reduce its price to drive out rivals. This is called predatory pricing policy. The price reduction policy is adopted to signal to the rivals that its costs are low, even if they are not. The incumbent may adopt extensive advertising or product differentiation activities for convincing the rivals that it enjoys economies of scale in undertaking such activities. Once the rivals are convinced that the incumbent firm possesses such advantages, they will calculate the future expected profits and may take exit. It should be clear that techniques employed for studying limit pricing can as well be used to study predatory pricing policy.

Appendix 13.1

NASH EQUILIBRIUM

In this appendix we first define Nash equilibrium rigorously. Let $\pi_A(r_A, r_B)$ be the pay-off function of player A, where r_A (r_B) is the strategy adopted by player A (B). Similarly, the pay-off function for player B is $\pi_B(r_A, r_B)$. A strategy combination $(\bar{r}_A, \bar{r}_B)$ is a Nash equilibrium if $\pi_A(\bar{r}_A, \bar{r}_B) \geq \pi_A(r_A, \bar{r}_B)$ for all strategies r_A of player A, and if $\pi_B(\bar{r}_A, \bar{r}_B) \geq \pi_B(\bar{r}_A, r_B)$ for all strategies r_B of player B.

THE MATCHING PENNIES GAME

Next, we find the mixed strategy Nash equilibrium for the matching pennies game. Since under Nash equilibrium the beliefs of the players are correct, we have $r_b = q_b$, $1 - r_b = 1 - q_b$, and $r_a = q_a$ and $1 - r_a = 1 - q_a$. Therefore, player A's expected payoff becomes

$$E_A = q_a(2 - 4q_b) + (2q_b - 1).$$

Similarly, player B's expected pay-off will be

$$E_B = q_b(4q_a - 2) + (1 - 2q_a).$$

Since E_A is increasing in q_a if $2 - 4q_b > 0$ and decreasing in q_a if $2 - 4q_b < 0$, A's best response is $q_a = 1$ (that is, heads) if $q_b < 1/2$ and $q_a = 0$ (that is, tails) if $q_b > 1/2$. When $q_b = 1/2$, A is indifferent between the pure strategies heads and tails. Furthermore, since E_A is independent of q_a if $q_b = 1/2$, A is indifferent among all mixed strategies $(q_a, 1 - q_a)$ when $q_b = 1/2$. That is, when $q_b = 1/2$, the mixed strategy $(q_a, 1 - q_a)$ is the best response to $(q_b, 1 - q_b)$ for any value of q_a between zero and one. Thus, $f_A(1/2)$, player A's best response as a function of $q_b = 1/2$, is the entire interval $[0, 1]$.

We can demonstrate similarly that if $q_a < 1/2$, then B's best response is tails; likewise, if $q_a > 1/2$, then B's best response is heads. For $q_a = 1/2$, B is indifferent not only between heads and tails but also among all mixed strategies $(q_b, 1 - q_b)$, so that $f_B(1/2)$ is the entire interval $[0, 1]$.

The intersection of the best responses yields the Nash equilibrium in matching pennies: if player i plays $(1/2, 1/2)$ then $(1/2, 1/2)$ is the best response for player j , as required for Nash equilibrium, where $i \neq j = A, B$. Therefore $(q_a, 1 - q_a) = (1/2, 1/2)$ and $(q_b, 1 - q_b) = (1/2, 1/2)$ is a mixed strategy Nash equilibrium here.

EXERCISES

Multiple Choice

1. Consider a two-person game in which one person is not playing Nash strategy. Then the other person should play his Nash strategy:
 - a) Never.
 - b) Always.
 - c) Sometimes.
 - d) Nothing can be concluded unambiguously.
2. The number of subgame perfect equilibria of the infinitely repeated prisoner's dilemma game is
 - a) Zero.
 - b) One.
 - c) Two.
 - d) Three.
3. Which of the following statements is true about the finitely repeated Bertrand duopoly game?
 - a) The only subgame perfect equilibrium is one in which firms charge the competitive price in each period.
 - b) The tit-for-tat strategy works.
 - c) The only subgame perfect equilibrium is one in which firms charge the monopoly price in each period.
 - d) All of the above.
4. Subgame perfection requires that
 - a) There be exactly one Nash equilibrium in each subgame.
 - b) The Nash equilibria in different subgames be different.
 - c) The strategy choices of each player involve credible threats.
 - d) All of the above.

Quantitative Problems and Questions

1. Prove or disprove the following.
 - a) In a zero-sum game any pay-off distribution among the players of the game is Pareto efficient.
 - b) A Nash equilibrium is always a dominant strategy equilibrium.

- c) In the prisoner's dilemma game the outcome is Pareto dominated.
 - d) A dominant strategy equilibrium is Pareto efficient.
 - e) There is no subgame perfect equilibrium in the finitely repeated prisoner's dilemma game.
 - f) Tit-for-tat is not a good policy if players can make mistakes in their actions.
2. Show that a dominant strategy equilibrium is a Nash equilibrium.
 3. What is a game in normal form? What is a dominated strategy in a normal form game? What is a Nash equilibrium in a normal form game?
 4. Two persons are bargaining over how to share a cake. Both persons simultaneously name the shares they would like to have s_1 and s_2 , where $0 \leq s_1, s_2 \leq 1$. If $s_1 + s_2 \leq 1$, the persons receive the shares they named; if $s_1 + s_2 > 1$, then both receive zero. What is the Nash equilibrium of this game?
 5. What is a mixed strategy in a normal form game? Define a mixed strategy Nash equilibrium in a normal form game.
 6. Consider a two-firm homogeneous good industry. Suppose that at a price of Rs. 2 to customers, the industry sales are 1,000 units. At a price of Rs. 3, the industry sales are 900 units. If both firms charge the same price, then the sales are split between the two. If one firm charges the higher price, then all sales go to the other firm. Present the game in normal form. Show that we have a prisoner's dilemma type situation. Determine the Nash equilibrium of this game. Is it Pareto efficient? Why?
 7. What is the folk theorem? Use the infinitely repeated Bertrand game to explain the central idea underlying this theorem.
 8. Consider the following normal form game:

		<i>Player B</i>	
		S_1	S_2
<i>Player A</i>	S_1	a_1, a_2	a_3, a_4
	S_2	a_5, a_6	a_7, a_8

where S_1 and S_2 are the strategies available to the players and entries in the table are pay-offs. Determine the inequalities for (a_1, a_2) to be a dominant strategy equilibrium. Do you need all of these inequalities for (a_1, a_2) to be a Nash equilibrium?

9. Consider the following normal form game. What are the Nash equilibria of this game?

		<i>Player B</i>	
		S_1	S_2
<i>Player A</i>	S_1	5, 5	0, 5
	S_2	5, 0	4, 4

10. The demand for the output of a two firm industry is $X = 5,000 - 100p$. The firms practise Bertrand behavior. Each firm has a constant average and marginal cost of 10. How does the interest rate r determine the maximum number of firms that can successfully collude in this problem? What is the maximum if $r = 0.2, 0.1$?
11. Why are not separate actions Nash equilibria in the battle of the sexes game?
12. Consider the following pay-off matrix for a Stackelberg game:

		<i>B's strategies</i>	
		Leader ($x_B = 50$)	Follower ($x_B = 25$)
<i>A's strategies</i>	Leader ($x_B = 50$)	(0, 0)	(1,600,800)
	Follower ($x_A = 25$)	(800, 1,600)	(900, 900)

where $x_A(x_B)$ is firm A 's (B 's) output. How would this game evolve if it were played many times?

13. Suppose that in a two-firm industry each firm must choose either half the monopoly quantity $x_M/2$ or the Cournot equilibrium quantity x_C . No other quantities are feasible. Show that this two-action game is equivalent to the prisoner's dilemma: both are worse off in equilibrium than they would be if they cooperated. Show that as long as the discount rate is not too high the joint profit maximizing solution will be a Nash equilibrium of the infinitely repeated game.
14. Suppose that player B rather than player A makes the initial move in the sequential game considered in the chapter. Show the extensive form of the new game. What will be the equilibrium for this game?
15. Two oligopolists operate in a market with the demand curve given by $p = \alpha_1 - x_1 - x_2$, where x_i is the output of firm i , $i = 1, 2$. Each firm has a constant marginal cost c . The firms choose their quantities as follows:

firm 1 chooses $x_1 > 0$; firm 2 observes x_1 and then chooses x_2 . Show how this game can be solved by the backward induction method.

16. There are n firms in a Cournot oligopoly. The demand function is of the type $p = a - x_1 - x_2 - \dots - x_n$, where x_i = firm i 's output. Consider an infinitely repeated game based on this stage game. What is the minimum discount rate such that the firms can use trigger strategies to sustain the monopoly output as a subgame perfect Nash equilibrium? How does the answer vary with n and why?
17. What is a strategy in a repeated game? What is a subgame in a repeated game? What is a subgame perfect Nash equilibrium?
18. Consider a set of persons uniformly distributed along the spectrum from left ($d = 0$) to right ($d = 1$). These persons will elect a person for an office. The candidates simultaneously choose a campaign platform (that is, a point on the line between $d = 0$ and $d = 1$). Each voter will vote for the candidate whose platform is closer to the voter's position on the spectrum. If there are only two candidates, what is the Nash equilibrium?
19. Determine a mixed strategy Nash equilibrium for the battle of the sexes game.
20. Game theorists define firms' activities as strategic substitutes if an increase in the level of an activity (say output, price, etc.) by one firm is met by a decrease in that activity by its rival. Similarly, activities are strategic complements if an increase in an activity by one firm leads to an increase in that activity by its rival. The firms are profit maximizers.
 - a) Show that quantities are strategic substitutes in the Cournot duopoly game with linear demand.
 - b) For the special case of constant marginal cost and linear demand, develop a condition under which prices are strategic complements in the Bertrand duopoly game.
21. Represent the matching pennies game in extensive form. Is your representation unique?
22. Demonstrate rigorously that Cournot and Bertrand equilibria can be interpreted as Nash equilibria.

14

Perfectly Competitive Factor Markets

One of the major objectives of this chapter is to investigate how factor price and employment levels are determined when competition prevails in both product and factor markets. Competition in factor markets implies several things. No firm takes enough of any factor to be able to influence its price. No factor supplier can place enough of a given factor on the market to be able to influence its price. Variable factors are mobile among different employments, and their market prices are flexible. On the basis of these we shall analyze the simultaneous employment of several variable factors by a firm, and also the pricing and employment of any given variable factor. We begin our presentation with a discussion of the 'rent' earned by a factor of production.

ECONOMIC RENT

The concept of rent has played an important role in the theory of factor pricing. The rent (more precisely, economic rent) earned by a factor of production is the total amount by which payments to the factor exceed the minimum amount necessary to keep it in its present employment.

An example of economic rent is the profit earned by a monopolist in the long run. We have seen in Chapter 11 that this profit will accrue to that factor which provides the basis for monopoly power and it will exceed the amount that the factor could earn elsewhere, since such an opportunity cost has already been reflected in the firm's costs. In the following figure we explain the concept in a general case. The factor supply curve is considered to be positively sloped on the assumption that higher factor prices will induce suppliers to offer more factor services in the market place. Analogously, a negatively sloped factor demand curve reflects the assumption that demanders (mainly firms) will hire a smaller quantity at higher prices. (The formal derivations of these curves are presented afterwards in this chapter.) The point of intersection of these two curves is an equilibrium price at which the total quantity of factor services supplied equals the quantity demanded. At a price above the equilibrium price, the quantity supplied is higher than the quantity demanded, and consequently, unemployment arises. On the other hand, if the price is below the equilibrium price, there is excess demand for the factor. Since the supply curve SS' depicts the factor services that would be supplied at different prices, the total amount of money required to retain the level $\bar{x}$ in the present occupation is given by the area $SO\bar{x}E$. Evidently if firms are able to discriminate perfectly among factor suppliers by hiring one unit at a time, then total factor payment becomes $SO\bar{x}E$. However, perfectly competitive markets do not work in this manner. In this case all units of a factor are paid the same price, and this price ($\bar{v}$) is determined by what the firms have to pay for the last unit hired. Because all other units receive a return of $\bar{v}$, but settle for somewhat less and still retain the current employment, these 'intramarginal' suppliers enjoy an economic rent. Total factor payments in the competitive set-up are given by $\bar{v}O\bar{x}E$, and hence total rents are given by $\bar{v}SE$.

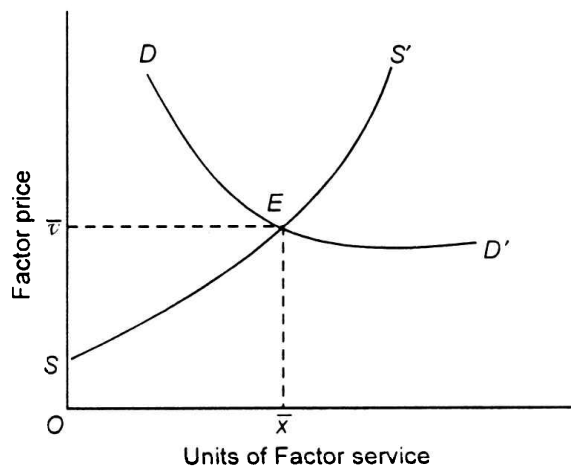


Fig. 14.1 Economic Rent

It is easy to check that the more elastic the supply curve for a factor, the smaller the area that represents economic rent. For an infinitely elastic supply curve (a horizontal line at the prevailing price), there would be no economic rent. On the other hand, for a perfectly inelastic supply curve (the factor is in fixed supply) all of the returns to a factor will be in the form of economic rent. The factor payment in this case is determined by demand only. The supply of the factor is independent of factor prices. Any return that is received is just a consequence of the accident of where the demand curve happens to be.

We refer to short run earnings of any reproducible factor of production (e.g., heavy machinery) as quasirent. We call it quasirent because this rent disappears in the long run (as the factor becomes variable) unlike rents which persist in the long run. In the long run the rental earnings of a reproducible factor of production represent the payments required to bring further the services of the factors. For non-reproducible inputs, economic rent is both a long and short run phenomenon.

Quite often rents exist because of government restrictions on entry into the industry. An example is liquor license. In this case the number of licenses is fixed by law, which creates a legal barrier to entry into the industry and generates economic rents. Firms exert effort (e.g. lobbying in the corridors of power, employing lawyers) to achieve such rents and in the process of this effort resources are used up to obtain such rents. Efforts for achieving or keeping claims on factors of production which are fixed in supply are called rent seeking (since such factors generate rents). From the viewpoint of society the expenditures incurred for securing rents are wasteful, since they do not create any additional output: they simply change the market value of factors of production.

ECONOMIC AND OPPORTUNITY COSTS

It will be interesting to look at the relationship between the concept of rent and that of opportunity cost. A factor of production which has many alternative uses will have a very elastic supply curve to any employment. If an employer reduces its price even slightly, the quantity supplied will go down drastically, (since this factor can receive a higher price elsewhere). In such a case economic rents will be small because the factor earns slightly in excess of what it may earn elsewhere. For instance, clerical workers who have numerous employment opportunities earn almost the same amount in any employment. Therefore, the earnings of a clerical worker in any employment will practically represent no economic rent, since he can earn almost the same amount in some other employment. In contrast, some factors are of limited use in the sense

that they are suited to one particular employment and have lower productivity elsewhere. In this case the supply curve of the factor to this employment is perfectly inelastic—a lower price would not reduce the quantity supplied sharply. For such a factor economic rent would be high and is given by the difference between the earnings of the factor and its opportunity cost in next-best employment. The high amount of money paid to football players can be regarded as economic rent since their alternative earning will be quite low.

LAND RENT

The best example of economic rent is the payment to a factor whose supply is limited. For example, the total area of land of a country is limited. We represent this situation in the following figure. The land supply curve is the vertical line SS' . No matter what the level of demand is, the supply will be fixed. If the demand is represented by the curve D_1 , the rental price of land (per unit) is v_1 and the total return to the owner is $v_1 OSE_1$. On the other hand, if demand is given by the curve D_2 , the equilibrium rental price becomes v_2 which in turn generates a return given by $v_2 OSE_2$. The increase in demand has no effect other than making the landowner richer. Note that this change in the rent of land is absolutely due to the forces of demand. Thus, rent is price-determined, not a price-determining factor for the economy as a whole.

For the individual farmer, the value of the land (more precisely, rent) is a cost of production, since the farmers must pay rents to the factors that create them in order to keep the inputs, which otherwise would be taken away by other firms in the industry. Therefore, from the viewpoint of an individual farmer the value of land is a cost that enters into the pricing of his produce.

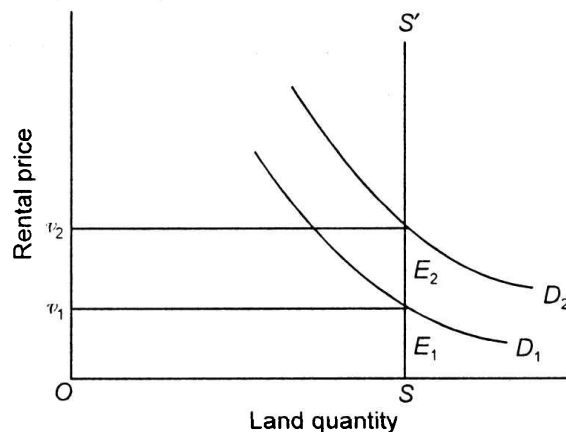


Fig. 14.2 Land Rent

Any factor of production for which alternative uses are slight will have a highly inelastic supply curve to its most favored employment. Since the factor has slight alternative uses, its opportunity cost will be almost zero. Hence most of the total return to this factor will represent economic rent. Since this factor has almost exclusively one use, reduction (or taxation) of the payment does not reduce the quantity supplied.

RICARDIAN THEORY OF RENT

This theory is concerned with rents earned on different parcels of land. In his celebrated work *The Principles of Political Economy and Taxation*, David Ricardo argued that more fertile land tended to command a higher rent. He carefully analyzed the way in which differential rents were related to the land's fertility and to the demand for the crops grown on the land. In the Ricardian set-up, additional land of inferior quality would be cultivated up to the point at which the last acre planted earns exactly zero economic rent. However, economic rent on more fertile land would be positive and these would represent a return based on the land's higher quality. Now, the market price for any particular crop is determined by the costs of the marginal producer and the rent for the marginal producer is zero. In view of this, Ricardo argued that economic rent cannot be regarded as a determinant of market price. Instead prices are determined by the market demand for output and by the availabilities of fertile land.

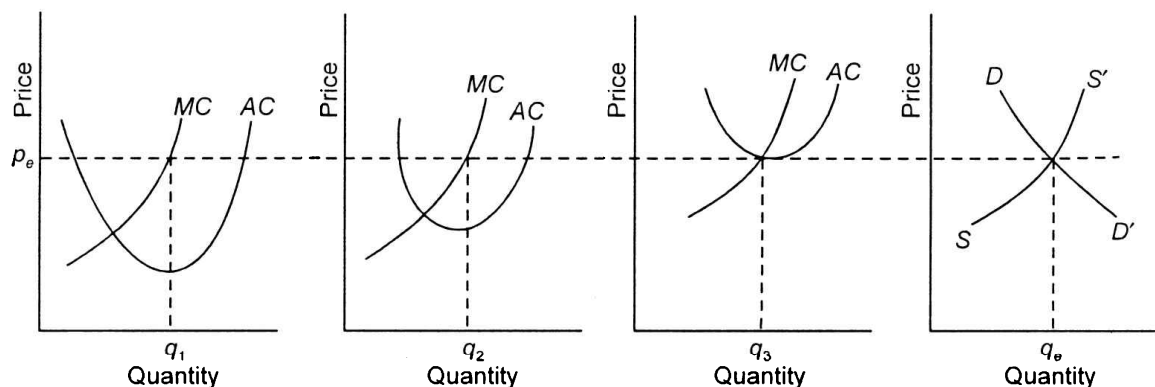


Fig. 14.3 Ricardian Rent

We may give a graphical exposition of Ricardo's analysis. Assume that there are many parcels of land suitable for growing some crop. These parcels range from very fertile (low cost of production) to poor quality (higher cost). We construct the long run supply curve of the crop as follows.

At low prices, the best quality land is cultivated. With increase in price, production continues on fertile land. At still higher prices it will be profitable to make use of low quality land. At the equilibrium price p_e , owners of low-cost farms earn high profits (rent); the extent of rent earned decreases with the quality of land and the marginal farm earns zero rent. If it were possible to earn any rent on additional pieces of land, then these would be brought into production. Evidently, the unplanted land must be of lower quality than that of the marginal farm. The equilibrium described above is stable in the long run since it will be impossible for entrants to earn a profit. Therefore, land quality becomes a natural barrier to entry here. This analysis shows that the demand for land is derived from the product market. The position of the market demand curve will determine how much land can be profitably cultivated and how much rent will be generated.

The Ricardian theory of rent is applicable more broadly to any type of input which earns a surplus. By a surplus we mean here any payment in excess of the amount necessary to have the input in question supplied. In the Ricardian framework the supply of land is fixed. Any payment for the use of land meets this criterion.

Such rent payments arise not only because of heterogeneity of input quality; rent will arise from any source of increasing marginal costs as output increases. Whatever be the source of rents, they will arise alike and will follow an equivalent pattern. We illustrate this by an example in the figure below.

In Fig. 14.4 (a), if one unit is produced the marginal cost is Rs. 6, but with the production of a second unit the marginal cost increases to Rs. 8. Since competition does not allow unequal returns on different units of output, the difference Rs. $(8 - 6) = \text{Rs. } 2$ is a rent on the first unit (shown by area I). If a third unit is added to output, the marginal cost increases to Rs. 12, and so the rent on the first unit rises to Rs. 6, and a Rs. 4 rent is earned on the second unit. The total rent is now shown by the area $(I + II)$.

But this rent is a transfer payment, not a real cost in the sense of using labor and raw materials. The resource cost of three units of output together is Rs. $(6 + 8 + 12) = \text{Rs. } 26$. In terms of Fig. 14.4 (a), this cost is the area $(III + IV + V)$. From Fig. 14.4 (a) it follows clearly that the rent does not represent any resource cost.

We show this in a more familiar way in Fig. 14.4 (b). Here the total resource cost of output level $\bar{y}$ is the area under the marginal cost curve, $0\bar{y}CB$. It is the sum of marginal costs—the equivalent of the sum $(III + IV + V)$ in Fig. 14.4 (a).

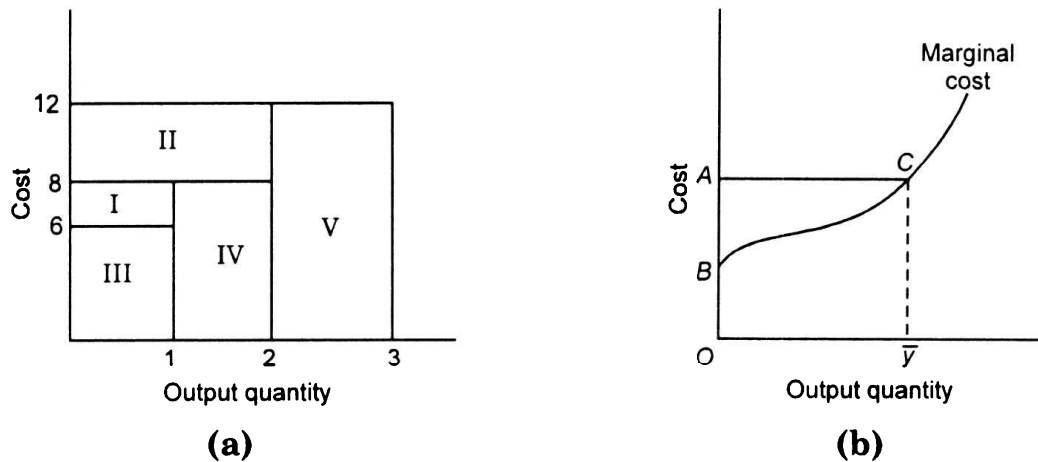


Fig. 14.4 Calculation of Rent

The area *ABC* is the producer's rent or surplus which goes to the suppliers of the firm's productive resources. This area is the equivalent of the sum (I + II) in Fig. 14.4(a). Area *ABC* is the producer's surplus, which together with the consumer's surplus is often taken as the approximate maximal in the analysis of welfare.

The following statement is true at any level of output: The average cost of output y including rent = marginal cost of y excluding rent. To see this, let $y = 3$. Then the marginal cost excluding rent is Rs. 12. But the cost of the first unit of output including rent is Rs. 6 + Rs. 6 (rent). Similarly, from this figure the cost for the second unit of output is also Rs. 12 = Rs. (8 + 4). Therefore, average cost including rent is Rs. $(12 + 12 + 12)/3 = \text{Rs. } 12$. It should be evident that the result is independent of the figures chosen.

MARGINAL PRODUCTIVITY THEORY OF FACTOR DEMAND

The Ricardian theory of rent is an important predecessor of the development of 'Marginalist' economics. Ricardo's hypothesis that price is determined by the cost of the marginal producer can be regarded as one of the bases of modern price theory. One major generalization of this hypothesis is the marginal productivity theory of factor demand which we discuss below.

One-Input Case

We have noted in Chapter 10 that profit maximization implies that output should be chosen such that marginal revenue equals the marginal cost. A similar condition will apply to hiring of inputs. We assume that a firm uses only

labor (L) as the variable input to produce some output Q . Assume also that this output is sold in a perfectly competitive market at price p and the firm can hire labor from a perfectly competitive labor market at the prevailing wage rate w .

Now, if L^* is the profit maximizing choice of labor, then p times the marginal product of labor equals the unit price of labor, that is, the wage rate w . Symbolically,

$$p.MP_{L/L=L^*} = w. \quad (14.1)$$

To understand this, suppose L were chosen less than L^* such that $p.MP_L > w$. Note that since the marginal product of labor MP_L is the increase in the firm's output from adding one more unit of labor, and since the firm sells output in a competitive market at the common price level p , $p.MP_L$ is the firm's increase in revenue due to the hiring of one more unit of labor. We refer to $p.MP_L$ as the marginal revenue product of labor (MRP_L). Thus, $p.MP_L > w$ means that the addition of one more unit of labor increases marginal revenue by $p.MP_L$, which is greater than the increase in cost w . Hence profit will increase if we increase labor hiring and this continues to hold as long as $p.MP_L > w$. Similarly, if $p.MP_L < w$, profit will increase with the decrease in labor hiring. Thus, profit maximization will require that $p.MP_L = w$.

Given the law of diminishing marginal productivity, the MP_L curve is downward sloping. Since $p > 0$ is a constant, the MRP_L curve will also be negatively sloped.

Now as argued in Chapter 8, decreasing MP_L implies that AP_L is also decreasing. However, the AP_L curve lies above the MP_L curve. By multiplying AP_L by the product price p , we get the average revenue product of labor ARP_L , which will also be downward sloping and its curve will lie above the MRP_L

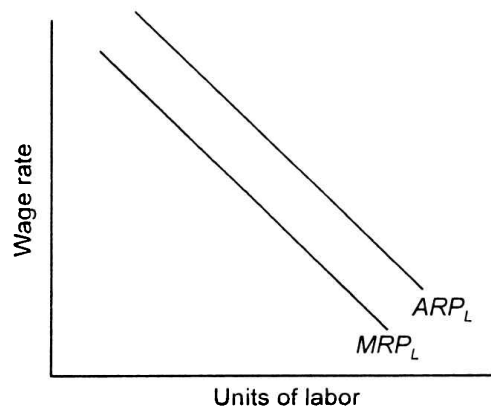


Fig. 14.5 Demand Curve for Labor
(labor is the only variable input)

curve. The MRP_L curve is in fact the labor demand curve. With only one variable input, the marginal revenue curve of the input is the input demand curve.

Two-Input Case

We assume now that there are two variable inputs (the familiar capital K , and labor L) that go into producing the output Q [$Q = f(K, L)$]. Suppose that output Q , capital K and labor L are all traded in perfectly competitive markets at prices p , r and w respectively. The firm has two different types of choices to make. First, it must choose K and L so that any output is produced at minimal cost. Second, it must choose the profit maximizing level of output.

Applying arguments similar to those employed in the single-input case, we can show that the conditions describing the optimal choices of inputs are:

$$pMP_K = r, \quad (14.2)$$

$$pMP_L = w. \quad (14.3)$$

That is, a firm should hire any input up to the point at which the value of its marginal product is equal to its cost. These two conditions give us two equations in two unknowns K and L . Since p , r and w are arbitrarily given, we can solve the optimal choice of each factor as a function of these prices. The resulting equations are called profit maximizing factor demand functions. They are different from conditional or derived factor demands discussed in Chapter 9. Evidently, while the derived demands give cost-minimizing choices for a given level of output, the profit maximizing demands give profit-maximizing choices for a given output price.

The derivation of the factor demand curves in this case is considerably more complex. The assumption of the diminishing marginal product of labor can be misleading here. The problem is that in the single-input case the demand curve depicts the quantity demanded of the input as a function of the input price, given that the output price and prices of the other inputs are constant. But in the case of two or more variable inputs, a change in the price of one input will not only change the quantity demanded of the input, but also of the other inputs. Thus, if w falls, there will not only be a change in L , but also a change in K as a new cost-minimizing combination of inputs is chosen. When K changes, the MP_L function will change (labor has now a different amount of capital to work with), and an argument used in the single-input case will not apply.

Throughout the following illustration of deriving the demand curve for labor, we assume that both capital and labor are variable factors of production. We shall examine the effects of hiring L as w changes, with r and p held constant.

When the price of labor w falls, we can decompose the total effect on the quantity of labor services L hired into two components. The first of these can be called the substitution effect. If output is held constant at Q_1 , there will be a tendency to substitute L for K (since w has fallen, say from w_1 to w_2) in the productive process. Since the cost minimizing condition for producing Q_1 requires that $MP_L/MP_K = w/r$, a decrease in w will necessitate a movement from the original least-cost-input combination A to a new least-cost-input combination B (see Fig. 14.6). Given that the production isoquant exhibits diminishing $MRTS$, it is evident from Fig. 14.6 that this substitution effect is negative; a decrease in w will cause an increase in L if output is held constant.

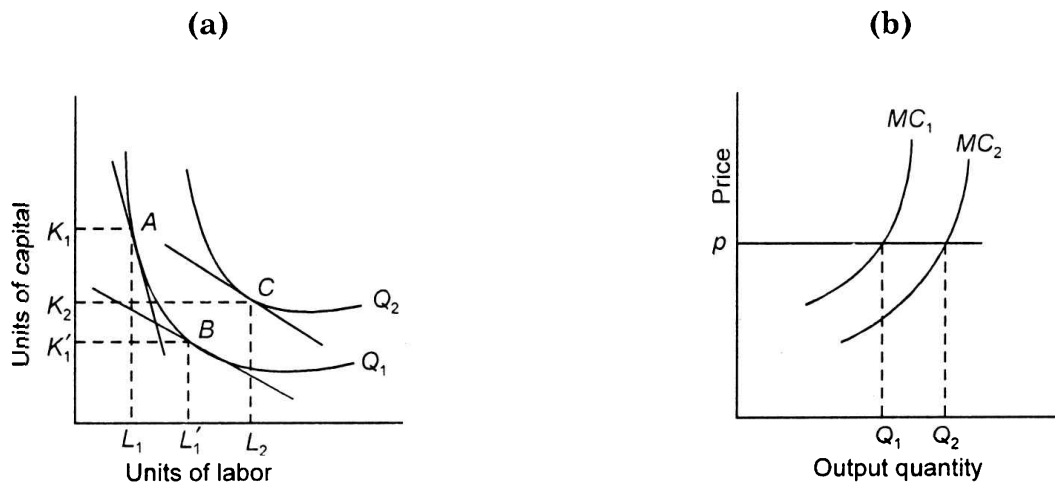


Fig. 14.6 Substitution and Output Effects of Decrease in Price of a Factor

Observe that the reduction of w from w_1 to w_2 will flatten the isocost line. (The isocost line passing through B is flatter than that passing through A .) This will shift the firm's expansion path. Consequently, all the cost curves of the firm will be shifted. In the normal situation the firm's marginal cost curve will shift downward to MC_2 , leading to the choice of a new profit maximizing output level Q_2 . The profit maximizing condition is now satisfied at a higher output level ($p = MC$ holds at Q_2). This increase in output will cause even more labor services to be demanded. The cost minimizing input bundle that can produce Q_2 is given by C . We refer to the movement from B to C as due to output effect. The result of both the substitution and output effects is the movement of the input choice to the point C on the firm's isoquant map $Q = Q_2$ from the input choice A on the isoquant $Q = Q_1$ (through the point B). Both effects work to increase L in response to a decrease in w . But in the case of capital usage, substitution and output effects work in opposite directions and,

in general, no definite conclusion can be drawn about how capital usage responds to the wage change.

Now, the points (w_1, L_1) and (w_2, L_2) lie on the demand curve for labor services D_L , given that r and p are fixed. Note that the D_L curve is not the marginal revenue product curve of labor, because the amount of capital services hired changes along D_L as w changes.

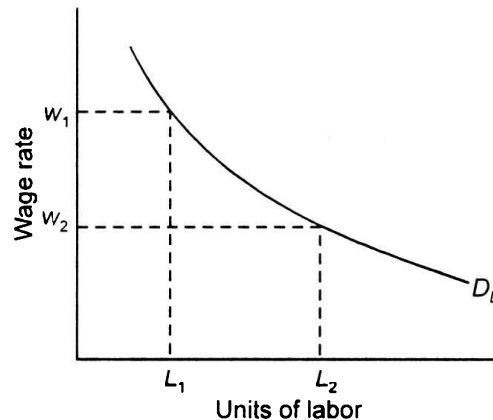


Fig. 14.7 Demand Curve for Labor
(both labor and capital variable inputs)

We now look at the degree of responsiveness of input demand to changes in input prices. Suppose, for instance, that wage rate increased. We consider the substitution effect initially. The decrease in labor hiring will depend on how easy it is for firms to substitute other factors for labor, that is, on the elasticity of substitution of labor for other inputs. In addition to the elasticity of substitution the magnitude of the substitution effect will depend on the length of time that is allowed for adjustment. In view of this and the fixed nature of some inputs in the short run, the substitution possibilities are very slight. In the long run, a firm may be able to adapt its machinery so as to use less labor per machine and the substitution possibility is high.

Next, an increase in wage rate will increase the firm's costs. We have, however, assumed that output price is fixed. This is appropriate if only one firm experiences a rise in labor cost. But if (which is likely) the rise is industry-wide, all firms marginal cost curves would shift upward and hence the industry supply curve would shift also. For a negatively sloped demand curve, this will increase output price which in turn reduces the demand. This is the output effect. The likely magnitude of the output effect will depend on (1) how large the increase in costs resulting from wage increase is, and (2) reduction in output demand due to price increase. The size of the first will depend on the importance

of labor in the total production cost, whereas the size of the second will depend on the price elasticity of demand for the product. In industries where labor costs constitute a significant portion of total costs and for which output demand is highly elastic, the output effect will be large. For instance, an increase in wage of restaurant workers will generate a high output effect. In contrast, since direct labor costs are a small fraction of drug production costs and the demand for drugs is highly inelastic, an increase in wage of pharmaceutical workers will induce a small output effect.

MARKET DEMAND CURVE

A first approximation to the market demand curve for a factor is the horizontal sum of different firms' demand curves. However, a straightforward horizontal sum leaves out what we call the market or external effect of a resource price change. We explain this effect in the figure. Suppose the firm in the figure and every other firm that uses the factor Y are in equilibrium and the price of Y is v_1 . By summing the amounts that all firms employ at v_1 , we determine the total amount Y_1 taken off the market at v_1 . Thus, (v_1, Y_1) is a point of the market demand curve for Y .

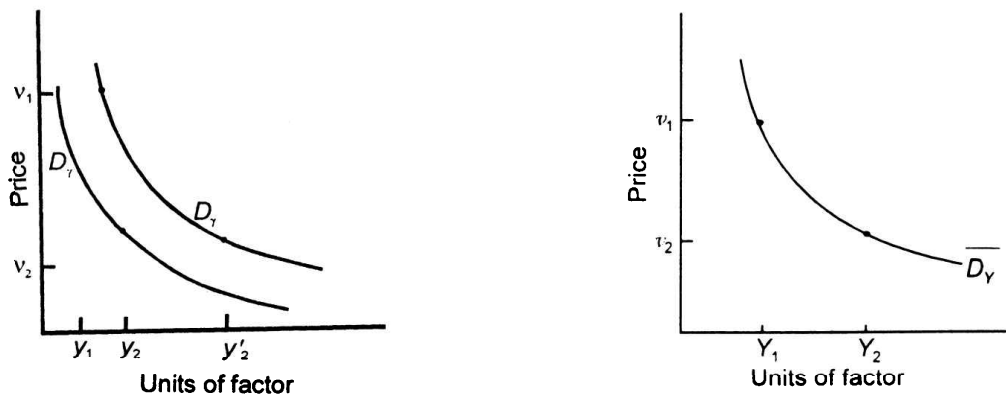


Fig. 14.8 Market Demand Curve for a Factor

Now, suppose that the price of Y decreases to v_2 . Each firm will now employ more of Y . The increased hiring of this factor leads to an increase in total output. The market supply curve of the product shifts downward to the right and the price of the product (given its demand) falls. Since output price is one argument of the demand function of a factor, this implies that individual firms' demand curves for Y shift towards positions such as D'_y . Thus, the individual firms' employment levels of Y will increase towards such quantities as y_2 rather than towards y_2' . When each firm has made the necessary adjustment to choose the least-cost input combination and a profit maximizing product output, and

when each firm's level of hiring is one such as y_2 , the amounts that all firms employ together can be totalled to obtain the point Y_2 , establishing that (v_2, Y_2) is a point on the market demand curve for Y . Similarly, we can obtain other points on this curve and trace out the resulting curve D_Y .

MARKET SUPPLY OF A VARIABLE FACTOR OF PRODUCTION

We have discussed the supply of a fixed factor of production and its implication for its market price in the section 'Ricardian Theory of Rent'. In this section we consider the derivation of the market supply curve of a variable factor of production. The consumers' choices of the variable factor, that is, how much of the factor to provide to firms, when aggregated, generate the supply side of the factor. For illustrative purposes we take up the problem of deriving the labor supply curve.

The market supply of labor is the summation of labor supply by individuals. We therefore begin by deriving the labor supply curve for an individual. The theory of labor supply provides a useful application of the model of individual choice that we have developed in Chapter 3. In that chapter we have analyzed the way in which an individual will choose to allocate a fixed amount of income optimally among a variety of available goods. Individuals must make similar choices in deciding how they will spend their time. Given the fixed amount of time in a day (or in a week, year, etc.), an individual must decide how many hours to work, and how many hours to enjoy as leisure.

We assume that there are only two uses to which an individual may devote his time—either engaging in market work at wage rate w per unit of time or not working. We refer to non-work time as leisure. Suppose that the goods and services consumed by an individual are represented by a 'composite good' denoted by C (consumption), with a per unit price p . We write L for the number of labor hours per day. Then $H = 24 - L$ is the number of hours of leisure per day. The consumer has an indifference map of the usual type, defined over consumption and leisure. Thus, the consumer maximizes his utility

$$U(C, H)$$

subject to the budget constraint

$$pC = M + wL$$

where M is the consumer's non-labor income. This constraint says that his consumption must exhaust his total (non-labor and labor) income. We can rewrite it as

$$pC + wH = M + 24w.$$

Any individual has a 'full labor income' given by $24w$. That is, an individual when worked all the time would have a command that can be denoted by pC over consumption goods. This equation shows that the opportunity cost of consuming leisure is w per hour: it is equal to earnings foregone per hour for not working. It also shows that if the individual does not work at all, then his amount of consumption is M/p .

Using arguments employed for utility maximization subject to budget constraints, we can demonstrate that given the real wage w/p , the individual should choose to work that number of hours for which the marginal rate of substitution of consumption for leisure is w/p .

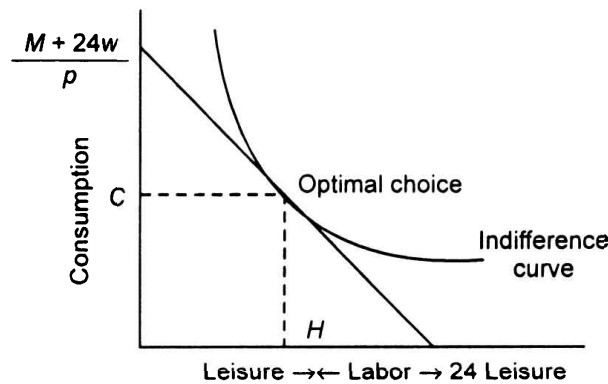


Fig. 14.9 Labor-Leisure Choice

We now consider comparative statics analyses of labor supply. First, suppose a consumer's non-labor income increases. Assuming that price and wage rate are held fixed, it is expected that labor supply would decrease (people would consume more leisure). This expectation implicitly assumes that leisure is a normal good. In fact there exists a fair amount of evidence that supports this assumption. We will maintain this assumption throughout our discussion on labor supply.

Next, we examine the implication of increase in wage rate on consumption-leisure choice. When w increases, the price of leisure becomes higher—the individual gives up more in lost wages for each hour of leisure. Therefore the substitution effect of an increase in w on hours of leisure will be negative. As leisure becomes expensive, it is reasonable to consume less of it. However, the income effect would be positive. Given that leisure is a normal good, the higher income resulting from increase in w will increase the demand for leisure. Therefore, the two effects work in opposite directions; which of these two effects is larger is an important empirical question and cannot be decided by theory alone. Thus, as the wage rate increases people may work more or less. This should be

contrasted with the cases discussed in Chapter 4. There we have noted that only for inferior goods, do income and substitution effects work in opposite directions. For normal consumption goods the two effects work in the same direction. However, in this case the two effects always work in different directions, though leisure is a normal good.

The two possible reactions to an increase in w are illustrated in the figures below. In both the figures the initial wage rate is w_1 and the optimal choices of C and H are (C_1, H_1) . When w_1 increases to w_2 , the optimal point becomes (C_2, H_2) . The movement from (C_1, H_1) to (C_2, H_2) can be broken down into two components: movement from (C_1, H_1) to S (substitution effect) and movement from S to (C_2, H_2) (income effect). In panels (a) and (b) of Fig. 14.10 these two effects combine to produce different results. In panel (a) of Fig. 14.10 the substitution effect of a higher wage outweighs the income effect, and the individual demands less leisure ($H_2 < H_1$). That is, in this case wage increase induces him to work more.

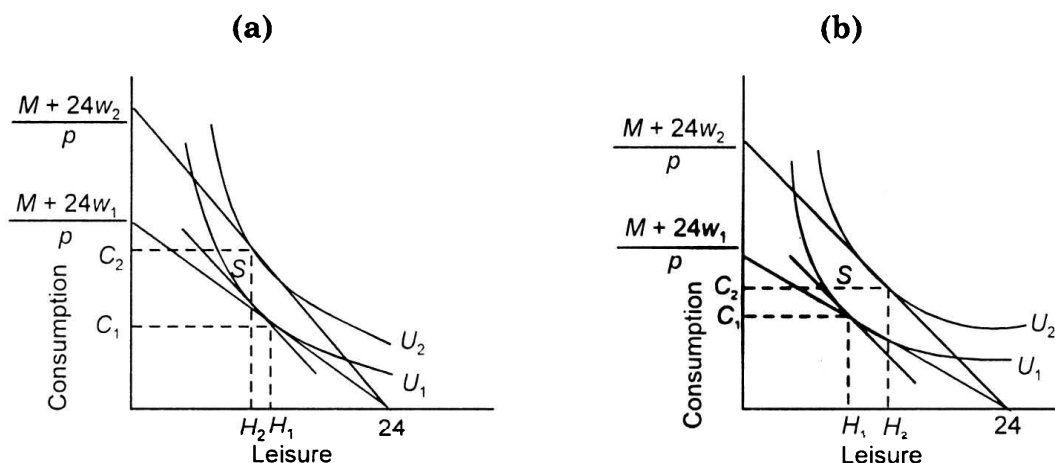


Fig. 14.10 Income and Substitution Effects of Increase in Wage Rate

The opposite situation arises in panel (b) of Fig 14.10. The income effect of higher wage outweighs the substitution effect and causes the individual to demand more leisure ($H_2 > H_1$). Thus, in this case the individual works less as w rises. In this case the consumer's supply curve of labor becomes backward bending as shown in panel (b) of Fig. 14.11. Using the above analysis we can now construct the labor supply curve of an individual. In panel (a), at higher wage the individual prefers to work more. The labor supply curve is upward sloping. (This corresponds to the situation in Fig. 14.10 [a].) In Fig. 14.11 (b), once wage rate exceeds a certain level, even higher wage rate induces the individual to work less. (This corresponds to the situation in Fig. 14.10 [b].)

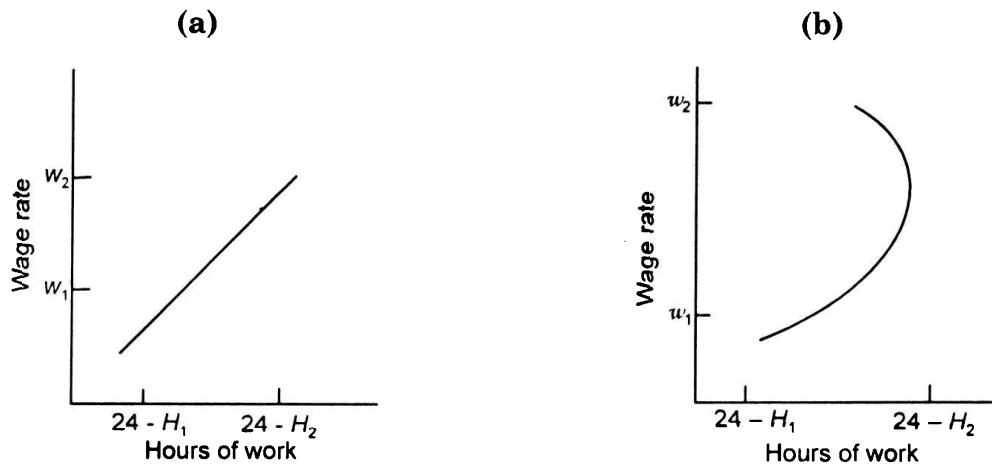


Fig. 14.11 Two Shapes for an Individual's Labor Supply Curve

As stated in the beginning of this subsection, at each possible wage rate we would add together the quantity of labor offered by each individual in order to arrive at a market total. Although there is general agreement that the labor supply curve of an individual may be either positively sloped or backward bending, there are disagreements regarding the shape of the market supply curve. It is argued that in the short run the market supply curve of skilled workers may be backward bending after a certain level of wage. However, since young people may get attracted by high wages, in the long run the entry of such workers into the job market may make the labor supply curve positively sloped everywhere. Even for affluent nations the entry of new workers generates positively sloped labor supply curves. There is evidence that in some situations individual labor supply curves are also positively sloped everywhere. An example is the case of extra time remuneration or overtime wage. Given that the extra time remuneration is higher than the usual wage, its increase will motivate workers strongly to increase labor supply. However, an increase in the usual wage could decrease the supply of labor.

DETERMINATION OF FACTOR PRICES

We have already indicated in the context of discussion of economic rent that given the market supply and market demand for an input, its price is determined by the intersection of these two curves. Equilibrium occurs in the market when it is cleared. In the market for labor services since all firms and consumers face the same wage, the following conditions hold: (i) the equilibrium wage rate equals the marginal revenue product of labor, and (ii) labor services are supplied by each consumer up to the point where the marginal rate of substitution between leisure and consumption equals the real wage rate.

As stated earlier, the demand for factors is a derived demand. The supply of a factor e.g. labor is not cost-determined, like the supply of commodities. Labor supply involves the attitudes of individuals toward leisure and work. Thus, the difference between commodity pricing and factor pricing lies in the determinants of the demand for variable factors and techniques adopted to generate the supply of factors.

Example 14.1

A perfectly competitive firm uses labor L as its only factor of production. Its production function is $x = -L^3 + 5L^2 + 210L$, where x is the firm's output. The wage rate is Rs. 10.00 and the per unit price of the product is Re. 1.00. The firm maximizes profit. We want to determine the number of workers the firm wishes to employ.

From the production function we see that the marginal product of labor is $dx/dL = -3L^2 + 10L + 210$. The value of the marginal product of labor is then $-3L^2 + 10L + 210$ (since the product price is Re. 1.00). Therefore, to maximize profit the firm should choose L such that $-3L^2 + 10L + 210 = 10$. Solving this quadratic equation we get $L = 10$.

Applied Example 14.1 Job Competition from Undocumented Workers

Suppose citizens of country A enter some states of country B illegally along the lengthy border between the two countries, seeking and getting such menial jobs as field workers/helpers, in unorganized sectors or dish washers in hotels and restaurants. We wish to see the impact of this on the employment of low-skilled workers of country B .

We distinguish between agricultural market (panel [b]) and the non-agricultural market (panel [a]) for labor. In the non-agricultural market the demand curve is D_n , the supply of domestic workers is S_d , and the supply of workers of both countries A and B is S_n . The horizontal difference between S_n and S_d shows the amounts that aliens are willing to supply at each wage rate. We denote the demand and supply curves of aliens in the agricultural sector by D_a and S_a respectively.

In the absence of any barriers to immigration, the equilibrium wage rate would be w_1 in both the markets. The employment level in the agricultural sector is L_4 and that in the non-agricultural sector is L_3 . In panel (a), the employment level of country B workers would be L_1 and that of country A workers would be L_1L_3 .

Now suppose a law is enforced prohibiting undocumented aliens from working in country *B*. The aliens are quite easily found in the non-agricultural sector. Enforcement policies can block them from non-agricultural employment. In such a case the effective supply curve in panel (a) is S_d . The equilibrium wage rate for domestic workers increases to w_2 and their employment level increases to L_2 . Therefore, non-agricultural workers of country *B* gain but their employers lose.

Now, many aliens blocked from working in non-agricultural markets shift to the agricultural sector, where they can hide easily. The supply curve in that sector will shift to the right, say to the position S_b . The wage rate falls to w_0 and the employment level rises to L_5 . Workers lose, but growers of agricultural products gain.

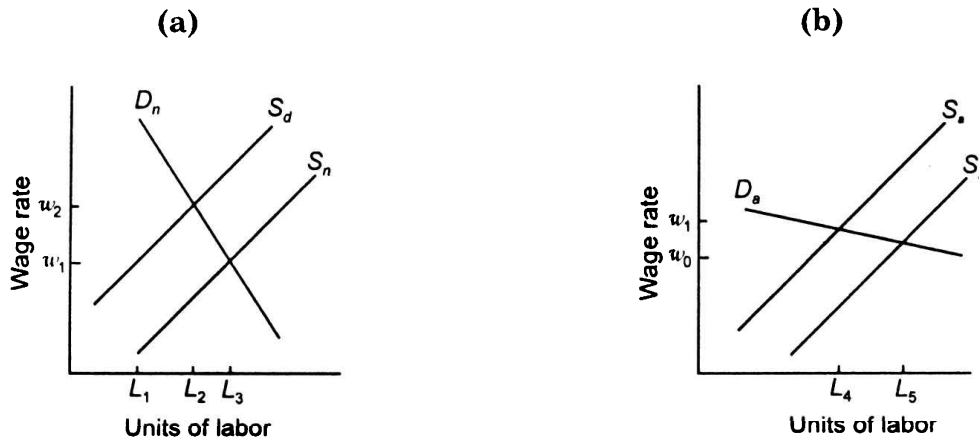


Fig. 14.12 Job Competition

Appendix 14.1

FACTOR DEMANDS BY A FIRM

Let us use calculus to derive factor demand curves rigorously. When labor is the only variable factor, we can write the production function as $h(L)$. The profit π of a competitive firm as a function of L is then given by

$$\pi = Qp - wL = ph(L) - wL.$$

To maximize profit the firm should choose L such that $d\pi/dL = 0$. Consequently,

$$\frac{d\pi}{dL} = p \cdot \frac{dh(L)}{dL} - w = 0$$

or $p \cdot MP_L - w = 0,$

or $p \cdot MP_L = w,$

which is Eq. (14.1). The second order condition for profit maximization requires that $d^2\pi/dL^2 < 0$. Now,

$$\frac{d^2\pi}{dL^2} = p \frac{d^2h(L)}{dL^2},$$

which is negative under the assumption that the production function meets the law of diminishing marginal productivity. Therefore, the labor demand curve derived in Eq. (14.1) is downward sloping.

When the two inputs are variable the profit maximization problem of the firm is

$$\max_{K,L} pf(K, L) - rK - wL,$$

which has the first order conditions

$$\begin{aligned} w - p \cdot MP_L &= 0, \\ r - p \cdot MP_K &= 0, \end{aligned} \tag{14.4}$$

$$\begin{aligned} \text{or} \quad w &= p \cdot MP_L, \\ r &= p \cdot MP_K. \end{aligned} \tag{14.5}$$

These are exactly the conditions derived in Eqs. (14.2) and (14.3).

EXERCISES

Multiple Choice

1. If A is the only variable factor of production for a competitive firm, then the firm's demand curve for factor A is derived from the
 - a) Marginal product curve of A .
 - b) Total product curve A .
 - c) Marginal revenue product curve of A .
 - d) None of the above.
2. Quasirent of a firm is

- a) Greater than the firm's profit.
 - b) Equal to the firm's profit.
 - c) Less than the firm's profit.
 - d) Any one of the above is possible.
3. Suppose all markets are perfectly competitive. If the price of an input falls, the quantity demanded of that input rises due to
- a) The substitution effect.
 - b) The output effect.
 - c) Both (a) and (b),
 - d) Neither (a) nor (b).
4. The price elasticity of demand for any input by a competitive firm will be greater
- a) The larger the elasticity of substitution for the inputs.
 - b) The larger the share of total cost represented by expenditure on that input.
 - c) The higher the price elasticity of demand for the good being produced.
 - d) All of (a), (b) and (c).
5. If three units of output have, respectively, marginal costs of Rs. 5, Rs. 9 and Rs. 11, then the total Ricardian rent is
- a) Rs. 8.
 - b) Rs. $25/3$.
 - c) Rs. 6.
 - d) None of the above.

Quantitative Problems and Questions

1. Prove or disprove the following.
- a) In a competitive labor market, a minimum wage above the equilibrium wage will generate unemployment.
 - b) A skilled worker can produce k times as much as an unskilled worker, where $k > 1$; if all markets are perfectly competitive, then the skilled worker will earn k times the wage rate of the unskilled worker.
 - c) Suppose all markets are perfectly competitive. Then a worker's wage will be equal to the value to the consumer of what his work adds to the firm.

- d) If all markets are perfectly competitive, the demand curve for a factor is that part of the marginal revenue product curve that lies below the maximum point of the average revenue product curve.
- e) If the last worker employed by a competitive firm adds to its output more than its average productivity of labor, then the firm is not maximizing profit.
2. A firm faces perfectly competitive markets for both factors and products. Its production function is $x = \sqrt{PL}$, where x is output, L is the number of workers employed and P is the number of units of plant used. The per unit price of the product is Rs. 10.00. The wage rate and the rental price of plant are both equal to Re. 1. In the short run plant is a fixed factor; the firm has one unit of plant in the short run.
- a) Calculate the following in the short run.
- The equation of the firm's demand curve for labor. If there is an industry consisting of 1,000 identical firms of the type stated above, can you calculate the market demand curve?
 - The number of workers employed by the firm.
 - The elasticity of the firm's demand curve for labor.
 - The firm's profit.
- b) In the long run the price of the product and the rental price of plant remain constant at the values specified above. Derive the firm's long run demand curve for labor.
3. A firm operates in competitive markets for inputs (capital $[K]$ and labor $[L]$) and outputs (x and y). The production functions are given by
- $$x = (0.2K^{-1} + 0.8L^{-1})^{-1},$$
- and
- $$y = K^{1/3}L^{2/3}.$$
- Is x or y more capital-intensive?
4. Suppose that capital and labor are used in fixed proportions for producing a good, and production exhibits constant returns to scale. The wage rate is Rs. 2.00, the rental price of capital is Rs. 3.00 and the price elasticity of demand for the good is -2 .
- What is the wage elasticity of demand for workers?
 - What is the elasticity of demand for capital with respect to its rental price?

- c) What is the cross-elasticity of demand for capital with respect to the wage rate?
5. A person behaves perfectly competitively in both labor and product markets. His utility function is of the form $U = \sqrt{C(24 - l)}$, where C stands for units of consumption per day and l is the number of labor hours supplied per day. Let p be the price of C and w be the hourly wage rate.
 - a) Derive his labor supply as a function of the hourly wage rate.
 - b) Calculate the elasticity of labor supply at $p = 1$, $w = 10$.
 - c) Derive his demand function for C .
 - d) Determine the elasticity of demand for C at $p = 1$, $w = 10$.
 - e) The person is required to pay a daily income tax of $T = 0.25(m - 60)$ where m is his pre-tax income. How many hours per day will he work at $p = 1$, $w = 10$?
6. What is meant by marginal product and marginal revenue product? What does the law of diminishing returns tell us about the marginal product?
7. For a competitive firm, increasing marginal costs result from the operation of the law of diminishing product. Analyze this statement.
8. A purely competitive manufacturer of a product uses two variable factors to turn out its product. State and explain the set of conditions that must be met if profits are to be maximized.
9. Use opportunity cost of time to discuss the following.
 - a) Some persons are ready to pay higher fares for travelling by air than by train.
 - b) Some persons are likely to stand in long queues to purchase tickets for a cricket match.
10. Assuming given preferences and a given wage rate, use an indifference curve approach to show how the individual who can supply labor will choose between leisure and income. What happens to the amount of leisure taken as wage increases?
11. Suppose that the demand curve for labor is flat at some wage rate w_0 . The supply curve of labor is upward sloping. Consider a tax on labor income at the rate t . Show that for tax revenue to decrease as tax rate increases, we require the elasticity of labor supply with respect to wage rate to exceed $t/(1 - t)$. (The graphical representation of the tax revenue as a function of the tax rate is called the Laffer curve. The interesting feature of this curve is

that when the tax rate is quite high, an increase in the tax rate will reduce tax revenue. The reason behind this is that at very high tax rates labor supply reduces, which reduces the tax revenue. This is called the Laffer effect.)

12. Suppose that a competitive firm that uses capital and labor as inputs chooses input vector (K^1, L^1) for producing x^1 amount of output when the prices of inputs and outputs are (r^1, w^1, p^1) . If the price vector is (r^2, w^2, p^2) then the firm's input-output vector is (K^2, L^2, x^2) . Formulate the 'Weak Axiom of Profit Maximization', and show that the corresponding factor demand curves are negatively sloped and the output supply curve is positively sloped.
13. Clearly bring out the distinguishing features between derived factor demands and profit maximizing factor demands.

15

Imperfectly Competitive Factor Markets

Conditions of imperfect competition can arise in any factor market. These conditions affect factor pricing in three different ways. First, a firm hiring a particular input may sell its output in an imperfectly competitive market. Under the assumption of perfect competition in factor markets, this aspect of imperfect competition can be handled as an extension of the analysis of the previous chapter. Second, imperfect competition may prevail in a factor market if there are only a few demanders of the factor. This is the case of oligopsony (monopsony, if there is only one demander) in the factor market. Third, imperfect competition can arise if the sellers of a factor are able to form an effective monopoly.

For the sake of concreteness, in this chapter we will analyze the above-mentioned departures from the perfectly competitive model in the context of a labor market. Note also that the three types of imperfect competition are found in labor markets quite frequently. It would be a simple matter to apply the formal analysis presented in the chapter to any other factor of production.

MONOPOLY IN THE OUTPUT MARKET

Assume that the monopolist hires labor from the perfectly competitive labor market. Assume also for simplicity that labor is the only factor of production. Since the monopolist faces a perfectly competitive labor market, the demands for labor by the monopolist are only an insignificant fraction of the total labor demand. Recall that the profit maximizing monopolist will hire labor up to the point at which the extra revenue from hiring one more labor unit is equal to the extra cost. Given that the labor market is perfectly competitive, the extra cost for an additional labor unit is the wage rate—the firm faces a horizontal labor supply curve. However, since the firm is a monopolist in the product market, the analysis of labor demand differs from that presented in the previous chapter.

To see this, we write the production function as $X = h(L)$, (since labor L has been assumed to be the only factor of production). The revenue $R(X)$ that the firm receives will depend on its output so that $R(X) = p(X)X$, where $p(X)$ is the firm's inverse demand function. The marginal revenue product of labor, the extra revenue from hiring one more unit of labor, is given by

$$\begin{aligned} MRP_L &= MR.MP_L \\ &= \frac{\Delta R \Delta X}{\Delta X \Delta L}, \end{aligned} \quad (15.1)$$

where ΔX is the increase in output resulting from ΔL increase in labor and ΔR is the increase in total revenue resulting from ΔX increase in output. Adding one more unit of labor increases output by the amount given by MP_L . By adding one more unit of output the monopolistic firm increases total revenue by the amount MR . Hence MRP_L for the monopolist is $MR.MP_L$. Since the monopolistic firm takes the price of labor as given (by the market wage w), profit maximization requires that labor be hired up to the point at which

$$MRP_L = w. \quad (15.2)$$

Since we can write the monopolist's marginal revenue as $p(X) [1 - 1/|e|]$, where e is the elasticity of demand, MRP_L in Eq (15.2) becomes

$$MRP_L = p(X) \left[1 - \frac{1}{|e|} \right] MP_L. \quad (15.3)$$

It is now easy to see that the analysis for the competitive firm is really a special case of (15.2). The absolute elasticity of demand that an individual firm faces in a competitive market is infinite. Therefore, for a competitive

firm marginal revenue is just equal to the price and the condition of Eq (15.2) would reduce to the condition derived in Eq (14.1), namely the wage will be equal to the marginal product of labor times the output price.

Given that the monopoly firm faces a downward sloping demand curve, for every X we have $MR(X) < p(X)$. Therefore, at any level of employment of the factor, the marginal value of an extra unit will be less for a monopoly firm than for a competitive firm. (Compare Eq. (14.1) with Eq (15.2).) Since MRP_L for a monopoly firm is smaller than that for a competitive firm, the value of L (where $MRP_L = w$ holds) for the competitive firm will be to the right of the value of L (where $MRP_L = w$) holds for the monopoly firm. Thus, the monopoly firm's demand for labor (L_m in Fig 15.1) will be smaller than the labor demand (L_c in the figure) by the same firm if it behaves competitively.

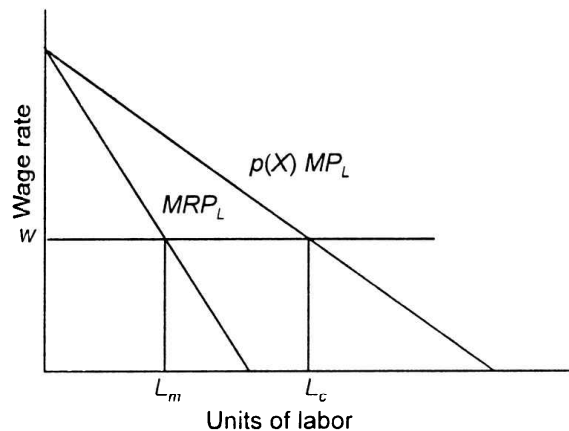


Fig 15.1 Factor Demand by a Monopolist

In the case where the firm has both capital and labor as variable inputs, its demand curve for labor can be constructed in much the same way as that used in the previous chapter. Again a change in w will induce both substitution and output effects. The analysis of the substitution effect is similar to that presented in the previous chapter. An analysis of output effect is somewhat different because now we have a negatively sloped demand curve and the firm recognizes that its marginal revenue is less than the market price. The two effects work in the same direction: an increase in w compels the firm to demand less of L . Therefore, a monopolist who acts as a price taker in the labor market will have a negatively sloped demand curve.

MONOPSONY

In many situations firms are not price takers for the inputs they buy. That is, the supply curve of the factor faced by the firm is not a horizontal line at

the prevailing factor price. Such a situation may arise if there are few buyers of a particular resource. One buyer will take a large proportion of the total supply of the resource to influence the market price. This is the case of oligopsony. A polar case of this is the situation that there is a single buyer of the particular resource. This resource market structure is called one of monopsony. Another situation may arise in which there are many buyers of a particular kind of resource, but there is differentiation within the resource category that causes specific buyers to prefer the resource of one seller to that of another. This is the case of monopsonistic competition. Our analysis of factor markets here will centre around monopsony, but it can also be applied to oligopsony and monopsonistic competition.

As above we will assume that the firm produces output x using a single factor, say, labor, according to the production function $x = h(L)$. Since the firm is a monopsonist, it dominates the labor market. The amount of factor it demands will influence the price it has to pay for the factor. This relationship is summarized by the labor supply curve $S_L = w(L)$, which is assumed to be upward sloping. That is, the more labor the firm wants to employ, the higher must be the wage rate. Thus, while a firm in a competitive factor market is a price taker, a monopsonist is a price maker. (In a perfectly competitive factor market a firm is a price taker.)

We assume that while the firm is a monopsonist in the labor market, its product is sold in a perfectly competitive market at price p . The total cost of labor to the firm is $w(L)L$. Therefore, the firm's profit maximization problem is

$$\max_L ph(L) - w(L)L. \quad (15.4)$$

The profit maximization condition requires that the firm hires the number of units of labor such that the marginal revenue product of labor equals the marginal cost. Because the firm sells its output in a perfectly competitive market, the marginal revenue product of labor is pMP_L . Since the firm is a monopsonist in the labor market, the marginal cost of labor is no longer w , as in the competitive case. In order to increase its hiring of labor by one more unit, the firm must move to a higher point on the labor supply curve. This will involve paying not only a higher wage to the marginal worker but also additional wages to those workers already employed. The marginal factor cost (MC_L) or the marginal expenditure of the extra unit therefore exceeds its wage rate. To see this, note that the total change in labor cost from hiring ΔL units more of labor will be $\Delta(w(L)L) = w(L)\Delta L + L\Delta w(L)$, so that the change in labor costs per unit change in ΔL is

$$\begin{aligned}
\frac{\Delta[w(L)L]}{\Delta L} &= MC_L = w(L) + \frac{\Delta w(L)}{\Delta L} L \\
&= w(L) \left[1 + \frac{L}{w(L)} \cdot \frac{\Delta w(L)}{\Delta L} \right] \\
&= w(L) \left(1 + \frac{1}{e_s} \right)
\end{aligned} \tag{15.5}$$

where $\Delta w(L)/\Delta L$ is the slope of the supply curve of labor and e_s is the elasticity of labor supply. Since labor supply curve is positively sloped, $e_s > 0$. If the supply curve is perfectly elastic, e_s is infinity and MC_L becomes the wage rate, the case of a firm facing the competitive factor market.

We may represent the above analysis graphically. D_L and S_L are the labor demand and supply curves respectively. Since S_L is positively sloped, the MC_L curve lies above the supply curve S_L everywhere. The curve S_L is an 'average cost of labor' curve and MC_L is marginal to S_L . The profit maximizing level of labor for the monopsonist is given by L_m , for at this level of L , $MC_L = pMP_L$ holds. (A calculus proof of the condition $pMP_L = MC_L$ is given in Appendix 15.1.) At L_m , the wage rate is w_m . Note that L_m is smaller than that which would be hired in a perfectly competitive wage rate (w_c). The difference ($w_m - w_c$) is a measure of monopsonistic exploitation. We thus derive the conclusion that in the labor market the monopsony results in a lower wage rate and fewer laborers being hired than perfect competition would. Although we derive this conclusion under the assumption of perfect competition in the output market, clearly the same conclusion holds if the firm is a monopolist in its output market and a monopsonist in its factor market.

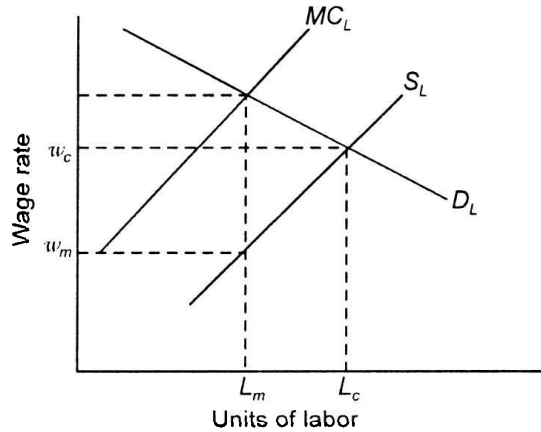


Fig 15.2 Pricing in Monopsonistic Labor Market

The formal similarities between monopsony and monopoly should be clear now. Monopsony restricts input demand by virtue of a monopsonistic position in the factor market, whereas monopoly restricts the level of output. The demand curve for a monopsony consists of a single point (L_m, w_m) , whereas a monopolist's output supply curve consists of one point only. Thus, while the monopoly operates in a Pareto inefficient point in the output market, the monopsony operates at a similar point in the factor market.

There are many examples of monopsony. For instance, in many small towns a local firm may be the only source of employment. Since moving costs are high for local people, the monopsonistic model may be appropriate. Some monopsonistic positions may arise because only one firm hires a particular type of labor that has a unique technical knowledge. For example, we may consider the market for astronauts. In this case the government is a monopsonistic buyer. A monopsonistic model may be appropriate for the market for professional football and hockey players. (See Applied Example 1 of this chapter.)

As in the case of monopoly power, we can talk about measures to counteract monopolistic exploitation of resources. One measure can be the use of administered or fixed minimum resource prices. (See Applied Example 2 of this chapter.) Second, measures successful in increasing resource mobility will reduce the monopsonistic power of particular resource users. For instance, with regard to labor resources, mobility among geographical areas and firms, horizontal mobility among occupations at the same skill level, and vertical occupation mobility to higher skill classifications will be of use in counteracting monopsony.

DISCRIMINATION IN HIRING BY MONOPSONIST

We have already noted some formal similarities between a monopoly firm and a monopsony firm. Another interesting similarity deserves a discussion at some greater length. In the context of monopoly we have noted that if two markets are separate, a monopolist can maximize profits by selling products at different prices in two markets. Analogously, if a monopsony can segregate the market of a factor into two or more distinct markets, it may be able to increase profits. To see this, consider the labor market. A monopsony may be able to discriminate in hiring between men and women. Since it is possible to identify the market to which a particular employee belongs, the firm will find it profitable to pay different wages in two markets. We assume that men and women are equally productive and that the firm has a constant

marginal revenue product of labor. (Given that the firm sells its product in a competitive market at price p , the latter assumption means that MP_L is constant, that is, the production function $x = h(L)$ is linear. Thus, x may be taken as $x = a + bL$, $b > 0$.) The labor supply curves for men and women are given by S_M and S_W respectively. Women have a more inelastic supply curve because they have relatively few employment alternatives. Given these supply curves, the firm will choose that quantity of labor in each market for which the marginal cost equals the marginal revenue product of labor (MRP_L). Therefore, the firm will hire L_M and L_W from the men's and women's markets respectively. The wage rates in these two markets will be w_M and w_W respectively. Men's wage exceeds that of women's, though men and women are equally productive. Thus, by segregating the labor market, the firm is able to pay different wages to different groups of employees and increase its profits.

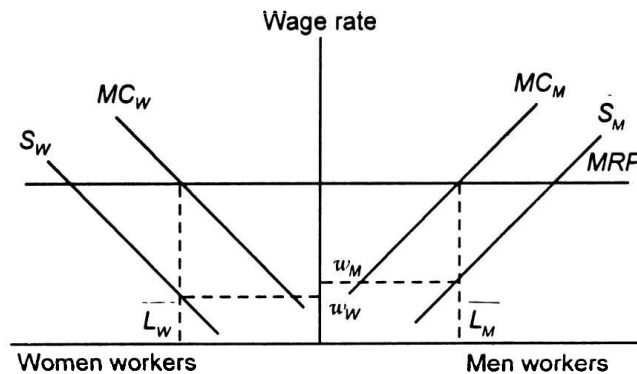


Fig. 15.3 Discrimination in Hiring

MONOPOLY IN SUPPLY OF INPUTS

A third way in which departures from perfect competition may occur in input markets is that the suppliers of the input may be able to form a monopoly. Examples of such monopolies are labor unions in industries, production cartels for certain types of capital equipment and firms that control unique supplies of some resources. The analysis of such situations is quite similar to that used for monopoly. The monopoly supplier will choose a point on the input demand curve it faces. Since this choice will result in input prices in excess of opportunity costs, monopoly rents will be generated. These rents will continue to exist as long as entry into the input market can be restricted.

Of particular interest among the monopolies in input markets are labor unions and bilateral monopoly. We analyze these cases below.

Monopoly Power and Labor Unions

Labor unions, the best known monopolists on the supply side of the labor market, usually restrict entry into the labor markets they control and bargain with the employers for wages and other benefits to their members. The justification for encouraging growth of labor unions has been that monopsony power by firms can be counteracted by labor unions only.

Consider the case of a labor union facing a competitive industry. Union members' preferences between leisure and income are reflected in the labor supply curve S_L . D_L stands for the labor demand curve by the competitive industry. It is derived by adding up the individual firms' demand curves. This curve is also the average revenue curve for the union, from which the marginal revenue curve for the union, MR_L , can be constructed by the usual method. At any value of L , MR_L is the increase in labor expenditure when an additional laborer is employed.

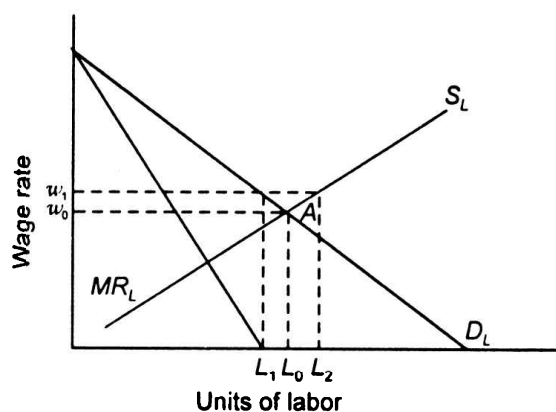


Fig. 15.4 Labor Union and Income Maximization of its Members

In the absence of labor union, the competitive market solution occurs at $A = (L_0, w_0)$. In contrast, if the union decides to maximize income (revenue) of its members, then maximization occurs at the point where $MR_L = 0$, that is, at $L = L_1$. D_L is of unitary elasticity at this point. Union control increases the wage rate from w_0 to w_1 and decreases labor hiring from L_0 to L_1 as compared to the competitive case. Thus, more revenue is now available for laborers and fewer workers will share this revenue. The amount of unemployment is $L_2 - L_1$. If the labor demand curve becomes more and more elastic, not only the wage rate but total employment also will decline, and the union members as a whole will be worse off. Clearly, monopoly power exercised by a union results in a Pareto inefficient situation. The loss of efficiency occurs in part, since some workers who would like to enter and work at a lower wage than the union wage cannot because of restricted entry.

Bilateral Monopoly

If both the demand and supply sides of the market are monopolized, then the market outcome is not determinable. In the following figure we present a market in which a monopolistic supplier of some input (say, a rare metal to produce an alloy) faces a monopsonistic buyer of the input (the only producer of the alloy). The monopolist will prefer the point $E_1 = (v_1, F_1)$ since at this point his marginal cost (MC) equals the marginal revenue (MR) associated with the demand for the rare metal. The monopsonist would prefer the point $E_2 = (v_2, F_2)$ since his profit is maximized at this point. Neither point E_1 nor the point E_2 are equilibrium points and the market outcome must be settled through bargaining.

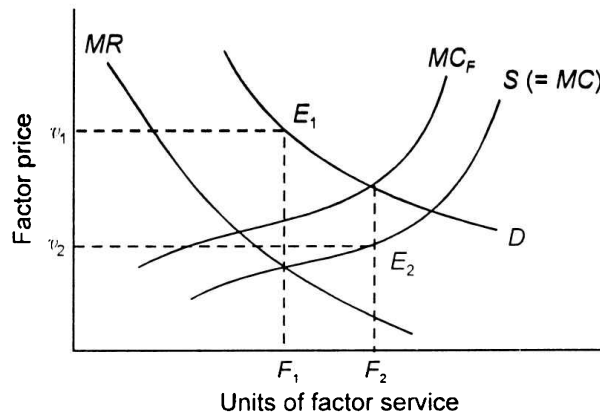


Fig. 15.5 Bilateral Monopoly

UPSTREAM AND DOWNSTREAM MONOPOLIES

Next, instead of considering variations of the models discussed above, following Varian (1993), we analyze a market structure in which a monopoly produces some product which is used as a factor of production by another monopolist. This analysis will enable us to compare the performances of integrated and non-integrated monopolists.

A monopolist produces output F at a constant marginal cost c . This upstream monopolist sells F to another monopolist, the downstream monopolist, at a price v . The downstream monopolist uses F to produce X according to the production function $X = f(F)$. The demand function for the downstream monopolist's output is assumed to be linear: $p(X) = a - bX$.

For simplicity of exposition, the production function f is assumed to be an identity mapping, that is, $X = f(F) = F$.

We begin the analysis with maximization of the downstream monopolist's profit.

$$\max_x p(X)X - vX.$$

The first order condition for maximum profit requires that marginal revenue equal marginal cost, that is,

$$a - 2bX = v,$$

which in turn gives

$$X = \frac{a - v}{2b}.$$

Since by assumption $X = F$, the factor demand function of the downstream monopolist is

$$F = \frac{a - v}{2b}. \quad (15.6)$$

Now, the upstream monopolist will choose F that will maximize his profit. From Eq. (15.6) we can express v as a function of F :

$$v = a - 2bF.$$

The marginal revenue corresponding to the above factor demand function is $(a - 4bF)$. Therefore, the upstream monopolist's profit maximization requires that $a - 4bF = c$, that is,

$$F = \frac{a - c}{4b}.$$

Thus, the total output produced in this set-up is

$$X = \frac{a - c}{4b}. \quad (15.7)$$

Now, suppose the upstream and downstream monopolies merge so that the merged monopolists' demand function is $p(X) = a - bX$ and their marginal cost is c . Equating marginal revenue with marginal cost, the profit maximizing output in this case becomes

$$X = \frac{a - c}{2b}. \quad (15.8)$$

Thus, the merged monopolist produces twice as much as the upstream-downstream monopolists.

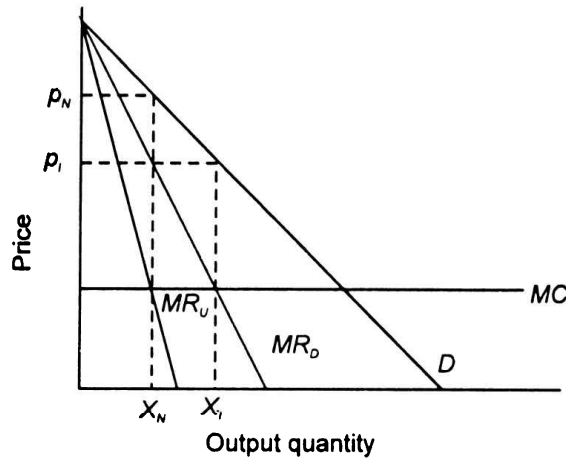


Fig. 15.6 Upstream and Downstream Monopolies

In the above figure D is the downstream monopolist's demand curve and $MR_D(MR_U)$ stands for marginal revenue of the downstream (upstream) monopolist. The integrated (non-integrated) monopoly produces $X_1(X_N)$. In the non-integrated case the upstream monopolist raises his price above the marginal cost and then the downstream monopolist raises *his* price above the already marked-up cost. Thus, there is a double mark-up in the non-integrated case. If the two firms merge, the price goes down from p_N to p_1 and profits also increase. Hence in the non-integrated case price (p_N) is too high from the point of view of aggregate profit.

Example 15.1

Suppose that the supply curve for labor to a firm has the form $L = 100w$, where w is the market wage. Suppose also that the firm's demand curve has the form $L = -50w + 450$. Assume the firm is a monopsonist. We want to determine the number of workers that the firm will hire in order to maximize profit.

From the equation $L = 100w$, the total cost of labor to the firm is $wL = \frac{L^2}{100}$.

The marginal factor cost of an extra unit of labor is $d(wL)/dL = 2L/100 = L/50$. We rewrite the firm's demand curve as $w = -L/50 + 9$. The profit maximizing labor input for the monopsonist is obtained by solving $L/50 = -L/50 + 9$. The solution is 225. At this level of labor input the wage rate w satisfies $100w = 225$, which means $w = 2.25$.

Let us now compare this to the competitive solution. In the competitive case the labor demanded is given by the intersection of the demand and the supply curves. Thus, the competitive solution is obtained from $L/100 = -L/50 + 9$, which gives $L = 300$. This in turn generates the competitive wage rate $w = L/100 = 3$. Therefore, the monopsonist hires less labor and pays less wage than would be if the labor market were perfectly competitive.

Applied Example 15.1 Monopsony in Market for Football Players

If firms as hirers of inputs can coordinate their hiring, they can behave monopsonistically, even though each firm individually has limited market power. Usually this type of coordination may be difficult since it is costly to enforce such agreements and there may exist incentives to cheat. (Here cheating takes the form of raising the factor price secretly by demanders of the factor.) However, sometimes powerful cartels may achieve successful monopsonistic positions. An example is the case of football clubs which, if they wish, can cartelize their activities (since the only job market for football players is the football market). The clubs, by suppressing wage competition, could capture for themselves some of the economic rents that would otherwise have gone to players.

Applied Example 15.2 Minimum Resource Price

Minimum resource prices (say, wage rate for laborers) can be established by the government or by organized groups of resource suppliers. The typical monopsonistic situation is depicted in panel (b) of the Fig. 15.7. The level of employment of factor F is F_1 . Its price is v_1 which is less than the marginal revenue product m_F and the resource is being exploited. (We assume that the monopsonist sells his output in a perfectly competitive market at price p . Therefore, the factor demand curve D_F is given by p times the marginal product of the factor F .)

Panel (a) shows that if the factor market is competitive, the level of employment of F is F_c and its price is v_c . Suppose under monopsony the government sets the minimum factor price at v_c . Then the monopsonist can hire workers at a constant price v_c . The firm will now use the quantity F_c of the factor F . At F_c the marginal resource cost equals the marginal revenue product of F . Thus, we see that the minimum price not only eliminates monopsonistic exploitation of the resource, it also increases the level of employment.

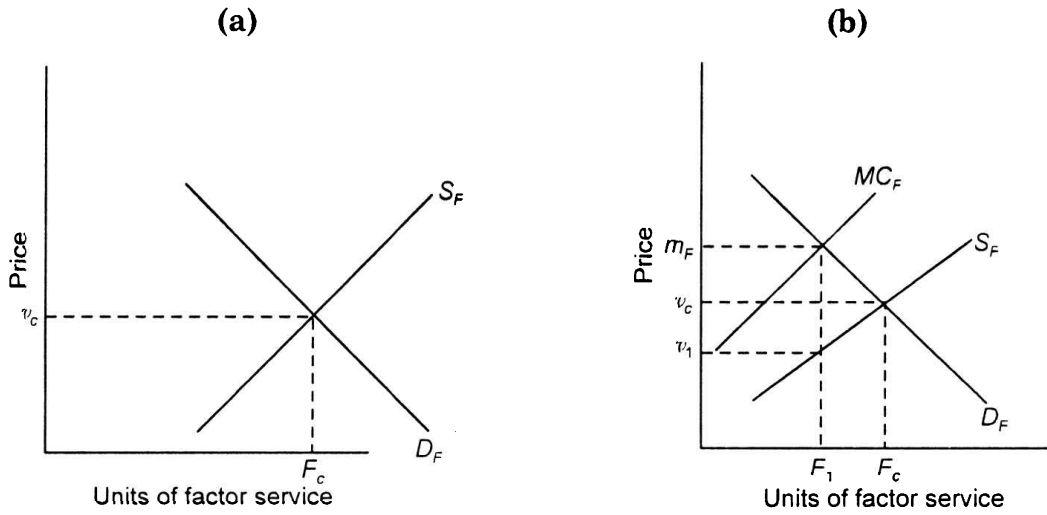


Fig. 15.7 Minimum Resource Price

Appendix

MONOPSONY DEMAND FOR A FACTOR

The profit function $\pi = ph(L) - Lw(L)$ is to be maximized with respect to L . Differentiating π with respect to L and setting the derivative equal to zero, we have

$$\frac{d\pi}{dL} = p \frac{dh(L)}{dL} - L \frac{dw(L)}{dL} - w(L) = 0.$$

Since $dh(L)/dL$ is the marginal productivity of labor, which we denote by MP_L , we can rewrite the above equation as

$$pMP_L - w(L) \left[1 + \frac{L}{w(L)} \frac{dw(L)}{dL} \right] = 0.$$

This on simplification becomes

$$pMP_L - w(L) \left(1 + \frac{1}{e_s} \right) = 0,$$

where e_s is the elasticity of labor supply. It is now possible to rearrange the above equation as

$$p.MP_L = w(L)\left(1 + \frac{1}{e_s}\right),$$

which is the desired result.

EXERCISES

Multiple Choice

1. In a monopsonistic labor market, a profit maximizing monopolist would pay a wage rate
 - a) Equal to the marginal revenue product.
 - b) Greater than the marginal revenue profit.
 - c) Less than the marginal revenue product.
 - d) Any one of the above is possible.
2. If the equilibrium wage for female workers is lower than that for male workers, even though they are equally productive, then the prohibition of wage discrimination will
 - a) Reduce the employment of females.
 - b) Increase the employment of females.
 - c) Not change the level of employment of females.
 - d) Any one of the above is possible.
3. When factor A is the only variable factor for an imperfect competitor in the product market, the firm's demand for A is determined from its
 - a) Marginal product curve.
 - b) Marginal revenue product curve.
 - c) Marginal factor cost curve.
 - d) Total product curve.
4. Suppose that the marginal revenue product of a factor A is greater than the competitive price of A . Then we have
 - a) Monopolistic exploitation.
 - b) Monopsonistic exploitation.
 - c) Both monopolistic and monopsonistic exploitation.
 - d) None of the above.

5. A monopolist hires one variable factor and the market for the variable factor is perfectly competitive. This variable factor is the only factor of production. Supposing that the marginal product of the factor is equal to its average product, then the monopolist will earn
- Positive profit.
 - Negative profit.
 - Zero profit.
 - Non-negative profit.

Quantitative Problems and Questions

- A firm uses labor (L) as the only variable factor of production. The firm is a monopsonist in the labor market. The firm's production function is $x = -L^3 + 20L^2 + 72L$, where x is the output. The product price is Re. 1.00 and the total fixed cost is Rs. 16.00. The profit maximizing input of labor is $L = 2$ and with this level of employment the firm's profit is Rs. 100.00.
 - Calculate the elasticity of the labor supply curve at the equilibrium point.
 - Derive the equation for the labor supply curve under the assumption that it is linear.
 - A minimum wage of Rs. 50.00 is now imposed in the labor market. Using the labor supply equation derived in (b), calculate the new level of employment and the new level of profit.
- A firm uses two factors of production, labor (L) and plant (P). It buys both factors in perfectly competitive markets, but sells the product as a monopolist. Its production function is $x = \sqrt{PL}$, where x is the output. The demand for the product is $x = p^{-3}$, where p is the price per unit of the product.
 - In the short run, the plant is fixed at 1 unit. The wage rate is Re. 1/3. The firm maximizes profit. Calculate:
 - The number of workers employed;
 - The elasticity of demand for labor;
 - The price of the product.
 - In the long run, the plant is a variable. The rental price of plant is Re. 1. The firm maximizes profit. Calculate:
 - The number of workers employed;
 - The elasticity of demand for labor.

3. Suppose that the supply curve for labor to a firm has the form $L = 100w$, where w is the market wage rate. Suppose also that the firm's demand curve has the form $L = -50w + 450$. Determine the possible range of wage rates and employment levels if a monopsonistic firm faced a monopolistic union. Explain your result.
4. Explain with diagrams the following concepts:
 - a) Monopolistic exploitation of a factor.
 - b) Monopsonistic exploitation of a factor.
5. What is a monopsony? Explain what is meant by the marginal factor cost curve of a monopsonist. Why does a profit maximizing monopsonist produce where marginal factor cost equals marginal revenue product?
6. Prove or disprove the following:
 - a) If a skilled worker can produce k times ($k > 1$) as much as an unskilled worker, then an employer who is a monopsonist in both the skilled and unskilled labor markets will pay the skilled worker k times as much as the unskilled worker.
 - b) A monopolist's demand curve for labor will be downward sloping even if the marginal product of labor is constant.
 - c) Suppose male and female workers are equally productive. Then a monopsonist would not employ some of each at different wages, since it would always be profitable to replace high wage workers with low wage workers.
 - d) When female productivity is the same as male productivity, opportunity cost differences explain wage differentials.
7. A firm uses labor (L) as the only factor of production to turn out a single product. It buys labor as a monopsonist and sells the product as a monopolist. The firm's production function is $x = \sqrt{L}$, where x is the output. The demand for the product is $x = p^{-3}$, where p is the price per unit. The supply curve of labor to the firm is $L = w^3$, where w is the wage rate. The firm is a profit maximizer. Calculate:
 - a) The number of workers employed;
 - b) The wage rate;
 - c) The output level of the firm;
 - d) The price of the product.
8. Suppose that a firm is a monopsonist in a labor market. Assume now that a labor union sets a fixed wage at which workers can be hired. Show that

there can be an increase in the wage rate and more workers are hired than under monopsony.

9. Suppose that a labor union faces a demand curve for labor that would prevail if the hirer acted as a competitor in the labor market. If the union desires to maximize the total wage bill, what wage rate will it demand? How would your answer change if the union desires to maximize the total amount of labor services?
10. What kinds of policies other than price fixing may be used to counteract monopsonistic exploitation of resources?
11. Consider a monopoly firm which faces competitive input markets (capital and labor). Derive the firm's labor demand curve when both capital and labor are variable.

16

Exchange

In previous chapters of this book we have studied the operations of single markets in isolation. The equilibrium of an individual market with respect to given data is called partial equilibrium. Partial equilibrium analysis becomes necessary if we wish to examine the determinants of supply and demand in a detailed manner. However, the partial equilibrium method of analysis has one major limitation: it never helps us to conceptualize how the system works as a whole. Evidently, an increase in the demand in one market will affect prices not only in that market but also in the entire economy. For instance, while an increase in the demand for vehicles will affect pricing in the vehicle market, it will also generate increased demand in factor markets (steel, glass, etc.), and this in turn affects demands in other markets for goods *those* factors use. We can extend this chain indefinitely. But the essential point is clear: the markets in an economy are interconnected.

In this chapter we study the general equilibrium approach to the economic system. The essential concept in this analysis is that of simultaneity. More precisely, general equilibrium for the entire economy refers to the simultaneous equilibrium of all individual markets. This concept of equilibrium recognizes the interdependence of all segments of the economy with one another.

There are two variants of the general equilibrium analysis. In the first one, following Walras (1874), economists discuss general equilibrium in mathematical terms. The interdependence of economic segments is shown through a system of simultaneous equations that determine the quantities and prices of all those goods and services that are traded among individuals. Just as the solutions to two unknowns depend simultaneously on both equations, so do the equilibrium prices and quantities determined by the economic system depend simultaneously on the workings of all markets. The Walrasian approach to general equilibrium provides the essential theoretical apparatus for understanding the interrelationships of the various sectors of the economy.

The second variant of general equilibrium is Leontief's (1951) input-output analysis. This approach is an empirical descendant of the Walrasian system. We relegate the discussion of the Leontief input-output system to Chapter 19.

In this chapter we discuss the Walrasian approach in a pure exchange economy. No production takes place in an exchange economy; the only economic activity is the exchange of goods, one for another, among the consumers. Once we have developed the pure exchange model explicitly, we will incorporate production in the general equilibrium set-up

THE EDGEWORTH BOX

By suppressing production, it is possible to use a simple graphic device known as the Edgeworth box, to show how a pure exchange economy with two consumers and two goods operates. The two consumers are referred to as A and B , with the two goods being 1 and 2. We denote A 's consumption bundle by (x_A^1, x_A^2) , where x_A^1 (x_A^2) represents A 's consumption of good 1(2). Similarly B 's consumption bundle is (x_B^1, x_B^2) . A pair of consumption bundles $[(x_A^1, x_A^2), (x_B^1, x_B^2)]$ is called an allocation of the two goods to A and B respectively. The initial endowments (before trade takes place) are labelled $[(w_A^1, w_A^2), (w_B^1, w_B^2)]$. This is the allocation that shows the amount of each good that consumers bring to the market for exchange. An allocation is feasible if the total amount of each good consumed equals the total amount available:

$$x_A^1 + x_B^1 = w_A^1 + w_B^1 \quad (16.1)$$

$$x_A^2 + x_B^2 = w_A^2 + w_B^2. \quad (16.2)$$

The Edgeworth box has dimensions given by the total quantities of the two goods. The horizontal dimension of the box represents the total quantity of good 1 available, whereas the vertical height of the box is the total quantity of good 2. The point O_A is considered to be the origin for A . Quantities of good 1 for A are measured along the horizontal axis rightward from O_A ; quantities of good 2, along the vertical axis upward from O_A . Any point in the box can then be regarded as some consumption bundle of A . For instance, at the point P , A gets (P_A^1, P_A^2) . The useful property of the Edgeworth box is that the quantities received by B are also recorded by P , B simply gets the amount which is left over. More precisely, we can treat B 's quantities being measured from the origin O_B . The point P corresponds to the quantities P_B^1 and P_B^2 for B . Observe that the quantities allocated to A and B in this manner exactly exhaust the total quantities of goods 1 and 2 available. Thus, any point in the Edgeworth box is a feasible allocation.

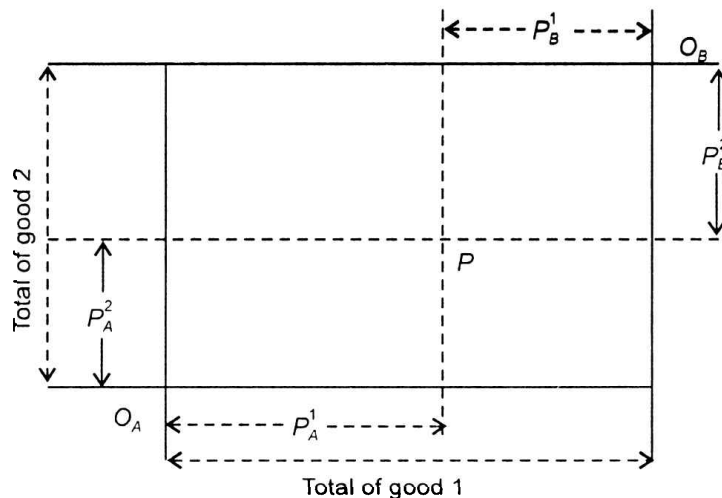


Fig. 16.1 Edgeworth Box

PARETO EFFICIENT ALLOCATIONS

Any point in the Edgeworth box represents a distribution of the available goods between A and B and all possible distributions are contained within the box. However, not all points in the box will be acceptable to consumers. That is, some distributions are inefficient. In general, a situation is inefficient if some beneficial activity can be increased without requiring a cutback in any other activities. In inefficient situations, conditions can be improved unambiguously. In contrast, if we wish to move from one efficient situation to another, reduction of at least one activity in order to increase other activities is required. To know whether or not conditions have improved, it becomes necessary to compare these activities.

In the context of an exchange economy we define efficiency as follows. An allocation of the fixed quantities of goods in an exchange economy is Pareto efficient or Pareto optimal, if through reallocation of goods no individual can be made better off without making at least one individual worse off. That is, in an efficient situation conditions cannot be made unambiguously better. The term Pareto efficiency or Pareto optimality is named after the Italian economist, engineer and sociologist Vilfredo Pareto whose contributions during 1900–1930 had a major influence on welfare economics. We will consider the notion of Pareto efficiency in greater detail in Chapter 18.

To identify the set of allocations that are efficient we must introduce tastes. Indifference curves for A and B are assumed to be of the usual type. In Fig. 16.2 A 's indifference map is drawn with origin O_A . Movements in the north-east direction indicate a higher level of utility for A . In the same figure B 's indifference map is plotted with the corner O_B as the origin. In fact, we have taken B 's indifference map, turned it upside down, and fit it into the north-east corner of the Edgeworth box. Movements in the south-west direction represent increases in B 's utility.

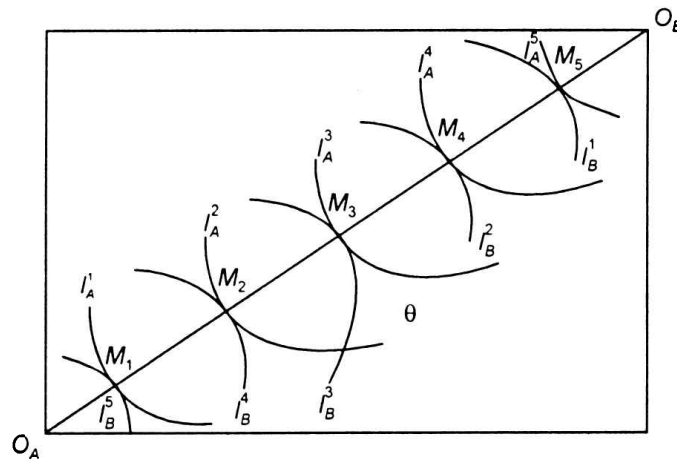


Fig. 16.2 Pareto Efficient Allocations

Using these superimposed indifference curves it is possible to find the efficient points. Consider any fixed utility level for A , say, U_A^4 , that corresponds to the indifference curve I_A^4 . The definition of efficiency implies that B 's utility be maximized for this given level of A 's utility—not to do so would mean inefficiency. The point of maximum utility for B in this case is the point where B 's indifference curve (I_B^2) is just tangent to the curve I_A^4 . If the two indifference curves are not tangent at an allocation inside the Edgeworth box, then they must cross. In such a case we can improve B 's utility without worsening A 's position. (Note that by assumption A 's utility is fixed at U_A^4 .)

However, if we have corner solutions, that is, if Pareto efficient allocations appear on the sides of the box then the tangency condition does not hold. (The corner solutions arise if some individuals choose not to consume all goods.) We ignore the corner solutions here, since they are not important for the current discussion.

Now, the choice of the utility level U_A^4 was quite arbitrary. In fact, given any indifference curve for A , we can find the best point for B . By definition this is a Pareto efficient point and therefore the two indifference curves must be tangent at this point. We label a few of these points (M_1, M_2, M_3, M_4 and M_5) in Fig. 16.2. Note also that for any specific level of B 's utility, A 's utility is as large as possible at these points. The locus of all these efficient allocations is called the contract curve. It is so called because it is the locus of all Pareto optimal contracts. In the figure the contract curve is given by the line running from O_A to O_B . At O_A , A has nothing and B has everything. This point is Pareto efficient because if we wish to make A better off we must make B worse off. In fact, this is true for every allocation on the contract curve. As we move up along the contract curve, A gets more and more and B gets less and less of each good. The utilities of A and B are directly in conflict on the contract curve. This is not true for points off this curve. For instance, the allocation θ which is off the contract curve is clearly inefficient in the sense that both persons can be made better off by moving to a point on the contract curve. At the point M_3 both A and B are better off than they were at θ . Therefore, θ is an inefficient allocation. It may be interesting to note that the contract curve does not depend upon the position of the initial endowment. However, since the total amounts of both goods determines the dimensions of the Edgeworth box, the spread of the curve depends on the initial endowment.

GAINS FROM EXCHANGE

In an exchange economy the individuals, through negotiations, try to determine some mutually advantageous trade. Clearly, no individual would engage in a trade that would leave him worse off than would be the case without trading. That is, the point determined through exchange should be preferred by each consumer to his initial endowment. Hence only a subset of the Pareto efficient set can be regarded as a set of efficient allocations that might result from exchange.

To illustrate this idea more explicitly, let us again consider the Edgeworth box of A and B . As before, the dimensions of the box are given by the total

quantities of goods 1 and 2. The initial endowments of A and B are represented by E . The contract curve of Pareto efficient allocations is given by the line $O_A O_B$. We denote the indifference curves of A and B passing through E by I_A^E and I_B^E respectively. Note that at E , the individuals' indifference curves are not tangent, therefore E is not an efficient allocation. Neither A nor B will accept trading outcomes that generate a utility level less than U_A^E or U_B^E , where U_A^E and U_B^E correspond respectively to I_A^E and I_B^E . It would be preferable for a person to abstain from trading rather than accept such an inferior outcome. Therefore, only those efficient outcomes between M_1 and M_2 on the contract curve can emerge as possible outcomes of mutual trade starting from the initial endowment shown in Fig. 16.3. Movement from the point E to the line segment $M_1 M_2$ is an unambiguous improvement for both A and B . If the initial distribution of goods favors A , any final allocation will also favor A , since A will refuse any allocation that provides less utility. However, it is clear that both individuals can benefit from trading.

We interpret the segment $M_1 M_2$ from a somewhat different perspective, as follows.

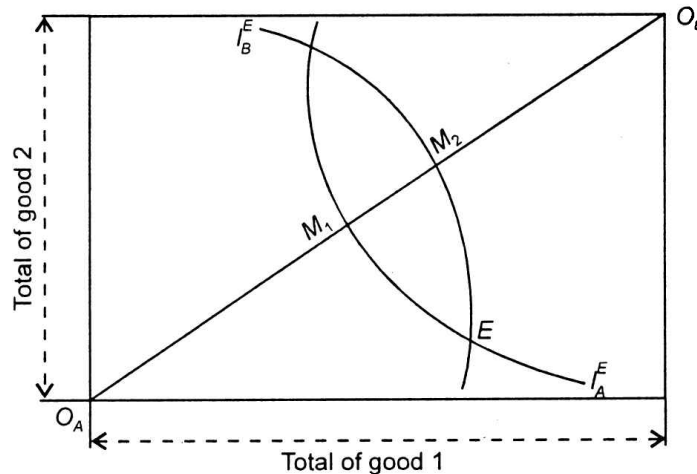


Fig. 16.3 Exchange with Initial Endowments

CORE OF AN EXCHANGE ECONOMY

In this approach emphasis is placed on the bargains that individuals in an exchange economy might make and whether these bargains are acceptable in the sense that the parties involved will not have an incentive to block them. The story goes as follows. Each trader starts with an initial endowment. A new distribution of goods is proposed by someone. The traders discuss the proposal among themselves. If some group of traders finds that it can do better on its own with its given initial endowments, it objects to the

proposal, and the proposal is blocked. Then a new proposal regarding allocation of goods is made. The process continues until a distribution of goods is found to which no group of traders objects. The set of all such allocations is called the core of an exchange economy. More precisely, the core of an exchange economy is a set of all those allocations which cannot be blocked by any coalition of traders.

We can express the idea of core allocations using simple algebra. Let us suppose that in the exchange economy under consideration there are m people and n different goods. We let x_i^j be person i 's quantity of good j and x_i be his bundle or vector of goods. Thus, $x_i = (x_i^1, \dots, x_i^n)$. Suppose also that w_i^j stands for person i 's starting or initial quantity of good j , and w_i stands for his initial endowment bundle.

Consider a group of traders S . S may contain only one, several or all traders. Assume without loss of generality that $S = \{1, 2, \dots, k\}$, $1 \leq k \leq m$. Let $(s_1, s_2, \dots, s_k)$ be a vector of bundles of goods, with bundle s_i being for the i th member of S . We say that $(s_1, s_2, \dots, s_k)$ is feasible for S if for each j , $\sum_{i=1}^k s_i^j = \sum_{i=1}^k w_i^j$. That is a vector of bundles is feasible for the coalition S if the total of any good in those bundles is the same as the total of the corresponding good the members of S initially possess. Note that this definition of feasibility is consistent with our definition presented in Eqs. 16.1 and 16.2.

If $x = (x_1, \dots, x_k, x_{k+1}, \dots, x_m)$ is an allocation and $S = (1, 2, \dots, k)$ is a coalition, we say that S blocks x if there is a feasible vector of bundles $(s_1, s_2, \dots, s_k)$ such that

$$U_i(s_i) \geq U_i(x_i) \text{ for all } i \text{ in } S, \quad U_i(s_i) > U_i(x_i) \text{ for at least one } i \text{ in } S, \quad (16.3)$$

where U_i is person i 's utility function. An allocation is in the core if no coalition, including the whole set of traders, blocks it. An allocation is Pareto efficient if the whole set of traders does not block it.

It should be clear that all core allocations are Pareto efficient. With inefficient allocations there is always the possibility that at least one person can be made better off without making anybody worse off. Therefore, the persons who can be made better off have always an incentive to seek the efficient alternative. Hence inefficient allocations cannot be in the core.

However, not all efficient allocations are in the core. We have already stated that an allocation will not be acceptable to a trader if it provides a lower utility level than is provided by his initial endowment. In Fig. 16.3 only the allocations on the segment M_1M_2 might result from such a constrained trade.

In the context of the core, A will block any allocation on the contract curve south-west of M_1 whereas B will block allocations north-east of M_2 . Hence in a two-person exchange economy illustrated in Fig. 16.3, the core consists of allocations along the curve M_1M_2 . With more individuals, the core may be even smaller than that.

MARKET TRADE AND COMPETITIVE EQUILIBRIUM

It is difficult to predict exactly which particular allocation in the core (the segment M_1M_2 in Fig. 16.3) will be chosen by the traders in an exchange economy. The reason behind this is that the trading process we have described is quite general. If we have some specific exchange program we can have a more precise idea about the equilibrium outcome. We describe an economic exchange, exchange governed by prices, that characterizes equilibrium in a competitive economy, as follows.

Each trader starts with his initial endowment. There exists a market in which people buy and sell according to given prices. Someone in the market, may be an auctioneer, announces the prices of the goods. Each trader, on the basis of the announced prices, decides how much of each good to buy and sell. The things a trader buys are paid for by the things he decides to sell. Therefore, the value of different goods bought must be equal to the value of the goods sold, where both the values are calculated at the announced prices.

Since in an exchange economy there is no production, what people can buy is limited by what is actually available in the economy. That is, for each good, total supply must equal total demand, in order to exist an equilibrium. Now, the first set of announced prices may not yield an equilibrium. In such a case we say that the market is in disequilibrium. The auctioneer then changes the prices. If for some good aggregate demand exceeds aggregate supply, then its price is increased. On the other hand, the price of a good is lowered if its aggregate demand falls short of its aggregate supply. This adjustment process continues until the total demand for each good equals the total supply. When this list of prices is obtained, transactions are actually made and the equilibrium allocation is established.

This is the price-governed exchange economy model introduced by Walras (1874). In equilibrium in this perfectly competitive exchange economy model, each individual buys the utility maximizing bundle and all choices are consistent in the sense that equality must hold between total demand and total supply for each good.

We can now formally define a competitive or Walrasian equilibrium. In a perfectly competitive pure exchange economy, each consumer treats the prices as given. The budget constraint for A is determined by his initial endowment (w_A^1, w_A^2) ; he can choose any bundle (x_A^1, x_A^2) that costs no more than the value of his initial bundle. Hence the budget constraint for A becomes

$$p_1 x_A^1 + p_2 x_A^2 \leq p_1 w_A^1 + p_2 w_A^2, \quad (16.4)$$

where p_1 and p_2 are the prices of goods 1 and 2 respectively. Similarly, B 's budget constraint becomes

$$p_1 x_B^1 + p_2 x_B^2 \leq p_1 w_B^1 + p_2 w_B^2. \quad (16.5)$$

With more of the goods preferred to less, the utility maximizing bundles of A and B will satisfy the budget constraints as equalities.

A market equilibrium, competitive equilibrium or Walrasian equilibrium occurs in the pure exchange economy at the prices $\bar{p}_1, \bar{p}_2$ and the final holdings $[(\bar{x}_A^1(\bar{p}_1, \bar{p}_2), \bar{x}_A^2(\bar{p}_1, \bar{p}_2), \bar{x}_B^1(\bar{p}_1, \bar{p}_2), \bar{x}_B^2(\bar{p}_1, \bar{p}_2)]$ under the following conditions:

1. $[\bar{x}_A^1(\bar{p}_1, \bar{p}_2), \bar{x}_A^2(\bar{p}_1, \bar{p}_2)]$ maximizes utility for A over bundles satisfying his budget constraint.
2. $[\bar{x}_B^1(\bar{p}_1, \bar{p}_2), \bar{x}_B^2(\bar{p}_1, \bar{p}_2)]$ maximizes utility for B over bundles satisfying his budget constraint.
3. The markets clear, that is, the total demand of each good should be equal to the total supply,

$$\bar{x}_A^1(\bar{p}_1, \bar{p}_2) + \bar{x}_B^1(\bar{p}_1, \bar{p}_2) = w_A^1 + w_B^1. \quad (16.6)$$

$$\bar{x}_A^2(\bar{p}_1, \bar{p}_2) + \bar{x}_B^2(\bar{p}_1, \bar{p}_2) = w_A^2 + w_B^2. \quad (16.7)$$

We can rewrite the equations stated in condition 3 as follows:

$$\begin{aligned} [\bar{x}_A^1(\bar{p}_1, \bar{p}_2) - w_A^1] + [\bar{x}_B^1(\bar{p}_1, \bar{p}_2) - w_B^1] &= 0, \\ [\bar{x}_A^2(\bar{p}_1, \bar{p}_2) - w_A^2] + [\bar{x}_B^2(\bar{p}_1, \bar{p}_2) - w_B^2] &= 0. \end{aligned} \quad (16.8)$$

The first third-bracketed term on the left hand side of the first equation in (16.8) is the difference between the quantity demanded of good 1 by

consumer A and his initial endowment of the good. This difference is called A 's net or excess demand for good 1. Therefore, Eqs. (16.8) say that for each good the net amount A wishes to demand (supply) must be equal to the net amount B wishes to supply (demand).

We can also state the equilibrium condition in terms of aggregate excess demand. The net or excess demand $x_A^1(p_1, p_2) - w_A^1$ for good 1 by consumer A at prices (p_1, p_2) is denoted by $e_A^1(p_1, p_2)$. We define $e_B^1(p_1, p_2)$ in a similar manner and write $Z_1(p_1, p_2)$ for $e_A^1(p_1, p_2) + e_B^1(p_1, p_2)$. $Z_1(p_1, p_2)$ is the aggregate excess demand for good 1 at prices (p_1, p_2) . Similarly we denote aggregate excess demand for good 2 at prices (p_1, p_2) by $Z_2(p_1, p_2)$. We can then say that $(\bar{p}_1, \bar{p}_2)$ is an equilibrium price vector if the aggregate excess demand for each good is zero at $(\bar{p}_1, \bar{p}_2)$:

$$\begin{aligned} Z_1(\bar{p}_1, \bar{p}_2) &= 0, \\ Z_2(\bar{p}_1, \bar{p}_2) &= 0. \end{aligned} \quad (16.9)$$

Before we investigate how the competitive equilibrium looks in a two-person economy and the question whether the competitive equilibrium is a core allocation, let us establish the Walras law—a law concerning the aggregate value of excess demands.

Walras' Law: According to the Walras law at any set of prices, the aggregate value of excess demands is zero. In terms of the notation introduced above, this law becomes

$$p_1 Z_1(p_1, p_2) + p_2 Z_2(p_1, p_2) = 0. \quad (16.10)$$

We can demonstrate this law using the budget constraints of A and B . Given any prices p_1 and p_2 , with more preferred to less, the budget constraints of A and B will be satisfied as equalities. That is, each consumer spends the entire income. Thus,

$$p_1 x_A^1(p_1, p_2) + p_2 x_A^2(p_1, p_2) = p_1 w_A^1 + p_2 w_A^2 \quad (16.11)$$

$$\text{and} \quad p_1 x_B^1(p_1, p_2) + p_2 x_B^2(p_1, p_2) = p_1 w_B^1 + p_2 w_B^2. \quad (16.12)$$

Adding the two budget constraints, we obtain the Walras law

$$p_1 Z_1(p_1, p_2) + p_2 Z_2(p_1, p_2) = 0. \quad (16.13)$$

Walras' law, it should be stressed, holds for any set of prices, not just for equilibrium prices. This law asserts that if demand equals supply in one

market, then demand must equal supply in other market also. That is, if excess demand in one market is zero, it must be zero in other market as well. In general, if there are markets for n goods, and if $(n - 1)$ of the markets are in equilibrium, then the market for good n will also be in equilibrium. Similarly, if there are markets with positive excess demands, then there must exist markets with negative excess demands.

One implication of the Walras law is that equilibrium conditions in n markets are not independent, (since if we have equilibrium in all markets but one, then the last market is also in equilibrium). That is, we do not have n independent equations in n unknowns (the prices). Rather, we have $(n - 1)$ independent equations and hope to determine $(n - 1)$ of the prices. But this is expected from the homogeneity of degree zero of excess demand functions. To see the homogeneity property, consider the two-person exchange economy with two goods. Note that if we multiply prices p_1, p_2 by any positive constant, we do not change the budget constraints of A and B in Eqs. (16.4) and (16.5) respectively. Hence A and B choose the same utility maximizing bundles under any positive multiples of p_1, p_2 as they do under p_1, p_2 . (Note that preference maps of the consumers do not depend on prices.) Since initial endowments are exogenously given constants, the excess demand functions are homogeneous of degree zero as well. That is, gross as well as excess demand functions depend on the relative price p_2/p_1 .

Now, in our two-good economy if $(\bar{p}_1, \bar{p}_2)$ is an equilibrium set of prices, then for any $k > 0$, $(k\bar{p}_1, k\bar{p}_2)$ are also equilibrium prices. Therefore, absolute price levels are indeterminate. Only the relative price p_2/p_1 is determinable. In the general case we can hope to determine only $(n - 1)$ equilibrium relative prices. We can conveniently set one of the prices equal to 1 so that all other prices can be regarded as being measured relative to it. Such a price is called a numeraire price. Thus, in the two-commodity pure exchange world, final holdings are determined by consumers' preferences, initial endowments and the relative price p_2/p_1 which becomes p_2 if the first price p_1 is taken as the numeraire price.

EXISTENCE OF COMPETITIVE EQUILIBRIUM

The problem we wish to investigate here is whether the markets will be able to generate prices such that demand equals supply in every market. This is the question of the existence of a competitive equilibrium.

The question of the existence of general equilibrium prices was first investigated rigorously by Walras in the late nineteenth century. He tried to

establish that for the n -good case the n equilibrium conditions were sufficient to ensure that such a set of prices would in fact exist, and hence the exchange model had a consistent theoretical framework. One may think that the existence of equilibrium prices might be assured by a simple counting of the equations and the unknowns. The market equilibrium conditions are described by $(n - 1)$ independent equations in $(n - 1)$ relative prices. Hence the elementary algebra of solving simultaneous linear equations suggests that an equilibrium might exist.

Unfortunately Walras recognized that the problem of solving for equilibrium prices is not so simple. First, the equations are not necessarily linear. Consequently, the conditions for existence of solutions to simultaneous linear equations are not relevant. Second, we require non-negativity of all prices. To avoid these two difficulties, Walras developed a very difficult proof, which requires solving for equilibrium prices in a series of successive approximations.

More recent works by Kenneth J. Arrow (1951), Gerard Debreu (1959) and Lionel W. McKenzie (1959) and others utilize relatively simple tools of advanced mathematics to prove the existence of equilibrium prices in an elegant way. The crucial assumption that ensures existence of competitive equilibrium is the continuity of the aggregate excess demand function in prices. Continuity means that small changes in one or more prices generate a small change in aggregate demand. Thus, aggregate demand will not be over-sensitive to minor changes in prices. One sufficient condition to guarantee continuity of aggregate demand is that individual demands are continuous in prices. Another sufficient condition is that all consumers are small relative to the size of the market, which is quite a reasonable assumption for a competitive economy. This latter requirement is quite general in the sense that even if consumers have discontinuous demand behavior, the aggregate excess demand will be continuous under this condition.

Assuming that the competitive equilibrium exists, we can now see how it looks graphically in a two-person exchange economy. Note that the budget line of A must pass through the initial endowment, and this holds for B 's budget line also. This has to be true because at any price level one alternative available to the consumer is that he can decide not to get involved in trade. Since both the consumers face the same prices p_1, p_2 , the slopes of the two budget lines are the same. We show the budget constraints of the two consumers A and B in Fig. 16.4, where $m_A = p_1 w_A^1 + p_2 w_A^2$ and $m_B = p_1 w_B^1 + p_2 w_B^2$.

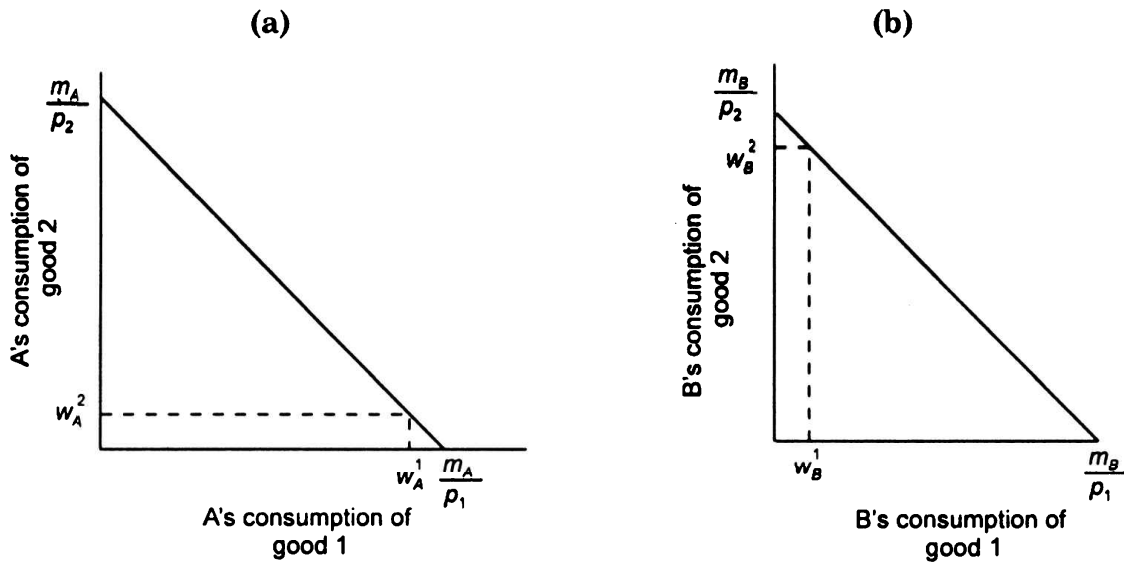


Fig. 16.4 Budget Constraints

Let us now transfer the budget lines of A and B to the Edgeworth box (Fig. 16.5). By turning the Edgeworth box upside down, we see that the budget line of A coincides with that of B . The initial endowments are represented by E . At C a competitive equilibrium occurs, because at this point both A and B maximize utility subject to their budget constraints. The markets are cleared, because A gives up $(w_A^1 - \bar{x}_A^1)$ units of good 1 to obtain $(\bar{x}_A^2 - w_A^2)$ units of good 2, while B gives up $(w_B^2 - \bar{x}_B^2)$ units of good 2 to obtain $(\bar{x}_B^1 - w_B^1)$ units of good 1.

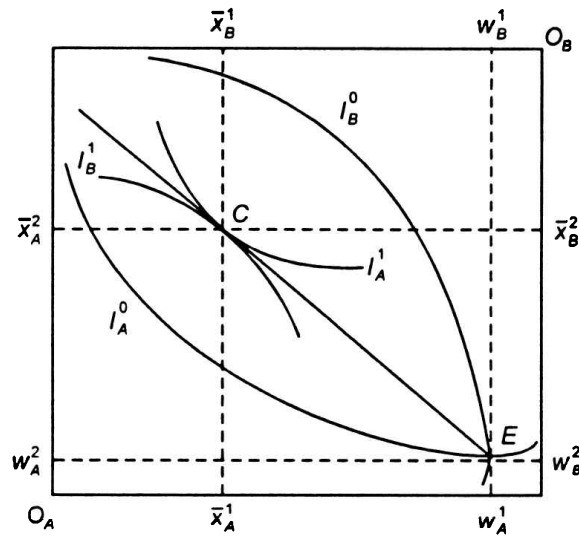


Fig. 16.5 Competitive Equilibrium in Exchange Economy

To have equality between aggregate demand and aggregate supply for each good, the final allocations (competitive equilibrium) must occur where the indifference curves of A and B are tangent to the common budget line at a common point. That is, the marginal rate of substitution, the rate of trade-off, between goods 1 and 2 should be the same for both A and B at a competitive equilibrium. It should be clear that this common rate of trade-off is $\bar{p}_1/\bar{p}_2$, the ratio of price of good 1 to the price of good 2. (Note that the slope of the common budget line is— $\bar{p}_1/\bar{p}_2$.)

Finally, observe that at the competitive equilibrium each individual is better off than at his initial endowment. In a competitive equilibrium exchanges are voluntary: no one will engage in them unless he becomes better off through exchange.

FIRST FUNDAMENTAL THEOREM OF WELFARE ECONOMICS

We will now investigate whether the competitive mechanism in a pure exchange economy distributes good among people in an optimal way. That is, we wish to answer the following question: After people have traded to a competitive equilibrium, will there be any more trades that they will desire to carry out? This is just a different way of asking the question whether a competitive equilibrium is a core (hence Pareto efficient) allocation.

The answer is provided by the first fundamental theorem of welfare economics—the most basic theorem of welfare economics.

FIRST FUNDAMENTAL THEOREM OF WELFARE ECONOMICS: *A competitive equilibrium allocation is in the core (and is, therefore, Pareto efficient as well.)*

Proof: Suppose that there were a coalition which chose to break away from a competitive equilibrium under the belief that it could do better by trading within itself. Then, for each individual in the coalition, the value (measured at the competitive equilibrium prices) of the good received through such bargaining must not be lower than the value of that individual's initial endowment, with the restriction that for at least one individual the value of good must exceed the value of the endowment. If not, no side-bargaining would be necessary and the new allocation could be obtained through the market. Therefore, the aggregate value of the new allocation for all the members of the coalition must exceed the aggregate value of the members' endowments. But this is impossible, since, we know that the coalition's aggregate final allocation must be identical to its aggregate endowments, so

that the aggregate values also are the same for the members of the coalition. Hence this is a contradiction, and no such coalition can exist. Therefore, a competitive equilibrium must be in the core and cannot be blocked by any side coalition. (An algebraic proof of the theorem is presented in Appendix 16.1.)

In general it is not possible to determine whether or not an allocation is in the core through an examination of all possible blocking coalitions. The First Fundamental Theorem shows us one way of arriving at a core allocation: use the competitive mechanism; use the free market. This theorem shows that a competitive market will exhaust all of the gains from trade. Note that competitive markets also economize on information—the consumers are required to know only the prices of the goods. Each consumer can determine his demand if he knows the prices, and if the market functions to determine the prices, we are guaranteed an efficient outcome. This is another point in favor of competitive mechanisms. However, the distribution of goods brought about by competitive mechanisms depends heavily on the initial allocation. For example, if a person owned everything initially, he would own everything after trade. This would be efficient, but would not probably be fair. However, efficiency has its own merits.

It may be noted that the theorem relies on several assumptions. First, the agents only care about their own consumption of goods, and not about what other agents consume. That is, consumption externality is absent. Next the people behave competitively—they are price takers. Finally, we need to ensure that the aggregate excess demand functions are continuous. This is necessary for the existence of a competitive equilibrium.

OFFER CURVES AND EFFICIENCY IN EXCHANGE

A graphic proof of the efficiency in an exchange economy can be presented for the two-person (say, A and B) two-good (say, 1 and 2) case. For this let us consider the price offer curves of the consumers. A price offer curve of a consumer represents his optimal choices at different prices. That is, this curve shows the quantities that the consumer will buy at different prices. If we draw a budget line for the consumer, then the point of intersection of the budget line and the price offer curve gives his optimal consumption.

To construct the price offer curve of, say, A , consider the set of his indifference curves shown in the Fig. 16.6. Suppose that the initial endowment (w_A^1, w_A^2) of A is denoted by P . If the market quotes a price ratio p_1/p_2 to the

individuals, then the trades that can be made from the initial endowment are along a straight line with slope $-p_1/p_2$ passing through P . This follows from the fact that A can afford any combination of goods 1 and 2 such that $p_1 x_A^1 + p_2 x_A^2 = p_1 w_A^1 + p_2 w_A^2$. Hence the budget constraint which passes through (w_A^1, w_A^2) has the slope $-p_1/p_2$. Given this budget constraint, consumer A 's choice is Q , where the tangency conditions for maximum utility are satisfied. To reach this point by trading in the market, the person will trade some of his initial endowment of good 1, $w_A^1 - Q_A^1$, for an additional amount of good 2, $Q_A^2 - w_A^2$. The ratio in which these two goods are traded is given by the ratio of the prices of the goods.

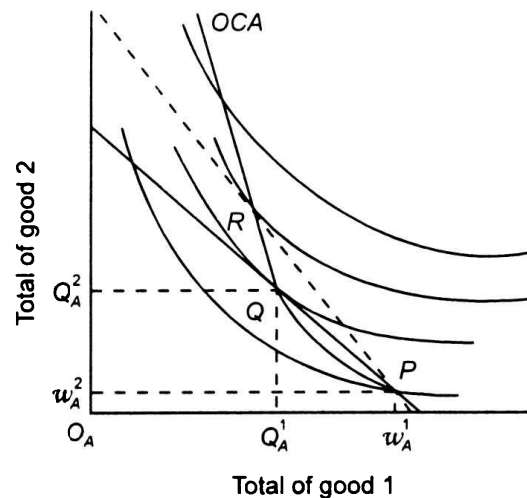


Fig. 16.6 Consumer's Offer Curve in Exchange Economy

The consumer's price offer curve can now be constructed as the locus of trade points, such as Q , for all possible price ratios. For example, an alternative budget constraint arising from a different price ratio is represented by the broken line, and the utility maximizing point here is R . The locus of points such as P , Q and R is A 's price offer curve. This curve is labelled OCA in the figure. Note that since there is some price ratio at which one choice of the consumer is to abstain from trade, the initial endowment P is included in his price offer curve. Note also that the price offer curve never lies below the indifference curve passing through P .

By using price offer curves for the two individuals A and B in an Edgeworth box, it is possible to demonstrate how a competitive price system leads to efficiency in exchange. In Fig. 16.7 the indifference curves of A and B are drawn with their price offer curves (OCA and OCB). These curves record the responses of A and B to all possible price ratios. At the points of intersection

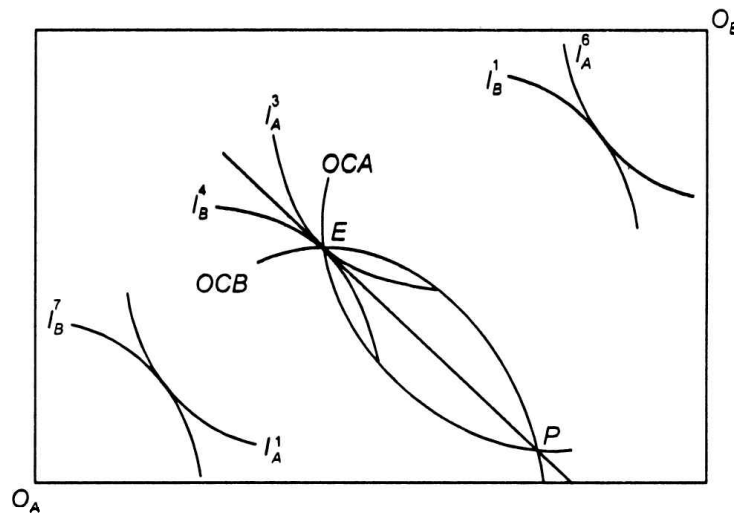


Fig. 16.7 Efficiency in Exchange Economy

of OCA and OCB , the demands of A and B exhaust the total quantities of the two goods available. This is clearly shown by the initial endowment point P .

If the price offer curves intersect at a point other than P , say, E , it can be shown that this point is on the contract curve and hence is an efficient allocation. Consider the straight line joining P and E . The slope of this line ($-\bar{p}_1/\bar{p}_2$) represents an equilibrium price ratio for the following reason. At E , PE is tangent to A 's indifference curve I_A^3 , since this is the way the price offer curve was constructed. Similarly, PE is also tangent to an indifference curve of $B(I_B^4)$ at E . Hence A and B 's indifference curves I_A^3 and I_B^4 must be tangent to each other at E . This point of tangency will be on the contract curve and therefore represents an efficient allocation of resources between the individuals. The market's generation of equilibrium prices has ensured that for each good what is demanded will in fact be supplied in aggregate and that the resultant trade is Pareto optimal. Therefore, if individuals take prices as given by the market, an equilibrium set of prices can generate an efficient allocation of goods in an exchange economy.

SECOND FUNDAMENTAL THEOREM OF WELFARE ECONOMICS

The First Fundamental Theorem says that a competitive market brings about a distribution of goods that is desirable in the sense that no group of traders can improve its position on its own. Further, it brings about the desirable distribution in automatic ways: prices tend to rise in response to excess demand and tend to fall in response to excess supply; the prices tend to adjust by themselves to solve the allocation problem.

However, the ideal distribution chosen by the price system on the contract curve is necessarily related to the initial endowments with which individuals start. Different initial endowments would result in different points on the contract curve. If the initial endowment is very unequal, so will the competitive allocation be. (For instance, as stated earlier, the extremely unequal point O_A [or O_B] on the contract can be a competitive equilibrium output.) This is an important objection to complete reliance on the competitive mechanism: it may generate very high inequalities. Therefore, the efficient allocation achieved in a competitive economy may not be optimal from a social welfare point of view.

The obvious solution to the inequality problem involves the direct transfer of quantities of various goods from one person to another. A government desiring to achieve some particular Pareto optimal allocation on equity grounds, by means of a price system, would have to ensure that the initial endowments permit the desired point to be achieved. This might necessitate some reallocation of initial endowments prior to the start of trading. That is, although the price system can generate an efficient allocation of goods, it may need some help to yield a specific allocation that is regarded as socially preferable. This discussion brings us to the second fundamental theorem of welfare economics, which is formally stated as follows.

SECOND FUNDAMENTAL THEOREM OF WELFARE ECONOMICS: *If all individuals have indifference curves that are strictly convex to the origin, then any Pareto optimal state of the economy that assigns positive quantities of every good to every person can be achieved as a competitive equilibrium for an appropriate reallocation of initial endowments.*

Proof: We demonstrate the theorem by geometrical arguments. Note that any Pareto optimal state of the economy in the interior of the Edgeworth box diagram occurs at a tangency of the indifference curves of A and B . Because the indifference curves are tangent at the Pareto optimal state, a price line can be drawn through the Pareto optimal state such that each consumer, in maximizing utility, chooses to stay at that state. (Given that the indifference curves of each consumer are strictly convex to the origin, the price line will be tangent to the indifference curves of the consumers at a unique point.) The slope of this line gives us the relative price. We can therefore say that any initial endowment that puts the two consumers on this line will make the given Pareto optimal state a competitive equilibrium allocation. The endowment that puts the consumers on this line can be obtained from the original endowment through some appropriate redistribution. This completes the proof of the theorem.

The Second Fundamental Theorem, which asserts that under certain conditions a particular Pareto optimal state can be achieved as a competitive equilibrium, introduces a new dimension into our discussion of the performance of an economic system. This theorem plays important roles in evaluating policy proposals to achieve objectives that are desired on equity grounds.

Direct interference with the market mechanism by using taxes or subsidies on specific goods to distribute income among individuals more equitably involves a loss of efficiency in the Pareto sense, that is, the society will end up in a non-Pareto optimal state. In this case loss of efficiency arises because the taxes depend on the consumer's choices. More precisely, consumer's marginal choices get directly affected by taxes. Taxes of this kind are distortionary taxes.

However, as long as the taxes are based on the value of the consumer's endowment of goods there will be no loss of efficiency. Here one consumer is taxed on the basis of the value of his endowment and this money is transferred to another. That is, the purchasing powers of endowments get redistributed. It is evident that a tax on endowments will change consumers' behavior. But the first fundamental theorem states that from the initial endowment, the free exchange contract achieved by bargaining will result in a Pareto optimal allocation. This means that no matter how a policy maker decides to redistribute the endowments among the consumers, the equilibrium allocations, as determined by market forces, will be Pareto efficient. Thus, any desired equity goals can be achieved through redistribution of wealth, without sacrificing efficiency. Hence we can separate equity considerations from efficiency considerations in evaluating policy proposals.

The society can implement any desired standard of equity by redistributive lump sum taxes on the wealth holdings of different individuals. (In a lump sum tax program the government takes away some fixed amount of money, irrespective of the individual's behavior. A lump sum tax can be incentive preserving—it need not destroy peoples' incentive to earn higher incomes.) The market forces then ensure that the pattern of consumption associated with the distribution of wealth will be Pareto efficient. There is no need to achieve equity at the cost of efficiency.

Finally, we make a discussion on the assumption that consumers' indifference curves should be strictly convex to the origin for the Second Fundamental Theorem to hold. In Fig. 16.8, *A*'s indifference curves are strictly convex to the origin, but *B*'s indifference curves do not satisfy this assumption. In this figure *P* is a Pareto optimal allocation. But there do not exist prices at which *A* and *B* will attempt to consume at *P*. *A* will choose the bundle *P*

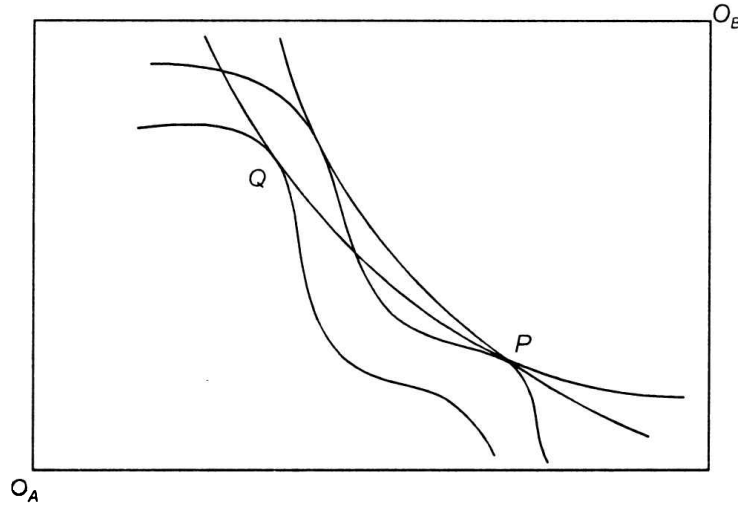


Fig. 16.8 Non-Convexity of Indifference Curves to the Origin and Pareto Optimality

while for B the obvious choice is Q . At the corresponding prices, demand is not equal to supply. This therefore shows the requirement for strict convexity of indifference curves to the origin.

Example 16.1

There are two consumers A and B with the following utility functions and endowments:

$$\begin{aligned} U_A(x_A) &= \frac{1}{3} \log x_A^1 + \frac{2}{3} \log x_A^2, \\ U_B(x_B) &= \frac{1}{2} \log x_B^1 + \frac{1}{2} \log x_B^2, \\ w_A &= (1, 0) \text{ and } w_B = (0, 1). \end{aligned} \tag{16.14}$$

We want to determine the competitive equilibrium allocation and market clearing prices.

Person A 's marginal rate of substitution of good 2 for good 1 is

$$\frac{\delta U_A / \delta x_A^1}{\delta U_A / \delta x_A^2} = \frac{(1/3)x_A^2}{(2/3)x_A^1} = \frac{x_A^2}{2x_A^1} \tag{16.15}$$

Person A 's budget line is $p_1 x_A^1 + p_2 x_A^2 = p_1 \cdot 1 = p_1$, which can be rewritten as $x_A^1 + \frac{p_2}{p_1} x_A^2 = 1$, where p_1 and p_2 are the prices of goods 1 and 2. At the

utility maximizing choice we have $x_A^2/2x_A^1 = p_1/p_2$. Substituting the value of p_2/p_1 from this equation in the budget line, we have $3x_A^1 = 1$, from which it follows that $x_A^1 = 1/3$. This is person A 's demand function for good 1. Substituting $x_A^1 = 1/3$ in the budget line we have $x_A^2 = \frac{2p_1}{3p_2}$. Similarly, we have $x_B^1 = \frac{p_2}{2p_1}$ and $x_B^2 = 1/2$.

At the equilibrium price total demand for each good must equal total supply. The total demand for good 1 is $1/3 + p_2/2p_1$. Since the total supply of good 1 is one, we have $1/3 + p_2/2p_1 = 1$ from which it turns out that $p_1/p_2 = 3/4$. In view of Walras' law for determining the price ratio it is sufficient to consider only one market where aggregate demand equals aggregate supply. Therefore, the market clearing price ratio is $3/4$. We can in fact check that at $p_1/p_2 = 3/4$; the value of aggregate excess demand is zero. Putting $p_1/p_2 = 3/4$ in demand functions, the competitive equilibrium becomes $x_A = (x_A^1, x_A^2) = (1/3, 1/2)$, $x_B = (x_B^1, x_B^2) = (2/3, 1/2)$. Finally, note that in equilibrium only relative price p_1/p_2 is determined.

Applied Example 16.1 Gains from Trade

In Chapter 21 on International Trade, Applied Example 21.1 shows how there can be potential gains to all parties from free trade. Since free trade between two nations is analogous to free exchange between two persons, applied example 21.1 (Chapter 21) can be regarded as an applied example of market exchange and optimality.

Appendix 16.1

PROOF OF THE FIRST FUNDAMENTAL THEOREM

We will prove the First Fundamental Theorem of welfare economics for the general case of m consumers and n goods. We write w_i^j for person i 's initial quantity of good j and $w_i = (w_i^1, \dots, w_i^n)$ for his initial endowment bundle.

Similarly, $x_i = (x_i^1, \dots, x_i^n)$ is person i 's bundle of different goods. An allocation is a vector of the type $x = (x_1, \dots, x_m)$.

Let y be a competitive equilibrium allocation. Denote the competitive equilibrium prices by $p = (p_1, \dots, p_n)$. Suppose that U_i stands for the utility function of person i .

The proof of the fundamental theorem relies on two facts stated and proved below.

Fact 1: If for some bundle x_i , $U_i(y_i) < U_i(x_i)$, then $\sum_{j=1}^n p_j w_i^j < \sum_{j=1}^n p_j x_i^j$.

Proof: If this were not the case, then y_i would not be the bundle that maximizes U_i subject to i 's budget constraint, which would contradict the assumption that y is a competitive equilibrium allocation.

Fact 2: If for some bundle x_i , $U_i(y_i) \leq U_i(x_i)$, then $\sum_{j=1}^n p_j w_i^j \leq \sum_{j=1}^n p_j x_i^j$.

Proof: Suppose that the claim is not true. Then we must have $\sum_{j=1}^n p_j w_i^j > \sum_{j=1}^n p_j x_i^j$. Now, define a bundle v_i by adding a small quantity c of every good to the bundle x_i :

$$v_i^j = x_i^j + c \text{ for all } j. \quad (16.16)$$

Choose c sufficiently small so that

$$\begin{aligned} \sum_{j=1}^n p_j v_i^j &= p_1 x_i^1 + p_2 x_i^2 + \dots + p_n x_i^n + p_1 c + p_2 c + \dots + p_n c \\ &= \sum_{j=1}^n p_j x_i^j + c \sum_{j=1}^n p_j \leq \sum_{j=1}^n p_j w_i^j. \end{aligned} \quad (16.17)$$

This is possible since we have assumed that $\sum_{j=1}^n p_j x_i^j < \sum_{j=1}^n p_j w_i^j$.

Now, given that a consumer prefers more to less, we have

$$U_i(\bar{v}_i) > U_i(x_i) \geq U_i(y_i). \quad (16.18)$$

Therefore, $U_i(v_i) > U_i(y_i)$ while $\sum_{j=1}^n p_j v_i^j < \sum_{j=1}^n p_j w_i^j$, which contradicts that y is a competitive equilibrium allocation.

Proof of the Theorem We are given that y is a competitive equilibrium allocation and p is the competitive equilibrium price vector. Suppose to the contrary that y is not in the core. Then some coalition, say S , will block y . Assume without loss of generality that $S = \{1, 2, \dots, k\}$, $1 \leq k \leq m$. Let us say that S blocks y with the bundles $x_1, x_2, \dots, x_k$. This will mean that

$$\sum_{i=1}^k x_i^j = \sum_{i=1}^k w_i^j \text{ for all goods } j, \text{ (i)}$$

$$U_i(x_i) \geq U_i(y_i) \text{ for all } i \in S, \text{ (ii)}$$

$$U_i(x_i) > U_i(y_i) \text{ for some } i \in S, \text{ (iii)}$$

where U_i is person i 's utility function.

We combine (ii) with Fact 2 to get

$$\sum_{j=1}^n p_j w_i^j \leq \sum_{j=1}^n p_j x_i^j \text{ for all } i \in S. \quad (16.19)$$

Combining (iii) with Fact 1 we have

$$\sum_{j=1}^n p_j w_i^j < \sum_{j=1}^n p_j x_i^j \text{ for some } i \in S. \quad (16.20)$$

We can add these inequalities over all the traders in S to get

$$\sum_{i=1}^k \sum_{j=1}^n p_j w_i^j < \sum_{i=1}^k \sum_{j=1}^n p_j x_i^j. \quad (16.21)$$

We rewrite the above inequality as

$$\sum_{i=1}^k (p_1 w_i^1 + p_2 w_i^2 + \dots + p_n w_i^n) < \sum_{i=1}^k (p_1 x_i^1 + p_2 x_i^2 + \dots + p_n x_i^n). \quad (16.22)$$

Rearranging, we have

$$p_1 \sum_{i=1}^k w_i^1 + p_2 \sum_{i=1}^k w_i^2 + \dots + p_n \sum_{i=1}^k w_i^n < p_1 \sum_{i=1}^k x_i^1 + p_2 \sum_{i=1}^k x_i^2 + \dots + p_n \sum_{i=1}^k x_i^n, \quad (16.23)$$

from which it follows that

$$p_1 \left[\sum_{i=1}^k (x_i^1 - w_i^1) \right] + p_2 \left[\sum_{i=1}^k (x_i^2 - w_i^2) \right] + \dots + p_n \left[\sum_{i=1}^k (x_i^n - w_i^n) \right] > 0. \quad (16.24)$$

Now, by (i), each term in brackets is zero. Therefore, the assumption that y is not in the core leads to a contradiction. Hence the theorem follows.

EXERCISES

Multiple Choice

1. In a two-person (A and B) two-good (1 and 2) exchange economy, general equilibrium is reached when
 - a) Persons A 's marginal rate of substitution of good 2 for good 1 is equal to the ratio of prices of goods 1 and 2.
 - b) Person B 's marginal rate of substitution of good 2 for good 1 is equal to the ratio of the prices of goods 1 and 2.
 - c) The marginal rates of substitution of good 2 for good 1 is the same for A and B .
 - d) Any one of the above.
2. The contract curve is the locus of
 - a) All contracts in an exchange economy,
 - b) All Pareto efficient contracts in an exchange economy,
 - c) The sum of utility levels of the consumers,
 - d) None of the above.
3. Which of the following statements is not true about a Pareto efficient allocation?
 - a) The individuals are on their contract curve.
 - b) One individual cannot be made better off without making the other individual worse off.
 - c) It may be an allocation where both are worse off than they are at an allocation that is not Pareto efficient.
 - d) It may be an allocation where one is worse off than he is at an allocation which is not Pareto efficient.

4. From an initial endowment somewhere within an Edgeworth box, a free exchange Pareto efficient contract will lie
 - a) Above the contract curve.
 - b) Below the contract curve.
 - c) On the contract curve.
 - d) Any one of the above.
5. If the value of excess demand in $(n - 2)$ out of n markets is zero, then by Walras' law
 - a) The value of excess demand in the remaining two markets must sum to zero.
 - b) The ratio between the values of excess demands in the remaining two markets must be equal to one.
 - c) The excess demand functions in the remaining two markets are homogeneous of degree one.
 - d) (b) and (c).

Quantitative Problems and Questions

1. Prove or disprove the following.
 - a) In a two-person two-good economy the price offer curves will intersect the contract curve.
 - b) A Pareto inefficient allocation will never yield a higher level of welfare than a Pareto efficient allocation.
 - c) We can determine only relative prices in a general equilibrium system.
 - d) The Walras law asserts that in a world with n goods if there is at least one good for which excess demand is positive, then there must be at least one good for which excess demand is negative.
 - e) If we know the contract curve in a two-person two-good exchange economy, then we know the exact allocation in the economy.
 - f) In a pure exchange economy if the initial endowments of two persons are not on the contract curve, free exchange between the consumers will induce them to move to a point on the contract curve.
2. Consider a two-person (A and B) two-good (1 and 2) exchange economy. Let p_1 and p_2 be the prices of the goods 1 and 2 respectively. Then answer the following questions.

- a) Can $E_1(p_1, p_2) = |\log p_1 - \log p_2|$ and $E_2(p_1, p_2) = |p_1 - p_2|$ be two examples of an excess demand function?
 - b) If $E(p_1, p_2) = k(p_1/p_2)^2 + b$, where k and b are constants, then for E to be an excess demand function, what should be the value of b ?
 - c) Suppose $E_1(p_1, p_2) = -p_1/p_2$ is the excess demand function for good 1. Under what condition $E_2(p_1, p_2) = p_1/p_2$ is an excess demand function for good 2?
3. In an exchange economy with two people (A and B) and two goods (1 and 2), suppose that the utility functions of A and B are given by $U_A(x_A) = x_A^1 x_A^2$ and $U_B(x_B) = x_B^1 + x_B^2$. The initial endowments are given by $w_A = (10, 0)$ and $w_B = (0, 10)$.
 - a) Graph the Edgeworth box to represent the economy.
 - b) Find the contract curve.
 - c) Solve for the competitive equilibrium.
 - d) Illustrate the core graphically.
 4. Consider an economy with two persons (A and B) and two goods (1 and 2). The utility functions of the two persons are given by $U_A(x_A) = x_A^1 x_A^2$ and $U_B(x_B) = 2x_B^1 + x_B^2$. The initial endowments are $w_A = (1, 0)$ and $w_B = (0, 1)$.
 - a) Represent the economy by an Edgeworth box.
 - b) Solve for the set of Pareto optimal allocations. Where is the core?
 - c) Solve for the competitive equilibrium.
 - d) Calculate the tax-transfer vector (T_1, T_2) and the price vector (p_1, p_2) such that the modified competitive equilibrium is $x_A = (1/4, 1/2)$ and $x_B = (3/4, 1/2)$.
 5. What is partial equilibrium analysis? What does general equilibrium analysis accomplish? Clearly distinguish between the two types of analyses.
 6. Explain why as an economy gets large, the core shrinks.
 7. Consider a two-person two-good exchange economy. Let C be a Pareto efficient allocation lying on a boundary of the Edgeworth box. The allocation C cannot be supported by competitive prices. Which assumption of the Second Fundamental Theorem of welfare economics is violated?
 8. Clearly explain the central ideas underlying the two fundamental theorems of welfare economics.

9. Is a Pareto efficient allocation necessarily equitable?
10. What is the contract curve for a pure exchange economy in which consumer A has no use for good 1 and consumer B has no use for good 2?
11. What is Walras' law? Discuss its implications.
12. Consider the exchange economy given by

$$U_A(x_A) = \log x_A^1 + \log x_B^2, \quad U_B(x_B) = x_B^1 + x_B^2.$$

The initial endowments are $w_A = (7, 3)$ and $w_B = (3, 7)$.

- a) What are the Pareto optimal allocations?
- b) What are the core allocations?
- c) What is the competitive equilibrium?

17

Production

In the previous chapter we focussed on models of exchange. In those models quantities of various goods were assumed to be fixed—nothing was produced. In this chapter, we will focus on a model of production and describe how the model fits into the general equilibrium framework. When we consider production, the amounts of the goods are not fixed but will respond to market prices. If the exchange model constitutes half of the story, the production model gives the other half. Near the end of chapter we will see how the two models can be merged.

PRODUCTION POSSIBILITIES

Suppose that there are two firms F and G in an economy. Assume that each firm uses two inputs, capital (K) and labor (L), to produce two different goods (1 and 2). We denote firm F 's output vector by (x_F^1, x_F^2) , where x_F^1 represents F 's output of good 1 and x_F^2 represents F 's output of good 2. Then G 's output vector is of the type (x_G^1, x_G^2) . By a production vector for the economy we mean a vector of the type $(x_F^1, x_F^2, x_G^1, x_G^2)$. The total amounts of capital and labor available to the economy are given by K_0 and L_0 respectively. We assume that all output and factor markets are perfectly competitive.

In Fig. 17.2 we represent the production possibilities set for a firm, say, F . This set gives the alternative combinations of goods 1 and 2 that the firm can produce using different amounts of capital and labor available to the firm. That is, every point in the set is technologically feasible for the firm. The boundary of the production possibilities set is called the production possibilities frontier. The points on this frontier satisfy the property that with fixed resources the output of one good cannot be increased unless the output of another good is decreased. In other words, the production possibilities frontier shows those alternative combinations of goods 1 and 2 that can be efficiently produced by the firm with a fixed amount of capital and labor. Thus, along the frontier output of good 1(2) is maximized for any output of good 2(1). Output combinations lying off the frontier involve inefficient allocation of resources. Given any such interior point, there are feasible output combinations such that more of both goods can be produced by the firm. At O_1 , no resources are devoted to production of good 1; consequently, the output of good 2 is as large as possible with the available resources. Similarly, at O_2 the output of good 1 is as large as possible.

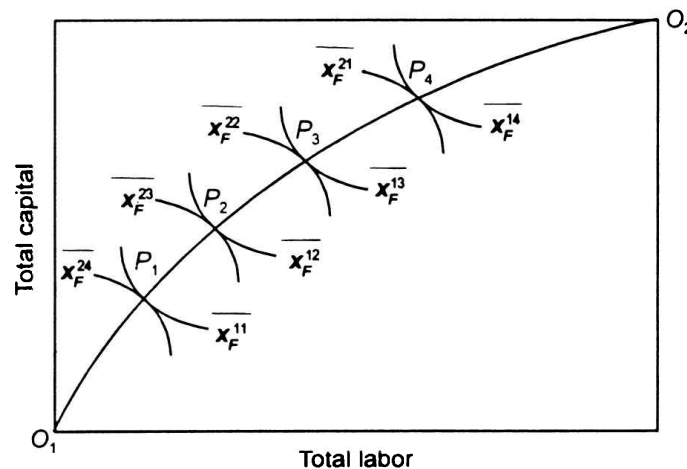


Fig. 17.1 Edgeworth-Bowley Box

The production possibilities frontier of a firm can be derived from its Edgeworth-Bowley box (Fig. 17.1). The dimensions of the box are given by the total quantities of capital and labor available to the firm. The south-west corner of the box, O_1 , is taken as the origin for an isoquant map showing alternative quantities of good 1 that might be produced by using various combinations of labor and capital. For instance, along the isoquant passing through P_1 the constant output level of good 1 produced by firm F is x_F^{11} and so on. Similarly, the north-east corner of the box, O_2 is taken as the origin for an isoquant map for good 2. Any point within the box represents a (fully employed) allocation of

labor and capital between the production of goods 1 and 2. It should be noted that the firm has identical rates at which it can trade labor for capital (the marginal rate of technical substitution) in all those outputs that it produces. This is assured by the existence of perfectly competitive factor markets for inputs. In minimizing costs the firm will equate the marginal rate of technical substitution between labor and capital to the ratio of competitive rental prices (w/r). This will be true for any output level the firm produces, and hence it will be equating all its marginal rates of technical substitution to the common price ratio (w/r). In this way the firm will be led to adopt efficient input proportions in a decentralized way.

The locus of efficient allocation of labor and capital is given by the points of tangency of various isoquants. Along this locus the inputs are allocated efficiently—the output of good 1 cannot be increased without reducing the output of good 2. Along the efficiency locus the marginal rate of technical substitution (of capital for labor) in the production of good 1 is equal to that in the production of good 2. By transferring the efficiency locus O_1O_2 on a separate graph with output of goods 1 and 2 on the axes, we get the production possibilities frontier shown in Fig. 17.2.

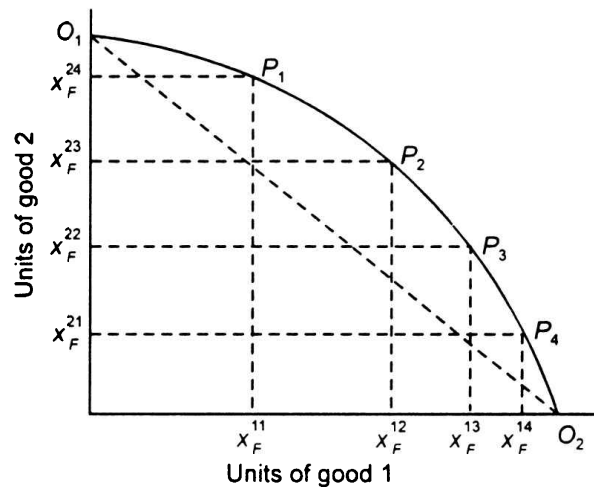


Fig. 17.2 Production Possibilities Set

The production possibilities frontier of a firm should be contrasted with the production function which gives the maximum possible output that the firm gets from a given amount of inputs. (In some treatments, both inputs and outputs are considered as part of the production possibilities set. In such treatments for any point in a firm's production possibilities set, if an entry is positive that good is an output for the firm; if an entry is negative, that good is an input for the firm. For instance, suppose a firm can produce 5 cars from 100

tons of steel and 1,000 man-days of labor. Suppose also that cars are good 1, steel is good 2 and labor is good 3. This combination of inputs (steel and labor) and output (car), which we write as $(5, -100, -1,000)$, is a point in the firm's production possibilities set. Given the complex nature of these types of treatments, we do not wish to pursue them here.)

The slope of the production possibility frontier in Fig. 17.2 shows how good 1 can be a substitute for good 2 when total resources are held constant. For instance, for points near O_1 on the production possibility frontier, the slope is a small negative number, say $-1/5$, implying that by reducing good 2 by one unit, good 1 can be increased by five. Near O_2 , on the other hand, the slope is a large negative number, say, -6 , implying that good 2 must be reduced by six units to permit the production of one more unit of good 1. The slope of the production possibilities frontier, then, clearly shows that possibilities exist for trading good 2 for good 1 in production. The negative of this slope is called firm F 's marginal rate of transformation (MRT) of good 2 into good 1. That is

$$MRT \text{ (of good 2 into good 1)} = - \frac{\Delta x_F^2}{\Delta x_F^1}. \quad (17.1)$$

The MRT records how firm F can technically trade good 2 for good 1 while continuing to keep the available productive inputs efficiently employed.

An important point that needs to be recognized here is that MRT can be interpreted in terms of the ratio of the marginal cost of good 1 to the marginal cost of good 2. Intuitively, this result is obvious. Observe that along the production possibilities frontier, the total cost of producing any output combination will be constant since the inputs are in fixed supply. Now, assume, for simplicity that the goods are produced only by labor. If it takes two labor-hours to produce one more unit of good 1, we might say that the marginal cost of good 1 is 2. Similarly, if it takes only one labor-hour to produce an extra unit of good 2, then the marginal cost of good 2 is 1. But in this situation it is clear that MRT is 2: two units of good 2 must be foregone to provide enough labor so that good 1 may be increased by one unit. Hence MRT is indeed equal to the ratio of the marginal costs of the two goods.

The production possibilities frontier illustrated in Fig. 17.2 exhibits an increasing MRT . For output levels near O_1 , MRT is small, whereas it is large near O_2 . Increasingness of MRT means that the production possibilities set is convex. (Recall that a set is called convex if the line segment connecting any two points of the set is entirely contained in the set.) We will now discuss why a convex production possibilities set might be expected to characterize most production situations.

First, the production possibilities frontier clearly demonstrates that there are many possible efficient combinations of goods and that production of more of one good necessitates cutback of the production of the other good. This is clearly what we mean by the term opportunity cost. The cost of producing more of good 1 can be most readily measured by the reduction in output of good 2 required. The cost of one more unit of good 1 is therefore best measured as the *MRT* (of good 2 into good 1) at the prevailing point on the production possibilities frontier.

Next, if some inputs are more suitable for production of good 1 than for good 2 production (and vice versa), the convex production possibilities set can be explained. In that case increases in good 1 output would require drawing progressively less suitable inputs into production of that good. The marginal cost of good 1 would therefore rise. Marginal costs for good 2 would fall as smaller output levels for good 2 would use only those inputs that are most suitable for its production. Therefore, *MRT* will rise, which means convexity of the production possibilities set. Such an argument will apply, for instance, to a farmer who has different types of land under cultivation. In trying to increase the production the farmer would be forced to grow it on increasingly unsuitable types of land.

A third point that will justify the convexity of the production possibilities set is the use of differing factor intensities in the production of goods 1 and 2 even if the production functions exhibit constant returns to scale. For instance, in Fig. 17.1 good 1 is capital intensive relative to good 2. That is, at every point along O_1O_2 the capital-labor ratio is higher for good 1 production than for good 2 production. This made the production possibilities set in Fig. 17.2 convex. To see intuitively how differing factor intensities make the production set convex, consider the dotted line in Fig. 17.2 that joins O_1 and O_2 . One way of producing an output combination between O_1 and O_2 would be to produce the output combination $\left(\frac{O_1}{2}, \frac{O_2}{2}\right)$ that lies on the midpoint of the dotted line. Because of constant returns to scale assumption, this combination is feasible and it would fully utilize both factors of production. Constant returns to scale imply that the firm can do at least as well as the output combinations on the dotted line between O_1 and O_2 . Now, suppose good 1 is relatively capital intensive and good 2 is relatively labor intensive. Then in moving from the point O_1 we can efficiently increase the output of good 1 while decreasing the output of good 2 by withdrawing proportionately more capital than labor from the production of good 2. The capital-labor ratio decreases in the production of both goods 1 and 2. Since good 1 is capital intensive, this change increases the marginal cost of

good 1. On the other hand, since good 2 is labor intensive, its marginal cost falls. Hence *MRT* rises. Allocating disproportionate amounts of labor in the production of good 2 and capital in the production of good 1 allows the firms to achieve output combinations superior to those on the dotted line. Therefore, an efficient curve must lie above the dotted line representing equal factor intensities for the products.

EFFICIENCY IN PRODUCTION

A society with a fixed amount of productive resources will be said to have allocated those resources efficiently in production if no more of one good can be produced without having a cutback in the production of something else. That is, if a production vector is efficient, for the economy there is no way to get more of one output without producing less of some other output. Given that the economy produces only two goods, an allocation of resources under which more of good 1 could be produced without having to reduce good 2 output would be clearly inefficient. There is no reason why resources should not be reallocated so as to yield more of good 2. On the other hand, if increasing output of good 2 required cutdown of good 1 output, the allocation would be efficient since output cannot be unambiguously increased. By using this definition of efficiency, it is never necessary to compare goods 1 and 2 directly to decide whether or not resources are allocated efficiently. Rather, good 1 is compared only to good 1 (to see if its output can be increased), and good 2 is compared only to good 2.

A discussion on the conditions that enable us to achieve efficiency for the economy is somewhat more complex than the exchange case because the economy is divided into firms which use many inputs and may produce many outputs. We therefore must not only study the allocation of resources within a firm but also analyze how resources should be allocated among firms. We have already analyzed in the previous section how resources should be allocated within a single firm. We will now examine what amounts of resources each firm should have.

If an economy has a fixed amount of resources, these must be allocated in some efficient way among firms in order to ensure productive efficiency for the economy as a whole. Intuitively, resources should be allocated to those firms where they can be used efficiently. More precisely, resources should be allocated so that the marginal product of any resource in the production of a particular good is the same no matter which firm produces that good.

To see this, consider the marginal productivity of labor curves for firms *F* and *G*, both producing the same output (say, good 1). Suppose that the initial

allocation of labor between the firms has L_F workers for firm F and L_G workers for firm G , where $L_F + L_G = L_0$. In Fig. 17.3 we see the marginal productivity of resources for the firms. The marginal productivity of labor is higher in firm F than that in firm G .

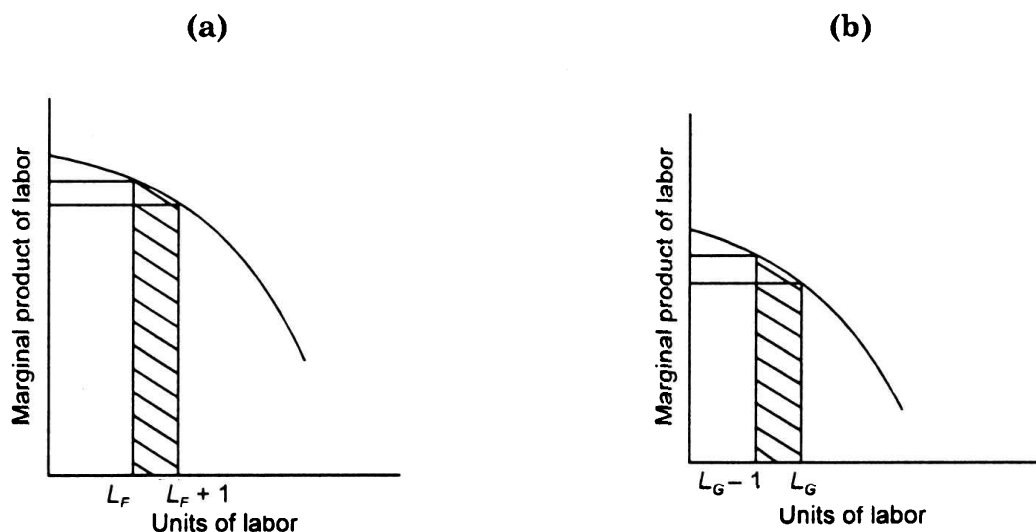


Fig. 17.3 Efficient Allocation of Resources between Firms

The labor should therefore be transferred from firm G to firm F . Suppose that only one worker is transferred. The effects of this are shown in Fig. 17.3. The reduction in output of firm G due to decrease in its labor force from L_G to $L_G - 1$ and the increase in the output of firm F resulting from expansion of its labor force from L_F to $L_F + 1$ are given respectively by shaded areas in panels (b) and (a) of Fig. 17.3. Clearly, output has gone up as a result of this transfer: The shaded area in panel (a) of the figure is larger than that in panel (b). The transfer of labor also serves to equilibrate the marginal product of labor between the two firms. The additional worker in firm F causes the marginal product of labor to fall, whereas there are fewer workers in firm G and marginal productivity has gone up. The transfer of labor should be continued until equality is achieved. That is, the available stock of labor should be reallocated such that marginal productivities do not differ among firms.

In fact, the above efficient rule of resource allocation is brought about by the competitive market. We know that a profit maximizing firm will hire additional labor up to the point at which the marginal revenue product of labor equals competitive wage rate (w). Thus, if we write MP_L^{F1} for firm F 's marginal productivity of labor in the production of good 1, then $p_1 \cdot MP_L^{F1} = w$, where p_1 is the price of good 1 and w is the wage rate. Since both p_1 and w are given by

the market, each firm will equate its marginal productivity of labor in production of good 1 to $\frac{w}{p_1}$. Consequently, every firm will have the same marginal productivity of a resource in the production of a good.

COMPETITIVE EQUILIBRIUM

Even though resources may be efficiently allocated both within a firm and among all firms, there is still another condition that must be obeyed. This is the behavioral assumption, the motive. We will assume that they act competitively, that is, they take prices of goods and factors of production as given by the markets.

The profit is given by revenue less cost. Revenue is the aggregate value of the goods that the firm sells. Cost is the aggregate value of the inputs that the firm employs. We write p_1 and p_2 for the prices of goods 1 and 2 respectively. The wage rate is w and the rental price of capital is r . Therefore, the profit of firm F is $p_1 x_F^1 + p_2 x_F^2 - wL_F - rK_F$, where L_F and K_F are firm F 's employment of capital and labor. The profit maximization problem is subject to the technology constraints described by firm F 's production possibilities set.

$$\max_{x_F^1, x_F^2, L_F, K_F} (p_1 x_F^1 + p_2 x_F^2 - wL_F - rK_F) \quad (17.2)$$

Our analysis in the previous section shows how resources can be efficiently allocated across firms. Suppose firm F has $\bar{L}_F$ units of labor and $\bar{K}_F$ units of capital. This in turn implies that firm G has $K_0 - \bar{K}_F$ units of capital and $L_0 - \bar{L}_F$ units of labor. We now want to determine how profit maximization chooses the output level. Let $\bar{C}_F = w\bar{L}_F + r\bar{K}_F$ be the total cost of production. Clearly, $\bar{C}_F$ is the constant cost for producing any output combination along firm F 's production possibilities frontier.

We can write the profits of the firm π , as

$$\pi = p_1 x_F^1 + p_2 x_F^2 - \bar{C}_F. \quad (17.3)$$

Eq. (17.3) is the equation for the locus of production vectors for which profit equals the constant amount π . This locus of points is a straight line with slope $-p_1/p_2$. Such straight lines are called isoprofit lines.

Fig. 17.4 shows the production set and isoprofit lines for firm F . In the figure l_1 , l_2 and l_3 are three different isoprofit lines. Clearly, the lines farther to the right represent higher profits (or higher π 's in Eq. [17.3]). Consequently, l_3 represents the highest profit level of the three isoprofit lines drawn. The firm's assumed behavior is to choose the production vector which puts it on the highest possible isoprofit line. But this will imply that the isoprofit line must be tangent to the production possibilities set; that is, the slope of the production possibilities frontier should be equal to the slope of the isoprofit line $-p_1/p_2$.

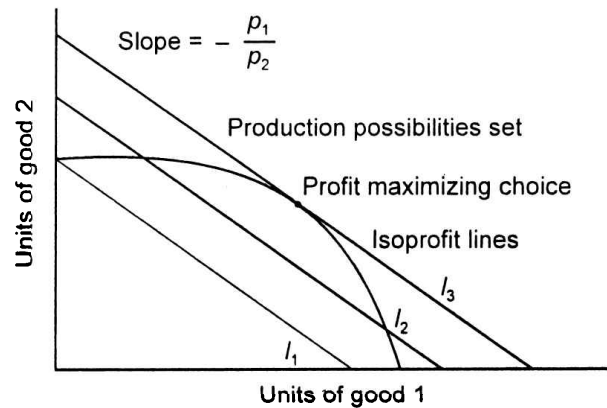


Fig. 17.4 Profit Maximization

We have derived the profit maximizing condition for firm F ($MRT = p_1/p_2$). But it holds for firm G also. This holds even if the firms have quite different production possibilities sets, given that they face the same prices for the two goods.

In a model of exchange economy a competitive equilibrium is an allocation in which all individuals maximize their utilities subject to their budget constraints. In the production model we define competitive equilibrium as follows. A production vector $(\bar{x}_F^1, \bar{x}_F^2, \bar{x}_G^1, \bar{x}_G^2)$ is said to constitute a competitive equilibrium for the production economy if $(\bar{x}_F^1, \bar{x}_F^2)$ and $(\bar{x}_G^1, \bar{x}_G^2)$ maximize the profits of firms F and G respectively.

Parallels with the exchange economy model are clear. In both the models there is only one price vector and everybody takes it as given. While in the exchange framework consumers maximize utility, in the production framework firms maximize profits. Utility maximization is done subject to a budget constraint and profit maximization is done subject to a constraint that the output vector is in the production set. In the exchange model the total quantities of all goods are fixed. In the production model quantities of outputs are not fixed; only input quantities are fixed which in turn make the production sets fixed.

FIRST FUNDAMENTAL THEOREM OF WELFARE ECONOMICS (PRODUCTION)

We are now in a position to state the production version of the First Fundamental Theorem. This theorem says that a competitive equilibrium production vector brings about perfect coordination of firms' activities. In other words, in a competitive equilibrium it is not possible to increase the output of one good without reducing the output of another good. That is, there is no slack in a competitive equilibrium. We state that theorem formally.

THEOREM: *A competitive equilibrium production vector is efficient for the economy as a whole.*

We need to make some important observations about the theorem. First, the theorem is meaningful only if a competitive equilibrium exists. Therefore, production functions with increasing returns to scale are ruled out. In the case of increasing returns to scale the firm's production set is not convex. Second, the theorem is concerned with efficiency considerations only: it has got nothing to do with distribution. Third, the output decisions of one firm do not affect the production possibilities of other firms, that is, production externalities are absent.

SECOND FUNDAMENTAL THEOREM OF WELFARE ECONOMICS (PRODUCTION)

Generally there are many efficient production vectors. Some of them might involve large outputs of certain goods. Some might involve large outputs of the remaining goods and so on. To say that a production vector is efficient might not be enough. We may prefer one efficient production vector to another on distributional considerations. Therefore, the First Fundamental Theorem leaves questions unanswered. These bring us to the Second Fundamental Theorem.

THEOREM: *Assume that all production sets are convex and closed. Then any production vector which is efficient for the economy can be achieved through the use of competitive markets.*

JOINT PRODUCTION POSSIBILITIES AND COMPARATIVE ADVANTAGE

In panels (a) and (b) of Fig. 17.5 we show the production possibilities sets of firms F and G respectively. Panel (c) shows the joint production possibilities—how much of each good could be produced in total by both firms. Firm F has a

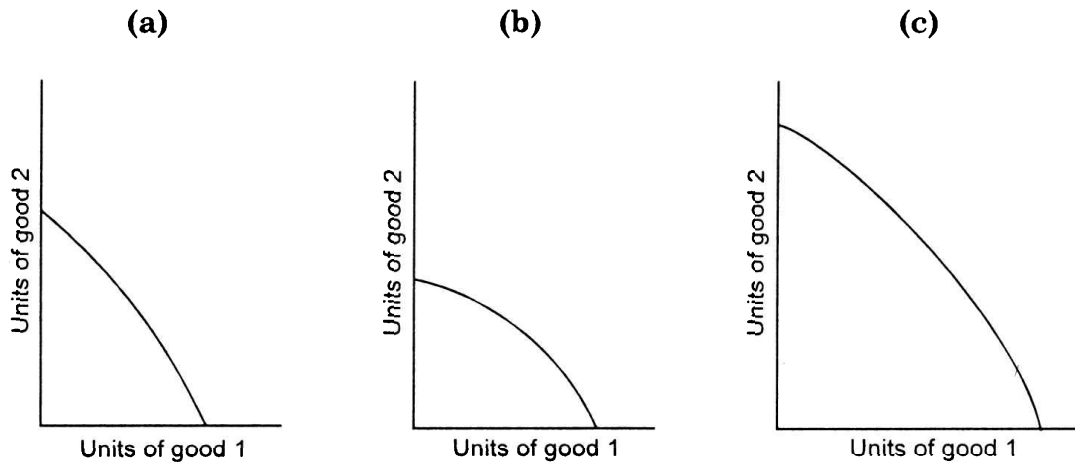


Fig. 17.5 Comparative Advantage

comparative advantage in the production of good 2—as we move up along the production possibilities frontier of firm *F*, we get higher amounts of good 2 when good 1 output is reduced by one unit. Similarly, firm *G* has a comparative advantage in the production of good 1. That is, as we move down along firm *G*'s production possibilities frontier, higher amounts of good 1 are obtained in exchange of one unit of good 2. Joint production possibilities combine the production possibilities of both firms. Its frontier combines the best of both firms. This curve shows the maximum amount of one good that the economy can obtain, given the quantity of the other good.

Given that the aggregate production possibilities are the sum of individual firms' production possibilities, if each firm maximizes profits, then aggregate profits must be maximized. Conversely, if aggregate profits are maximized, then each firm's profits must be maximized.

COMBINING PRODUCTION AND EXCHANGE

The discussion above leaves certain important issues unclear. Where do the inputs of the firms come from? Where do the outputs of the firms go? The model is incomplete.

One way to make it more complete is to assume that there are certain fixed amounts of inputs available to the economy, and that these amounts are owned by consumers. That is, initial endowments of consumers are now the factors of production. Private ownership of resources brings us to a more complex and fuller model. Each consumer *i* owns more than the initial endowment at the start. He also owns shares of the firms. A share of a firm gives him a right to

a fraction of the profit of that firm. In addition to profits from the firms, consumers get income from selling their endowments also.

The consumers use the money to purchase consumption goods. Each person chooses the best bundle of goods that he can afford at the prevailing prices. The optimal consumption bundle for the consumer must satisfy the condition that the marginal rate of substitution between the two goods must be equal to the common price ratio. But the profit maximization condition by the firm implies that this common price ratio equals the marginal rate of transformation. Since prices are common, the marginal rate of substitution for each consumer must be equal to the marginal rate of transformation. In fact, this is the necessary condition for Pareto efficiency in this mixed economy which is defined as follows. An allocation of goods and resources among both firms and individuals is Pareto efficient if no one can be made better off without making someone else worse off. The condition for Pareto efficiency in the mixed economy is demonstrated below graphically. A rigorous proof is given in Appendix 17.1.

We consider a one-consumer, one-firm, two-good economy. The traditional name for this economic model is the Robinson Crusoe economy. The two goods being produced by the firm are denoted by 1 and 2. Those combinations of goods 1 and 2 that can be produced efficiently are given by the production possibilities frontier PP . By superimposing the individual's indifference map on the figure we see that only one point on PP provides maximum utility. This point of maximum utility is denoted by M , where PP is tangent to the individual's highest attainable indifference curve I_2 . At this point of tangency the marginal rate of substitution (between good 1 and good 2) equals the marginal rate of transformation (between good 1 and good 2) and hence this is the required condition for overall efficiency. That is, allocation M is Pareto efficient.

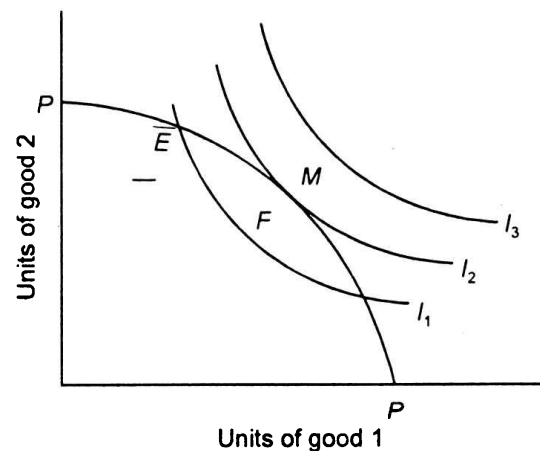


Fig. 17.6 Efficiency of Production and Exchange in Robinson Crusoe Economy

Observe that M is preferred to every other point on PP . In fact for any point (other than M) such as E , on PP , there are points such as F that are inefficient in the productivity sense but are preferred to E . It would be preferable from the consumption point of view to produce inefficiently rather than be forced to produce the wrong combination of goods in an efficient way.

We can also illustrate the efficiency condition numerically. Suppose a consumer's MRS is 1; the consumer is willing to trade one unit of good 1 for one unit of good 2. Suppose MRT is 2, which means that one unit of good 1 could be traded for two units of good 2 in production. If we reduce the production of good 1 by one unit, we get two units of good 2. Now, the consumer is indifferent between giving up one unit of good 1 and getting one unit of good 2 in exchange. Therefore, he certainly becomes better off by getting one extra unit of good 2. Therefore, for efficiency, equality of MRS with MRT is required.

THE TWO FUNDAMENTAL THEOREMS IN COMBINED MODEL

With the above formulation of the production/exchange economy model the two fundamental theorems can be stated and discussed. These two theorems establish the relationship between an individual's private goals of utility maximization and the social goals of efficient use of resources. To state these theorems precisely, let us first define a competitive equilibrium for this mixed economy. As before, assume that there are two consumers A and B in the economy. Assume also that A owns an endowment (L_A, K_A) and B owns an endowment (L_B, K_B) of labor L and capital K . There are two firms F and G . Firm F 's employments of resources are L_F and K_F , and firm G 's employment are L_G and K_G . Firm F 's profit is $\pi_F = p_1 x_F^1 + p_2 x_F^2 - wL_F - rK_F$ and firm G 's profit is $\pi_G = p_1 x_G^1 + p_2 x_G^2 - wL_G - rK_G$. The consumption bundles of A and B are (x_A^1, x_A^2) and (x_B^1, x_B^2) respectively. Person A 's shares of different firms' profits are s_A^F and s_A^G . The corresponding shares for person B are s_B^F and s_B^G . Clearly, $s_A^F + s_B^F = s_A^G + s_B^G = 1$.

A competitive equilibrium occurs in this mixed economy at commodity prices $(\bar{p}_1, \bar{p}_2)$, resource prices $(\bar{w}, \bar{r})$, final holdings $(\bar{x}_A^1, \bar{x}_A^2)$, $(\bar{x}_B^1, \bar{x}_B^2)$ and output vectors $(\bar{x}_F^1, \bar{x}_F^2)$, $(\bar{x}_G^1, \bar{x}_G^2)$ under the following conditions:

- I. $(\bar{x}_A^1, \bar{x}_A^2)$ maximizes A 's utility subject to $\bar{p}_1 \bar{x}_A^1 + \bar{p}_2 \bar{x}_A^2 \leq \bar{r} K_A + \bar{w} L_A + s_A^F \pi_F + s_A^G \pi_G$; $(\bar{x}_B^1, \bar{x}_B^2)$ maximizes B 's utility subject to $\bar{p}_1 \bar{x}_B^1 + \bar{p}_2 \bar{x}_B^2 \leq \bar{r} K_B + \bar{w} L_B + s_B^F \pi_F + s_B^G \pi_G$.

- II. $(\bar{x}_F^1, \bar{x}_F^2)$ maximizes firm F 's profits; $(\bar{x}_G^1, \bar{x}_G^2)$ maximizes firm G 's profits.
- III. $K_F + K_G = K_A + K_B$, $L_F + L_G = L_A + L_B$; $\bar{x}_F^1 + \bar{x}_G^1 = \bar{x}_A^1 + \bar{x}_B^1$; $\bar{x}_F^2 + \bar{x}_G^2 = \bar{x}_A^2 + \bar{x}_B^2$.

Condition I in the above definition means that each consumer's utility is maximized subject to his budget constraint. Condition II states profit maximization problems of firms. Finally, condition III means that the supply of each resource/good is equal to its demand.

With this rather general definition of the competitive equilibrium the two fundamental theorems can now be stated. The first theorem says that a competitive equilibrium in the production/exchange economy is Pareto efficient. The second theorem says that virtually any Pareto efficient allocation can be achieved as a competitive equilibrium, if the initial endowments including the ownership of resources are suitably redistributed.

According to the first theorem if the behavioral assumptions (utility and profit maximization) of our model are satisfied, then the market equilibrium is Pareto efficient. The only connections between consumers and firms are the competitive prices of goods. The competitive market ensures that the social problem of utilizing resources efficiently can be decentralized at individual levels. The second theorem can be interpreted as follows. We can think of the state as seizing the consumers' original endowments of resources and redistributing the endowments in some way compatible with the desired distribution. This redistribution may involve a redistribution of profit shares and goods also.

EXAMPLE 17.1

Consider a one-person economy in which the utility function of the individual for consumption, x , and leisure, l , is given by $U(x, l) = \log x + \log l$. The person is endowed with 24 hours of labor/leisure. The production function of the economy is given by $x = L$, where L stands for labor. We want to determine the competitive equilibrium of this economy, that is, the values of x and l and the price ratio w/p , where w is the wage rate and p is the price of x .

The profit maximization problem is

$$\max_{x, L} (px - wL), \quad (17.4)$$

which we rewrite as

$$\max_L (pL - wL).$$

Letting $\pi = pL - wL$ and putting $\frac{d\pi}{dL} = 0$, we get $w = p$. That is, the equilibrium wage rate must be equal to the marginal revenue product of labor. Thus, at equilibrium the price ratio w/p is 1.

The maximization problem of the consumer is

$$\max_{x, l} (\log x + \log l),$$

subject to $px = wL$. Since $l + L = 24$, we can rewrite the constraint $px = wL$ as $px + wl = 24w$.

$$\text{Let } \mathcal{L} = \log x + \log l - \lambda(px + wl - 24w), \quad (17.5)$$

where λ is the Lagrange multiplier. Putting $\frac{\delta \mathcal{L}}{\delta x} = 0$, $\frac{\delta \mathcal{L}}{\delta l} = 0$, and $\frac{\delta \mathcal{L}}{\delta \lambda} = 0$, we get

$$\begin{aligned} \frac{1}{x} &= \lambda p \\ \frac{1}{l} &= \lambda w, \quad px + wl = 24w. \end{aligned} \quad (17.6)$$

Dividing the first of these three equations by the second, we have $l/x = p/w$, which gives $px = wl$. Substituting this in the third equation, we have $2wl = 24w$, which implies that $l = 12$. It is then easy to find that the value of x will also be 12. Thus, the equilibrium supply of labor is $(24 - 12) = 12$ hours and the equilibrium output is 12 units.

Applied Example 17.1 General Equilibrium Model for India

Narayana, Parikh and Srinivasan (1991) constructed an applied general equilibrium model of the Indian economy and employed the model to investigate several policy questions. Their model has some desirable properties for policy analysis in an open (where trading with other countries is allowed) and mixed (where both private and public sectors exist) economy like India's. The various consistency conditions incorporated in the model ensured that all the feedbacks have been taken into account and there were no unaccounted supply sources or demand sinks in the system. Several policy issues including growth promoting

policies, policies to promote agricultural growth and redistributive policies have been addressed. The redistributive issues considered can be interpreted in the framework of the Second Fundamental Theorem. In spite of various limitations of the model structure, the policy conclusions arrived at are defensible.

Appendix 17.1

MARGINAL RATE OF TRANSFORMATION

We show rigorously that the marginal rate of transformation is equal to the ratio of the marginal cost of the two goods. Let $C(w, r, x_F^1, x_F^2)$ be firm F 's cost of production of output combination (x_F^1, x_F^2) at the given factor prices (w, r) . Along the production possibilities frontier, C will be constant since the factors are in fixed supply. Hence we can write the total differentiation of the cost function as

$$dC = \frac{\delta C}{\delta x_F^1} \cdot dx_F^1 + \frac{\delta C}{\delta x_F^2} \cdot dx_F^2 = 0 \quad (17.7)$$

for changes in x_F^1 and x_F^2 along the frontier. From this equation we have

$$-\frac{dx_F^2}{dx_F^1} = \frac{\delta C / \delta x_F^1}{\delta C / \delta x_F^2} = \frac{\text{marginal cost of good } \bar{1}}{\text{marginal cost of good } \bar{2}}. \quad (17.8)$$

Now, $-dx_F^2/dx_F^1$ is the slope of the production possibilities frontier. Since the negative of this slope is the marginal rate of transformation of good 2 into good 1, we have marginal rate of transformation (of good 2 into good 1)

$$= \frac{\text{marginal cost of good } \bar{1}}{\text{marginal cost of good } \bar{2}}. \quad (17.9)$$

Hence to demonstrate the reasons why the marginal rate of transformation may rise for movement along the production possibility frontier from the direction of good 2 to good 1, we can proceed by showing why the ratio of marginal cost of good $\bar{1}$ to the marginal cost of good $\bar{2}$ should rise as good $\bar{1}$ output expands and good $\bar{2}$ output contracts.

PARETO EFFICIENCY

We derive now the calculus conditions for Pareto efficiency in an economy with production. For this assume that there are two goods (1 and 2) and one individual in the society (Robinson Crusoe). Let X^1 and X^2 represent the total amounts of good 1 and good 2 produced and consumed. The utility function of the only individual is $U(X^1, X^2)$.

To describe the production possibilities frontier, consider the implicit function $T(X^1, X^2)$. The output combination (X^1, X^2) is on the production possibilities frontier if and only if $T(X^1, X^2) = 0$. Robinson Crusoe's problem is to maximize utility subject to this production constraint. The Lagrangian expression for this problem is

$$\mathcal{L} = U(X^1, X^2) - \lambda[T(X^1, X^2)], \quad (17.10)$$

where λ is Lagrange's multiplier. The first-order conditions for a maximum are

$$\begin{aligned} \frac{\delta \mathcal{L}}{\delta X^1} &= \frac{\delta U}{\delta X^1} - \lambda \frac{\delta T}{\delta X^1} = 0, \\ \frac{\delta \mathcal{L}}{\delta X^2} &= \frac{\delta U}{\delta X^2} - \lambda \frac{\delta T}{\delta X^2} = 0 \\ \frac{\delta \mathcal{L}}{\delta \lambda} &= T(X^1, X^2) = 0. \end{aligned} \quad (17.11)$$

Combining the first two of these equations we get

$$\frac{\delta U / \delta X^1}{\delta U / \delta X^2} = \frac{\delta T / \delta X^1}{\delta T / \delta X^2}. \quad (17.12)$$

The right hand side of the above equation is the marginal rate of transformation. The slope of the production possibilities frontier is the negative of the marginal rate of transformation. Note also that the left hand side of the above equation is Robinson's marginal rate of substitution between good 1 and good 2. Therefore, we have the following condition for Pareto efficiency: the marginal rate of substitution (between good 1 and good 2) = the marginal rate of transformation (between good 1 and good 2). This shows that only when an individual's preferences are taken into consideration will resources be allocated in an overall efficient way. Relating preferences to production in this way ensures that the utility-maximizing mix of output will be produced.

EXERCISES**Multiple Choice**

1. The production possibilities frontier of any economy is derived from
 - a) The social indifference curve.
 - b) The production contract curve.
 - c) The consumption contract curve.
 - d) None of the above.
2. A production possibilities frontier which is concave to the origin exhibits
 - a) An increasing marginal rate of transformation.
 - b) A non-decreasing marginal rate of transformation.
 - c) A decreasing marginal rate of transformation.
 - d) A non-increasing marginal rate of transformation.
3. A firm's marginal rate of transformation of good 2 into good 1 can be interpreted as
 - a) The ratio of the marginal cost of good 1 to the marginal cost of good 2.
 - b) The ratio of the marginal cost of good 2 to the marginal cost of good 1.
 - c) $|\text{Marginal cost of good 1} - \text{Marginal cost of good 2}|$.
 - d) $\text{Marginal cost of good 1} + \text{Marginal cost of good 2}$.
4. In an economy with two goods (1 and 2) and two factors (K and L), at general equilibrium of production
 - a) The marginal rate of technical substitution between K and L equals the ratio of prices of K and L .
 - b) The marginal rate of transformation between goods 1 and 2 equals the marginal rate of substitution between goods 1 and 2.
 - c) The marginal rate of technical substitution between L and K for good 1 is the same as that for good 2.
 - d) The marginal rate of technical substitution between L and K equals the marginal rate of substitution between A and B .
5. Pareto efficiency in production and consumption requires that
 - a) Each individual's marginal rate of substitution be equal to the marginal rate of transformation.
 - b) The sum of the marginal rates of substitution across individuals equals the marginal rate of transformation.

- c) The ratio of each individual's marginal rate of substitution to the marginal rate of transformation is nonconstant.
- d) The marginal rate of transformation is inversely proportional to the marginal rate of substitution of each individual.

Quantitative Problems and Questions

1. Suppose that 50 units of labor L and 50 units of capital K are available to an economy. The production function for good 1 is given by $X_1 = K$ and the production function for good 2 is given by $X_2 = K + L$. Determine the production possibilities frontier for this economy.
2. In a country there are two regions I and II. Two goods (1 and 2) are produced in both regions. The production functions for region I are given by

$$x_I^1 = \sqrt{L_I^1}, x_I^2 = \sqrt{L_I^2}$$

where L_I^1 and L_I^2 are the quanta of labor devoted to the production of goods 1 and 2 respectively. The total labor available in region I is 100 units. That is, $L_I^1 + L_I^2 = 100$. Using a similar notation for region II, the production functions are given by

$$x_{II}^1 = \frac{1}{2}\sqrt{L_{II}^1},$$

$$x_{II}^2 = \frac{1}{2}\sqrt{L_{II}^2}.$$

There are 100 units of labor available in region II.

- a) Calculate the production possibilities frontiers for regions I and II.
- b) What condition must hold if production in this country is to be allocated efficiently between regions I and II?
- c) Calculate the economy's production possibilities frontier.
3. Why is the production possibilities frontier of an economy concave to the origin? What does the slope of the frontier represent?
4. Prove or disprove the following.
 - a) The production possibilities frontier for an economy is linear if there are constant returns to scale in all production functions.
 - b) If the production possibilities frontier of an economy is concave to the origin, then the economy's production will be subject to a diminishing marginal rate of transformation.

- c) The complete equilibrium of an economy is given at the point at which the marginal rate of transformation is tangent to the price line.
 - d) Consider a one-person two-good economy. Suppose that the individual's marginal rate of substitution between goods 1 and 2 is 3 and the marginal rate of transformation between the goods is 1. Then he should increase the production of good 1 to increase utility.
 - e) Consider a two-good economy. The competitive prices of goods 1 and 2 are Rs. 10 and Rs. 5 respectively. If society decides to give up 1 unit of good 1, 2 additional units of good 2 can be produced.
 - f) A convex production set requires that the underlying production functions have the same factor intensities.
 - g) For efficient production, resources should be allocated so that the marginal product of a resource in the production of a given good is the same across firms.
5. An economy produces two goods 1 and 2 from two factors of production, land and labor. All factors are privately owned; perfect competition prevails in all markets. The production functions are $x^1 = l_1^{.75} L_1^{.25}$, $x^2 = l_2^{.25} L_2^{.75}$, where $l_1(l_2)$ is the number of square miles of land used in the production of good 1(2), and $L_1(L_2)$ is the number of workers employed in the production of good 1(2). All persons have identical utility functions given by $U(x^1, x^2) = (\sqrt{x^1 x^2})$. There are 324 square miles of land and 2,500 workers. The price of good 1 is Rs. 100. Calculate the equilibrium values for (i) the price of good 2, (ii) the rent on one square mile of land, and (iii) the wage rate.
 6. Suppose that there are decreasing returns to scale in producing two goods 1 and 2. Explain why the production possibilities frontier will become concave even if factor intensities are the same in two industries.
 7. What would happen to the equilibrium of a two-good economy if technological change occurs in the good 1 industry? (Good 1 industry is more capital intensive than the good 2 industry.) What happens to the marginal rate of transformation and prices of these goods?
 8. Robinson Crusoe obtains utility from the quantities of fish and coconuts he consumes. Robinson's production of fish and coconuts is determined by two well-defined production functions. Labor is the only factor of production. Explain how the full equilibrium provides answers to three central questions (i) 'what to produce' (given by preferences and tastes, shown on the indifference curve), (ii) 'how to produce' a given level of output (determined by the technical characteristics of the economy (marginal rate of transformation,

shown on the production possibilities frontier), and (iii) 'for whom to produce' (determined by the income from ownership of the factors of production which is reflected on the price line).

9. An economy produces two goods, 1 and 2. The production of good 1 is given by $x_1 = (\sqrt{K_1 L_1})$, where $K_1(L_1)$ is the capital (labor) input devoted to the production of good 1. The production of good 2 is given by $x_2 = K_2^{1/3} L_2^{2/3}$, where $K_2(L_2)$ is the capital (labor) input devoted to the production of good 2. It is given that $K_1 + K_2 = 100$, $L_1 + L_2 = 200$. That is, the supply of capital (labor) is fixed at 100(200) units.
 - a) Show that efficient production requires that the capital-labor ratio in good 1 production be related to the corresponding ratio in good 2 production
 - b) Let $\alpha_1 = L_1/200$. Then show that $\alpha_1 r_1 + (1 - \alpha_1) r_2 = 1/2$, where $r_1 = K_1/L_1$ and $r_2 = K_2/L_2$.
 - c) Using information from parts (a) and (b), determine r_1 for all values of α_1 .
 - d) What is the Edgeworth-Bowley box for the economy? Give a sketch of the production contract curve.
 - e) Determine the production possibilities frontier.
 - f) Which good is capital intensive in this economy?
10. Robinson Crusoe obtains utility from the quantities of fish F and coconuts C he consumes each day according to the utility function $U(C, F) = \sqrt{CF}$. Robinson's production of fish is given by $F = \sqrt{L_F}$ and his production of coconuts is given by $C = \sqrt{L_C}$ where $L_F(L_C)$ is the time he spends on fishing (picking coconuts). He is bound by the time constraint $L_C + L_F = 16$. (Robinson has already decided to consume 8 hours as leisure.)
 - a) How should Robinson allocate his time to maximize utility?
 - b) Graph his production possibilities frontier. Show his optimal choices of these goods.

18

Welfare Economics

Our analysis in Chapters 16 and 17 was concerned primarily with exploring the relationship between Pareto efficiency and a perfectly competitive price system. We showed that there is a clear connection between these concepts. In addition we mentioned that nothing within the competitive model ensures that even efficient outcomes will be regarded as fair or equitable. In choosing among efficient allocations of resources, it is inevitable that some individuals will prefer one specific allocation whereas others will prefer some alternative. The competitive model provides no method for making such choices. For instance, when we come to the problem of applying the second fundamental theorem of welfare economics, we may ask: How is the particular Pareto efficient allocation (that the theorem says we can reach) to be chosen?

In this chapter we will attempt to tie up these loose ends by examining the theory of welfare economics. That is, we will briefly describe the way economists have tried to conceptualize the problem of how societies should make choices among various feasible allocations. More precisely, this chapter is concerned with evaluation of alternative economic situations from the viewpoint of society's well-being in terms of some explicitly stated moral or ethical criteria.

PRELIMINARIES

In this chapter we will discuss social states labeled x , y , z , etc., which might be allocations in an exchange economy, or which might be more general arrangement like a listing that shows: (i) for each consumer the amount of each good or service he receives and the amounts of labor services he supplies; and (ii) for each firm the amount of each input used up in production and the amount of each output produced. Or, they might be simply mutually exclusive possibilities, like 'build a bridge' or 'do not build a bridge'.

PARETO EFFICIENCY PRINCIPLE RECONSIDERED

We have already considered the Pareto efficiency criterion as a tool for selecting different allocations in some earlier chapters. To discuss this criterion in greater detail let us consider the unanimity principle first. This principle states that between any two social states x and y , x is judged better than y if everyone finds x at least as good as y and at least one individual finds x better than y .

Suppose that x is regarded as better than y under the unanimity principle. Then we also say that x is Pareto superior to y , or y is Pareto dominated by x .

Using the notion of Pareto superiority, we define a Pareto optimal or Pareto efficient state as follows.

Given a set of social states $\{x, y, z, \dots\}$ suppose there is a state v with the property that no other state in the set is Pareto superior to v . Then v is said to be a Pareto efficient or Pareto optimal state for the set $\{x, y, z, \dots\}$.

We can illustrate the Pareto criterion very easily in an Edgeworth box diagram. In Fig. 18.1 both people prefer allocation x to allocation y . Consequently, x is Pareto superior to y . But neither x is Pareto superior to z nor is z Pareto superior to x . Therefore, the unanimity principle does not apply between x and z . The same is true for comparison between y and z . When the unanimity principle does not apply, we say that the two states are Pareto non-comparable.

In moving from y to x , both the individuals are made better off. In moving from x to z , one person is made better off and the other person is worse off. These are special cases of the following general characterizations of Pareto efficient states.

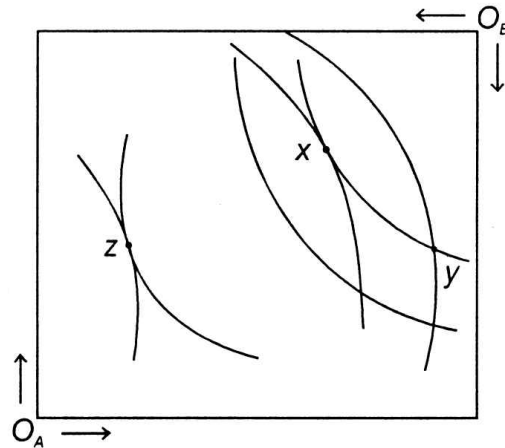


Fig. 18.1 Pareto Superiority and Pareto Non-Comparability

A change from a Pareto efficient state to any other attainable state for making one or more persons better off cannot be done without making some other person(s) worse off.

If the society is in a non-Pareto efficient state, then there is an attainable state such that in that state no one is worse off and at least one person is better off than the non-Pareto efficient state.

In the Edgeworth box example we have considered here, there are many Pareto efficient allocations (such as x , z , etc.). For essentially all cases of interest there will be several and often many Pareto efficient allocations—Pareto efficient states are not unique. Note also that except under pathological conditions, for any set of states we can always find a state that is Pareto efficient for the set.

We can now summarize the qualities of the Pareto efficiency condition as follows. This condition allows us to rank many states of the economy as inferior on social welfare grounds because other states dominate them in utility terms by allowing some or all individuals to be better off while making none worse off. That is, we can use the criterion to identify a set of states as the proper subset of the set of all feasible states which cannot be dominated by other feasible states. Note that the dominance criterion does not require utility comparisons among different individuals: it requires only that any individual's utility be ordinally measurable. We can also easily determine the conditions for a state to be Pareto optimal. (See Chapters 16 and 17.) Finally, the two fundamental theorems provide a two-way link between the set of Pareto efficient states and the set of resource allocation outcomes of perfect competition. (See Chapters 16 and 17.)

The major limitation of the Pareto criterion is that Pareto non-comparability makes its applications limited. This criterion does not rank every Pareto efficient allocation superior to a non-efficient allocation. Thus, its weakness lies in the fact that it yields an incomplete ordering. The Pareto criterion can therefore rarely be used. So economists have developed other criteria to decide whether one alternative is socially preferable to another.

COMPENSATION CRITERIA

In 1939 Nicholas Kaldor devised a social improvement criterion which we discuss below.

Suppose that a policy maker is faced with a choice between two alternatives which are mutually exclusive e.g. build a road from some point, or do not build it. The construction of the road will benefit some (gainers) and hurt some (losers). The remaining individuals are indifferent between road and non-road situations. Let x be the state (allocation of goods among people) when the road is built. On the other hand, if the road is not built, a particular state (allocation of goods) occurs. Let us refer it to as y . We then say that x is strictly preferred to y by the Kaldor compensation criterion (x is Kaldor superior to y , for short), if it is hypothetically possible to undertake costless (lumpsum) redistribution and achieve an allocation z from x such that z is Pareto superior to y . Since z is Pareto superior to y , either the losers are at least as well-off as in y and the gainers are better off than in y , or the losers are better off than in y and the gainers are at least as well-off as in y , or both the gainers and losers are better off than in y . In the first case the gainers are at least able to compensate the losers and still have some net gain. In the second and the third cases the losers are paid more than their losses (since they are better off in z than in y). In all these three cases the gainers are able to more than compensate the losers. This is what the Kaldor compensation criterion demands. That is, according to the Kaldor criterion a change is socially better if those who benefit from it could more than compensate those who are losers. Clearly, compensation is brought about by costless redistribution.

We can also think of hypothetical compensation in terms of money. In this case we can ask the gainers how much money they could be prepared to pay in order to have the road built, and the losers how much money they would be prepared to pay to prevent the construction of the road. If the amount of money of the gainers is greater than the amount of the losers, then clearly, the construction of the road is desirable according to the Kaldor criterion. The compensation the losers get from the gainers for not blocking the

construction of the road can be regarded as a bribe to the losers for accepting the construction.

We formally state the Kaldor compensation principle as follows: Policy I is to be judged better than policy II if the gainers under I can more than compensate those who lose in changing from II to I.

Clearly, for our road/no-road example, policy I is 'build the road' and policy II is 'do not build it'.

Hicks (1939) suggested a social improvement principle which is equivalent to the Kaldor principle. We state the Hicks criterion formally as: Policy I is to be judged better than policy II if those who would lose from I could not profitably bribe the gainers into not making the change from II to I.

We will now discuss certain shortcomings of the Kaldor principle. Hence these criticisms apply to the Hicks principle as well. First, to assert that policy I is better than policy II, the Kaldor criterion implicitly compares individual utilities and assigns a higher weight, in terms of welfare, to the gainers under I than to the losers. But we do not have any a priori criteria for determining such weights. A second shortcoming of the Kaldor principle is that it does not require that compensations actually be paid (in which case policy I leads to a state that is Pareto superior to that achieved under II), but only that they could be paid. Another shortcoming of the Kaldor criterion was pointed out by Scitovsky (1941) which is discussed below.

Suppose that under policy I, the economy produces 10 units of good 1 and 5 units of good 2, and that under policy II the economy produces 5 units of good 1 and 10 units of good 2. There are two consumers *A* and *B* whose preferences are as follows (the first component shows units of good 1 and the second is units of good 2):

A: (5, 5) preferred to (10, 0) which is preferred to (5, 0);

B: (5, 5) preferred to (0, 10) which is preferred to (0, 5).

Assume the following distributions of output under I and II:

<i>Policy</i>	<i>I</i>	<i>II</i>
Total output :	(10, 5)	(5, 10)
Consumer <i>A</i> :	(10, 0)	(5, 0)
Consumer <i>B</i> :	(0, 5)	(0, 10)

We refer to the distributions under I as state x and under II as state y . A is better off under I and B is better off under II. Now, given I, suppose A gives 5 units of good 1 to B . This hypothetical transfer (compensation) generates a new state z in which A has the bundle (5, 0) and B has the bundle (5, 5). In state z , B is better off than in y , while A is just as well off as in y . Therefore, z is Pareto superior to y . Applying the compensation principle, x is better than y . (The gainer [A] can more than compensate the loser [B] in change from I to II).

Next, given II, consider a transfer of 5 units of good 2 from B to A . We obtain a new state v in which the bundles of A and B are (5, 5) and (0, 5) respectively. Since A is better off in v than in x , whereas B is just as well off as in x , y is regarded as better than x by the compensation principle.

The above example shows that I is Kaldor superior to II, and II is Kaldor superior to I. Thus, the Kaldor compensation principle is inconsistent. This combined with the fact that so long as compensations are not actually paid we generally have Pareto non-comparable states, makes the compensation principle an imperfect guide to social policy.

Scitovsky devised a social improvement criterion which necessarily avoids Kaldor's inconsistency. According to the Scitovsky criterion, policy I should be counted superior to policy II if I is Kaldor superior to II, and II is not Kaldor superior to I. That is, I is strictly preferable to II by the Scitovsky criterion (or, I is Scitovsky superior to II) if the move from I to II is a Kaldor improvement but the move from II to I is not a Kaldor improvement. If this double test is not satisfied and if the corresponding double test of superiority of II over I is also not satisfied, then the states are Scitovsky non-comparable.

Though the Scitovsky principle avoids Kaldor's inconsistency, it shares two other shortcomings of the Kaldor criterion. Further, in the case when all states can be reached from each other, as in an exchange economy model, the Scitovsky criterion does not yield any information beyond what is inherent in the notion of Pareto efficiency.

NON-ENVY AND FAIRNESS

The previous sections described some of the properties of the social improvement criteria of Pareto, Kaldor, Hicks and Scitovsky. The question of when one alternative ought to be unambiguously socially better than another remains largely unresolved. In this section we will examine a new criterion for

judging social states, which is explicitly egalitarian in its viewpoint. This criterion is fairness, in the sense of non-envy, where envy occurs when a person derives more utility from another person's consumption bundle than from this own (Foley [1967]). Thus, under fair distribution of goods no person would prefer any other's bundle of goods to his own. On the other hand, under unfair distribution some people would prefer others' bundles to their own. We will investigate here whether this distinction can be fruitfully employed to decide if one distribution is socially better than another.

For discussion on fairness we revert to the exchange economy model. This discussion will abstract from production, although production could be explicitly introduced into the fairness issue (Varian [1974]). The analysis of production in the context of fairness is beyond the scope of this book.

Let $x = (x_1, x_2, \dots, x_m)$ be an allocation of n goods among m persons where $x_i = (x_i^1, x_i^2, \dots, x_i^n)$, with x_i^j being the amount of good j possessed by person i , $i = 1, \dots, m$; $j = 1, 2, \dots, n$. Suppose U_i stands for person i 's utility function. Foley's definition of fairness is that no person should envy another. Person i is said not to envy person k if $U_i(x_k) \leq U_i(x_i)$, that is, if i likes his own bundle as well as he would like k 's. If $U_i(x_i) \geq U_i(x_k)$ for every pair of persons (i, k) , we say that the allocation x is fair.

To illustrate a fair allocation graphically let us assume that there are two persons (A and B) and two goods (1 and 2). The bundles of A and B are $x_A = (x_A^1, x_A^2)$ and $x_B = (x_B^1, x_B^2)$. In Fig. 18.2, the allocation $\bar{x} = (x_B, x_A)$ reverses or swaps the bundles assigned by the allocation $x = (x_A, x_B)$. That is, $\bar{x}$ gives person A person B 's bundle, and it gives person B person A 's bundle. In the Edgeworth box, $\bar{x}$ can be obtained by taking the mirror image of x through the centre of the box c . Therefore, x , c and $\bar{x}$ lie on a straight line. The points x and $\bar{x}$ are equidistant from c .

In Fig. 18.2, A 's indifference curve I_A through x lies above $\bar{x}$. Therefore, A prefers x to $\bar{x}$, and, consequently $U_A(x_A) > U_A(x_B)$. Similarly B 's indifference curve I_B through x lies above $\bar{x}$. Therefore, $U_B(x_B) > U_B(x_A)$. This means that the allocation x is fair. Since x is not Pareto efficient, we can say that fairness need not imply Pareto efficiency. The converse is also true: Pareto efficient allocations are generally not fair.

The existence of a fair and Pareto efficient allocation in an exchange economy can be guaranteed if we assume that the initial endowment w is an equal allocation. For simplicity let us consider the two-person two-good case

again. The initial endowment is $w = (w_A, w_B)$, where $w_A = (w_A^1, w_A^2)$ and $w_B = (w_B^1, w_B^2)$ satisfies the condition $w_A^1 = w_B^1$ and $w_A^2 = w_B^2$. Let $y = (y_A, y_B)$ be a competitive equilibrium allocation in the economy. Then by the First Fundamental Theorem y is Pareto efficient. In fact y is fair also. To see this, recall from the definition of a competitive equilibrium that y_A must maximize U_A subject to $p_1 x_A^1 + p_2 x_A^2 \leq p_1 w_A^1 + p_2 w_A^2$, where p_1 and p_2 are the prices of the goods 1 and 2. Similarly y_B will maximize U_B subject to $p_1 x_B^1 + p_2 x_B^2 \leq p_1 w_B^1 + p_2 w_B^2$. Now, the right hand sides of the above inequalities are the same. That is, each person maximizes his utility subject to the same budget constraint. Therefore, it is impossible for A to end up with a bundle which he finds inferior to B 's and vice-versa. That is, we have $U_A(y_A) \geq U_B(y_A)$ and $U_B(y_B) \geq U_A(y_B)$, or the allocation y is fair.

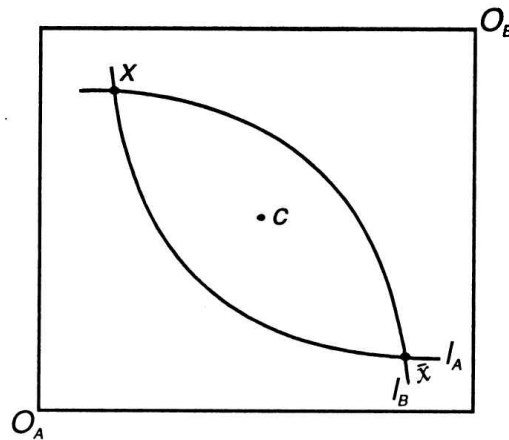


Fig. 18.2 Fair Allocation

Therefore, we have a way of achieving both Pareto efficiency and fairness at the same time. We start at the equal initial endowment and take a move to a competitive equilibrium allocation, which will be both fair and Pareto efficient. Consequently, we have one way of choosing among Pareto efficient allocations. If the choice is between a Pareto efficient allocation that is not fair and an allocation that is Pareto efficient and fair, we choose the fair one.

SOCIAL PREFERENCE AND ARROW'S IMPOSSIBILITY THEOREM

We have looked at several ways to answer the question "When is one state socially better than another?" Each of the answers has been disappointing to some extent. The Pareto criterion is incomplete. The compensation principles are invoked to extend the power of the Pareto principle for ranking

states on the basis of their potential for Pareto improvements through hypothetical compensation. But the Kaldor and the Hicks compensation criteria are possibly inconsistent. The Scitovsky criterion reduces to a question of optimality versus non-optimality. Finally, the fairness criterion may not be consistent with the Pareto principle.

Therefore it becomes necessary to make stronger assumptions than those implicit in the Pareto principle, the compensation criteria and fairness. Since our objective is to have a complete ordering of social states, we need either to have more information or to invoke more value judgements or to have both. (A value judgement is a statement of ethics which cannot be found to be true or false on the basis of factual evidence. Examples of value judgements are the Pareto criterion, the compensation criteria and fairness.)

Our objective is now to investigate the consequences of adding more information on individual utilities as well as invoking somewhat stronger value judgements. More precisely, we consider the problem of 'aggregating' individual utilities into a 'welfare function' that can rank different distributions of utilities among individuals in a complete and consistent manner. Before we discuss the concept of a welfare function it will be worthwhile to investigate how the notion of social preferences can be derived in principle from a group of differing individual preferences.

Initially we assumed that a consumer's preference is defined over his own bundle of goods. We now expand the domain of definition of a consumer's preference and assume that each consumer has preferences over the entire allocation of goods among consumers. Each individual can rank the allocations according to their desirability. The question is: can we use this information to develop a social preference or ranking of the various allocations? This is the most general formulation of social decision making.

Therefore, we need to investigate whether there are social decision mechanisms, that is, ways of aggregating the individual preferences into a social preference. Now, a social preference relation should satisfy certain reasonable requirements. Arrow (1951a) assumed that a social decision rule should meet the following seemingly unobjectionable postulates.

1. Completeness, reflexivity and transitivity: Given any set of complete, reflexive and transitive individual preferences, the social decision rule should result in social preferences that satisfy these properties.
2. Universality or unrestricted domain: Any logically possible set of individual preferences is admissible in determining the social preference.

That is, the same mechanism must be used to aggregate individual preferences regardless of what the individual preferences happen to be.

3. Independence of irrelevant alternatives: The social preference between any two allocations x and y must be independent of individual orderings on other pairs of allocations. Thus, if people's feelings change about some set of irrelevant allocations, but do not change about the pair of allocations x and y , the social preference rule must preserve the social ordering of x and y .
4. Pareto consistency: For any pair of allocations x and y , if everybody strictly prefers x to y , then x must be socially better than y .
5. Non-dictatorship: The social preference relation must be non-dictatorial. That is, one person's preferences should not determine society's preferences.

Arrow was able to show that these five conditions are not compatible with one another: no general social preference relation obeying postulates 1–5 exists. To put the matter another way, the only social preference rule which satisfies properties 1–4 is the rule that makes one person a dictator. This is popularly known as Arrow's impossibility theorem. We state the Arrow's theorem formally.

ARROW'S IMPOSSIBILITY THEOREM: *Any social decision rule which satisfies the requirements of (1) completeness, reflexivity and transitivity, (2) universality or unrestricted domain, (3) Pareto consistency, and (4) independence of irrelevant alternatives, makes one person a dictator. Therefore, there is no rule which satisfies all five requirements.*

To illustrate Arrow's theorem let us assume that there are three social states (x , y and z) and two individuals in the society (A and B). Suppose that A strictly prefers x to y (we will denote this by xP_Ay , where P_A represents the words 'is strictly preferred by A to') and y to z . These preferences are written as xP_Ay , and yP_Az . If the individual is 'rational', then it should be the case that xP_Az , that is, the individual's preferences should be transitive. Suppose also that among the three states, B has preferences zP_Bx , xP_By and zP_By .

Using the preferences of persons A and B , it is possible to see the kind of inconsistencies that can arise in social preferences. Since yP_Az and zP_By , it must be the case that society is indifferent between y and z . Otherwise society's preferences would be in accord with only one individual (and against the other), and this would violate postulate 5 requiring non-dictatorship.

Since A and B strictly prefer x to y , Pareto consistency requires that society strictly prefer x to y . Given that society is indifferent between y and z , the transitivity of the social preference relation demands that society strictly prefer x to z . But, again, this is a violation of property 5 since xP_Az , but zP_Bx . Therefore, in this simple case an inconsistency arises in the attempt to construct a social preference relation. Though this example is a bit contrived, it clearly illustrates the problem of trying to aggregate individual preferences into some reasonable social preference.

A famous example of a social rule which does not work is the principle of majority voting. To see this suppose that there are three allocations x , y and z and three persons A , B and C . We know the order in which the individuals prefer allocations. For example, A 's preference of allocations is (from highest to lowest) x , y , z ; B 's is y , z , x and C 's is z , x , y . Given these preferences, a vote between x and y will declare x as the winner since A and C would vote for x and B would vote for y . Similarly, y wins in a vote between y and z and z wins when there is a vote between x and z . The social ordering is, thus: x is strictly preferred to y , y is strictly preferred to z and z strictly preferred to x . This is not a social ordering at all—it does not satisfy transitivity, that is, x is strictly preferred to y , y is strictly preferred to z and z is strictly preferred to x . The reason for this circularity is that there is too much diversity of preferences among the voters.

So to have social ordering of alternative allocations, we need more information. In fact, we must be able in some way to compare utilities of different people. We have stated in chapter 2 that an individual's utility function is a representation of his preference ordering. (Recall that in this chapter we have defined a consumer's preference ordering over the entire allocation of goods among the consumers rather than on his own bundle of goods. Therefore, the utility function will also be defined over the entire allocation.) Hence a social ordering (if it exists) will also be representable by a social welfare function that assigns a number to each allocation. A social welfare function depends on individual utility functions.

Let us now look at the question of comparability of utilities.

COMPARABILITY REQUIREMENTS

Comparability refers to the sense in which the real numbers attached to different individuals' utility levels can be meaningfully compared. To understand this, suppose that we adopt a policy that will benefit the worst-off person and makes others less happy in a society. In terms of information,

this policy requires only that we can compare the levels of utility of different people in different situations.

We will also be willing to see that others suffer if there were some substantial gains to the worst-off person. That is, we should be able to measure in a comparable way the changes in happiness accruing to different people. More conclusively, for a given individual we must be able to compare the magnitude of that person's change in happiness in moving from allocation 1 to allocation 2 with the change in happiness in moving from allocation 2 to allocation 3. That is, here we need measures of changes in utility that are both cardinal and interpersonally comparable.

Discussions on existence of social welfare functions under alternative comparability assumptions about utility have been made, among others, by Sen (1977), Blackorby, Donaldson and Weymark (1984), Boadway and Bruce (1984) and Chakravarty (1990). We now turn to a brief treatment of social welfare functions.

SOCIAL WELFARE FUNCTIONS

A social welfare function is a function of the utility functions of all individuals in a society. It is assumed to be increasing in individual utility levels. Clearly, this increasingness property is the Pareto improvement criterion. A social welfare function enables us to rank alternative allocations in a society.

Given the preferences of each individual i over the set of allocations, we can construct individual utility functions U_i that reflect the individuals' value judgements. More precisely, person i regards allocation x as better than allocation y if and only if $U_i(x) > U_i(y)$. Clearly, these utility representations are not unique.

For any allocation x we denote the social welfare function by $W[U_1(x), U_2(x), \dots, U_m(x)]$, where m is the number of persons in the society. We say that allocation x is socially better than allocation y if $W[U_1(x), U_2(x), \dots, U_m(x)] > W[U_1(y), U_2(y), \dots, U_m(y)]$. This is certainly guaranteed if everybody in the society strictly prefers x to y and if W is increasing in individual utilities.

Analytically this is all well and good, but how can we utilize such a framework in practice? For practical purposes we require specific functional forms of W . Perhaps the oldest and the best-known form is the classical utilitarian or Benthamite social welfare function:

$$W(U_1, U_2, \dots, U_n) = \sum_{i=1}^m U_i \quad (18.1)$$

In this case, social welfare is the unweighted sum of individual utilities. A slightly less restrictive form is the generalized utilitarian or weighted-sum-of-utilities welfare function:

$$W(U_1, U_2, \dots, U_m) = \sum_{i=1}^m a_i U_i \quad (18.2)$$

where a_i , $i = 1, 1, \dots, m$, are positive constants. The positive weights $a_1, a_2, \dots, a_n$ indicate the importance of each consumer's utility in the social welfare function.

Other specific forms are the Bernoulli-Nash social welfare function:

$$W(U_1, U_2, \dots, U_m) = \prod_{i=1}^m U_i \quad (18.3)$$

and the generalized Bernoulli-Nash social welfare function:

$$W(U_1, U_2, U_m) = \prod_{i=1}^m U_i^{a_i} \quad (18.4)$$

where the utilities U_i 's are positive and a_i 's are positive weights. In this case, social welfare is the product of (unweighted or weighted) utilities of the individuals. Note that the Bernoulli-Nash and the generalized Bernoulli-Nash welfare function turn out respectively to be the utilitarian and the generalized utilitarian forms in logarithms of utility.

Another interesting welfare function is the Rawlsian or maximin social welfare function:

$$W(U_1, U_2, \dots, U_m) = \min \{U_1, \dots, U_m\}. \quad (18.5)$$

Here social welfare is identified with the utility of the worst-off person. This social welfare function is termed 'maximin' because in maximization it involves maximizing the minimum value of the utility vector $(U_1, \dots, U_m)$ (Rawls [1971]).

The social indifference curves for the utilitarian, Bernoulli-Nash and maximin welfare functions are shown in panels (a), (b) and (c) respectively of Fig. 18.3. The social indifference curves are also called isowelfare curves since they depict those distributions of utility that have constant welfare.

In panel (b) of Fig. 18.3 the indifference curves exhibit diminishing marginal rate of substitution of person 2's utility for person 1. This assumption would seem to be reasonable if society is basically egalitarian and progressively less willing to make person 2 better off at the expense of making 1 worse off. The straight line social indifference curve in panel (a) means that the marginal rate of substitution between the utilities of persons 1 and 2 is

the same along the entire indifference curve. In contrast, no substitution is possible between the utilities of the persons when the indifference curve is of the type given by Fig. 18.3 (c).

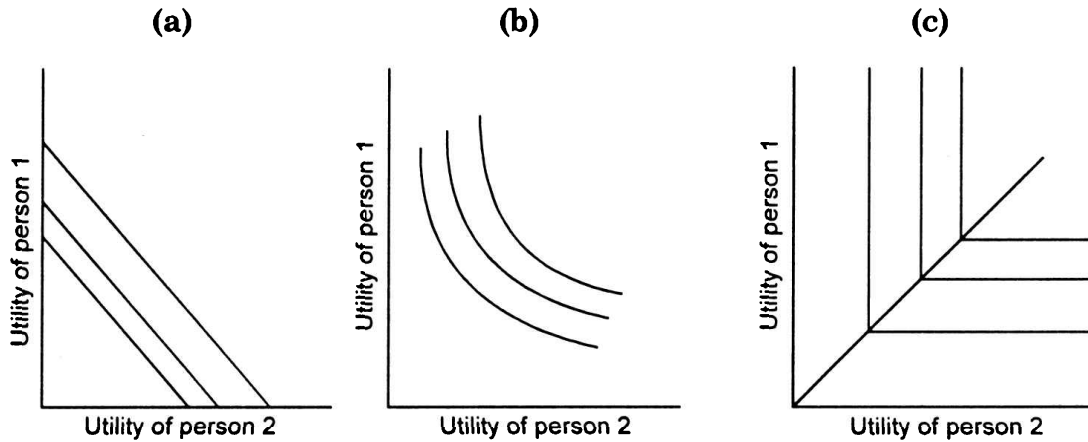


Fig. 18.3 Social Indifference Curves

It therefore turns out that we do not have a unique method of ordering allocations, that is, a unique social welfare function. Each of these functions represent some value judgements about comparability of different individuals' utility functions.

MAXIMIZATION OF SOCIAL WELFARE

Given a social welfare function, we can examine the problem of welfare maximization of society. For simplicity let us go back to the model considered in the section 'Combining Production and Exchange' of Chapter 17. The assumptions of the model are as follows. There are two consumers (A and B) and two goods (1 and 2). Each of these two goods is produced by two firms F and G using capital (K) and labor (L) as factors of production. These factors are homogeneous and perfectly divisible. The production functions of the firms give rise to smooth isoquants that are strictly convex to the origin. These production functions are independent—there are no production externalities. The firms maximize their profits and consumers maximize their utilities. The utility functions of the consumers represent indifference curves which are strictly convex the origin. Utilities are assumed to be independent—there are no consumption externalities. A consumer's income is determined by his initial endowments (if any) of the two goods, the ownership of factors of production and shares of this firms owned. Finally, factor as well as output markets are perfectly competitive.

Let us use the notation x_i^j to indicate how much of good j ($j = 1, 2$) is possessed by individual i ($i = A, B$). Then the allocation $x = (x_A; x_B) = (x_A^1, x_A^2; x_B^1, x_B^2)$ consists of the list of how much each person has each of the goods.

If we have a total amount X^1 and X^2 of goods 1 and 2 to distribute among the consumers, we can pose the following welfare maximization problem:

$$\max W[U_A(x_A), U_B(x_B)] \quad (18.6)$$

such that $x_A^1 + x_B^1 = X^1$ and $x_A^2 + x_B^2 = X^2$. Thus, our objective is to find the feasible allocation that maximizes social welfare.

We note that in the above maximization problem each individual's utility function is defined on his own bundle of goods rather than over the entire allocation. The welfare function directly is a function of individual utility levels but indirectly is a function of the consumption bundles of different individuals. This special form of social welfare function, which is less general than the general $W(U_1(x), U_2(x), \dots, U_n(x))$ considered in the earlier section, is known as the Bergson (1938)-Samuelson (1956) social welfare function or individualistic social welfare function.

Now, let us turn to the problem of welfare maximization. We begin by observing that a welfare maximizing allocation is a Pareto efficient allocation. To prove this, suppose it were not. Then we can move to a Pareto improving situation that gives everyone at least as large a utility, and someone higher utility. Given that the welfare function is increasing in individual utility levels, the Pareto improving situation generates higher welfare level than the welfare maximizing allocation. A contradiction! Therefore, a maximal welfare allocation is Pareto efficient.

We can illustrate this situation graphically using the utility possibilities frontier which is constructed as follows. If we take a particular point on the economy's production possibilities frontier, we specify a particular combination of goods 1 and 2 produced. (Recall that the economy's production possibilities set is obtained by combining the production possibilities sets of the firms F and G .) Given this particular combination of the two goods, we can construct an Edgeworth box diagram and derive the corresponding contract curve. In view of our results in the section 'Combining Production and Exchange' of Chapter 17 we know that the economy will then be simultaneously in general equilibrium of production and exchange if the marginal rate of transformation between goods 1 and 2 is the same as the marginal rate of substitution between the goods.

Suppose that in Fig. 18.4 the output of the two goods produced by the economy is given by the point M . By dropping perpendiculars from M on the axes, we construct the Edgeworth box diagram for the persons A and B . Every point on the contract curve $OCDEM$ in the Edgeworth box is a point of general equilibrium of exchange. However, the economy will be simultaneously in general equilibrium of production and exchange at D , where the common marginal rate of substitution between the two goods for the two persons equals the marginal rate of transformation.

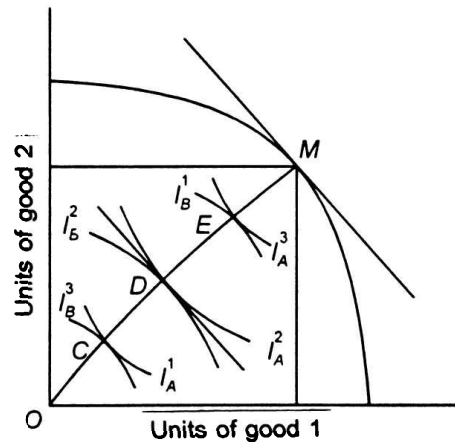


Fig. 18.4 Production Possibilities Frontier and Edgeworth Box

By mapping the contract curve of the figure from the output space to the utility space, we get the corresponding utility possibilities curve, the locus of various combinations of utility obtained by persons A and B when the economy is in general equilibrium of exchange.

Note that the points C' , D' and E' on the utility possibilities curve correspond respectively to the points C , D and E on the contract curve. At D' , the economy is simultaneously Pareto efficient in consumption and production.

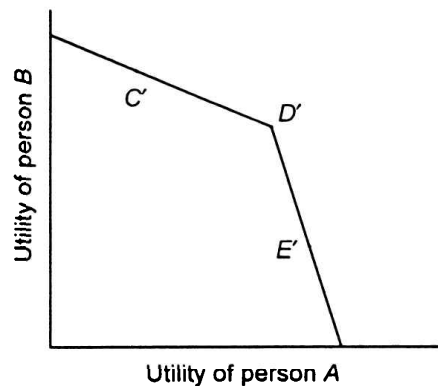


Fig. 18.5 Utility Possibilities Curve

By taking another point on the production possibilities frontier, we can construct a different Edgeworth box diagram and contract curve. This in turn will generate a different utility possibilities curve (given by $S'V'T'$ in Fig. 18.6) and another point of Pareto efficiency in consumption and production (given by T'). This process can be repeated any number of times. By joining the resulting Pareto efficient points of production and consumption (that is, points like D' , T'), we derive the utility possibilities frontier (sometimes referred to as the grand utility possibilities curve).

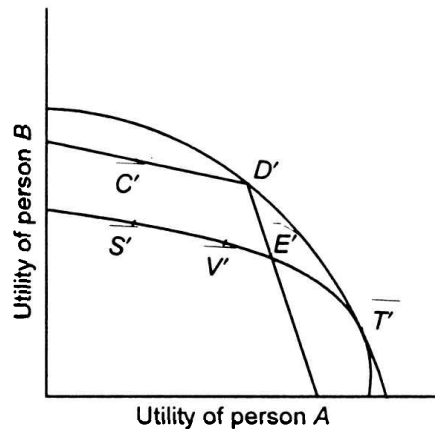


Fig. 18.6 Utility Possibilities Frontier

To determine the point of maximum social welfare we superimpose social indifference curves on the utility possibilities frontier. The maximum social welfare is attained at the point where the utility possibilities frontier is tangent to a social indifference curve. In Fig. 18.7 point E is the optimal point of social welfare, because it gives the highest level of W achievable with the given utility possibilities frontier. In order to determine the socially optimal allocation of goods, it is necessary to go back from point E to the Edgeworth box. It is clear from the above discussion that the maximal welfare point is Pareto efficient.

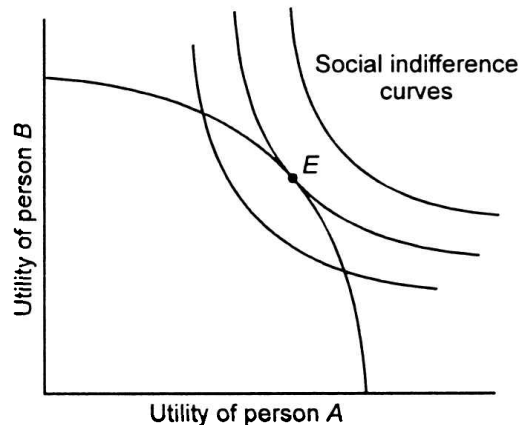


Fig. 18.7 Welfare Maximization

Next, we know that any point on the utility possibilities frontier is Pareto efficient. We can evidently have a social welfare function whose indifference curve will be tangent to a point on the utility possibilities frontier. That is, every Pareto efficient allocation is a maximal welfare allocation.

Now, given that all markets are perfectly competitive we have an intimate relationship between Pareto efficient allocations and market equilibria: a competitive equilibrium is Pareto efficient, and, under appropriate assumptions, any Pareto efficient allocation is a competitive equilibrium.

Since we know the relationship between Pareto efficiency and welfare maxima, we can conclude that all welfare maximizing allocations are competitive equilibria and that all competitive equilibria are welfare maxima for some welfare function.

SAMUELSON SOCIAL IMPROVEMENT CRITERION

Samuleson (1950) developed a social improvement principle which in terms of the utility possibilities frontier can be defined as follows. An allocation $x = (x_A, x_B)$ is said to be strictly preferred to another allocation $y = (y_A, y_B)$ by the Samuelson criterion (or, x is Samuelson superior to y) if the utility vector $U(x) = (U_A(x_A), U_B(x_B))$ corresponding to x lies on a higher frontier than the utility vector $U(y) = (U_A(y_A), U_B(y_B))$ corresponding to y . Fig. 18.8 illustrates the situation.

Consider now the utility vector $[U_A(z_A), U_B(z_B)]$ corresponding to the allocation $z = (z_A, z_B)$. Suppose $U(z) = [U_A(z_A), U_B(z_B)]$ lies on the lower utility possibilities frontier. The allocation z is accessible from y . We can clearly have an allocation $w = (w_A, w_B)$ accessible from x that provides the utility

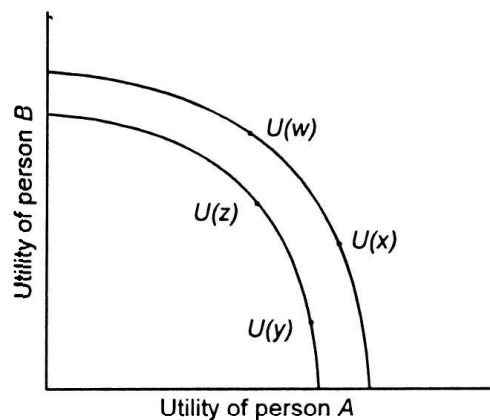


Fig. 18.8 Samuelson Criterion

vector $U(w) = [U_A(w_A), U_B(w_B)]$, where at least one person enjoys a higher utility level than z and neither with a lower utility level. That is, the utility vector $U(w)$ lies to the north-east of the vector $U(z)$. Using this we can now restate the Samuelson criterion as follows:

Allocation x is Samuelson superior to allocation y if, for any z accessible from y , there is a w accessible from x that is Pareto superior to z .

Our figure should make it clear that the Samuelson criterion cannot be inconsistent: it is impossible to have x Samuelson superior to y and y Samuelson superior to x . However, in case the utility possibilities frontier is unique (that is, when all states are mutually accessible), the Samuelson criterion becomes undefined. No state is Samuelson superior to another.

Example 18.1

Suppose that the production possibilities frontier of an economy for its two goods (1 and 2) is given by $T = X^1 + 2X^2 - 1 = 0$. There is only one person in the economy whose utility function is given by $U(X^1, X^2) = \log X^1 + \log X^2$. The economy has fixed factor supplies. Find the values of X^1 and X^2 that will maximize the utility function of the individual.

Welfare or utility maximization requires that the person's marginal rate of substitution between two goods be equal to the marginal rate of transformation. That is,

$$\frac{\delta U / \delta X^1}{\delta U / \delta X^2} = \frac{\delta T / \delta X^1}{\delta T / \delta X^2}, \quad (18.7)$$

which gives $X^2/X^1 = \frac{1}{2}$, or $2X^2 = X^1$. The welfare maximizing choice of X^1 and X^2 will lie on the production possibilities frontier. Therefore, substituting $2X^2 = X^1$ in T we get $4X^2 = 1$, which gives $X^2 = 1/4$. This in turn shows that $X^1 = \frac{1}{2}$. Therefore, the welfare maximizing choice of (X^1, X^2) is $\left(\frac{1}{2}, \frac{1}{4}\right)$.

Applied Example 18.1 Welfare Comparisons

Comparison of real national income is one of the most frequently performed exercises in applied economics. The significance attached to comparisons of real national income depends heavily on the implicit welfare content. Sen

(1976) presented a general approach to real income comparisons which makes distribution an integral part of such comparisons. Sen applied his procedure of real income comparison to study regional differences in the rural standard of living in India. The comparisons are made in terms of the Gini social welfare function of the rural community of any given state putting that community in position of the rural population of every other state respectively.

To define the Gini social welfare function let us consider the Gini index G of inequality of the income distribution of a population of size n , where

$$G = \frac{1}{2n^2\bar{Z}} \sum_{i=1}^n \sum_{j=1}^n |Z_i - Z_j|. \quad (18.8)$$

Here Z_i is the income of person i ($i = 1, 2, \dots, n$) and $\bar{Z} = \sum_{i=1}^n Z_i / n$, the mean income of the society. The index G can be interpreted as follows. In any pair-wise comparison, the individual with lower income may suffer from depression upon discovering that his income is lower. The average of all such depressions in all pair-wise comparisons, when divided by the mean income, becomes the Gini index if the extent of depressions is given by the differences in incomes. The index G is thus a scalar representation of interpersonal income differences within a given population. It is bounded between 0 and 1, where the lower bound is achieved whenever incomes are equal, that is, $Z_1 = Z_2 = \dots = Z_n$. If we assume that incomes are ordered non-decreasingly, that is $Z_1 \leq Z_2 \leq \dots \leq Z_n$, then G can be written as

$$G = 1 - \frac{1}{n^2\bar{Z}} \sum_{i=1}^n [2(n-i) + 1]Z_i \quad (18.9)$$

from which we have

$$G_w = (1 - G)\bar{Z} = \frac{1}{n^2} \sum_{i=1}^n [2(n-i) + 1]Z_i \quad (18.10)$$

G_w is the Gini social welfare function. It is increasing under rank preserving increase in individual incomes. It increases under a transfer of income from a rich person to a poor person where the transfer is so small that it does not alter the ranks of the persons in the income distribution (see Chakravarty [1990] for further discussions on the Gini social welfare function).

Sen shows how comparison of the Gini social welfare function squeezes out as much juice as possible from constant-price-real-income comparison. In his numerical illustration Sen covered 15 states and since welfare com-

parisons are done for each state, $15 \times 14 = 210$ pair-wise comparisons are involved. The calculations are based on the per capita expenditure data for rural India thrown up by the Indian National Sample Survey Organization for the accounting period 1961-62. (Incomes are approximated by expenditures for calculating the Gini social welfare function.) Values of Gini social welfare functions for each state at prices of every other state were calculated. A comparative picture of the welfare values reveals that at one end there is Punjab which has a higher level of welfare than all other states and at the other end there is Bihar without a single 'victory'.

Appendix 18.1

WELFARE MAXIMIZATION

In this appendix we consider the problem of welfare maximization technically. Recall that we can write society's production possibilities frontier as $T(X^1, X^2) = 0$, where X^i is the total amount of good i , $i = 1, 2$. The Lagrangian expression for this problem is

$$\mathcal{L} = W[U_A(x_A^1, x_A^2), U_B(x_B^1, x_B^2)] - \lambda[T(X^1, X^2)] \quad (18.11)$$

The first order conditions for a maximum are

$$\frac{\delta \mathcal{L}}{\delta x_A^1} = \frac{\delta W}{\delta U_A} \cdot \frac{\delta U_A(x_A^1, x_A^2)}{\delta x_A^1} - \lambda \frac{\delta T(X^1, X^2)}{\delta X^1} \cdot \frac{\delta X^1}{\delta x_A^1} = 0,$$

$$\frac{\delta \mathcal{L}}{\delta x_A^2} = \frac{\delta W}{\delta U_A} \cdot \frac{\delta U_A(x_A^1, x_A^2)}{\delta x_A^2} - \frac{\lambda \delta T(X^1, X^2)}{\delta X^2} \cdot \frac{\delta X^2}{\delta x_A^2} = 0,$$

$$\frac{\delta \mathcal{L}}{\delta x_B^1} = \frac{\delta W}{\delta U_B} \cdot \frac{\delta U_B(x_B^1, x_B^2)}{\delta x_B^1} - \frac{\lambda \delta T(X^1, X^2)}{\delta X^1} \cdot \frac{\delta X^1}{\delta x_B^1} = 0,$$

$$\frac{\delta \mathcal{L}}{\delta x_B^2} = \frac{\delta W}{\delta U_B} \cdot \frac{\delta U_B(x_B^1, x_B^2)}{\delta x_B^2} - \lambda \frac{\delta T(X^1, X^2)}{\delta X^2} \cdot \frac{\delta X^2}{\delta x_B^2} = 0,$$

$$\frac{\delta \mathcal{L}}{\delta \lambda} = T(X^1, X^2) = 0. \quad (18.12)$$

From the first four of these equations we have

$$\begin{aligned} \frac{\delta U_A / \delta x_A^1}{\delta U_A / \delta x_A^2} &= \frac{\delta T / \delta X^1}{\delta T / \delta X^2}, \\ \frac{\delta U_B / \delta x_B^1}{\delta U_B / \delta x_B^2} &= \frac{\delta T / \delta X^1}{\delta T / \delta X^2}. \end{aligned} \quad (18.13)$$

Therefore, welfare maximization requires that each person's marginal rate of substitution between two goods must be equal to the marginal rate of transformation. This is also the first order condition for the Pareto efficiency problem. The intuitive reasoning behind the equivalence between these two have already been discussed in the main text of the chapter.

EXERCISES

Multiple Choice

1. In constructing the utility possibilities frontier, we make interpersonal comparisons of utility:
 - a) Always.
 - b) Often.
 - c) Sometimes.
 - d) Never.
2. In a two-person two-good economy, the maximum social welfare is reached at
 - a) Any point on the utility possibilities frontier.
 - b) Any point on the social indifference curve.
 - c) The point where the common marginal rate of substitution between the two goods for the two persons equals the marginal rate of transformation.
 - d) The point of tangency of the utility possibilities frontier and a social indifference curve.

3. The locus of Pareto efficient points in production and exchange is the
 - a) The production possibilities frontier.
 - b) The social indifference curve.
 - c) The utility possibilities frontier.
 - d) The utility possibilities curve.
4. Which of the following statements is true about the Samuelson criterion for welfare improvement?
 - a) It requires interpersonal comparisons of utility.
 - b) It is possibly inconsistent.
 - c) For an allocation x to be Samuelson superior to another allocation y we require the utility possibilities frontier of y to lie strictly inside that of x .
 - d) It is well-defined even if all alternatives are mutually accessible.
5. Which of the following statements is not true about the Kaldor criterion for welfare improvement?
 - a) It is possibly inconsistent.
 - b) It requires interpersonal comparisons of utility.
 - c) It does not require that the compensations actually be paid, but only that they *could* be paid.
 - d) None of the above.

Quantitative Problems and Questions

1. Prove or disprove the following:
 - a) A welfare maximizing allocation is Pareto efficient.
 - b) A utilitarian has the ethical objective of seeking to maximize the sum of individual utilities, and so if a fixed sum of money is to be distributed between two individuals he would favor complete equality.
 - c) Suppose all states in an economy are mutually accessible. Then a state x is Scitovsky superior to another state y if and only if x is Pareto efficient and y is not.
 - d) A Pareto efficient allocation is always fair.
 - e) One Pareto efficient state of an economy can be the allocation in which one person gets all the goods and services and the rest of the population gets nothing.

2. Explain why a point off the contract curve is a Pareto dominated allocation.
3. What is Kaldor's compensation criterion? Discuss the difficulties with relying upon the Pareto efficiency principle augmented by the compensation criterion.
4. Suppose that the linear production possibilities frontier for an economy is given by $X_1 + 2X_2 = 180$, where X_1 and X_2 represent quantities of goods 1 and 2 respectively. There are two individuals (A and B) in this economy, each with an identical utility function of the form $U(x_j^1, x_j^2) = \log x_j^1 + \log x_j^2$, where $x_j^1(x_j^2)$ is person j 's consumption of good 1(2), $j = A, B$.
 - a) Suppose that production of good 2 is set at 10. What would the utility possibilities frontier for this economy be?
 - b) Suppose that the production of good 2 is set at 30. What would the utility possibilities frontier be?
 - c) How should the production of good 2 be chosen so as to ensure the best utility possibilities frontier?
5. a) Refer to the example considered in the text for demonstrating inconsistency of the Kaldor compensation criterion. Are the two distributions x and y Pareto comparable?
 - b) Consider the following economic states.

State 1 : $x_A^1 = 20, x_B^1 = 20, x_A^2 = 20, x_B^2 = 20;$

State 2 : $x_A^1 = 18, x_B^1 = 26, x_A^2 = 26, x_B^2 = 18;$

State 3 : $x_B^1 = 18, x_B^1 = 26, x_A^2 = 18, x_B^2 = 26;$

where x_A^1 is person A 's consumption of good 1 and so on. The utility functions of the persons A and B are given by $U(x_A^1, x_A^2) = x_A^1 + x_A^2$ and $U(x_B^1, x_B^2) = x_B^1 + x_B^2$ respectively. What can you say about the ranking of the three states by the Pareto and the Kaldor criteria?

6. Consider the following conditions which Arrow imposed on a social welfare function: unlimited domain, Pareto consistency, independence of irrelevant alternatives and non-dictatorship.
 - a) Take any set of 3 of these 4 conditions. Construct a social welfare function which satisfies these three axioms. Explain why your rule

satisfies these conditions. Also explain why the fourth condition is violated.

b) Repeat (a) for another combination of 3 conditions.

7. Suppose that there are two individuals in an economy. The utilities of these individuals under three possible social states are shown below.

<i>State</i>	<i>Utility A</i>	<i>Utility B</i>
1	25	25
2	35	20
3	22.5	27

Individuals do not know which utility number (*A* or *B*) they will be assigned. Which social state will be preferred if an individual adopts the following strategies in his voting behavior to deal with uncertainty?

- Choose the Rawlsian maximin strategy.
 - Assume that there is 50-50 chance of getting either of the two numbers. Choose that strategy which maximizes expected utility.
 - Assume that there is 60% chance of getting the lower utility and 40% chance of getting the higher utility. Choose that state which maximizes expected utility.
 - Choose the maximax criterion. (The maximax criterion maximizes the welfare of the best off person.)
8. Consider a two-person two-good economy. The two persons have the same increasing utility function whose indifference curves are strictly convex to the origin. Explain why the two persons will receive the same bundle at any fair allocation.
9. Show that if transitivity is dropped, majority voting criterion can generate an acceptable social preference ordering.

19

Linear Models

Models in which mathematical relations have the form of linear equations or linear inequalities are called linear models. Optimization (maximization or minimization) of linear economic functions requires tools which are different from the ones employed for optimization within the calculus framework. This follows from the fact that neither the first order nor the second order conditions within the calculus framework can be satisfied when the economic function to be optimized is linear. Thus, linear models require a separate discussion. Two of the most important examples of linear models are input-output analysis and linear programming.

Input-output analysis, a creation of Professor W.W. Leontief, is an attempt to take account of general equilibrium in the empirical analysis of production (Leontief, [1951], Leontief and others, [1953]). Its first important characteristic is that it deals almost exclusively with production. The purpose of the model is to determine what can be produced, and the quantities of different intermediate products which must be used up in the production process, under the existing state of technology and given quantities of available resources. The problem is thus essentially technological in the sense that output level is mainly controlled by the state of technology and resource availability. Demand theory does not come into the picture. Strictly speaking,

this is true for the open model considered here. In an open input-output model, the final demand sectors taken to be outside the production economy. Consequently, final demands are not influenced by any change inside the production economy. (In contrast, in a closed input-output model, labor is regarded as a produced commodity and consumption as the raw materials used up in the production of labor. Therefore, some demand theory must enter into the analysis to show how labor supply is related to consumption demands.) The open model we consider here is a flow or statical model. Dynamic input-output models which involve time and capital stocks are not discussed here.

The second important feature of the considered model is that it deals with empirical issues. The empirical investigation with the help of statistical measurement has been possible because of the simplified form of the model.

The third important feature of the model is its general equilibrium character. The input-output approach divides the economy into a number of sectors or industries including the government and households as industries of final demand. Each industry is viewed as selling its product to other industries; these products become raw materials for purchasing industries. Likewise, each industry employs the output of other industries as its inputs. Therefore, the interdependence of each industry on the others is established. Statistical data gathered around the basic framework of the system gives a very useful picture of the interindustry flows of goods and services. Input-output analysis thus provides a detailed breakdown of macroaggregates and money flows.

The mathematical formulation of an input-output model is developed with linear equations. Hence, as stated earlier, an input-output model is a linear model. The theory of input-output can be regarded as a peculiarly simple form of linear programming. This is another distinctive feature of input-output analysis. In the open static model, the simplest Leontief system, in which no substitution of inputs is allowed, the optimal solution turns out to be a uniquely efficient possible solution.

The empirical implementability of the input-output model requires certain oversimplifying assumptions. The first assumption is that each industry produces only one homogeneous output. That is, there is no joint production in any industry in the economy. But this assumption can be relaxed to some extent by interpreting this good as a composite commodity which is made of items produced in fixed proportions. For instance, such a composite good can consist of packages of chocolates and sugar in which there are always twelve chocolates and one kilogram of sugar.

Perhaps a more serious requirement of the model is that in any production process in the economy all the inputs are employed in rigidly fixed proportions, and as the level of output increases the amounts of inputs used increase proportionately. In addition to this, Leontief also assumed that the production functions exhibit constant returns to scale and that the isoquant surface is convex. The first requirement is far more restrictive than the constant returns to scale assumption. As discussed earlier, constant returns to scale means that an equiproportionate variation in all inputs changes the output by the same proportion. This assumption is perfectly consistent with the substitution of one factor for another. A production function that fulfills the constant returns to scale condition allows both labor-intensive and capital-intensive processes. Suppose that we have a plot of land in which we can employ 10 persons with 10 ploughs to get an output level of 30 units. Assuming that the underlying production function exhibits constant returns to scale, if we want to get an output of 60 units we must employ 20 persons with 20 ploughs. If 20 persons are not available, the same 10 persons can be employed with 10 tractors instead of 20 ploughs under the assumption that the work done by a person gets doubled if he uses a tractor in place of a plough. However, such a possibility is ruled out under the Leontief fixed proportion technology. Under the Leontief technology if exactly one machine is required for each person at any level of output, then this 1:1 ratio will be used no matter how much the size of the firm increases. Finally, a third limitation of the Leontief input-output model lies in its exclusive emphasis on the production side of the economy.

We have already stated that one feature of the input-output model is its linear programming interpretation. Given that linear programming is used widely in economics, it seems worthwhile to discuss a bit more of the economic applications of linear programming. We have therefore divided this chapter into two parts. While in the first part, we present the input-output system, the subject of the second part is a brief discussion on the applications of linear programming.

INPUT-OUTPUT ANALYSIS

Numerical Illustration

Leontief considered an economy in which goods A , B , C , etc., are produced in their respective industries with the help of a primary resource such as labor and with the help of other inputs like A , B , C , etc. Thus, in the Leontief

formulation we do not find an industry which is selling to another industry but buying nothing from its buyer. For instance, we require coal to produce steel and steel is used for machines required to dig coal mines to get coal.

To illustrate this, let us consider an economy consisting of two industries, steel and coal. Each of these industries employs the product of the other in its productive process. Further, each industry directly uses some of a primary factor called labor in its manufacture.

A simplified picture of the economy under consideration is shown in Table 19.1.

Table 19.1 Numerical Input-Output

		<i>Producer of input</i>		
		<i>Steel</i>	<i>Coal</i>	<i>Labor</i>
User of output	Steel	20	30	8
	Coal	150	15	32
	Final demand	30	45	0
	Total output	200	90	40

Column 1 shows that the total output of the steel industry is 200 units per year. Of these 200 units, 20 units are required as inputs by the steel industry itself, 150 units are required as inputs in order to make coal production possible, and 30 units go directly to final consumption, that is, to households and the government. A similar explanation holds for column 2 of the table. The third column shows that society has 40 units of labor endowment. The amounts of labor inputs of the steel and the coal industries are given by 8 and 32 units respectively.

We can explain the rows of the table in an analogous manner. The first row shows that 200 units of steel are produceable using 20 units of steel, 30 units of coal and 8 units of labor. A similar interpretation holds for the second row. The third row shows that the households and government as final demanders require 30 units of steel and 45 units of coal. Note that we have assumed that labor is not directly consumed.

From Table 19.1 it should be clear that the economy under consideration can be regarded as a machine that uses up labor and produces final consumption.

Analytical Formulation

To translate Table 19.1 from a descriptive device into an analytical form, let us introduce some notation. The two industries in the economy will be referred to as industry *C*(coal) and industry *S* (steel) respectively. Let x_i denote the total output of industry i , $i = C, S$ and let x_{ij} denote the output of industry i that is used as an input in industry j , $j = C, S$. Suppose f_i stands for the output of industry i that appears as final demand, $i = C, S$. We write l_{oi} ($i = S, C$) for the labor input of industry i . If l_0 denotes the total labor endowment, then $l_0 = l_{0s} + l_{0c}$. We can present these figures in Table 19.2.

Table 19.2 Analytical Input-Output

		<i>Producer of input</i>		
		<i>Industry S</i>	<i>Industry C</i>	<i>Labor</i>
User of output	Industry S	x_{SS}	x_{CS}	l_{0s}
	Industry C	x_{SC}	x_{CC}	l_{0c}
	Final demand	f_S	f_C	0
	Total output	x_S	x_C	l_0

Using arguments analogous to those employed in the context of Table 19.1, we say that industry *S* produces x_S units of output with the help of x_{SS} units of good *S*, x_{CS} units of good *C* and l_{0s} units of labor. A similar statement can be made for the output level x_C . It should be clear that these statements describe the technological possibilities available to the economy.

Let F^S and F^C stand for the production functions of the industries *S* and *C* respectively. Then

$$\begin{aligned} x_S &= F^S(x_{SS}, x_{CS}, l_{0s}) \\ x_C &= F^C(x_{SC}, x_{CC}, l_{0c}). \end{aligned} \quad (19.1)$$

We also know that

$$\begin{aligned} x_{SS} + x_{SC} + f_S &= x_S \\ x_{CS} + x_{CC} + f_C &= x_C \\ l_{0s} + l_{0c} &= l_0. \end{aligned} \quad (19.2)$$

Clearly, with the help of the formulation of Eqs. (19.1), Table 19.1 gives us $200 = F^S(20, 30, 8)$ and $90 = F^C(150, 15, 32)$.

As stated earlier, Leontief assumed constant returns to scale, convexity of isoquant surfaces, and fixed coefficients of production; that is, certain minimal input of each commodity (possibly zero) is necessary per unit of output of each commodity. To describe the fixed coefficient production function analytically, let a_{ij} be the required minimal input of commodity i per unit of output of commodity j ($i = C, S, O$ [labor], and $j = C, S$). Then

$$\begin{aligned} x_S &= \min \left(\frac{x_{SS}}{a_{SS}}, \frac{x_{CS}}{a_{CS}}, \frac{l_{OS}}{a_{OS}} \right); \\ x_C &= \min \left(\frac{x_{SC}}{a_{SC}}, \frac{x_{CC}}{a_{CC}}, \frac{l_{OC}}{a_{OC}} \right). \end{aligned} \quad (19.3)$$

If the quantities x_{SS} , x_{CS} and l_{OS} are multiplied by a constant $k > 0$, then x_S also gets multiplied by k . Similarly, x_C changes equiproportionally, if we change x_{SC} , x_{CC} and l_{OC} equiproportionally. This means that the Leontief production function in Eq. (19.3) meets constant returns to scale. The isoquants for this production function are L-shaped. Therefore, isoquant surfaces are convex. Suppose a_{ij} is zero for some i and j (that is, good i is not needed at all in industry j). Then we can either drop the term x_{ij}/a_{ij} as an argument of the production function, or interpret it as $+\infty$, in which case it will be larger than the remaining arguments of the production function. That is, whenever, a_{ij} is zero for some i and j , say $i = C$ and $j = C$, x_C in Eq. (19.3) cannot be equal to x_{CC}/a_{CC} .

From Eq. (19.3) it follows that x_S is the smallest of the ratios x_{SS}/a_{SS} , x_{CS}/a_{CS} and l_{OS}/a_{OS} . On the other hand, x_C is the smallest of x_{SC}/a_{SC} , x_{CC}/a_{CC} and l_{OC}/a_{OC} . We can therefore, write Eq. (19.3) alternatively as

$$\begin{aligned} a_{SS}x_S &\leq x_{SS}, \quad a_{CS}x_S \leq x_{CS}, \quad a_{OS}x_S \leq l_{OS}, \\ a_{SC}x_C &\leq x_{SC}, \quad a_{CC}x_C \leq x_{CC}, \quad a_{OC}x_C \leq l_{OC}, \end{aligned} \quad (19.4)$$

where equality will hold at least once in each row. However, if we assume that none of the goods is free, then equality will hold everywhere. To understand this, let us consider the following example. Suppose we require two tons of iron ore to produce one ton of iron. Clearly, one ton of iron can also be produced by using more than two tons of iron ore. But as long as iron ore is not freely available, it will be silly to use more than the absolutely required two tons.

From now on throughout the chapter we will assume that no good is free. Then under this realistic assumption we have $x_{ij}/x_j = a_{ij}$ where $i = C, S, O$

(labor) and $j = C, S$. The numbers a_{ij} are called input-output coefficients for the economy.

We can divide each item in the first row of Table 19.2 by the first column total x_S and each item in the second row by the second column total x_C , and since $x_{ij}/x_j = a_{ij}$, we get Table 19.3.

Table 19.3 Analytical Input-Output

		<i>Producer of input</i>		
		<i>Industry S</i>	<i>Industry C</i>	<i>Labor</i>
User of output	Industry S	a_{SS}	a_{CS}	a_{OS}
	Industry C	a_{SC}	a_{CC}	a_{OC}
	Final Demand	f_S	f_C	0
	Total output	x_S	x_C	l_O

Table 19.3 is therefore a translation of Table 19.2 in the sense that the gross input uses of the latter are transformed into input-output coefficient of the former.

Given an input-output table of the type of Table 19.3, the input-output coefficient a_{ij} can be combined with the specified final demands to determine the gross output x_j required to produce the set of final demands. This is shown below.

Solving an Input-Output System

We assume that the final demands f_S and f_C are given. The question that then arises is: how much each of the goods C and S will have to be manufactured, given that f_S and f_C are the final demands for the two goods? To answer this, let us consider the materials balance given by Eqs. (19.2). Since $x_j = x_{ij}/a_{ij}$, we can rewrite Eqs. (19.2) as

$$a_{SS}x_S + a_{SC}x_C + f_S = x_S$$

$$a_{CS}x_S + a_{CC}x_C + f_C = x_C \quad (19.5)$$

$$a_{OS}x_S + a_{OC}x_C = l_O. \quad (19.6)$$

From Eqs. (19.5) we have two equations in two unknowns: x_S and x_C . If we solve these two equations for the values of the variables x_S and x_C , we find what we sought—the total outputs of the two goods needed to meet the

stated final demands. To see the explicit solutions note that Eqs. (19.5) can be represented as

$$\begin{aligned}(1 - a_{SS})x_S - a_{SC}x_C &= f_S, \\ -a_{CS}x_S + (1 - a_{CC})x_C &= f_C.\end{aligned}\tag{19.7}$$

If we denote the solutions by $\bar{x}_S$ and $\bar{x}_C$, it turns out from Eqs. (19.7) that

$$\begin{aligned}\bar{x}_S &= \frac{(1 - a_{CC})f_S + a_{SC}f_C}{(1 - a_{SS})(1 - a_{CC}) - a_{SC}a_{CS}} \\ \bar{x}_C &= \frac{a_{CS}f_S + (1 - a_{SS})f_C}{(1 - a_{SS})(1 - a_{CC}) - a_{SC}a_{CS}}.\end{aligned}\tag{19.8}$$

We will show in the next section that $a_{SS} < 1$, $a_{CC} < 1$, $(1 - a_{SS})(1 - a_{CC}) > a_{SC}a_{CS}$. These conditions guarantee positivity of $\bar{x}_S$ and $\bar{x}_C$. One more step is required. The labor input at the solution $(\bar{x}_S, \bar{x}_C)$ is $a_{OS}\bar{x}_S + a_{OC}\bar{x}_C$. We have to compare this with l_O , the labor endowment. If the total labor input at $(\bar{x}_S, \bar{x}_C)$ does not exceed l_O , all is well—the targets are feasible. Otherwise, more modest targets will have to be specified. (We argue in the section on linear programming interpretation of the input-output model that the equality restriction in Eq. (19.6) can be relaxed to the inequality restriction $a_{OS}x_S + a_{OC}x_C \leq l_O$. This is why we have stated above that $a_{OS}\bar{x}_S + a_{OC}\bar{x}_C$ should not exceed l_O .)

In the two-good case considered above, we have solved two linear equations to determine two unknowns. This generalizes to an arbitrary number of variables, n . Therefore basically the input-output analysis consists of nothing more than the solution of a set of n simultaneous linear or fixed coefficients of production. With variable input proportions, it becomes necessary to deduce, from the available statistics and engineering information, functional relationships between the output level of each industry and the amount of each input required to produce it. The relevant equations would be tremendously complicated—some of them may be highly non-linear. Therefore, it may be convenient to work with production functions of the type of Eqs. (19.3).

Hawkins-Simon Conditions

The input-output coefficients a_{ij} give the direct requirement of good i to produce one unit of good j . But there are indirect requirements in the form

of the amount of input j required to produce input k which is then used to produce input i , and so on. For instance, we directly require steel to make steel, and indirectly need steel to make steel to make steel, steel to make coal to make steel, and so on.

A natural condition to be imposed on an input-output model is that direct and indirect requirements of input i to produce one unit of input i should be less than one unit. This means that we should not use as much as 1 unit of a good in production to produce 1 unit of the good, failing which the economy will not be able to produce any final goods. For instance, if 1 ton of iron contains more than 1 ton of iron directly and indirectly, self-contained production is not viable. In our two-good model, this condition particularly implies that $a_{ss} < 1$ and $a_{cc} < 1$, since a_{ss} (a_{cc}) is the amount of good S (C) required directly to produce 1 unit of good S (C).

If we set final demand f_c in the second equation of Eqs. (19.5) equal to zero, we can solve the resulting equation for the number of units of good C required directly and indirectly to produce one unit of good S :

$$x_c = \frac{a_{cs}}{1 - a_{cc}} x_s. \quad (19.9)$$

Thus, $a_{cs}/(1 - a_{cc})$ is the number of units of good C required directly and indirectly to produce one unit of good S . If we substitute x_c from Eq. (19.9) in the first equation of Eqs. (19.5), we get

$$a_{ss}x_s + a_{sc} \frac{a_{cs}}{1 - a_{cc}} x_s + f_s = x_s, \quad (19.10)$$

from which it follows that the total amount of good S required directly and indirectly to produce one unit of good S is $a_{ss} + a_{sc}a_{cs}/(1 - a_{cc})$. But this total requirement is less than 1, which on simplification gives $a_{sc}a_{cs} < (1 - a_{cc})(1 - a_{ss})$.

The inequality $a_{sc}a_{cs} < (1 - a_{ss})(1 - a_{cc})$ together with the inequalities $a_{ss} < 1$ and $a_{cc} < 1$ are called the Hawkins-Simon conditions (Hawkins and Simon [1949]). These conditions extend to more than two commodities. (See, Dorfman, Samuelson and Solow, [1958], for a discussion.)

The Hawkins-Simon conditions ensure positive solutions of the input-output system. (That is, the solutions given by Eqs. [19.8] are positive under these conditions.) When we are solving n simultaneous linear equations for

the output levels on n commodities, it may be possible that some of the values in the solution are negative. Negative output of a commodity means that more than 1 unit of the commodity is used directly and indirectly to produce 1 unit of it. This is an embarrassing situation—we require negative outputs of some commodities in order to achieve the final demands! This embarrassing situation is ruled out if the Hawkins-Simon conditions are fulfilled. Thus, these conditions give us mathematical requirements which must be satisfied by any acceptable input-output model.

Linear Programming Interpretation

The purpose of this section is to introduce the mathematical technique of linear programming and to interpret the input-output problem in terms of this technique. Of course, a detailed treatment of linear programming is beyond the scope of this book, and therefore our analysis must be brief. (For an interesting survey of linear programming techniques, see Dorfman, Samuelson and Solow, [1958]).

To see how this technique works here, let us consider the materials balance Eqs. (19.5) again. The left hand side of the first equation gives us the sum of alternative uses of the total output x_S of good S . (The alternative uses here are: inputs to industry S , inputs to industry C , and consumption as final demands.) Since commodity S is not freely available, the sum of alternative uses must be equal to the total amount x_S . A similar relation holds for x_C .

Society specifies a set of final demands f_S and f_C . Two questions arise: Are they achievable? Are the specified final demands within the society's net-production-possibility schedule? The first question can be restated as follows: Does the society have enough labor endowment to produce the specified final demands? To understand the second question, suppose that the two industries have capacity limits. This question then means that it is necessary to verify whether the specified final demands can be produced without gross output overstepping the available capacity. (Strictly speaking, the second question is outside this static model, because if there are capacities, there are ways of changing them. This is clearly an issue of the dynamic model. However, we are assuming here that the capacity limits are fixed.)

The left hand side of Eq. (19.6) gives the sum of labor use in the two industries. (Labor use in industry S is $a_{OS}x_S$ and in industry C is $a_{OC}x_C$.) Labor is not produced, but is available in amounts up to l_O . That is, labor supply is imposed as a constraint on the model. Therefore, the labor use relation of Eq. (19.6) can be relaxed to

$$a_{OS}x_S + a_{OC}x_C \leq l_O. \quad (19.11)$$

If a strict inequality holds in Eq. (19.11), the difference $l_O - a_{OS}x_S - a_{OC}x_C$ may be assumed to be consumed as leisure.

Let M_S and M_C denote the production capacities of the industries S and C respectively. Then the second question requires that

$$\begin{aligned} x_S &\leq M_S, \\ x_C &\leq M_C. \end{aligned} \quad (19.12)$$

Therefore, our problem is to determine the values of x_S and x_C from Eqs. (19.7) where x_S and x_C also fulfil the inequalities given by Eqs. (19.11) and (19.12). (Note that Eqs. [19.7] are rearrangement of Eqs. [19.5].) This is in fact a linear programming problem. We need to find here the values of two variables (x_C and x_S) from two equations which are linear in these variables (Eqs. [19.7]), where the variables in these equations are constrained in the values they can take on by other linear equations (Eqs. [19.11] and [19.12]). Thus, Eqs. (19.11) and (19.12) are constraints in our linear programming problem here.

To see the solution graphically, let us plot x_S on the horizontal axis and x_C on the vertical axis, as shown in Fig. 19.1. The constraint $x_S \leq M_S$ represents a vertical straight line and the area to the left, $x_C \leq M_C$ is a horizontal line and the area below it. The labor supply constraint $a_{OS}x_S + a_{OC}x_C \leq l_O$ is a downward sloping straight line (line L_3 in figure 19.1 and the area below L_3). The slope of L_3 is $-a_{OS}/a_{OC}$. The region of feasible gross outputs is the polygon OM_CDHM_S . (This polygon is called a feasible region because any point which lies within it or on its boundary satisfies the constraints of Eqs. (19.11) and (19.12).)

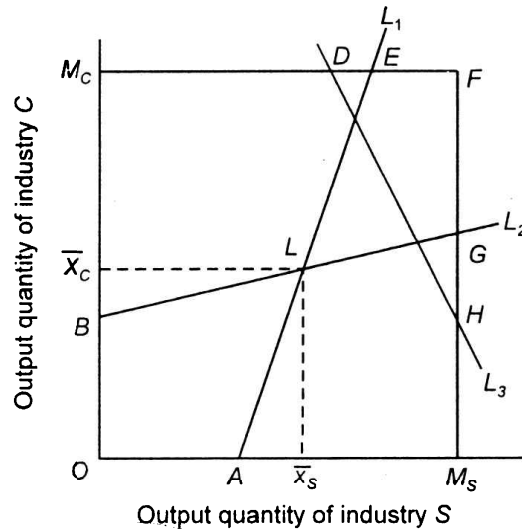


Fig. 19.1 Linear Programming Solution to Input-Output System

The solution to our linear programming problem is then that pair of (x_S, x_C) which lies inside or on the boundary of the region OM_CDHMS , and which also satisfies Eqs. (19.7). The line $(1 - a_{SS})x_S - a_{SC}x_C = f_S$ is drawn as L_1 . Since at A , x_C is zero, the distance OA is $f_S/(1 - a_{SS})$, which is positive in view of $a_{SS} < 1$. The slope of L_1 , $(1 - a_{SS})/a_{SC}$, is also positive. Next, we draw the line $a_{CS}x_S + (1 - a_{CC})x_C = f_C$ as L_2 . Clearly the distance OB is $f_C/(1 - a_{CC})$ and the positive slope of L_2 is $a_{CS}/(1 - a_{CC})$.

The gross outputs which will yield f_S and f_C are those that lie on the point of intersection L of L_1 and L_2 . Any other output in the region $LEFG$ has both outputs higher than at L . At L , the output levels of commodities C and S are denoted by $\bar{x}_C$ and $\bar{x}_S$ respectively. Now, at L , we have $\bar{x}_C < M_C$, $\bar{x}_S < M_S$. Therefore the capacity constraints of Eqs. (19.12) are fulfilled. At L , the labor input is $\bar{x}_S a_{OS} + \bar{x}_C a_{OC} < l_O$. (The line L_3 represents the equation $a_{OS}x_S + a_{OC}x_C = l_O$.) Therefore, the labor supply restriction of Eq. (19.11) is also met. Since L is the point of intersection of the lines L_1 and L_2 both of Eqs. (19.7) are satisfied. Thus, the program $L = (\bar{x}_S, \bar{x}_C)$ is feasible.

If the point of intersection of the lines in Eqs. (19.7) lies outside the feasible region OM_CDHMS then the society cannot produce such large final demands. In this case, the targets need to be revised. If, as in Fig. 19.1, the point of intersection lies strictly inside the feasible region, then both the outputs can be expanded and the society could increase both the final demands beyond f_S and f_C .

From Fig. 19.1 it follows that if the lines L_1 and L_2 were parallel, that is, if they had equal slopes, then there would be no point L . (Equality of slopes means $[1 - a_{SS}]/[1 - a_{CC}] = a_{SC}/a_{CS}$: violation of one of the Hawkins-Simon conditions.) If L_2 has a bigger slope than L_1 , there would also be no point L . In this case the two lines would intersect but not in the positive quadrant. A higher slope for L_2 than for L_1 means $(1 - a_{SS})(1 - a_{CC}) < a_{SC}a_{CS}$, which is again against one of the Hawkins-Simon conditions. In this case it would be impossible to meet Eqs. (19.7) for positive final demands. That is, no final demand would be producible.

Price and Income Determination

Let us apply the competitive condition that price equals unit cost for each industry. Suppose $p_S(p_C)$ stands for the price of good $S(C)$. We designate the wage rate by w . From Table 19.3 it follows that the unit cost for industry S is

$$p_S a_{SS} + p_C a_{CS} + w a_{OS}. \quad (19.13)$$

Therefore by the competitive condition we have

$$p_S = p_S a_{SS} + p_C a_{CS} + w a_{OS}. \quad (19.14)$$

This equation means that the value of the factors which are directly and indirectly required to produce one unit of good S is equal to its price p_S . We can rewrite this equation as

$$(1 - a_{SS}) p_S - a_{CS} p_C = a_{OS} w. \quad (19.15)$$

Similarly, for industry C we have

$$-a_{SC} p_S + (1 - a_{CC}) p_C = a_{OC} w. \quad (19.16)$$

Solving Eqs. (19.15) and (19.16) for p_S and p_C we have

$$p_S = \frac{a_{OS}(1 - a_{CC}) + a_{OC} a_{CS}}{(1 - a_{SS})(1 - a_{CC}) - a_{SC} a_{CS}} w, \quad (19.17)$$

$$p_C = \frac{a_{OC}(1 - a_{SS}) + a_{OS} a_{SC}}{(1 - a_{SS})(1 - a_{CC}) - a_{SC} a_{CS}} w. \quad (19.18)$$

Since the input-output model we have considered here has only one non-produced factor, labor, its price w may be set at unity, that is, labor may be chosen as the numeraire here. Income at market price equals the value of final demands. The value of final demands is

$$p_S f_S + p_C f_C. \quad (19.19)$$

The factor requirement of industry S is $x_S a_{OS}$. Similarly, for industry C this requirement is $x_C a_{OC}$. Therefore total factor payment is given by $(x_S a_{OS} + x_C a_{OC}) w$, the total factor requirement multiplied by the wage rate. Income at market price must equal income at factor cost, that is,

$$p_S f_S + p_C f_C = (x_S a_{OS} + x_C a_{OC}) w. \quad (19.20)$$

Eq. (19.20) also says that factor payments exhaust the value of net output.

Substitution Theorem

We may recall that a restrictive assumption of the input-output system is that there are fixed coefficients of production. That is, the same a_{ij} portion of the i th industry output would be required as input for the unit output of the j th industry, whatever may be the total output of industry j . There is no possibility of any other input proportion.

Samuelson (1951) demonstrated that in some circumstances this assumption may not be as serious as it appears. He showed that even if variation of input proportion is allowed, it would never be advantageous provided that there are constant returns to scale, only one scarce factor and no joint production. This is Samuelson's substitution theorem. The theorem says that input-output proportions may be fixed, but they will be fixed by productive efficiency requirements rather than by unchangeable technological considerations. For each commodity there is a unique most-efficient-input ratio (say, capital-labor ratio) and any change in the output level will not alter this ratio.

We state the theorem formally as

SUBSTITUTION THEOREM: *Suppose there is only one non-produced input to production, this input is indispensable to production, there is no joint production and the technology exhibits constant returns to scale. Then there is one preferred set of input ratios, which will continue to be preferred no matter what the desired levels of final demands happen to be.*

Proof: We give an outline of the proof of the theorem. A detailed proof, which requires considerable knowledge of linear algebra, can be found in Gale (1960). In Chapter 8 it has been noted that if a production function exhibits constant returns to scale, then all isoquants can be derived by radial expansion or contraction of the unit isoquant (by means of rays emanating from the origin). The slopes of the isoquants along any straight line through the origin are equal. (Also see exercise 16 of Chapter 8.) Therefore, if input prices are fixed it will never be advantageous to change input proportions, whatever the level of output may be.

Next, suppose that there is only one scarce factor of production, say labor. No other factor will prevent the expansion of the national output arbitrarily. The costs of the remaining inputs must be calculated in terms of the amount of labor required to produce them. For instance, suppose that the real price of input *A* (say, steel) is 10 man-hours per unit and that of input *B* (say, coal) is 15 man-hours. The relative input price therefore will depend on their labor content. For any wage rate this relative price remains unaltered. Consequently, any output should always be produced in a manner which uses the same input proportions. This completes the proof of the theorem.

The substitution theorem might be called a non-substitution theorem. Substitution is possible, but it will never take place in an economy with multiple production activities if there is only one non-produced scarce factor.

If there is more than one scarce factor of production, then the substitution theorem does not remain valid. To see this, note that we will have a consumption frontier: the set of final demands that can be produced for each factor taken separately. But only those final demands are achievable which lie on or under both frontiers. For the first scarce factor, the consumption frontier is ABD and for the second scarce factor it is EBF . The set of achievable final demands is EBD . Since EBD is not a straight line it does not have a constant slope. Therefore, the presence of two factors destroys the constant cost characteristic. Hence the substitution theorem does not work.

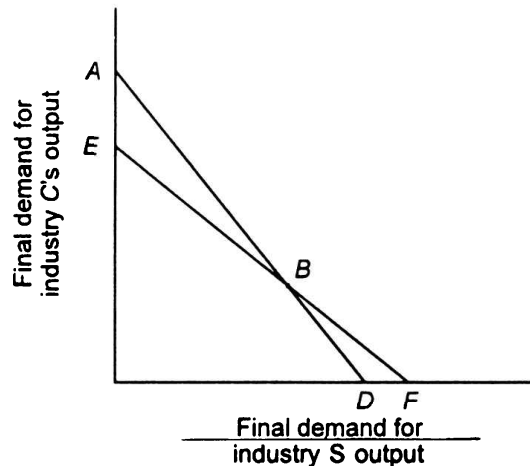


Fig. 19.2 Violation of Substitution Theorem

In the case of joint production, a shift in demand for some commodities may seek to change the aim of activities and hence to change the input proportion. So substitution will be advantageous for a joint product system.

LINEAR PROGRAMMING

A Linear Programming Problem and Its Graphical Representation

Linear programming interpretation of the input-output model involves determination of n gross output levels from n linear equations when the values of outputs are constrained by some other linear functions. The linear programming technique can also be applied to find the optimum (maximum/minimum) values of linear functions when the variables in these functions face some linear constraints. Examples are (i) choice of optimum product lines and production processes, (ii) transportation routing, (iii) meeting product specification and so on. An illustration of (i) can be the problem of allocation of fixed amounts of various productive resources available in an economy to the production of goods so that the total value of output is as

large as possible. For a micro-level illustration, consider a firm operating at high levels of outputs. The firm produces two commodities with several inputs. This firm is likely to face certain constraints like the amount of time available on different machines, factory size, and so on. The firm's problem, given its resources, is to choose the optimum product mix which maximizes the profit of the firm. To illustrate (ii), suppose that a firm has many plants and its processes require transshipment of items in different stages of production. The firm can make substantial savings if careful planning of commodity movements is made. The firm's objective here is to minimize transport costs. Situations described by example (iii) arise when the ingredients of a product are required to meet certain minimum specifications. The producer's problem is to choose the optimum combinations of ingredients so that his profit is maximized. Examples of other interesting applications of the linear programming technique are cutting of paper and cloth in a way that minimizes raw material waste.

We will examine only one specific example of linear programming techniques. In this example an economy is assumed to have fixed amounts of various productive inputs and must choose how to allocate these resources to the production of two goods, A and B , such that the total value of output is maximized. (It is assumed that the given prices of A and B are Rs. 45 and Rs. 55 respectively.) Thus, the goal is to choose outputs x_A and x_B of A and B so that

$$\text{The total value} = TV = 45x_A + 55x_B \quad (19.21)$$

is as large as possible, given the available resources.

In any production economy two types of constraints exist: the total quantities of various inputs are fixed, and certain technological rules (that is, production functions) must be employed in turning these inputs into outputs. For the purpose of our two-good example we will assume that there are only two inputs; labor and machines. Suppose, moreover, that the economy has available only 120 labor-hours and 180 machine-hours. The amounts of the two inputs that are required to produce 1 unit of each of the two goods A and B are as follows: it takes 6 labor-hours and 3 machine-hours to produce 1 unit of A , it takes 4 labor hours and 10 machine hours to produce 1 unit of B . Clearly, these production techniques are of the fixed proportions type: no substitution between inputs is possible. This kind of linear technology is one characteristic of most linear programming problems.

We are now in a position to examine the constraints that the availability of inputs impose on the combinations of goods A and B that can be produced. The quantity of labor employed in good A production plus the quantity employed in good B production cannot exceed the 120 labor-hours available. (Recall that we require 6 labor-hours for 1 unit of A and 4 labor-hours for 1 unit of B). Thus, all possible choices of x_A and x_B must satisfy

$$6x_A + 4x_B \leq 120. \quad (19.22)$$

Similarly, the machine-hour constraint, which states that no more machines can be used than are available, is

$$3x_A + 10x_B \leq 180. \quad (19.23)$$

Finally, we have the non-negativity requirements

$$x_A \geq 0, \quad x_B \geq 0. \quad (19.24)$$

Given the two input constraints, and non-negativity constraints for output levels, our problem is to find values of x_A and x_B that satisfy these constraints and make the total value of output as large as possible. This is our linear programming problem. It consists of three parts: (a) the function (that is, the total value) whose value is to be maximized, which is called the objective function, (b) the capacity constraints of Eqs. (19.22) and (19.23), and (c) the non-negativity conditions on the variables, that is $x_A \geq 0$, $x_B \geq 0$. As stated earlier, this problem can be called linear because the expression to be maximized and the inequalities involved are linear.

Let us now do some geometry, as has been done in the context of linear programming interpretation of input-output systems, to look at the set of possible solutions. Two resource constraints are plotted in Fig. 19.3. If the society decides to produce only good A , 20 units could be produced. In producing only good A , the society runs out of labor-hours first (there are enough machine-hours to produce 60 units of A). For good B , on the other hand, the binding constraint is machine availability. There are only enough machines to produce 18 units of good B if only good B is produced. Other combinations of good A and B on or inside the heavy line in Fig. 19.3 satisfy the two constraints. The area bounded by the thick lines is called the feasible region because every point on its boundary or within it represents a combination of values of the variables (the output levels of A and B) which does not violate the constraints of Eqs. (19.22) to (19.24). This is why every point in the feasible region represents what is called a feasible solution to the programming problem. The boundary of the feasible region, the thick

line in Fig. 19.3, is the production possibility frontier for the simple economy we have described. Observe that though the production functions used here are of a fixed proportion, constant returns to scale type, the production possibility frontier is concave. The reason behind this is that the two goods use the various inputs in different proportions.



Fig. 19.3 Production Possibility Frontier for Linear Programming Problem

We can now use the production possibility frontier derived in Fig. 19.3 to solve our maximum value problem. Fig. 19.4 shows the production possibility frontier with several isorevenue lines. (Any isorevenue curve represents combinations of x_A and x_B which yield the same revenue or the value of output. For instance, the revenue is the same at points R and K .)

The objective of a programming calculation is to determine the optimal (in our example the maximal revenue generating) among those feasible output combinations. It follows that the optimal point of a linear program will always lie on the boundary of the feasible region. (This follows from the result that a move to a higher isorevenue curve always increases revenue.) Let us look at the reasoning behind this more acutely. Any commodity whose production generates higher revenue will continue to be lucrative as its output increases because there will be neither diminishing returns to scale nor unfavorable effects on input and output prices. It will therefore be profitable to expand production until some capacity limit is reached, that is, until the boundary of the feasible region is attained. In view of the fact that the isorevenue curves are linear, there will be at least one optimal point which occurs at one of the corners O , T , K or U of the feasible region. Therefore, the optimal point will always be a corner tangency point, such as K or, if the isorevenue curve is parallel to a segment of the feasible region, such as TK , then the entire segment, including the corner points T and K , will be optimal.

Any values of the variables (x_A , x_B , x_L and x_M) which satisfy the constraints of Eqs. (19.26) and the non-negativity requirements is called a feasible solution. Any set of values of the variables (x_A , x_B , x_L and x_M) in which the number of non-zero-valued variables (either ordinary or slack) is equal to the number of constraints is called a basic solution. An example of a basic solution of the problem we have considered here is $x_A = 10$, $x_B = 15$, $x_L = 0$, $x_M = 0$.

It is possible to show that only the corner points of the feasible region represent basic feasible solutions. At point T , we have $x_B (= 18) > 0$, $x_L (= 48) > 0$, and in addition since this point lies on the boundary of the feasible region, $x_A = 0$ and $x_M = 0$. Hence the point T is basic. Similarly, the solution at U is $x_A (= 20) > 0$, $x_M (= 120) > 0$, $x_B = 0$ and $x_L = 0$, which is also basic. At K , $x_A (= 10) > 0$, $x_B (= 15) > 0$ and $x_L = 0$, $x_M = 0$. Therefore K is a basic solution. Finally, since the origin lies below the boundary of the feasible region, it leaves unused capacity, so $x_L > 0$, $x_M > 0$. That is, at point O , we have the basic feasible solution $x_A = 0$, $x_B = 0$, $x_L (= 120) > 0$ and $x_M (= 180) > 0$. These are the only four basic feasible solutions of our problem. No other feasible point represents a basic solution. For example, at any point in the interior of the feasible region, we have $x_A > 0$, $x_B > 0$, $x_L > 0$, $x_M > 0$, which is certainly not a basic solution.

We can summarize the above observations in the form of a theorem, which is known as the Basic Theorem of Linear Programming.

THEOREM: In any linear programming problem, the optimal solution will be found from among the basic solutions only. In other words, there will always exist an optimal solution in which the number of non-zero variables (both ordinary and slack) is equal to the number of constraints of the problem.

Simplex Method of Solving Linear Programming Problems

To understand a simplex, let us consider a triangle ABC . The three coordinates A , B , C of the triangle may be regarded as the points of intersection of three straight lines or linear equations. Likewise we can get the n -corner points as the points of intersection of n linear equations. The space specified by these n -corner points is called a simplex. Thus, loosely speaking, a simplex may be regarded as the n -dimensional analogue of a triangle. The simplex will represent the feasible region of the n variables which constitute the n linear equations.

The simplex method of computational technique is the most frequently used technique for dealing with linear programming problems. This method

relies on the basic theorem. Therefore, an optimal solution will be found from among the corner points of the feasible region, and all other points in the diagram can be ignored in the calculation.

The simplex method proceeds as follows,

1. Find any basic solution. It will be a corner point of the feasible region. If the origin is a corner of the feasible region, it is convenient to start from the origin. Suppose then that we start at O (Fig. 19.4).
2. Compute the total value of output at O and at the two adjacent corners T and U .
3. If one or both of T and U generate higher value than does O , move to the corner which yields more value. Let this corner of higher value be U .
4. Now, let us compute the total value at K to see if it is higher than that at U . If so, move to K , otherwise stay at U . In the course of this elimination process we will eventually reach the optimal point.

Simplex Method Applied

To apply the simplex method for solving a linear programming problem, let us consider our original problem and write it in a slightly modified form.

$$\begin{array}{ll}
 \text{Maximize} & TV = 0 + 45x_A + 55x_B, \\
 \text{subject to} & x_L = 120 - 6x_A - 4x_B, \\
 & x_M = 180 - 3x_A - 10x_B.
 \end{array} \tag{19.28}$$

'0' is introduced in the objective function to make the number of terms equal to four as in the equations which follow.

We write out the table or matrix of coefficients corresponding to the system of Eqs. (19.28). (A matrix is a tabular array of figures, set up in a special way for mathematical or numerical solution.) This table summarizes the numerical data in the system of Eqs. (19.28).

Our first basic solution is clearly, $x_L = 120$, $x_M = 180$, $x_A = 0$, $x_B = 0$. In Table 19.4 the zero-valued variables x_A and x_B appear at the top and positive-valued variables appear along the left side. The equality signs after the variables indicate that the table represents a system of equations each with terms involving a constant term (that is, a term in which the coefficient is multiplied by unity rather than by a variable), and the variables x_A and x_B .

Table 19.4

	1	x_A	x_B
$TV =$	0	45	55
$x_L =$	120	- 6	- 4
$x_M =$	180	- 3	- 10

The initial solution $x_L = 120$, $x_M = 180$, $x_A = 0$, $x_B = 0$ is far from being optimal, since if it were optimal the best alternative open to the economy would be to close down production altogether. (The total value of output, TV , is zero at this solution.) So we look for the next basic solution. For this, we must attach a positive value to one of the variables which was zero in the initial solution, and a zero value to one of the formerly non-zero variables. That is, in the next table we want to move one of the top variables, say x_B , over to the left side, and move one of the formerly non-zero left hand side variables, say x_M , over to the top. Clearly, this is a case of interchange of the roles of x_B and x_M . We say that in this case we are pivoting on the element which is at the intersection of the x_M row and x_B column in the table. Clearly, this element is - 10.

Now, let us look at the consequences of this interchange. For this, we consider first the equation in which element x_M appears,

$$x_M = 180 - 3x_A - 10x_B. \quad (19.29)$$

Given that the capacity constraints of the economy remain unchanged, we want this and the other constraint of Eq. (19.28) to continue to hold. We also want these constraints rewritten with a new arrangement of the variables, where x_M and x_B interchange their positions. Thus, we want to express x_B in terms of x_M in the preceding equation. This new equation becomes

$$x_B = 18 - \frac{1}{10}x_M - \frac{3}{10}x_A. \quad (19.30)$$

This is our new third equation, and its coefficients are the elements of the third row of our table 19.5.

To obtain the remaining two rows we must substitute the value of x_B from Eq. (19.30) into the objective function and the other constraint in the system. We then get

$$TV = 0 + 45x_A + 55 \left(18 - \frac{3}{10}x_A - \frac{1}{10}x_M \right) = 990 + \frac{57}{2}x_A - \frac{11}{2}x_M. \quad (19.31)$$

Table 19.5

	1	x_A	x_M
$TV =$	990	$\frac{57}{2}$	$-\frac{11}{2}$
$x_L =$	48	$-\frac{24}{5}$	$\frac{2}{5}$
$x_B =$	18	$-\frac{3}{10}$	$-\frac{1}{10}$

$$x_L = 120 - 6x_A - 4 \left(18 - \frac{3}{10}x_A - \frac{1}{10}x_M \right) = 48 - \frac{24x_A}{5} + \frac{2}{5}x_M \quad (19.32)$$

The coefficients in these equations will give us the first and the second rows of Table 19.5.

Table 19.5 gives us our next trial basic solution. We have for this solution $x_A = x_M = 0$, so that $x_B = 18$, $x_L = 48$. Clearly, this is a basic solution. Note that we also have $TV = 990$, which is a substantial improvement over the previous zero total value of Table 19.4.

But $TV = 990$ is not the optimal value of the total output, because the solution $x_A = x_M = 0$, $x_B = 18$, $x_L = 48$ is not optimal. The following theorem gives a necessary and sufficient condition for an optimality check.

THEOREM: *A basic solution is optimal if and only if none of the coefficients in the objective function corresponding to the basic solution is positive.*

To understand this necessary and sufficient condition, let us consider the objective function $TV = 100 - 5x'_A - 10x'_B$, where x'_A and x'_B represent any variable (either ordinary or slack). Now, as $x'_A \geq 0$, $x'_B \geq 0$, the maximum value of the total output will be obtained at the minimum values of x'_A and x'_B , which are zeros. At this stage we cannot increase the total value by increasing x'_A and/or x'_B , because if we increase x'_A and/or x'_B , a positive number will be deducted from 100. Since in Table 19.5 the coefficient of x_A is positive—it is equal to $57/2$ —the solution $x_A = x_M = 0$, $x_B = 18$, $x_L = 48$ is not optimal.

We are now ready to choose another pivot element and repeat the above procedure as many times as required until an optimal basic solution has been found. Once we state how we choose the pivot element, we finish describing the simplex method.

The choice of the pivot element is guided by the following rules.

Choice of the pivot column: The pivot element comes from that column which has the largest positive top element.

Choice of the pivot element: Take each negative element in the column selected, and use it to divide the corresponding first-column element. The element for which the resulting quotient is the smallest in absolute value should be selected as the pivot.

To understand the selection of the pivot column, let us refer to Table 19.4. In this table the pivot column must be the third column because its top element (55) is positive and is higher than the corresponding elements of the other columns. The intuitive reasoning behind this choice is as follows. The positive number 55 is the coefficient of x_B in the objective function. Therefore this number can be regarded as the marginal revenue of x_B . Consequently, we increase total revenue by a higher amount if we introduce x_B instead of any other formerly zero variable (for example, x_A , whose marginal revenue is 45).

Next, in Table 19.4 the two negative elements in the third column are -4 and -10 . If we divide the corresponding first-column elements by these numbers, the quotients become $120/-4 = -30$ and $180/-10 = -18$. Since the absolute value of -18 is smaller than that of -30 , the corresponding element -10 of the column in question will be selected as the pivot.

The reason behind this choice can be explained intuitively. When we introduce x_B into the solution, some of the economy's resource must be diverted from its current outputs to the production of x_B . This means that the quantities of other outputs will go down. That is, for increasing the output of x_B we must reduce the resource employed for producing other outputs. We must stop increasing the output of x_B whenever the production of some commodity—let us call it x_D —becomes zero. This is because any further reduction in x_D is not possible. Therefore, in the next basic solution the output which will be cut to zero is x_D .

In order to identify the commodity x_D we must find the output which can be cut to zero first by an expansion of x_B . Let us therefore try to observe the impact of an increase in x_B on each of the previously non-zero output x_L and x_M . From the first constraint Eq. in (19.28) we note that an increase in x_B by 1 unit reduces x_L by 4 units (the common element of the x_L - row, and x_B - column of Table 19.4, that is, the second element in the pivot column). Since in the initial solution there are 120 units of x_L (the common element of the

x_L – row and the first column of Table 19.4, it will take $120/4 = 30$ units of x_B to reduce x_L to zero.

From the second constraint Eq. in (19.28), we see that a 1-unit increase in x_B decreases x_M by 10 units (the common element of the x_M – row and x_B – column of table 19.4). Since there are 180 units of x_M to start with (the common element of the first column and the x_M – row), we will require $180/10 = 18$ units of x_B to reduce the production of x_M to zero. Thus, x_M can be reduced to zero before we require 30 units of x_B to eliminate the output of x_L . It is therefore impossible to increase the output of x_B beyond 18 units. (Verify this from Fig. 19.3 also.) Thus, x_M is the output which should be reduced to zero in the next basic solution. This means that -10 , the common element of the x_B – (column) x_M – (row), is the pivot element.

Our discussion of the simplex method of computation is now complete. However, we have not yet determined the optimal solution of the problem considered. In table 19.5 there is only one column (the second) whose top element is positive. The pivot will therefore be chosen from this column. The two remaining elements in this column are negative. Let us therefore look at the quotients obtained by dividing the corresponding first column elements by these two elements. They are

$$\frac{48}{-24/5} = -10, \quad \frac{18}{-3/10} = -60.$$

Since the absolute value of -10 is smaller than that of -60 , the common element of x_L – row and x_A – column, that is, $-24/5$ becomes the pivot. This means that we need to interchange the roles of x_A and x_L , where x_L will be zero and x_A becomes positive.

Given that x_A and x_L are to be interchanged, let us use Eq. (19.32) to express x_A in terms of x_L :

$$x_A = 10 - \frac{5}{24} x_L + \frac{1}{12} x_M \quad (19.33)$$

This is our (final) second equation and its coefficients are the elements of the second row of our table 19.6.

The remaining two rows can be obtained by substitution of x_A from Eq. (19.33) into the objective function in Eq. (19.31) and the constraint in Eq. (19.30). We then have

$$TV = 990 + \frac{57}{2} \left(10 - \frac{5}{24} x_L + \frac{1}{12} x_M \right) - \frac{11}{2} x_M,$$

Table 19.6

	1	x_L	x_M
$TV =$	1,275	$-\frac{95}{16}$	$-\frac{25}{8}$
$x_A =$	10	$-\frac{5}{24}$	$\frac{1}{12}$
$x_B =$	15	$\frac{1}{16}$	$-\frac{1}{8}$

$$= 1,275 - \frac{95}{16}x_L - \frac{25}{8}x_M \quad (19.34)$$

$$\begin{aligned}
 x_B &= 18 - \frac{1}{10}x_M - \frac{3}{10}\left(10 - \frac{5}{24}x_L + \frac{1}{12}x_M\right) \\
 &= 15 - \frac{1}{8}x_M + \frac{1}{16}x_L.
 \end{aligned} \quad (19.35)$$

The coefficients in these two equations give use the first and third rows of Table 19.6.

We note that all (except the first) of the elements in the top row of Table 19.6 are negative. These negative numbers are the marginal revenues of x_L and x_M , whose outputs are now zero. Therefore by our optimality check theorem the corresponding basic solution is optimal. This solution is $x_A = 10$, $x_B = 15$, $x_L = x_M = 0$, and it yields a total value of 1,275.

Duality and The Pricing of Inputs

So far we have not said anything about input prices in our linear programming problem. Given input quantities, the economy sets about the task of maximizing the value of total output. A basic problem of this type whose solution is attempted by the linear programming technique is called the primal problem. To every primal problem there corresponds a dual problem which yields some important information. The nature of the dual problem will be determined by that of the primal problem. For instance, if the primal problem is a maximization exercise, the dual is a minimization problem. Conversely, if the primal minimizes some objective function, the dual will be a maximization problem.

In the context of the example we have considered here, the dual linear programming problem is the determination of the shadow prices of the factors, that is, the input prices associated with the optimal choice of x_A and x_B . Formally, these dual input prices solve the following linear problem:

$$\text{Minimize} \quad E = 120w + 180r, \quad (19.36)$$

$$\text{subject to} \quad 6w + 3r \geq 45,$$

$$4w + 10r \geq 55, \quad (19.37)$$

where w and r are the (non-negative) per unit prices of labor and machinery.

The dual problem considered above can be given an interesting economic interpretation. We seek to determine the prices of the inputs that will minimize the total value of inputs (or expenditure on inputs) and they also have the property that the production of neither of the goods earns a pure profit (this is stated by inequality Eqs. [19.37]). For example, the first inequality means that the cost of producing one unit of good A (that is, the cost of 6 labor-hours and 3 machine-hours) should not be less than the per unit price of A . If constraints such as those of Eqs. (19.37) were not imposed, an obvious solution of the minimization problem would be $w = r = 0$.

It should be clear that the dual problem is related in many ways to the primal problem. All the parameters that appeared in the primal problem also appear in the dual problem but in different places. For instance, the quantities that appeared in the primal constraints now appear as coefficients in the dual objective function. Also note the difference between inequality signs in the primal and dual constraints.

The dual problem may be solved independently from the primal by a procedure similar to that we have described above. The solution to the dual problem turns out to be $r = 25/8$ and $w = 95/16$. These are the prices of machine and labor that minimize total expenditure and that obey the constraint Eqs. (19.37).

The following are some important features of this solution.

1. With these input prices the inequality Eqs. (19.37) become equalities. Neither of the two goods is produced at a loss. Hence both goods can be produced in the society.
2. The value of E with these input prices is 1,275. This is the same as the maximum value of TV we found in our primal problem. Such a relation holds between the primal and the dual solutions of all linear programming

problems. This equality clearly resembles the income-output identity in national income accounting. In our case this equality means that the total output value is the same as the total input cost.

3. The input prices w and r indicate how much extra value would be provided by one more unit of these inputs. (That is, they indicate how much total revenue would rise per unit increase in the use of inputs labor and machine).

We conclude this chapter with some general remarks. The problem we have studied here is about an economy as a whole. But as stated earlier, an optimization problem where linear programming is applicable may arise in other cases also. However, the technique is the same in all cases. The solution to the primal problem is equivalent to solving the dual problem. This relationship is widely used. For instance, the calculation of factor prices in a linear programming problem can be very useful for policy purposes. Such shadow prices show which factors are bottlenecks to the further expansion of the productive unit. (These are the factors with positive shadow prices. If the optimal solution contains a slack activity indicating that some quantity of the corresponding factor remains unemployed, this factor's shadow price is zero.) Shadow prices can be compared with the actual prices of inputs, and in some cases the differences are quite high. For instance, because of institutional reasons (labor unions, minimum wage, etc.) some workers have high wages. But the linear programming technique may suggest that the real value of labor is rather low. Therefore, in such a case planners could adopt productive techniques that utilize labor to a greater extent than would seem to be guaranteed by looking at market wages only.

Example 19.1

Consider the following input-output table for a two-industry (S and C) model.

Table 19.7

	<i>Industry S</i>	<i>Industry C</i>	<i>Labor</i>
Industry S	0.2	0.3	0.3
Industry C	0.4	0.1	0.2
Final demand (in millions)	10	2	0

We want to determine the gross levels of output of the two industries. We also want to verify whether they obey the labor constraint. The labor supply is limited up to an amount of 7 million units.

Let the gross output levels of industries S and C be x_S and x_C respectively. Then from the given input-output table we have:

$$x_S = 0.2x_S + 0.4x_C + 10,$$

$$x_C = 0.3x_S + 0.1x_C + 2.$$

We rewrite these equations as

$$0.8x_S - 0.4x_C = 10,$$

$$-0.3x_S + 0.9x_C = 2.$$

Solving these two equations we get $x_S = 49/3$, $x_C = 23/3$, which are the required gross output levels.

The total labor requirement for the above gross output is

$$(0.3) \frac{49}{3} + (0.2) \frac{23}{3} = 6.4 < 7$$

Therefore, the labor constraint is verified by the solution.

Applied Example 19.1 Multisector Plan Model for India

A multisector plan model in computable form relating to planning in India was developed by Ragnar Frisch who came to India as a guest of the Indian Statistical Institute in 1955 to work on planning methodological problems for India (see Frish, [1960]). Frisch considered 22 productive sectors with no joint production. Apart from considering constant input-output proportions, he also imposed certain upper and lower bounds in the form of inequalities. For instance, he proposed certain upper bounds on sectoral output level to take into account the capacity limits in the short run. Certain variables like total government current expenditure were treated exogenously.

For data, Frisch made use of an inter-industry table prepared at the Indian Statistical Institute in connection with studies relating to planning. Frisch considered the problem from both long-run and short-run perspectives. For the long run or asymptotic problem he assumed that output levels, investment allocations, consumption of each individual item, etc., grow at a constant rate. The short run or coming-year problem was required to satisfy certain short run objectives and was also sought to minimize the difference between the structure of capacities during the coming year from the structure that would be ideally conforming to the long term growth pattern.

A distinctive feature of the model is its attempt at generating a scalar measure of overall welfare by aggregating diverse measures of welfare. Another attractive feature of the model is its way of using capital stock from a long run problem in a short run model.

EXERCISES

Multiple Choice

1. Negative output of a commodity in an input-output system means that
 - a) More than 1 unit of the commodity is used directly to produce 1 unit of it.
 - b) Directly more than 1 unit and indirectly less than 1 unit of the commodity is used to produce 1 unit of it.
 - c) More than 1 unit of the commodity is used directly and indirectly to produce 1 unit of it.
 - d) Any one of the above is possible.
2. An input-output model with n commodities is nothing more than a consequence of
 - a) Increasing returns to scale.
 - b) Fixed coefficients of production.
 - c) Convexity of isoquant surfaces
 - d) None of the above.
3. Which of the following is not an assumption of the input-output system?
 - a) Fixed coefficients of production.
 - b) Possibility of joint production.
 - c) General equilibrium framework.
 - d) Convexity of isoquant surfaces.
4. Which of the following is not relevant to the Samuelson substitution theorem?
 - a) Constant returns to scale.
 - b) Only one scarce input.
 - c) No joint production.
 - d) Law of variable proportions.

5. In the two-commodity case, the Hawkins-Simon conditions assert that the two materials balance relations will have
 - a) Equal positive slope.
 - b) Equal negative slope.
 - c) Different positive slopes and will intersect in the positive quadrant.
 - d) Slopes which are equal in magnitude but opposite in sign.
6. Shadow price is
 - a) The change in the value of the objective function per unit of change in any one of the constraints.
 - b) The change in the value of the primal solution per unit of change in the dual solution.
 - c) The change in the value of one of the constraints per unit change in the objective function.
 - d) None of the above.
7. Which of the following is true?
 - a) The dual of profit maximization is cost minimization.
 - b) Reaching the highest possible isoprofit line is the dual of reaching the lowest possible isocost line.
 - c) The dual of cost minimization is profit maximization.
 - d) All of the above.
8. In a linear programming problem, all possible solutions that fulfil the constraints are called
 - a) Primal solutions.
 - b) Dual solutions.
 - c) Basic solutions.
 - d) Feasible solutions.

Quantitative Problems and Questions

1. Clearly explain why the Samuelson substitution theorem does not remain valid if:
 - a) There is more than one scarce factor;
 - b) There is joint production.
2. From the following input-output table determine the gross levels of output of the three industries *S*, *C* and *R*. What will be the total labor requirement?

Table E1

	<i>Industry S</i>	<i>Industry C</i>	<i>Industry R</i>	<i>Labor</i>
Industry S	0.2	0.5	0.1	0.2
Industry C	0.4	0.1	0.2	0.3
Industry R	0.5	0.1	0.1	0.3
Final demands	70	60	90	

3. A country's final consumption demands for three goods *S*, *C* and *R* are 50 million, 90 million and 120 million. Given the following table determine the gross levels of output of the three goods needed to meet the final demands. What would be the capital requirements of the three industries? Determine the amount of capital that must be imported.

Table E2

	<i>Industry S</i>	<i>Industry C</i>	<i>Industry R</i>	<i>Capital</i>
Industry S	0.2	0.5	0.1	0.2
Industry C	0.3	0.2	0.2	0.3
Industry R	0.4	0.1	0.2	0.3

4. Given the following information calculate the gross output levels of *S* and *C* required to meet the specified amounts of final consumption. The amounts of capital and labor endowments of the economy are 2 and 7. Determine the amount of capital that has to be imported and the amount of undemanded labor.

Table E3

	<i>Industry S</i>	<i>Industry C</i>	<i>Capital</i>	<i>Labor</i>
Industry S	0.2	0.3	0.2	0.3
Industry C	0.4	0.1	0.3	0.2
Final demands	10	2		

5. What are the possible implications of fixed coefficient production functions in an input-output model?
6. Explain the main features of the Leontief open static system.
7. Clearly explain the role of the Hawkins-Simon conditions in an input-output model.

8. A firm produces two goods A and B with unit profits of Rs. 4 and Re. 1 respectively. Each product requires two inputs, capital (K) and labor (L). The total available quantities of these two inputs are 5 and 4 units respectively. The amounts of different inputs required to produce one unit of each of the two goods are given in the following table.

Table E4

<i>Inputs required to produce</i>	
1 unit of A	1 unit of B
1 unit of K	2 units of K
3 units of L	2 units of L

Assume that the firm is a profit maximizer.

- Formulate a linear programming statement of the problem
 - Graph the firm's production possibilities frontier. Is it linear? If not, what are the reasons that do not make it linear?
 - What are the basic solutions of this linear programming problem?
 - What is the optimal solution?
 - What is the maximized level of profit?
 - What is the dual problem here? Does the optimal value of the dual objective function coincide with that of the primal objective function? Interpret the dual problem economically.
9. A firm uses three machines to produce three goods A , B and C . The available capacities of the three machines are 80, 60 and 130 hours respectively during a specified period. The profits per unit of output of the three goods are Rs. 800, Rs. 400 and Rs. 900 respectively. Each unit of product A requires 3 hours of time on the first machine, 5 hours of time on the second machine and 8 hours of time on the third machine. For each unit of goods B and C these times are given by (4 hours, 2 hours and 7 hours) and (5 hours, 1 hour and 3 hours) respectively. The firm is a profit maximizer.
- Formulate a linear programming statement of the problem.
 - Use the simplex method to determine the optimal solution.
 - What is the maximum profit?
10. Consider a two-industry (A , B) input-output model in which neither of the two industries uses up any of its own product.

- a) Then formulate the input-output equations.
- b) What are the Hawkins-Simon conditions here?

11. Prove or disprove the following.

- a) Fixed coefficients production functions do not exhibit constant returns to scale.
- b) The Hawkins-Simon conditions are necessary and sufficient for positivity of output levels in an input-output system.
- c) In a Leontief technology, relative prices can never change.
- d) The substitution theorem means that change in relative price will not induce changes in input proportions.
- e) Consider a two-variable linear programming problem where the values of the variables face two linear constraints. Then the feasible region is a triangle.

20

Theory of Distribution

The concern of this chapter is the theory of distribution. The distribution theories are designed to explain the levels of payments to the various factors of production—rents, wage rates and rates of profits, and hence the shares of various factors in the total product. In view of the fact that general equilibrium analysis seeks to account for the determination of every price in the economy, the analysis of distribution should be regarded as a segment of general equilibrium theory.

A great deal has been written by economists on this subject. The subject has also become an issue of debate between contributors associated with Cambridge, Massachusetts, and Cambridge in the United Kingdom, and a group of Italian economists.

The basic question here is how to tackle the problem of distribution through some analytical formulation. We may talk about two possibilities: (a) use of a sophisticated analytical (micro) model which will satisfy different consistency conditions and completeness, and whose complexity reasonably reflects the complexities of the real world; (b) use of a simplified macro model that will demonstrate an intuitive grasp of the main relationships and a reasonable basis for policy design.

The marginal productivity analysis is a theoretical micro model of distribution. When embodied within its general equilibrium relationships the model enables us to determine payments going to workers, capitalists, landlords, etc. We analyze the model in greater detail in the next two sections of this chapter. A somewhat related model is the one proposed by Bliss (1987). We analyze the Bliss model immediately after our discussion on marginal productivity theory.

Next, we consider the macromodels of distribution. In such models a large number of moderately diverse economic variables and relationships are lumped together and the resulting aggregates are treated as homogeneous economic elements. In this way the models involve only small numbers of variables and relationships. Two such models proposed by Ricardo (1871) and Kaldor (1955–56) have been subjects of considerable discussion. Accordingly, we present a review of these two models also.

ON THE MARGINAL PRODUCTIVITY THEORY

In Chapters 14 and 15 we have analyzed the partial equilibrium elements of marginal productivity theory. Under pure competition a profit maximizing firm will hire any input up to the point where its price equals the marginal revenue product of the input. When some degree of monopoly exists in the product market, the determination of the amount of input to be hired occurs in the same way. Like the purely competitive firm, the monopolistic firm employs those quantities of the resource at which its marginal revenue product equals the marginal resource cost. However, the monopoly firm's demand for input would be smaller than the quantity of input demanded if it behaved competitively.

The main point is that the marginal productivity theory enables us to determine a firm's derived demand for any given input. It demonstrates how the quantity of the input demanded by a firm will vary with the price of the input. The theory presumes that these prices are somehow determined outside the model and asks how much of an input will be demanded in response.

The determination of input prices requires one crucial step. Suppose that from marginal productivity analysis we have obtained a demand function for every resource by every firm. This information along with supply information for each input, and along with supply-demand information for every good in the economy, can be embodied into a Walrasian model of general

equilibrium. From this model a set of equilibrium prices and quantities will be obtained for every item in the economy, including wages for different types of labor, rents for different types of land, etc. This is the marginal productivity model of price determination. No partial equilibrium model will be sufficient to do this job because input markets get affected by interdependence of different activities. An increase in wages in industry I will affect the labor cost in industry II. The price of a resource is not determined by any single employer or by any single resource owner. It is determined by the interactions of all buyers and all sellers.

It may be worthwhile to point out some limitations of this theory. We need to aggregate large sectors of the marginal productivity model to give it a general equilibrium character. In the process we cut down the scope of the model to two to three homogeneous inputs. In particular, models have been constructed using labor and capital only. But this procedure requires aggregation of a heterogeneous set of capital inputs into an artificial single factor. The Cantab (Cambridge)-Italian school regarded this as an objectionable point.

MARGINAL PRODUCTIVITY ANALYSIS AND THE DETERMINANTS OF INPUT SHARES

The analysis of the determinants of the shares of total output accruing to alternative factors of production has been a major concern in the development of economic theory. Since capital, unlike land, is not in fixed supply, the Ricardian theory of rent is not appropriate for discussing this issue. The marginal productivity theory of factor demand provides a tool for this kind of analysis.

Although the marginal productivity theory of factor demand can be used to analyze the distribution of income among an arbitrary number of factors, for simplicity, we will assume that there are only two factors, labor and capital. We also assume that the economy is producing a homogeneous good Q using labor and capital as inputs. The production function is given by $Q = f(K, L)$ and the output is sold in a perfectly competitive market at price p . The total income received by capital from the productive process in one period is rK and the total income accruing to labor is wL . Since the total value of output (that is, the revenue) is given by pQ , labor's share of output is wL/pQ and capital's share is rK/pQ .

Since the economy can be regarded as a firm which is a profit maximizer and if the firm operates in a perfectly competitive market, then it will choose labor and capital so that the marginal revenue product of each factor is equal to its price. Therefore,

$$\text{labor's share} = \frac{wL}{pQ} = \frac{pMP_L L}{pQ} = \frac{MP_L L}{Q} = e_{Q, L} \quad (20.1)$$

$$\text{capital's share} = \frac{rK}{pQ} = \frac{pMP_K K}{pQ} = \frac{MP_K K}{Q} = e_{Q, K} \quad (20.2)$$

where $e_{Q, L}$ ($e_{Q, K}$) is the elasticity of output with respect of labor (capital). Therefore, the shares of labor and capital depend on the technical properties of the production function. If we know the form of the production functions, we can explicitly calculate these shares.

However, it is not clear whether the above relations will ensure that the total value of output equals the total cost of factor inputs. (Since we have been speaking of distribution of income among factors of productions, internal consistency needs to be verified.) This means that it is necessary to check whether the entire production will always be distributed among those who worked on it, or something is left over. This issue is referred to as the adding-up controversy in the literature. Technically we require that

$$\begin{aligned} wL + rK &= pQ, \\ \text{or } \frac{wL}{pQ} + \frac{rK}{pQ} &= 1, \end{aligned} \quad (20.3)$$

or labor's share + capital's share = 1. That is, the payments to the two factors of production must exhaust the total income (revenue). An important question is whether marginal productivity theory ensures this. It can be shown that marginal productivity theory is internally consistent (in the sense that factor shares add up to one, that is Eq. [20.3] holds) if and only if the production function exhibits constant returns to scale (see Appendix 20.1). In this case factor payments exhaust total income.

It is interesting to note that the elasticity of substitution can be quite helpful in analyzing the behavior of factor shares. Recall that the elasticity of substitution measures the percentage change in K/L relative to the change in the marginal rate of technical substitution (MRTS). This elasticity tells us how easy it is to substitute capital for labor and hence becomes helpful

in investigating the consequences of an increasing K/L . Since the firm in question is a cost minimizer, MRTS must equal w/r . Therefore, relative factor share wL/rK will be determined by relative marginal productivity w/r . We can write the elasticity of substitution as

$$\sigma = \frac{\text{percentage } \Delta(K/L)}{\text{percentage } \Delta(w/r)}. \quad (20.4)$$

If $\sigma = 1$, w/r will change in exactly the same proportion in which K/L changes. In this case the relative factor reward wL/rK will remain constant. An increase in the capital-labor ratio K/L will be counterbalanced by an increase in MRTS. For $\sigma > 1$, the percentage increase in K/L will be greater than that in (w/r) and the share of capital over time increases as the capital-labor ratio increases. The opposite situation arises if $\sigma < 1$.

A consequence of modern industrialization is that the capital-labor ratio has an increasing trend over time. While this seems to hold in many cases, an important question that can be asked is how this increased capital-labor ratio affects the distribution of income between capital and labor. As we have just seen, the elasticity of substitution can be regarded as an important tool for understanding the effect of changing input proportions on factor shares.

The Bliss Model

In this section we examine a static general equilibrium model considered by Bliss (1987). Factors and goods are distinct, and factors are not directly consumed. There are m factors available in given quantities and n goods producible from m factors. The number m need not be equal to the number n . Let y^i stand for the output of good i , $i = 1, \dots, n$. Then

$$y^i = f^i(x^i) \quad (20.5)$$

where f^i is the production function of good i , $x^i = (x_1^i, \dots, x_m^i)$ is the vector of factor inputs required for good i 's production. Clearly, x_j^i here indicates the amount of factor j employed in the production of good i ($j = 1, \dots, m$; $i = 1, \dots, n$). The function f^i is concave and exhibits constant returns to scale.

The cost function shows the unit cost of producing good i , given factor prices. Factor prices are a vector $w = (w_1, w_2, \dots, w_m)$, where w_i is the price of factor i ($i = 1, 2, \dots, m$). The unit cost of good i is $C^i(w)$. That is, $C^i(w)$ solves the optimization problem

$$\min_{x_1^i, \dots, x_m^i} (w_i x_1^i + w_2 x_2^i + \dots + w_m x_m^i) \quad (20.6)$$

such that $f^i(x^i) \geq 1$.

In words, $C^i(w)$ is the minimum resource cost required to produce one unit of good i ($i = 1, \dots, n$). Since in the model we are following marginal cost pricing, $C^i(w)$ is in fact the price of good i ($i = 1, \dots, n$). We write the vector $[C^1(w), C^2(w), \dots, C^n(w)]$ as $C(w)$.

There are H households and the h -th household owns factor $x^h = (x_1^h, \dots, x_m^h)$, where x_j^h is the quantity of factor j owned by household h ($j = 1, \dots, m$; $h = 1, 2, \dots, H$). Therefore, the income of household h is $I_h = \sum_{j=1}^m w_j x_j^h$. Each household is assumed to spend its entire income on goods. The vector of goods Z^h demanded by household h is a function of commodity prices and its income. That is,

$$Z^h = g^h[C(w), I_h], \quad (20.7)$$

where g^h is household h 's demand function. The i th component of the vector $Z^h = (Z_1^h, \dots, Z_n^h)$, Z_i^h , gives the quantity of good i demanded by household h .

The following line of reasoning shows that the factor prices w imply demand for factors. Given w , we have the household's income I_h and goods' prices $C(w)$. Hence we have total demand for goods

$$\sum_{h=1}^H g^h[C(w), I_h] = \bar{Z} \quad (20.8)$$

The i th component of the vector $\bar{Z} = (\bar{Z}_1, \bar{Z}_2, \dots, \bar{Z}_n)$ is the quantity of good i demanded by all the households, that is, $\bar{Z}_i = \sum_{h=1}^H Z_i^h$, $i = 1, 2, \dots, n$.

By Shepherd's lemma the amount of factor j used in the production of one unit of good i is the partial derivative of $C^i(w)$ with respect to w_j , denoted by α_j^i (see Chapter 14). In view of the constant returns to scale assumption these coefficients depend on w only. Note that the total household demand for good i is $\bar{Z}_i$. Now, the amount of factor j used in the production of $\bar{Z}_i$ units of good i is $\bar{Z}_i \alpha_j^i$. Therefore, the total demand for factor j is $\sum_{i=1}^n \bar{Z}_i \alpha_j^i$, $j = 1, \dots, m$. Clearly, the supply of factors is $\sum_{h=1}^H x^h = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m)$, where $\bar{x}_j = \sum_{h=1}^H x_j^h$, the total supply of factor j .

To prove the existence of factor prices such that factor demands and supplies are equal, we need to establish a continuity of relationship between factor prices and excess demand for factors. An analogous discussion has been made in the context of existence of equilibrium in the chapter on exchange. The reader is therefore referred to the detailed discussion there.

The Bliss model has been designed to illustrate the manner in which the determinants of the distribution of income may be regarded as the outcome of a general equilibrium model of supplies and demand for factors of production. Prices of factors are determined by the supply and demand for them. However, factor demands are derived demands depending on their use in the production of goods. Technology of production influences the solution for factor prices. The process of cost minimization ensures that factor prices measure marginal products of factors.

The model has certain limitations. It is less general than the standard general equilibrium model. It requires use of a constant returns to scale production function. There is no joint production in the model. Goods are not used as inputs to the production of goods. The introduction of these features will require introduction of fundamentally new principles. The simplicity and neatness of the model will be undermined if we wish to take care of these limitations without introducing new principles.

The Ricardian Model

One shortcoming of the marginal productivity theory is that it does not say anything about the magnitude of the income gap between the poor and the wealthy and the implications of that on the productive mechanism. The same criticism applies to the Bliss model also. We therefore briefly summarize two macro models which deal with this topic to some extent. These are Ricardo's analysis and Kaldor's model. We discuss the Ricardian theory in this section and the Kaldor model will be analyzed in the next section.

The major components of the Ricardo model are: diminishing returns to labor utilizing a fixed quantity of land, the theory of rent, the tendency of universal competition to equalize returns to investment, and the population principle advocated by Malthus. We will consider these issues separately.

The principle of diminishing returns has already been discussed in the chapter on technology. In the Ricardian model this is applied to the increased use of labor (guaranteed by population growth) on society's fixed supply of land. We have seen in Chapter 14 that the landlord will receive the shortfall

of investor's return on his least productive investment (that is, his return on the margin) from the higher return on all other units of investment as rent.

The theory of rent specifies the mechanism for the determination of one distributive share—that going to landlords. In the Ricardian model, profits are determined as what is left for capitalists after payment of rent and wages from total output. Therefore, we need to discuss the wage determination process only. This process relies on the Malthusian population theory which in its simplified form states that whenever wage exceeds some subsistence level people will get married early, and consequently will reproduce more rapidly. This in turn increases the labor force to a point where diminishing returns along with competition of workers will drive wages back to the subsistence level.

We can illustrate the process diagrammatically. Fig. 20.1 shows the behavior of population, wages, rents, and output with respect to time. It is assumed that the ratio between the size of the population and the size of the labor force remains constant over time. Under this assumption, we can present our discussion in terms of the population size. The curve OX shows how total output changes with respect to changes in the population size—output increases at a decreasing rate reflecting the law of diminishing returns.

Total rent payments increase steadily with the growth of population and the resultant increase in land use. Therefore, the curve $X-R$, that is, total output minus rent increases at a decreasing rate with respect to increase in population size. As stated earlier, $X-R$ is that part of output which is available for division between wages and profits. If we write N for population size, then the equation $g = sN$, where $s \neq 0$ is a constant, shows total subsistence wages as a function of the population size. This is shown by the line OS in the figure. Clearly, s is the subsistence wage received by a member of the population.

Suppose that the initial population size is N_0 and that the rate of accumulation is high so that wages are bid up to a point (point W_0) such that they absorb most of the non-rent output. Clearly, $W_0 > g_0$, the subsistence wage level at N_0 . This will encourage population to grow to N_1 . At N_1 , wage payments will cover no more than subsistence payments g_1 . At this point total profit will be high ($g_1 W_1$) which will induce increased capital accumulation. That is, we are implicitly assuming here that, the profit motive acts as an increment to capital accumulation. This is natural since capitalists stimu-

lated by high profits are induced to save a large portion of their earnings, which they invest to earn still more profits. This increased investment increases demand for labor, which in turn will increase total wage to W_1 . We thus have a wage-profit sequence $W_0 g_1 W_1 g_2 W_2 \dots$ which will converge to the point Q , where output minus rent coincides with total subsistence wages. Let us denote the population level corresponding to the point Q by N_Q . At N_Q , in the absence of new technology or other exogenous events, profits and population growth, remain at zero, wage bills remain steady at subsistence level, and rents remain at their maximum attainable level TQ .

Therefore, in this process the workers may gain little (they receive subsistence wage) and capitalists almost lose (profits are zero). The landlords are the only ones who benefit from this dynamic mechanism. The mechanism also shows a diminishing propensity to accumulate as the return from accumulation declines.

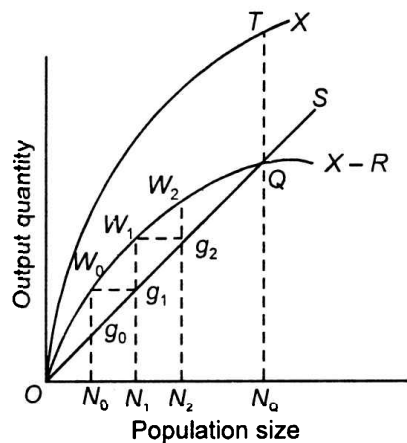


Fig. 20.1 The Ricardian Model

Certain interesting conclusions emerge from the Ricardian model. First, an increase in the standard of living will decrease population growth which in turn increases real wages. Thus one way to increase the welfare of workers is to increase the subsistence level of living. An alternative way of increasing workers' income is to encourage continued capital accumulation, which increases the demand for labor and hence wage bills go up. A third possibility is free trade. Under free trade the costs of the subsistence of workers and rent decrease because of reduction in demand for land. These two factors augment the rewards of capital, hence increase accumulation, and raise wages.

Clearly the Ricardian model is an over-simplification of the complex distribution process. Population growth cannot be determined so mechanically. Technological change leads to growth of real wages even as population grows. Wage determination cannot be done so smoothly by a dynamic process because of involvement of unions and other factors in the wage bargaining process. However, since Ricardo's model assigns distinctive roles to different economic classes, its implications may still have some relevance for some underdeveloped countries.

The Kaldor Model

Kaldor developed a model which derives its roots from the Ricardian framework and from Keynesian theory. He considered a one-sector economy in which there are two classes: one which derives income entirely from capital (the capitalists, who are not wage earners); a second one whose incomes are derived uniquely from wages (the workers). It is assumed that capitalists and workers save different proportions of their incomes. The proportion of income saved is higher for capitalists and lower for workers. Kaldor's model is based on the assumption of a long run full-employment equilibrium. Given the level of full employment investment and total income, the workers and capitalists will uniquely share the total income at which total output demand equals total supply of output.

The level of employment, as in Keynesian analysis, is a function of national output (Y). We denote this functional relationship by $f(Y)$. This level of employment $f(Y)$, when multiplied by the wage rate W gives us the total wage bill $Wf(Y)$. The remaining output $Y - Wf(Y)$ is the income which goes to the other sections of income earners.

By assumption $s_w < s_c$, where $s_w(s_c)$ is the constant proportion of income saved by workers (capitalists). Total desired saving is then given by $s_w Wf(Y) + s_c[Y - Wf(Y)]$. If I is the given level of investment, equilibrium is determined by the condition that total desired saving equals the level of investment, that is,

$$s_w Wf(Y) + s_c[Y - Wf(Y)] = I. \quad (20.9)$$

The constants s_w, s_c and I are given a priori. If we substitute the value of full employment output Y_e in the above equation, we can solve it explicitly for the equilibrium wage level W_e . The solution turns out to be

$$W_e = \frac{I - s_c Y_e}{s_w f(Y_e) - s_c f(Y_e)}. \quad (20.10)$$

The model has certain interesting implications. Suppose that at some wage rate total output is below the full employment level and $f(Y)$ does not depend on W . Then let us examine the consequences of an increase in wage level. Suppose W increases to $W + \Delta W$, where $\Delta W > 0$. Then workers' consumption expenditure becomes $(1 - s_w)(W + \Delta W)f(Y)$. The capitalists' consumption expenditure increases to $(1 - s_c)[Y - (W + \Delta W)f(Y)]$. This means that the change in total consumption expenditure is

$$\begin{aligned}\Delta C = & (1 - s_w)(W + \Delta W)f(Y) + (1 - s_c)[Y - (W + \Delta W) \\ & f(Y)] - (1 - s_w)Wf(Y) - (1 - s_c)[Y - Wf(Y)]\end{aligned}$$

$$\text{or,} \quad \Delta C = (1 - s_w)\Delta Wf(Y) - (1 - s_c)\Delta W[f(Y)], \quad (20.11)$$

which on simplification becomes

$$\Delta C = (s_c - s_w)\Delta Wf(Y) \quad (20.12)$$

which is positive in view of the fact that $s_w < s_c$. Thus, effective demand goes up, and hence employment and the level of national output increase. Thus, an increase in the wage level does not reduce the demand for labor. Increase in effective demand due to increase in wages thus means that wage rise is likely to produce at least part of the income necessary to pay for it.

A second implication of the Kaldor formulation is that capitalists can always increase their income share by decreasing their saving. To see this, note that from Eq. (20.10) we have total wage earnings $W_e f(Y_e)$ equal to

$$W_e f(Y_e) = \frac{I - s_c Y_e}{s_w - s_c} \quad (20.13)$$

This shows that total profit π is given by

$$\begin{aligned}\pi &= Y_e - W_e f(Y_e) \\ &= \frac{[Y_e(s_w - s_c) - I + s_c Y_e]}{s_w - s_c} \\ &= \frac{I - s_w Y_e}{s_c - s_w}.\end{aligned} \quad (20.14)$$

An increase in spending or a decrease in saving by capitalists can take place if their savings rate s_c decreases. A decrease in s_c clearly shows that π increases. Thus, this rather surprising implication of the Kaldor model shows that the more the capitalists spend, the more they have.

There are certain limitations of the Kaldor model. First, it is not clearly argued why in this model the economy has a tendency to approach the level of full employment. Next, the level of employment is assumed to depend on the output level only, not on wage level. However, the model offers some suggestive ideas and it is easy to analyze the model, particularly to study different policy implications.

Example 20.1

Consider the Cobb-Douglas production function

$$Q = AL^\alpha K^{1-\alpha} \quad (20.15)$$

where A and α are positive constants ($\alpha < 1$). Here Q represents the total national output, L is the quantity of labor input and K is the quantity of capital employed. We want to show that if labor is paid a wage equal to its marginal revenue product, this production function will yield a fixed share of wages relative to total income.

The marginal product of labor is

$$\frac{\delta Q}{\delta L} = \alpha AL^{\alpha-1} K^{1-\alpha} \quad (20.16)$$

This multiplied by the price of output p is the wage per worker, w . Therefore, the total wage bill wL is

$$wL = pL \frac{\delta Q}{\delta L} = p\alpha AL^\alpha K^{1-\alpha} = p\alpha Q. \quad (20.17)$$

Therefore the share of wages in total income is

$$\frac{wL}{pQ} = \alpha, \text{ a constant.} \quad (20.18)$$

Similarly the share of capital in total income turns out to be $(1 - \alpha)$. Thus, the ratio between total wage (capital income) and total income is $\alpha(1 - \alpha)$, the exponent of $L(K)$ in the Cobb-Douglas production function.

APPLIED EXAMPLE 20.1 Factor Shares in Indian Agriculture

In their applied general equilibrium model of the Indian economy, Narayana, Parikh and Srinivasan (1991), assumed that the share of wages remains

constant in real terms in agriculture. This implies that the production function can be assumed to be of Cobb-Douglas type and hence that wages are completely and instantaneously adjusted with respect to price changes. Since the national income statistics show a fair deal of stability in factor shares in agriculture, their assumption seems to be quite plausible. This can be regarded as an application of marginal productivity analysis.

Appendix 20.1

SHARES OF CAPITAL AND LABOR

We show that the sum of shares of capital and labor equals unity if and only if the underlying production function exhibits constant returns to scale.

Sufficiency: If $f(K, L)$ exhibits constant returns to scale, then

$$f(tK, tL) = tf(K, L) \text{ for all } t > 0. \quad (20.19)$$

Differentiating both sides of the above with respect to t , we get

$$\frac{\delta f(tK, tL)}{\delta(tK)} \frac{\delta(tK)}{\delta t} + \frac{\delta f(tK, tL)}{\delta(tL)} \frac{\delta(tL)}{\delta t} = f(K, L)$$

or,
$$K \frac{\delta f(tK, tL)}{\delta(tK)} + L \frac{\delta f(tK, tL)}{\delta(tL)} = f(K, L). \quad (20.20)$$

The right hand side of the above equation is independent of t . Any positive value of t is allowable on the left hand side. Putting $t = 1$ on the left hand side we get

$$KMP_K + LMP_L = f(K, L). \quad (20.21)$$

Multiplying this equation by p and using the marginal productivity relations $pMP_K = r$ and $pMP_L = w$, yields eq (20.3).

Necessity: Let us start from the equation

$$\frac{wL}{pQ} + \frac{rK}{pQ} = 1. \quad (20.22)$$

Using the marginal productivity relations $pMP_L = w$ and $pMP_K = r$ again, we get

$$LMP_L + KMP_K = Q. \quad (20.23)$$

Since $Q = f(K, L)$, $MP_L = \delta f(K, L)/\delta L$ and $MP_K = \delta f(K, L)/\delta K$, the above equation becomes

$$L \frac{\delta f(K, L)}{\delta L} + K \frac{\delta f(K, L)}{\delta K} = f(K, L). \quad (20.24)$$

From this by the converse of Euler's theorem for homogeneous function, it follows that f is homogeneous of degree one, that is $f(tK, tL) = tf(K, L)$ for all $t > 0$. Thus, f exhibits constant returns to scale.

EXERCISES

Multiple Choice

1. According to marginal productivity analysis the share of total output accruing to a particular factor is
 - a) Equal to the elasticity of output with respect to the factor.
 - b) Less than the elasticity of output with respect to the factor.
 - c) Greater than the elasticity of output with respect to the factor.
 - d) Anyone of the above is possible.
2. Which of the following conditions is not relevant to the Ricardian macromodel?
 - a) The theory of rent.
 - b) The Malthusian population principle.
 - c) The law of variable proportions.
 - d) The law of diminishing marginal productivity.
3. A sufficient condition to guarantee that the relative factor reward remains constant is that
 - a) The elasticity of substitution between the factors is equal to one.
 - b) The elasticity of substitution between the factors is greater than one.
 - c) The elasticity of substitution between the factors is less than one.
 - d) None of the above.

4. In the Bliss model of factor shares, to prove the existence of factor prices such that excess factor demands are zero, we require
 - a) Continuity of the excess factor demand functions in factor prices.
 - b) Concavity of the excess factor demand function in factor prices.
 - c) Monotonicity of the excess factor demand function in factor prices.
 - d) (b) and (c).
5. If the marginal product of capital K multiplied by the quantity of K used is always equal to the marginal product of labor L multiplied by the quantity of L used, and if the price of K is twice the price of L , then the quantity L used will be
 - a) $2K$.
 - b) $K/2$.
 - c) K^2 .
 - d) None of the above.

Quantitative Problems and Questions

1. Prove or disprove the following:
 - a) A sufficient condition to ensure that the marginal productivity theory is internally consistent is the law of diminishing marginal productivity.
 - b) In the Kaldor macro distribution model it is possible for capitalists to increase their income share by increasing their spending.
 - c) According to the Ricardian theory of rent, for any output y , the average cost of y including rent is smaller than the marginal cost of y excluding rent.
 - d) In the Bliss model of factor share distribution suppose that n production functions $f^1, \dots, f^n$ satisfy the assumption $f^i(0 \ 1^m) = 0$, where $i = 1, \dots, n$ and 1^m is the m -coordinated vector of ones. That is, if quantities of all the m inputs are zero, output is also zero. Then this assumption along with concavity will ensure constant returns to scale principle.
 - e) Suppose that the production function $f(K, L)$ exhibits constant returns to scale. If the marginal product of labor (L) exceeds the average product of labor, then the marginal product of capital (K) is positive.
2. Clearly examine the implications of the constant returns to scale assumption for the production functions in the Bliss model.

3. Consider the production function

$$Q = (aK^{-1} + bL^{-1})^{-1}$$

where $0 < a, b < 1$, $a + b = 1$; Q is the output, and K and L are capital and labor.

- a) Calculate the elasticity of substitution.
 - b) Suppose $K = 1/2$ and $L = 1$. Determine capital's and labor's share of output. What is the relative share?
 - c) Assume that the rental price of capital and the wage rate are both equal to 1. What K/L ratio will be used?
4. Suppose that the behavioural relations of a hypothetical economy have the following relations: given level of investment $I = 10$, full employment output $Y = 30$, s_c and s_w , the proportions of income saved by capitalists and workers are 0.5 and 0.4 respectively.
- a) Compute the total wage earnings and total profits.
 - b) What happens to total profits if the proportion of income saved by capitalists increases from 0.5 to 0.6?
5. In a hypothetical economy suppose that the given level of investment $I = 6$. The equilibrium output which is below full employment level is $Y = 16$. The proportions of income saved by capitalists and workers are 0.5 and 0.4 respectively. The employment function is of the type $f(Y) = \sqrt{Y}$.
- a) What is the wage rate?
 - b) Suppose wage rate increases by 1. What will be the impact of this on effective demand? Does it decrease or increase? What will happen to the level of employment and the wage bill?
6. Critically evaluate the assumptions underlying the Ricardian macrodistribution model.
7. Identify the falsity in each of the following statements.
- a) Marginal product is calculated as output per worker.
 - b) Distribution theory is no problem. You simply figure out how much each factor produced, and then give it that part of output.
8. Discuss the alternative ways of increasing workers' income in the Ricardian macromodel.

21

International Trade

Statistics show that the economies of the world are becoming more interdependent. Indian consumers are buying more foreign products and foreign consumers are buying more Indian products. Producers around the world are using more imported raw materials and parts in the products they manufacture. How should Indians react to this interdependence? Does the availability of Japanese automobiles and other foreign goods simply mean less demand for Indian products and fewer jobs in Indian domestic manufacturing industries? To respond intelligently to questions of this type, it is necessary to have some understanding of international economics. We need to understand why countries trade. What are the benefits of trade? Is trade mutually beneficial or does it help one country at the expense of another? What are the arguments for and against trade barriers?

The pure theory of international trade studies the basis for and the gains from trade. The theory of commercial policy examines the reasons for and the results of barriers to the free flow of trade. Since these topics can be discussed by treating each nation as a single unit and by dealing with individual (relative) commodity and factor prices, they can be regarded as microeconomic aspects of international economics.

Pre-Ricardian Economic Views on Basis for Trade

Why do countries trade? The various nations of the world are not equally blessed in terms of natural resources, capital and labor endowments. Therefore different nations have different production abilities. For instance, Kuwait may be self-sufficient in petroleum products because of its possession of oil. But Kuwait must depend on other nations for items like automobiles, electronic gadgets, tea, rice, coffee. One reason why countries trade, then, is to acquire the products they cannot produce themselves. But this is not the sole—or even the most important—reason for trade. In this section we briefly discuss the pre-Ricardian doctrines.

The mercantilist economic doctrine that prevailed during the sixteenth and the seventeenth centuries of the development of the modern national state maintained that the purpose of trading with foreign countries was to export as much domestic product as feasible to import the necessary quantities of precious metals (mostly gold) to keep the economy functioning. According to mercantilism the more gold a country could acquire, the richer and more powerful it became. Thus, mercantilists advocated that the government should encourage exports as long as they bring in gold, but restrict imports.

Reacting against this simplistic concept of international trade, in his famous book *The Wealth of Nations*, Adam Smith saw clearly that free trade is the best policy for the nations of the world. Smith argued that if country 1 can produce a good more efficiently, say with less labor, than country 2, and country 2 can produce some other good with less labor, they would both lose if they were to continue producing both goods. Each should concentrate on the production of the good that it can produce with less labor cost, that is, in the production of the good in which it has advantage, and then they should exchange their surplus production. Thus, both countries have more of both goods by trading. This explanation of the possible gain from trade is popularly called the theory of absolute advantage because the production of the good in which a country should specialize actually costs less in terms of labor units than in the other country.

We may give an example to illustrate the idea. Table 21.1 shows that country 1 has an absolute advantage over country 2 in the production of good *A* and country 2 has an absolute advantage in the production of good *B*. This will follow from the fact the labor cost of production for each unit of good *A* (*B*) is smaller in country 1 (2).

Table 21.1 Absolute Cost Advantage

<i>Labor cost of production of 1 unit (in hours)</i>		
	Good A	Good B
Country 1	8	12
Country 2	16	9

Suppose that the two countries agree to exchange the two goods in 1:1 ratio. That is, 1 unit of good *A* is exchanged for 1 unit of good *B*. Country 1 can then get 1 unit of good *B* by sacrificing 8 units of labor, whereas it has to give up 12 units of labor if it produces the good on its own. Similarly country 2 will have to give up 9 units of labor to get 1 unit of good *A*, whereas it will require 16 units of labor if it produces the good itself. Therefore, with a given resource, both countries can have more of both goods by trading. That is, a nation should specialize in the production of those goods in which it has an absolute advantage. This kind of international specialization would result in an increase in world output which would be shared by the trading nations.

Ricardo and the Principle of Comparative Advantage

Smith's theory of absolute advantage is certainly correct. Ricardo proved that absolute advantage is not necessary for international trade to bring gain to two countries. What happens if one country has absolute advantage over another country in all branches of production? Does it still benefit the countries to trade? Ricardo's answer was yes. He demonstrated that in order for the gain to arise it was sufficient that the countries should have what is called 'comparative advantage'.

To illustrate this let us consider the simple model described in Table 21.2. According to this model India has an absolute advantage in the production of tea as well as cloth, because the labor cost of production for each unit of the two commodities is smaller in India than in Japan. Now, to see how countries can benefit from trade in such a situation we need to make comparisons based on opportunity cost. Let us recall that in such a situation the opportunity cost of a good is what we have to give up to obtain one (additional) unit of the good. The opportunity cost of steel produced in India would be whatever of the domestic products we could have produced with the same resources that we would employ for producing one (additional) unit of steel. If we look at Table 21.2, then in India the opportunity cost of tea (cloth) is $8/9$ ($9/8$) cloth (tea) units. On the other hand, for Japan these costs turn to be $12/10$ cloth units and $10/12$ tea units respectively.

Table 21.2 Comparative Cost Advantage

<i>Labor cost of production of 1 unit (in hours) of</i>		
	Tea	Cloth
India	80	90
Japan	120	100

We say that a country enjoys a comparative advantage in producing a good if the opportunity cost of producing the good is lower at home than in the other country. Thus, India has a comparative advantage in the production of tea and Japan has a comparative advantage in the production of cloth.

We will now see how in our two-good two-country model when one country has a comparative advantage in the production of one good and the other country has a comparative advantage in the production of the other good, both countries gain from trade, regardless of the fact that one of the countries may have an absolute disadvantage in the production of both goods. Note that in India it takes 80 hours of labor to make 1 unit of tea and 90 hours of labor to make 1 unit of cloth, which implies that 1 unit of tea will cost $8/9$ units of cloth. Similarly in Japan 1 unit of tea will cost $12/10$ units of cloth. If trade opens between these countries and the international price of tea is somewhere between $8/9$ and $12/10$, then both countries will gain from trade. This follows from the fact that if Japan can import 1 unit of tea at a price less than $12/10$ units (the home price of tea), then it certainly gains. Similarly India gains from trade because of the fact that it is able to sell to Japan at a price higher than $8/9$ units of cloth, the home price of tea.

Under the assumption that in the international market 1 unit of tea can be exchanged for 1 unit of cloth, mutual trade benefits both the countries. (Clearly, a 1:1 exchange ratio ensures that the international price of tea lies between $8/9$ and $12/10$ units of cloth.) That is, in this case it is advantageous for India to export tea and import cloth, while Japan benefits from exporting cloth and importing tea. To see this note that Japan needs to give up 120 man-hours of labor for every unit of tea it wants. But under trade, with 100 hours of effort it can produce 1 unit of cloth for 1 unit of tea. Therefore, in autarky (or absence of trade), Japan would have to give up 120 man-hours of labor for every unit of tea, whereas under trade it is required to give up only 100 hours. Each remaining 20 hours could be employed for producing cloth or could be consumed as leisure, which in turn means that Japanese can now enjoy a higher level of consumption or a higher amount of leisure with a given amount of consumption. Using analogous arguments we can

demonstrate that India also gains from trade in the case when 1 unit of tea is exchangeable for 1 unit of cloth.

Thus, economic logic dictates that each country should specialize in the products it can produce at relatively low opportunity costs and trade those products for the items that other countries can produce more cheaply (in terms of opportunity costs). This logic is called **the principle of comparative advantage**.

We can see the gains from trade more clearly by using the production possibilities frontier introduced in Chapter 17. A production possibilities frontier shows the combinations of goods that an economy is capable of producing efficiently with its given stock of economic resources and existing techniques of production.

Let us assume that each day India and Japan have 72,000 and 60,000 working hours respectively to use in producing either tea or cloth or both. Given constant opportunity costs, the production possibility frontier will be a straight line rather than a curved line. This is so, because one always has to sacrifice $8/9$ ($12/10$) units of cloth in India (Japan) to get one more unit of tea. Now, in India 72,000 working hours can be used to produce either 900 kgs. of tea or 800 mtrs. of cloth or any other combination of tea and cloth found on India's production possibilities frontier. Similarly in Japan 60,000 working hours can be used to produce either 500 kgs. of tea or 600 mtrs. of cloth or any other combination of tea and cloth found on its production possibilities frontier. Recall that any combination of products we can plot either on or within the production possibilities is available to the society. (Any point inside the frontier is inefficient.) Combinations falling out the frontier are beyond the economy's production capability. Suppose that prior to trade India chooses to produce and consume 450 kgs. of tea and 400 mtrs. of cloth, whereas Japan chooses a combination of 375 kgs. of tea and 150 mtrs. of cloth (refer to Fig. 21.1). These combinations lie on the corresponding frontiers. (To produce 450 kgs. of tea India needs to give up $450 \times 80 = 36,000$ working hours. Therefore, when India produces 450 kgs. of tea on the frontier, it **must** produce $(72,000 - 36,000) / 90 = 400$ mtrs. of cloth. A similar calculation holds for the Japanese combination (375, 150).)

Suppose next, that the two countries decide to specialize along the lines of comparative advantage, agreeing to trade at our hypothetical exchange ratio: one unit of cloth = one unit of tea. India has a comparative advantage in tea, so it will use all its 72,000 working hours to produce tea and will trade for cloth. The Japanese will produce cloth—their product of comparative advantage—and will trade for tea.

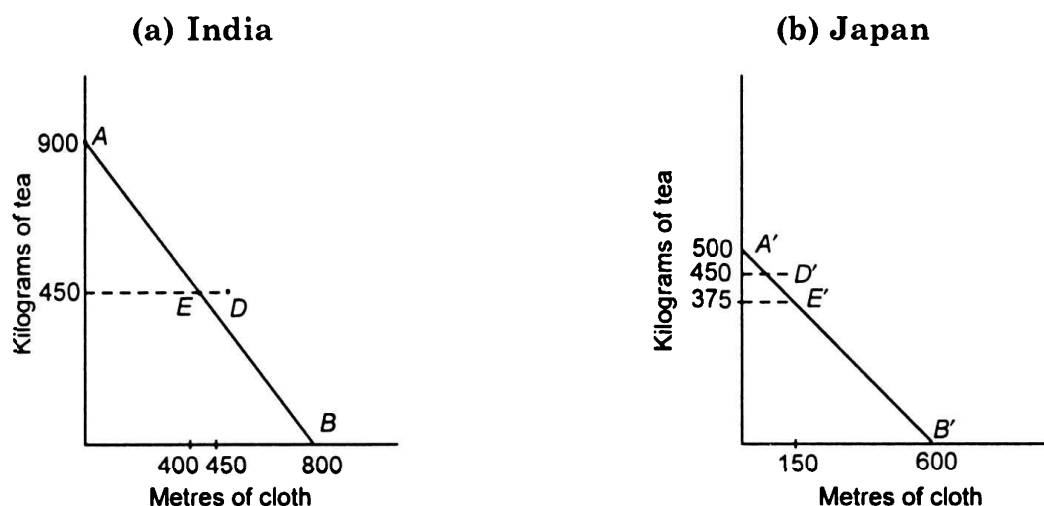


Fig. 21.1 Production Possibilities Frontier

We know that India is now producing 900 kgs of tea and Japan is producing 600 mtrs. of cloth. That is, India (Japan) moves to point A (B') from point E (E') in production. What will be the result if India trades 450 kgs. of tea for some cloth? With the given exchange ratio, India will receive 450 mtrs. of cloth in return for 450 kgs. of tea and still have 450 kgs. of tea left over. Thus, India ends up consuming at point D which involves a gain in consumption of $(450 - 400)$ mtrs. = 50 mtrs. of cloth over its no-trade consumption point E . That is, D lies above the production possibilities frontier AB . On the other hand, Japan will have 150 mtrs. of cloth left over plus 450 kgs. of tea. Therefore, Japan will end up consuming at point D' with a gain of $(450 - 375)$ kgs. = 75 kgs. of tea over point E' . Hence D' lies above the frontier $A'B'$.

Trade and specialization have made it possible for each country to move beyond its production frontier and enjoy a combination of products that it could not obtain on its own. That is, each country becomes better off under trade than it would have been under autarky. Moreover, the production of the two countries has increased from $(450 + 375)$ kgs. = 825 kgs. to $(450 + 450)$ kgs. = 900 kgs. of tea and $(400 + 150)$ mtrs. = 550 mtrs. to $(450 + 150)$ mtrs. = 600 mtrs. of cloth. The principle of comparative advantage has allowed each of the trading partners to obtain more goods from each resource unit and enjoy a higher standard of living.

Analytical Formulation of the Ricardian Theory

We will now reformulate the Ricardian theory using mathematical methods. Let us begin by stating the explicit and implicit assumptions on which

Ricardo's law of comparative advantage is based. These assumptions are: (i) there are two countries and two goods; (ii) there is only one factor of production—labor which is mobile within a nation but immobile internationally; (iii) the costs of production are constant—regardless of the quantity produced the costs per unit of production remain the same; (iv) there are no transport costs; (v) there is free trade between the countries—goods are exchanged for goods in barter trade; (vi) there is no technological change; and (vii) the value or price of a commodity is equal to the amount of labor going into the production of the commodity. Clearly, the first six assumptions are self explanatory. The last assumption, which is known as the labor theory of value, means that the price of a commodity is not influenced by its demand. We will refer to these seven assumptions as the Ricardian assumptions.

In view of assumptions (i) and (ii), let us consider a world consisting of two countries 1 and 2, where each country produces two goods C (say, cloth) and T (say, tea) with labor, which is the only factor of production. Suppose L^1 and L^2 stand for the labor endowments of country 1 and 2 respectively. We write l_C^1 and l_T^1 for the constant labor cost of production (in hours) of 1 unit of goods C and T respectively in country 1 (assumption [iii]). For country 2, the corresponding figures are given by l_C^2 and l_T^2 respectively. (If we refer to the model considered in the previous section and let country 1 be India and country 2 be Japan, then $l_C^1 = 90$, $l_T^1 = 80$, $l_C^2 = 100$, and $l_T^2 = 120$. Note also that there we had $L^1 = 72,000$ and $L^2 = 60,000$.)

Next, suppose C^1 and T^1 denote the outputs of goods C and T respectively in country 1. The amount of labor employed by this country in the production of C and T are given by L_C^1 and L_T^1 respectively. Then in view of assumption (vii), country 1's production functions are

$$\begin{aligned} C^1 &= L_C^1 / l_C^1, \\ T^1 &= L_T^1 / l_T^1. \end{aligned} \tag{21.1}$$

Similarly, for country 2 we have

$$\begin{aligned} C^2 &= L_C^2 / l_C^2, \\ T^2 &= L_T^2 / l_T^2, \end{aligned} \tag{21.2}$$

where $C^2(T^2)$ is country 2's outputs of good $C(T)$ and $L_C^2(L_T^2)$ is the amount of labor input of $C(T)$ industry in country 2.

The amount of goods produced in each country are constrained by total labor endowment of the country

$$C^1 l_C^1 + T^1 l_T^1 \leq L^1, \quad (21.3)$$

and
$$C^2 l_C^2 + T^2 l_T^2 \leq L^2. \quad (21.4)$$

We may rewrite these inequalities as

$$\frac{C^1}{L^1/l_C^1} + \frac{T^1}{L^1/l_T^1} \leq 1, \quad (21.5)$$

$$\frac{C^2}{L^2/l_C^2} + \frac{T^2}{L^2/l_T^2} \leq 1. \quad (21.6)$$

The set of combinations of (C^1, T^1) [(C^2, T^2)] satisfying Eqs. (21.5) [(21.6)] describes the production possibilities set of country 1 (2). If for any combination (C^1, T^1) [(C^2, T^2)] equality is satisfied in Eq. (21.5) [(21.6)] then it lies on country 1(2)'s production possibilities frontier. It may be noted that for our two-country model considered in the previous section, $L^1/l_C^1 = 800$, $L^1/l_T^1 = 900$, $L^2/l_C^2 = 600$ and $L^2/l_T^2 = 500$.

Let us suppose that country 1(2) has a comparative advantage in the production of good $T(C)$. That is,

$$\frac{l_T^1}{l_C^1} < \frac{l_T^2}{l_C^2} \quad (21.7)$$

or, equivalently,
$$\frac{l_T^1}{l_T^2} < \frac{l_C^1}{l_C^2}. \quad (21.8)$$

Clearly, while in Eqs. (21.5) and (21.6) we compare ratios between costs of producing different commodities within the same countries, in (21.7) we compare the ratios between costs of producing the same commodities in different countries. We will say that an exchange ratio r/s belongs to the comparative advantage region if $l_T^1/l_C^1 < r/s < l_T^2/l_C^2$. The exchange ratio r/s here means that r units of good T are exchangeable for s units of good C . Recall that in our numerical example $r:s$ has been assumed to be 1:1. There we had $l_T^1/l_C^1 = 8/9$ and $l_T^2/l_C^2 = 12/10$.

We can now formally state the Ricardian comparative advantage theorem.

THEOREM: *Under the Ricardian assumptions suppose that one country, say, country 1, has a comparative advantage in the production of one good, say good T , and the other country, country 2, has a comparative advantage in the production of the other good, good C . Then complete specialization by country 1(2) in the production of good T (C) and free trade with country 2(1) for good C (T) at an exchange ratio belonging to the comparative advantage region will make country 1(2) better off than it would have been under autarky. That is, under complete production specialization in the good of comparative advantage, the post-trade consumption point of each country, if free trade takes place in the comparative advantage region, will lie above the production possibilities frontier of the country.*

ELIMINATION OF RICARDO'S SIMPLIFICATIONS

Increasing Costs

The meaning of constant opportunity costs, or straight line production possibility frontier, is that factors of production are equally efficient in all lines of production. But this assumption seems to be quite unrealistic. For instance, in our two-good, one-factor (labor) case, we would expect some workers to be more efficient in the production of good C and the others to be especially efficient in the production of the other good T . In such a case we have variable opportunity costs. If we assume that the opportunity costs of obtaining an additional unit of a commodity increase as more units of the commodity are produced, then the production possibility frontier curves out from the origin. That is, the frontier is concave to the origin; equivalently, we say that the production set is convex.

To understand increasingness of opportunity costs suppose we have some factors, such as highly skilled labor especially trained to produce good C (say cloth), employed in clothing production. A further expansion in clothing production will rely on less skilled workers and they are released from good T -production. Now, it is highly likely that these workers who are less skilled in production of good C were quite efficient in producing good T . Therefore, the amount of good T -foregone in order to release these workers is likely to be quite high and naturally this amount is expected to increase as we employ more workers who are more efficient in production of good T .

With a concave production possibilities frontier, there is a basis for mutually advantageous trade whenever there is a difference in the pre-trade 'relative commodity price' between the two countries. To see this, suppose

that in the absence of trade, country 1(2) produces and consumes at point A (A') in Fig. 21.2. At A , the relative price r_C/r_T is higher than the corresponding relative price at A' , where $r_C(r_T)$ is the cost of producing one unit of $C(T)$, expressed not in terms of money amount, but in terms of labor content.

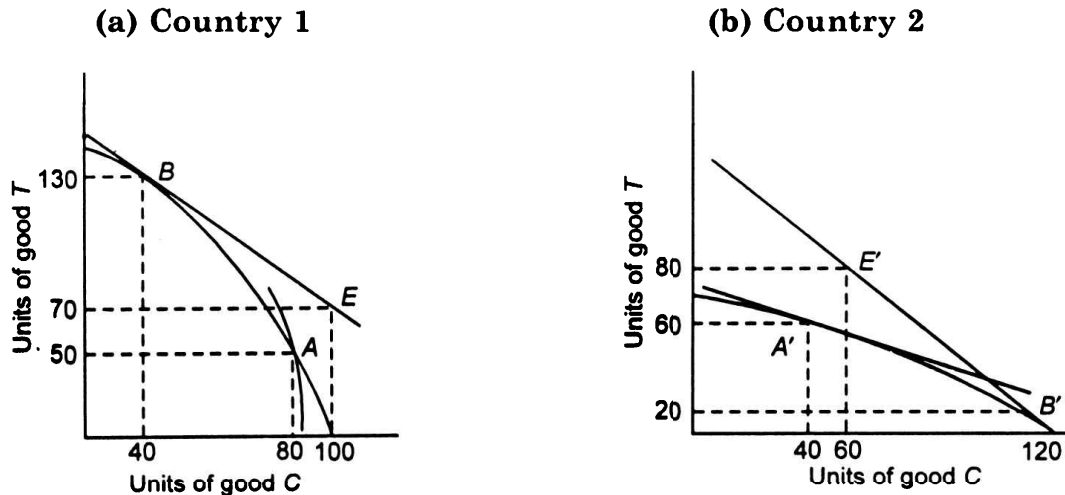


Fig. 21.2 Increasing Opportunity Costs and Gains from Trade

Clearly, r_C/r_T is nothing but the opportunity cost of producing one more unit of C . Suppose that $r_C/r_T = 1/4$ at A' and the corresponding ratio is 4 at A . Country 2(1) has comparative advantage in the production of good $T(C)$. Mutually advantageous trade is possible within the limits $1/4 < r_C/r_T < 4$. If r_C/r_T is stabilized at 1 with trade, then country 1(2) can move from $A(A')$ to $B(B')$ in production. Country 1 can exchange 60 units of its 130 units of T (produced at B) for 60 units of C from country 2 and end up consuming at point E (with a gain of 20 units of C and 20 units of T over the point A). Country 2, on the other hand, exchanges 60 units of T for 60 units of C with country 1, ends up at point E' (which involves 20 units of both C and T more than at A'). Note that as country 1 specializes in the production of good T it incurs increasing cost in the production of T . Country 2, on the other hand, incurs increasing costs in the production of good C as it specializes in good C . Specialization will continue until $r_C/r_T = 1$. Another interesting point to be noted here is that the two countries do not specialize completely in the production of their goods of comparative advantage (as in the case of constant costs).

Introduction of Demand

When only the supply side of international exchanges was discussed, the conclusion was drawn that international prices or factor rewards in trading

countries will have to be within certain limits. But a precise figure for factor reward in a country cannot be established so long as we do not incorporate the demand conditions for various goods. We now consider explicitly the tastes and demand preference of each country. These are introduced by community indifference curves.

A community indifference curve is an average of the indifference curves of all individuals in the country and as such may be expected to retain the same shape. Therefore, this curve shows the various combinations of two commodities which generate equal satisfaction to the community. Higher indifference curves represent higher utility levels. They are downward sloping and strictly convex to the origin. But averaging may lead to a serious problem. Suppose the marginal rate of substitution of good *C* for good *T* is 3. Then the nation's citizens will, on balance, be prepared to give up three units of *C* for an extra unit of *T*, but if the trade-off is actually made, some consumers will be happy and others unhappy. Thus there is a problem in using a community indifference curve since it implicitly involves interpersonal comparison of satisfaction, and especially so if trade changes the distribution of income. Nevertheless, the community indifference curve is a useful way to represent demand side considerations, relating them to underlying consumer preferences.

It is now possible to restate the meaning of gains from trade, combining supply and demand sides. In the absence of trade, a country is in equilibrium when it reaches the highest possible indifference curve given its production possibilities frontier. At equilibrium the community indifference curve is tangent to the production frontier. The common slope of the two curves at the tangency point gives the commodity price ratio in autarky. When this pretrade commodity price ratio differs in two countries, there is a basis for mutually advantageous trade between them.

In Fig. 21.3 the community indifference *I* is the highest indifference curve that country 1 can reach with its given production possibilities frontier. Thus, in autarky, country 1 produces and consumes at *A*. Similarly, country 2 produces and consumes at *A'* in isolation. The commodity price ratio at *A* is higher than the corresponding ratio at *A'*. Therefore country 1(2) benefits from specialization by comparative advantage in the production of *T*(*C*). The optimal output mix is at point *B*(*B'*) for country 1(2) where the commodity price ratios are equal. The consumers of country 1(2) will maximize utility by consuming at *E*(*E'*), the tangency point between the line whose slope is the negative of the common price ratio and the community indifference curve II(IV). Since this indifference curve lies on the right of the autarky indifference curve I(III), the society benefits from trade.

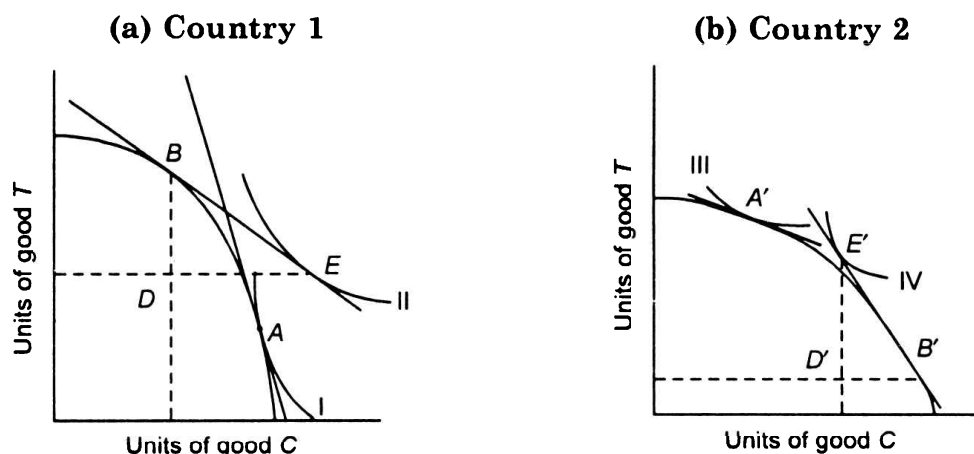


Fig. 21.3 Community Indifference Curves and Gains from Trade

The extent of trade for country 1 is measured by the triangle BDE . The shorter sides of the trade triangle BDE measure imports (DE) and exports (BD), the common commodity price ratio, which is also known as the world price ratio, or the terms of trade, is measured by the slope of the trade triangle's hypotenuse BD/DE . A similar analysis holds for the trade triangle $B'D'E'$ of country 2. It is important to note that in a two-country world model the trade triangles must be of equal size because each country's exports are the other country's imports and vice-versa. In a many-country world the same identity between exports and imports must characterize the equilibrium set of world prices.

A useful tool for identifying the common relative commodity price or the world commodity price ratio is the offer curve, which traces out the quantities of goods a country will want to import for various quantities of its exports. Offer curves can be derived for every trading nation, and their intersection will determine the set of world prices at which world demand and supply for each product are equalized.

In panel (a) of Fig. 21.4 country 1 starts at the autarky equilibrium point A . If trade takes place at some common price ratio, the country moves to D in production, exchanges DH units of T for HG units of C and reaches point G . This gives point G in panel (b). Now suppose we have a different common commodity price ratio and the country moves to B in production. In this case BF units of T are traded for FE units of C and the country reaches E . This gives point E in panel (b). If we join the origin with points G , E , and other similarly obtained points, we generate the offer curve for country 1 shown in panel (b). The offer curve of country 2, the trade partner of country 1, can be obtained in a similar manner.

The Heckscher-Ohlin theory is usually formulated in terms of a two-factor model, with capital (K) and labor (L) as the two factors of production. Let us consider a world consisting of two countries, each being able to produce two goods using the same two factors K and L . The endowments of the factors K and L are fixed for each country. We will refer to the countries as country 1 (say, India) and country 2 (say, Japan) respectively. The two goods can be called C (say, cloth) and T (say, tea).

The Heckscher-Ohlin theory of trade is based on certain assumptions. These assumptions are: (i) all production functions exhibit constant returns to scale with the property of diminishing returns; (ii) there is perfect competition in both commodity and factor markets; (iii) the factors are fully employed, mobile domestically with no impediments and with negligible transport costs but are completely immobile internationally; (iv) there is free trade and goods are mobile internationally as well domestically with no impediments and with negligible transport costs; (v) the production functions are such that the two commodities show different factor intensities, and there is no factor intensity reversal; (vi) the production functions differ between commodities, but are the same in both countries, that is, good C is produced with the same technique in both countries and good T is produced with the same technique in both countries; and, (vii) countries have equal taste or demand preferences.

The above assumptions are not difficult to understand. The first assumption means that the technology linking inputs and outputs is such that a given percentage increase in all factor inputs increases outputs by the same percentage. The meaning of the second assumption should be clear from Chapters 10 and 14. It implies that the factors of production will be allocated in an optimal way. In the third assumption international mobility of factors is ruled out either because of prohibitively high transports costs or other restrictions, such as immigration laws. The fourth assumption is an abstraction to facilitate analysis. It implies that commodity prices under trade will be the same in both countries.

According to the fifth assumption the feature that distinguishes production of one good (say C) from that of the other good T is the factor intensity that each requires; the ratio of capital to labor used in C differs from that adopted by the sector T . Given the two factors (capital K and labor L) and two commodities C and T , we say that a commodity (for example C) uses a factor more intensively or is more intensive in a factor (for example capital) relative to the other commodity if the K/L input ratio in the former commodity is greater than the K/L input ratio in the latter. Therefore, assumption (v) means that different techniques of production are used in the two industries.

Elimination of factor intensity reversal will imply that there is one-to-one correspondence between factor intensities and factor prices.

The behavior of capital-labor ratio in the two sectors in a country in the absence and presence of a factor intensity reversal are shown in Fig. 21.5. In both cases the capital-labor ratio K/L is an increasing function of the factor price ratio or relative price of factors (p_L/p_K), since producers will find it profitable to substitute capital for labor as the relative price of labor goes up. (Here $[p_L/p_K]$ is the price of labor relative to that of capital.) This can be demonstrated diagrammatically by considering the various points of tangency between the unit isoquant and isocost lines with varying slope. In the case of no factor intensity reversal, the two curves will never intersect; in the case of a factor intensity reversal, they will.

In panel (a) of Fig. 21.5, the curve showing K/L ratio in industry C i.e. curve $(K/L)_C$, lies above the $(K/L)_T$ curve since commodity C is always assumed to be more capital-intensive than commodity T . In panel (b), the curves under consideration intersect at I . Thus, to the left of I , that is, for lower (p_L/p_K) ratios, good T is capital-intensive relative to C , while the opposite is true to the right of I , that is, for higher (p_L/p_K) ratios.

The sixth Heckscher-Ohlin assumption is quite a strong one. It says that the best techniques of production are known to both the countries. The final assumption says that demand differences are non-existent between the countries because of the assumption of identical tastes. We will refer to these seven assumption as the Heckscher-Ohlin assumptions.

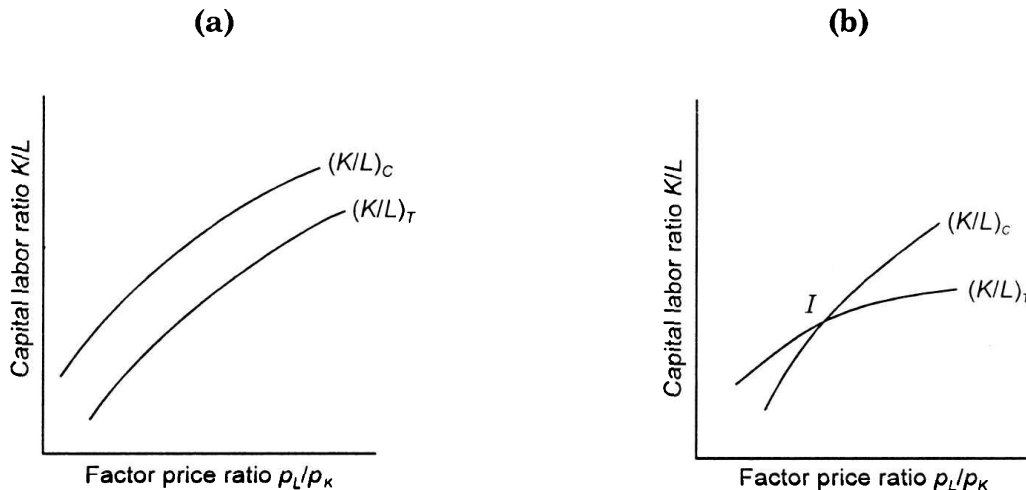


Fig. 21.5 Behavior of Capital-Labor Ratio
(in two sectors without and with factor intensity reversal)

In addition to factor intensity, the Heckscher-Ohlin trade theory is based on a concept known as relative factor abundance. A country is said to be relatively labor-abundant if the ratio of home overall labor supply to that of capital exceeds the labor/capital ratio of endowments abroad. We can now formally state the Heckscher-Ohlin theorem as follows.

THEOREM: *Under the Heckscher-Ohlin assumptions a country exports (imports) that good which is relatively more intensive in the factor which is relatively more (less) abundant in that country. That is, the relatively labor-abundant country exports the relatively labor-intensive commodity.*

The above theorem has a strong intuitive appeal. It means that each country exports the commodity that requires relatively intensive use of the factor of production in which it has relative abundance. For example, India which is presumably more labor-abundant than Japan, would export those goods which are more labor-intensive such as tea, and import more capital-intensive goods such as machinery.

We now give a graphical demonstration of the Heckscher-Ohlin theorem. (A rigorous proof of the theorem is presented in Appendix 21.1) Panel (b) of Fig. 21.6 shows the production possibilities frontiers of country 1 and country 2 on the same set of axes. Technology is assumed to be the same in both countries. Production is skewed toward the C-axis in country 2 because *C* is capital-intensive and country 2 has a relative abundance of capital. Production is skewed toward the T-axis in country 1 because *T* is labor-intensive and country 1 has a relative abundance in labor. The same indifference map refers to both countries because of the assumption of identical demand preferences.

In the absence of trade (panel [a] of Fig. 21.6), country 1 produces and consumes at point *A* on its indifference curve *I*. At *A* the common slope of the curve *I* and the production possibilities frontier gives commodity price ratio in country 1 under autarky. Similarly, in the absence of trade, country 2 produces and consumes at *A'*.

With trade (panel [b] of Fig. 21.6), country 1 produces at point *B* (that is, it specializes in the production of good *T*) and consumes at point *E* on the indifference curve *II*. The corresponding points for country 2 are given by *B'* and *E'* respectively. By specializing and trading each country can reach a community indifference curve which was unattainable under autarky. This completes the proof of the theorem.

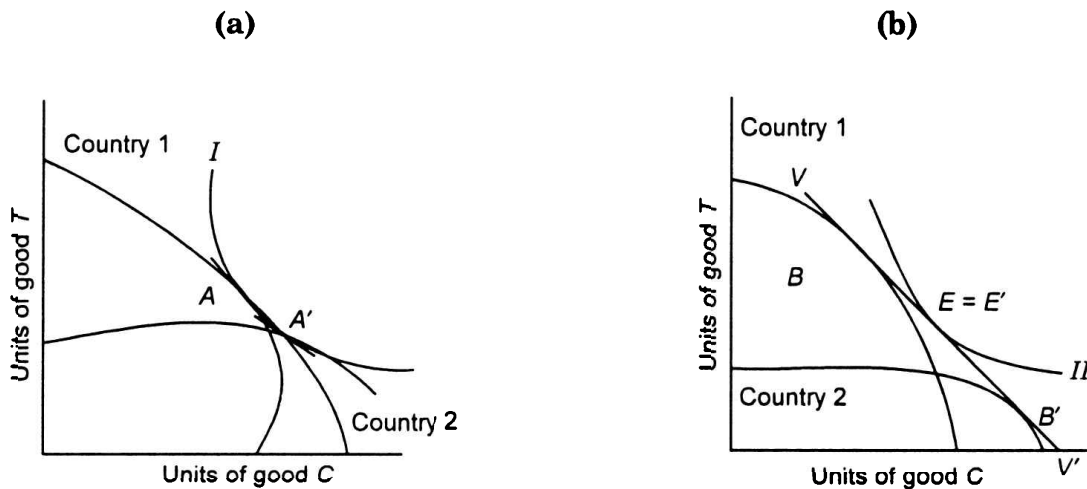


Fig. 21.6 Heckscher-Ohlin Theorem

Observe that as long as the commodity price ratio in country 1 differs from the corresponding ratio in country 2, mutually advantageous trade between the two countries is possible. Therefore, trade between the two countries will take place until the commodity price ratios are equalized. The common commodity price ratio is given by the absolute value of the slope of the line VV' .

THE FACTOR PRICE EQUALIZATION THEOREM

Under trade country 1 specializes in good T , that is, production of T expands (and C contracts) along its production possibilities frontier. Expansion of labor-intensive good T in country 1 raises wage and lowers rental price of capital. In country 2 the opposite changes are introduced by trade. Move to a capital-intensive good there drives up rental price of capital and lowers wages.

This in turn implies that trade allows country 1's labor to be exported (via good T) and relieves pressure on scarce labor in country 2. As we have observed, trade increases wages in country 1 and lowers wages in country 2. Therefore, the gap between wages in the two countries is reduced. Similarly, rental prices of capital in the two countries also get closer. Given that the two countries share the same technology, free trade establishes that wages and rental prices of capital are equalized between countries.

To see this result diagrammatically, refer to Fig. 21.6. As long as there is difference in the wage/rental ratio in the two countries, there will be a difference in the commodity price ratio and trade expands. With the expansion of trade the difference in the commodity price ratio and thus the difference

in the wage/rental ratio will go down. If trade stops expanding when the commodity price ratio has become identical across countries, the wage/rental ratio and wages and rentals will also be equalized. At $B(B')$ the negative of the common slope of the line VV' and the production frontier of country 1(2) is the wage/rental ratio in country 1(2). But the constant slope of the line VV' is the negative of the identical commodity price ratio. (At E , the point of tangency between the community indifference curve II and the line VV' , the common slope equals the negative of the commodity price ratio). Therefore the wage/rental ratio and wages and rentals for the two countries must be identical.

The above consequences of moving from autarky to free trade are summarized in the following theorem, which is a corollary to the Heckscher-Ohlin theorem. It is demonstrated rigorously in Appendix 21.1.

FACTOR-PRICE EQUALIZATION THEOREM: *Under Heckscher-Ohlin assumptions, international trade in commodities equalizes factor prices across countries. That is, free trade that equalizes commodity prices between countries sharing the same technology and producing the same two commodities must equate wages and rental prices of the capital of one country with those of another.*

This remarkable theorem implies that free migration of labor and mobility of capital are not necessary for efficiency in the world economy. Free movement of goods can perform the function of equating the marginal product of a given factor in all its uses.

THE LEONTIEF PARADOX AND FACTOR INTENSITY REVERSAL

Since after World War II, the United States emerged as the most capital-rich of the industrialist countries, US export goods should have been more capital-intensive than goods it produced in comparison with its imports. But using 1947 data, Leontief (1954) showed that US import substitutes were about 30% more capital-intensive than US exports. Given that the US is the most capital-abundant nation, this finding opposed the Heckscher-Ohlin theorem. This observation of Leontief is known as the Leontief paradox.

One possible source of the Leontief paradox could be factor intensity reversal. Factor intensity reversal here refers to a situation where a commodity is the labor-intensive one in relation to the other commodity in the labor-abundant nation and is the capital-intensive commodity in the capital-abundant nation. To understand this more explicitly, suppose, for example, that capital and labor can be substituted for each other in the

production of good *C*. Then production of good *C* could be capital-intensive in a capital-rich country and labor-intensive in a labor-rich country. But according to the Heckscher-Ohlin theorem the capital (labor) rich country should export capital (labor) goods. Since with factor intensity reversal the same good *C* is labor-intensive in both the countries, they have to export good *C* to each other. Therefore, the Heckscher-Ohlin theorem fails.

Other possible sources of the paradox could be: (i) not incorporating human capital in the calculation of capital; and, (ii) use of natural resources. Human capital refers to expenses on education, job training, health and so on, all of which increase labor productivity. Leontief's measure of capital included only physical capital (equipment, machinery) and ignored human capital. If the extraction and processing of natural resources demands more capital (relative to labor) than activities that make light use of natural resources, then import replacements might exhibit a higher capital/labor ratio than exports because the former are dominated by resource-using activities.

BARRIERS TO TRADE

Free trade benefits people but often may impose substantial costs on particular groups. International trade can benefit consumers by providing a wider variety of products from which to choose. Availability of foreign products limits pricing discretion of domestic producers and forces them to be more responsive to consumer demands. However, transition to greater or complete specialization under trade leads to reduction in size or elimination of domestic industries. Unemployed workers need time to search for other jobs, and factories have to be put to other uses. Some of these resources will never be reemployed. Even though the total output of both countries will be greater than before trade, not every group will be better off. Workers forced out of jobs by foreign competition are examples of persons who are worse off under free trade.

Virtually all countries of the world impose trade barriers of different types, partly in response to political pressure. The main objective of these barriers is to limit competitions from imports, although export restrictions are sometimes established. The most common tools for limiting import competition are tariffs and import quotas.

A tariff is a tax on the importation of a good from a foreign country. The purpose of a tariff is to protect domestic producers from foreign competition. A tariff on a foreign good increases its price and makes it less competitive in the market which in turn encourages consumers to buy domestic products.

Quotas restrict trade in a different manner. An import quota gives the maximum amount of a particular good that can be imported. Domestic producers usually prefer import quotas over tariffs. If people prefer foreign goods, they may continue to buy these products in spite of higher prices caused by tariffs. But a quota will prohibit imports completely once the limit is achieved. Tariffs are less damaging to consumer welfare than quotas. When demand causes prices of domestic products to rise, comparable tariff-bearing foreign goods become more attractive because the price differential between foreign and domestic goods is now lower. Therefore domestic producers may be restrained from raising prices further, otherwise they may lose their customers. Another point in favour of tariffs is that they generate revenue for the government. Other tools for restricting import competition can be: (i) making acquisition of an import license expensive, and (ii) compelling government agencies to buy domestic products to the extent possible.

Trade barriers may allow domestic producers to charge higher prices and be less responsive to consumers. In spite of such costs imposed by trade barriers, they exist. There are many popular arguments for protecting domestic industries from foreign competition. For instance, it is often argued that infant industries need protection from foreign competition, defence related industries must be protected to ensure our military self-sufficiency, industries producing goods of basic requirements need protection, and so on. Removal of trade barriers will hurt industries that are most vulnerable to competition from abroad. Therefore, efforts to reduce barriers should be accompanied by assisting those harmed by foreign competition.

IMPORT TARIFF AND THE STOLPER-SAMUELSON THEOREM

An important concern of international trade is the functional distribution of income (that is, factor rewards). Stolper and Samuelson (1941) considered the effects of imposition of a tariff on the distribution of income between the factors and concluded that import tariffs would increase the reward to the country's relatively scarce factor. This is known as the Stolper-Samuelson theorem.

The Stolper-Samuelson theorem retains all the seven assumptions of the Heckscher-Ohlin model. Now suppose that the two countries get involved in free trade. Then a country will export (import) a good relatively more intensive in the factor which is relatively more (less) abundant in the country. Now suppose that one country decides to impose an import tariff. We wish to investigate the effects of such a tariff on the distribution of income. Stolper

and Samuelson made two additional assumptions for analyzing the problem. These are:

- a) The imposition of a tariff will raise the domestic price of the import-competing good relative to that of the export good.
- b) The tariff will not alter the trade pattern. That is, the export good remains the good to be exported and the import-competing good remains the good to be imported after the tariff.

We now formally state the Stolper-Samuelson theorem.

STOLPER-SAMUELSON THEOREM: *Under the Heckscher-Ohlin assumptions and the additional assumptions (a) and (b), the price of the factor which is relatively scarce will increase in terms of the price of the other factor.*

This theorem, which shows the effects of a tariff on the income distribution, is quite appealing. Suppose in India capital is relatively scarce and it imports the capital-intensive good. Then an imposition of import tariff will increase the rental price of capital in terms of the price of labor. A rigorous proof of the theorem is presented in the appendix.

Example 21.1

Suppose that the production possibilities frontier for an economy that produces two goods T and G is given by

$$G_p^2 + T_p^2 = 100,$$

where $G_p(T_p)$ is the quantity produced of good $G(T)$. The community utility function is of the form

$$U = G_c T_c$$

where $G_c(T_c)$ is the quantity consumed of good $G(T)$. Let us denote the price of $G(T)$ by $p_G(p_T)$. We wish to determine the offer curve of this economy.

The marginal rate of transformation is

$$-\frac{dG_p}{dT_p} = \frac{T_p}{G_p}.$$

Therefore, if community consumption occurs at a point on a line which touches the frontier, then

$$-\frac{dG_p}{dT_p} = \frac{T_p}{G_p} = \frac{p_T}{p_G} = \theta \text{ (say).}$$

Therefore, $G_p = \frac{T_p}{\theta}$. Putting this value of G_p in the equation for the frontier, we have

$$T_p = \frac{10\theta}{(\sqrt{1+\theta^2})}.$$

Hence
$$G_p = \frac{10}{(\sqrt{1+\theta^2})}.$$

Now, the trading budget constraint is $\theta(T_c - T_p) = G_p - G_c$, which can be rewritten as

$$\begin{aligned}\theta T_c + G_c &= \theta T_p + G_p = \frac{10 + 10\theta^2}{(\sqrt{1+\theta^2})} \\ &= 10(\sqrt{1+\theta^2}).\end{aligned}$$

This budget constraint means that the value of exports equals the value of imports. The community marginal rate of substitution is given by

$$\frac{G_c}{T_c} = \theta,$$

which gives $G_c = \theta T_c$. Substituting this in the trading budget constraint, we have

$$T_c = \frac{5}{\theta}(\sqrt{1+\theta^2}),$$

and hence
$$G_c = 5(\sqrt{1+\theta^2}).$$

Therefore the traded quantities of the goods are

$$T_p - T_c = \frac{10\theta}{(\sqrt{1+\theta^2})} - \frac{5}{\theta}(\sqrt{1+\theta^2})$$

$$= \frac{5(\theta^2 - 1)}{\theta(\sqrt{1 + \theta^2})},$$

$$G_c - G_p = \frac{5(\theta^2 - 1)}{(\sqrt{1 + \theta^2})}.$$

If $\theta < 1$, then the exports (imports) of good $T(G)$ are positive. For $\theta > 1$, the reverse holds. For alternative values of θ , if we plot the exports of one good against the imports of the other good, we get the offer curve.

Applied Example 21.1 Gains from Trade

In Chapter 16 we have shown that there might be gains from trade when they occur between two individuals. The analysis of the Edgeworth box may be applicable to more complex problems. For instance, consider the case of trading among nations. This is also a situation in which there may be gains to all parties from trade. The precise source of these gains is the comparative advantage consideration. As stated in the text, this idea was put forward by Ricardo (1817) who argued that countries should specialize in producing those goods of which they are relatively more efficient producers. The countries should then get involved in trade with the rest of the world. If countries specialize in this way, aggregate production will be greater than it would be if each country tries to produce a balanced bundle of goods. Data reveal that there have been potential gains from free trade as a result of reductions in trade barriers agreed to under the 'Tokyo Round' of tariff negotiations in the late 1970s. (See Cline W.R., N. Kawanaabe, T.O.M. Kronsjo and T. Williams, *Trade Negotiations in the Tokyo Round*, Brookings, Washington D.C., 1978).

Applied Example 21.2 Factor Intensity Reversal

Minhas (1962) studied 24 industries in 19 countries, including India, Japan and the US to investigate how frequent factor intensity reversals are. He estimated the CES production function and found that factor reversals are quite common because the elasticity of substitution differs between industries, and that these factor reversals occur in the empirically relevant range of relative factor prices. He concluded that factor intensity reversals were

sufficiently frequent to explain the Leontief paradox. This finding would tend to minimize the practical relevance of the Heckscher-Ohlin theorem and the factor price equalization theorem. Leontief (1964) questioned Minhas' conclusion and recalculated some of his results using additional information. He found that the possibility of factor reversals was much less than that suggested by Minhas. The existence of factor intensity reversals is not disputed, but they are insufficient to explain the Leontief paradox.

Appendix 21.1

THE HECKSCHER-OHLIN THEOREM

The proof of the Heckscher-Ohlin theorem relies on the following lemma.

LEMMA 21.1. *At the same commodity price ratio, a country abundant in one factor has a production bias in favor of the commodity which uses that factor more intensively.*

Proof: In our two-country (1 and 2) two-good (C and T) model, country 1 has a relative abundance of labor, while country 2 has a relative abundance of capital. It is assumed that good C is capital-intensive and good T is labor-intensive.

Let us now consider the following full employment relations for country 1:

$$a_{KC}C^1 + a_{KT}T^1 = K^1,$$

$$a_{LC}C^1 + a_{LT}T^1 = L^1,$$

where $K^1(L^1)$ is country 1's endowment of capital (labor), a_{KC} is the quantity of capital required to produce 1 unit of good C in country 1, C^1 is country 1's production of good C , and so on. We rewrite the above relation as

$$a_{KC}C^1/L^1 + a_{KT}T^1/L^1 = K^1/L^1,$$

$$a_{LC}C^1/L^1 + a_{LT}T^1/L^1 = 1.$$

By solving this linear system we can express C^1/L^1 and T^1/L^1 in terms of the remaining quantities:

$$\frac{C^1}{L^1} = \frac{a_{LT}K^1/L^1 - a_{KT}}{a_{KC}a_{LT} - a_{KT}a_{LC}}, \quad \frac{T^1}{L^1} = \frac{a_{KC} - a_{LC}K^1/L^1}{a_{KC}a_{LT} - a_{KT}a_{LC}},$$

from which it follows that

$$\frac{C^1/L^1}{T^1/L^1} = \frac{C^1}{T^1} = \frac{a_{LT}K^1/L^1 - a_{KT}}{a_{KC} - a_{LC}K^1/L^1}.$$

The above equation expresses the output ratio C^1/T^1 in terms of the factor endowment ratio (K^1/L^1), given the technical coefficients a_{ij} . These coefficients depend on the factor price ratio. Under constant returns to scale they are constant for any output level. Consequently, we can differentiate C^1/T^1 with respect to the capital-labor ratio K^1/L^1 , treating a_{ij} 's as constants.

$$\begin{aligned} \text{Now, } \frac{d(C^1/T^1)}{d(K^1/L^1)} &= \frac{a_{KC}a_{LT} - a_{LC}a_{KT}}{(a_{KC} - a_{LC}K^1/L^1)^2} \\ &= a_{LC}a_{LT} \frac{\theta_C - \theta_T}{(a_{KC} - a_{LC}K^1/L^1)^2}, \end{aligned}$$

where $\theta_C = a_{KC}/a_{LC} = \frac{K_c^1/C^1}{L_c^1/C^1} = \frac{K_c^1}{L_c^1}$, with $K_c^1(L_c^1)$ being the amount of capital (labor) required to produce C^1 units of good C . Similarly, $\theta_T = \frac{a_{KT}}{a_{LT}} = \frac{K_T^1}{L_T^1}$. The ratios θ_C and θ_T are the factor intensities in the two sectors of country 1.

The absence of factor intensity reversal means that either θ_C is greater than θ_T or θ_T is greater than θ_C independently of the factor price ratio. If we assume, as in the main body of the chapter, that good T is labor-intensive, then the derivative under consideration will be negative. This means that the greater the factor endowment ratio K^1/L^1 , the smaller the output of C relative to T and vice versa. Thus, country 1 should increase the output of good T at the expense of good C . Since the production functions are assumed to be identical across countries, the above result holds for both countries. This completes the proof of the lemma.

Thus, country 1, the labor-abundant country, has a production bias in favor of the labor-intensive commodity T , while the capital-abundant country 2 has a production bias in favor of the capital-intensive commodity C . This means that each country can increase its production of the commodity which is intensive in the country's abundant factor, at a lower opportunity cost than the other.

Proof of the Theorem: We need to show that each country exports the commodity which uses the country's more abundant factor more intensively. This will follow from Lemma 21.1 and from the assumption that the structure of demand is identical across countries.

To see this, note that under free trade and with no transport costs, the commodity price ratio (terms of trade) is the same in both the countries. Now, by Lemma 21.1, at the same commodity price ratio, country 1 (labor-abundant country) will produce relatively more T (the labor-intensive commodity) and country 2 (the capital-abundant country) will produce more C (the capital-intensive commodity). Thus, the ratio T/C is greater in country 1 than in country 2. Now, under identical demand conditions, at the same relative price of goods, both countries wish to consume T and C in the same proportion. Therefore, country 1 will export T (and import C , which will be exported by country 2). Hence after trade the structure of the quantities of the two goods available—the quantity available is given by domestic output plus imports or minus exports—will be identical in both countries and equal to the structure of demand. This completes the proof of the theorem.

THE FACTOR PRICE EQUALIZATION THEOREM

The demonstration of the factor price equalization theorem relies on a consequence of the absence of factor intensity reversal. This consequence states that a unique factor price ratio corresponds to each commodity price ratio, and vice versa; that is, there is a one-to-one correspondence between relative price of goods and the relative price of factors.

To demonstrate this graphically, let us assume that the commodity price ratio p_T/p_C is 5, that is, 5 units of C exchange for 1 unit of T . Under perfect competition this implies that the production cost of 1 unit of T equals that of 5 units of C . In panel (a) of Fig 21.7 we have drawn isoquants $1T$ and $5C$. This figure applies to both countries because of assumption of internationally identical production functions. Since factor prices are the same in the two sectors and since the production cost of 1 unit of T equals that of 5 units of C , the minimum isocost will be identical for $5C$ and $1T$. Thus we must find an isocost line which is simultaneously tangent to the isoquants $1T$ and $5C$. The absolute value of the slope of this isocost line is the factor price ratio.

For panel (a) of Fig. 21.7 it turns out that only such an isocost (represented by the line VV') exists in the case of single crossing of the isoquants. Therefore, a unique factor price ratio corresponds to a given commodity price ratio.

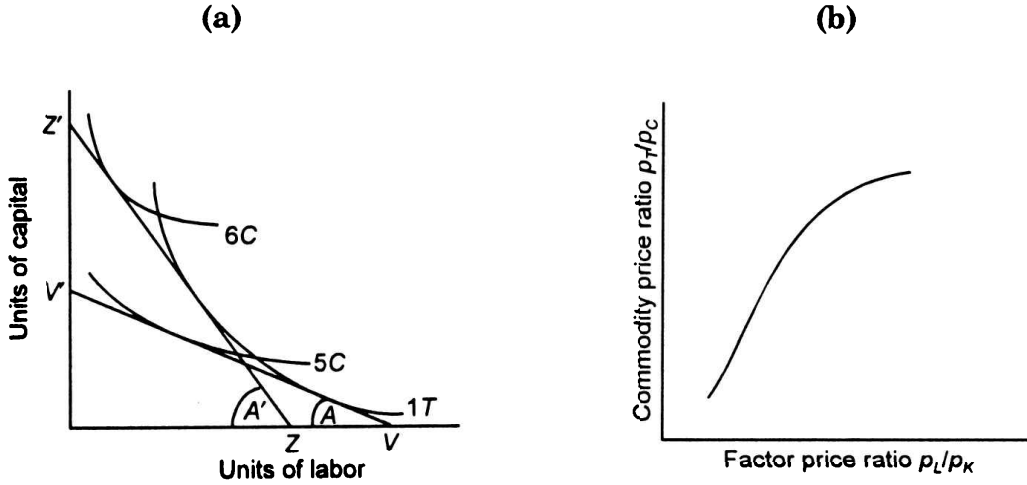


Fig. 21.7 Commodity Price and Factor Price Ratios

Now suppose the relative price p_T/p_C increases from 5 to 6. Therefore, we must determine the isocost line which is simultaneously tangent to the isoquants 1T and 6C. This isocost line is ZZ'. Since $\tan A' > \tan A$, a greater factor price ratio corresponds to the greater commodity price ratio $p_T/p_C = 6$. This relationship is shown in panel (b) of Fig. 21.7. The relationship is increasing since good T is labor-intensive. In the opposite case it would be decreasing. Thus, in either case the relationship is monotonic and hence one-to-one.

The above observation can be established analytically also. To see this let us consider the following equilibrium conditions for country 1:

$$a_{LC}p_L + a_{KC}p_K = p_C,$$

$$a_{LT}p_L + a_{KT}p_K = p_T,$$

where a_{LC} is the quantity of labor required to produce 1 unit of good C in country 1, p_L is the price of labor L, p_C is the price of good C, and so on. These equations give us the relationship between factor prices and commodity prices. If we divide the second equation by the first, we get

$$\frac{p_T}{p_C} = \frac{ra_{LT} + a_{KT}}{ra_{LC} + a_{KC}}, \quad r = p_L/p_K.$$

Therefore,

$$\frac{d(p_T/p_C)}{dr} = \frac{(a_{LT} + ra'_{LT} + a'_{KT})(ra_{LC} + a_{KC}) - (a_{LC} + ra'_{LC} + a'_{KC})(ra_{LT} + a_{KT})}{(ra_{LC} + a_{KC})^2},$$

where $a'_{LT} = d(a_{LT})/dr$, and so on. Note that the coefficients a_{LT} , etc., are functions of r through optimization.

From the equilibrium conditions

$$p_L d a_{LC} + p_K d a_{KC} = 0,$$

$$p_L d a_{LT} + p_K d a_{KT} = 0,$$

from which we have

$$r a'_{LC} + a'_{KC} = 0,$$

$$r a'_{LT} + a'_{KT} = 0.$$

Therefore, $d(p_C/p_T)/dr$ can be rewritten as

$$\begin{aligned} \frac{d(p_C/p_T)}{dr} &= \frac{a_{LT}a_{KC} - a_{LC}a_{KT}}{(ra_{LC} + a_{KC})^2} \\ &= a_{LC}a_{LT} \frac{\theta_C - \theta_T}{(ra_{LC} + a_{KC})^2}, \end{aligned}$$

where $\theta_C = a_{KC}/a_{LC}$ and $\theta_T = a_{KT}/a_{LT}$.

Therefore, the derivative of the relative price of commodities p_C/p_T with respect to the relative input price r has an unambiguous sign, that is, either a positive or a negative sign, if and only if θ_C is greater or less than θ_T . This means that the derivative is positive or negative if and only if factor intensity reversal does not occur. Therefore, the relative commodity price is monotonically related to relative factor price whenever factor intensity reversal is absent. Clearly, this claim is true for country 2 also.

We summarize these observations in the following lemma.

Lemma 21.2. In the case of no factor intensity reversal there is one-to-one correspondence between the relative price of goods (p_T/p_C) and the factor price ratio (p_L/p_K), the price of labor relative to that of capital. If good T is labor (capital) intensive, then p_L/p_K increases (decreases) as p_T/p_C increases.

Proof of the Factor Price Equalization Theorem: Let f and g stand for production functions of sectors C and T. These functions are the same for both the countries. Then using our earlier notation C^1 and T^1 for country 1, we have

$$C^1 = f(L_C^1, K_C^1),$$

$$T^1 = g(L_T^1, K_T^1),$$

where L_C^1 is the quantity of labor employed by country 1 in sector C, and so on. Since f and g are homogenous of degree one (guaranteed by constant returns to scale), we have

$$C^1 = L_C^1 f(1, K_C^1/L_C^1) = L_C^1 F(K_C^1/L_C^1), \text{ (say),}$$

$$T^1 = L_T^1 g(1, K_T^1/L_T^1) = L_T^1 G(K_T^1/L_T^1), \text{ (say).}$$

We now employ the equilibrium conditions that state the equality between the marginal productivity of a factor and its price. (This is equal in both sectors.) The marginal productivity of capital in sector C is

$$\frac{\delta C^1}{\delta K_C^1} = F'(K_C^1/L_C^1),$$

where F' is the derivative of F . This productivity in sector T is

$$\frac{\delta T^1}{\delta K_T^1} = G'(K_T^1/L_T^1).$$

The equilibrium condition then gives

$$F'(K_C^1/L_C^1) = \frac{p_T}{p_C} G'(K_T^1/L_T^1) = p_K, \quad (E1)$$

where p_K is the price of capital, and p_T/p_C is the price of good T relative to that of good C. The marginal productivity of labor in sector C is

$$\frac{\delta C^1}{\delta L_C^1} = F(K_C^1/L_C^1) - \frac{K_C^1}{L_C^1} F'(K_C^1/L_C^1),$$

and in sector T this productivity is

$$\frac{\delta T^1}{\delta L_T^1} = G(K_T^1/L_T^1) - \frac{K_T^1}{L_T^1} G'(K_T^1/L_T^1).$$

By the equilibrium condition we have

$$F(K_C^1/L_C^1) - \frac{K_C^1}{L_C^1} F'(K_C^1/L_C^1) = \frac{p_T}{p_C} \left(G(K_T^1/L_T^1) - \frac{K_T^1}{L_T^1} G'(K_T^1/L_T^1) \right)$$

$$= p_L. \quad (E2)$$

Dividing Eq. (E2) by (E1) we get

$$\frac{p_L}{p_K} = \frac{F(K_C^1/L_C^1) - \theta_C F'(K_C^1/L_C^1)}{F'(K_C^1/L_C^1)} = \frac{F(K_C^1/L_C^1)}{F'(K_C^1/L_C^1)} - \theta_C$$

and

$$\frac{p_L}{p_K} = \frac{G(K_T^1/L_T^1) - \theta_T G'(K_T^1/L_T^1)}{G'(K_T^1/L_T^1)} = \frac{G(K_T^1/L_T^1)}{G'(K_T^1/L_T^1)} - \theta_T,$$

where, as stated earlier, $\theta_C = K_C^1/L_C^1$ and $\theta_T = K_T^1/L_T^1$.

Let us now differentiate (p_L/p_K) with respect to θ_T .

$$\frac{d(p_L/p_K)}{d\theta_T} = \frac{(G'(K_T^1/L_T^1))^2 - G''(K_T^1/L_T^1)G(K_T^1/L_T^1)}{(G'(K_T^1/L_T^1))^2} - 1,$$

where G'' is the derivative of G' . Therefore,

$$\frac{d\theta_T}{d(p_L/p_K)} = - \frac{(G'(K_T^1/L_T^1))^2}{G''(K_T^1/L_T^1)G(K_T^1/L_T^1)}.$$

In view of the diminishing marginal productivity condition, $G'' < 0$. Since $G > 0$, $d\theta_T/d(p_L/p_K)$ is positive. Therefore, factor intensity in sector T is increasing in relative factor price. Clearly, the same conclusion will hold for sector C also. That is, there is a one-to-one relation between the relative price of factors p_L/p_K and θ_i ($i = C, T$).

In Lemma 21.2 we have established that there is a one-to-one relation between the commodity price ratio p_T/p_C and the factor price ratio p_L/p_K . This, combined with our demonstration that p_L/p_K is related to θ_i ($i = C, T$) in a one-to-one manner, shows that a unique value of θ_C and a unique value of θ_T will correspond to the relative price of commodities p_T/p_C , determined as a consequence of free trade. By substituting these values in Eqs. (E1) and (E2), the values of p_K and p_L can be uniquely determined.

Now, we are given the assumption of internationally identical production functions. Consequently, the single-valued relations between θ_C and p_T/p_C and between θ_T and p_T/p_C are identical in both countries. Similarly, Eqs. (E1)

and (E2) are also identical across countries. Therefore, since the ratio p_T/p_C is the same, given the assumptions of free trade and zero transport cost, the absolute factor prices will be equalized between countries. This completes the proof.

Proof of the Stolper-Samuelson Theorem: Since the Heckscher-Ohlin assumptions are satisfied, the capital-abundant country exports C (the capital-intensive commodity) and imports the labor-intensive commodity T . The reverse holds for the labor-abundant country. Let us suppose that the former country imposes an import duty. This increases the domestic price p_T/p_C . In view of our Lemma 21.2 p_L/p_K increases. From panel (a) of Fig. 21.7 it also turns out that the capital-labor ratio increases in both sectors.

Now, let C^2 and T^2 be the output of goods C and T of country 2, the capital-abundant country. Then

$$C^2 = f(L_C^2, K_C^2),$$

$$T^2 = g(L_T^2, K_T^2),$$

where f and g are production function of sectors C and T , L_C^2 is the labor of country 2 in sector C , and so on. Clearly $L_C^2 + L_T^2 = L^2$, country 2's total labor endowment. Similarly, this country's total capital endowment is $K_C^2 + K_T^2 = K^2$. Since both f and g homogeneous of degree one, we have

$$C^2 = L_C^2 f(1, K_C^2/L_C^2) = L_C^2 F(K_C^2/L_C^2) \text{ (say),}$$

$$T^2 = L_T^2 g(1, K_T^2/L_T^2) = L_T^2 G(K_T^2/L_T^2) \text{ (say).}$$

The marginal productivity of labor in sector C is

$$\frac{\delta C^2}{\delta L_C^2} = F(K_C^2/L_C^2) - \frac{K_C^2}{L_C^2} F'(K_C^2/L_C^2).$$

The marginal productivity of capital in this sector is

$$\frac{\delta C^2}{\delta K_C^2} = F'(K_C^2/L_C^2),$$

where F' stands for the derivative of F . Thus, both the marginal productivities are functions of the capital-labor ratio. In view of the diminishing marginal productivity, F' is decreasing in K_C^2/L_C^2 . Therefore the marginal productivity

of labor is an increasing function, and the marginal productivity of capital is a decreasing function of the capital-labor ratio. Clearly this holds for sector *T* also.

We have shown that the capital-labor ratio has gone up in both sectors. Therefore the marginal productivity of labor (and so its unit real reward, which under perfect competition coincides with the marginal productivity of labor) increases. This completes the proof of the theorem.

Exercises

Multiple Choice

1. Why are tariffs preferable to quotas?
 - a) Consumers may continue to buy imported products in spite of the tariff.
 - b) Tariffs do not really block trade; in fact they may increase trade.
 - c) As prices of domestic products increase, foreign products become more competitive.
 - d) (a) and (c).
2. Suppose that South Korea can produce either 20 bicycles or 100 calculators with one unit of some resource, and that India can produce either 10 bicycles or 80 calculators. Which of the following statements is correct?
 - a) India has a comparative advantage in calculators.
 - b) India has an absolute advantage in bicycles.
 - c) South Korea has a comparative advantage in calculators.
 - d) India has an absolute advantage in calculators.
3. Which of the following statements describe the impact of tariffs accurately?
 - a) They tend to increase the prices of imported products that are subject to the tariff.
 - b) They tend to increase the prices of domestically produced products that are comparable to those being taxed.
 - c) They are less damaging to consumer welfare than quotas.
 - d) All of the above.

4. With trade, specialization in production is likely to be
 - a) Complete with increasing costs, and incomplete with constant costs.
 - b) Complete with constant costs, and incomplete with increasing costs.
 - c) Complete with both constant and increasing costs.
 - d) Incomplete with both constant and increasing costs.
5. If a nation gains from trade, then its consumption point will be
 - a) Exactly on its production possibilities frontier.
 - b) Below its production possibilities frontier.
 - c) Above its production possibilities frontier.
 - d) Any of the above.
6. If a nation's terms of trade is 3, its trade partner's terms is
 - a) $1/3$.
 - b) 3.
 - c) $\sqrt{3}$.
 - d) 3^2 .
7. The offer curve of a nation bends towards the axis that measures its
 - a) Import commodity.
 - b) Export commodity.
 - c) Non-traded commodity.
 - d) Export or import commodity.
8. The Heckscher-Ohlin theory postulates that as a result of trade, the difference in factor prices between the countries
 - a) Increases.
 - b) Decreases.
 - c) Does not change.
 - d) Any of the above is possible.
9. Factor intensity reversal refers to the situation where
 - a) Both commodities are more labor-intensive in one nation than in the other nation.
 - b) One commodity is labor-intensive in one country and capital-intensive in the other country.
 - c) The same commodity is more labor-intensive in one nation than in the other nation.
 - d) Any of the above since they are all equivalent statements.

10. In the Heckscher-Ohlin set-up, the most important cause of the difference in relative commodity prices and trade between nations is a difference in
- a) Factor endowments.
 - b) Technology.
 - c) Tastes.
 - d) All of the above.

Quantitative Problems and Questions

1. Explain the difference between comparative advantage and absolute advantage. Why do economists prefer comparative advantage as a basis for trade?
2. How can specialization and trade allow countries to consume beyond their respective production possibilities frontiers?
3. How is opportunity cost related to the principle of comparative advantage?
4. How do you compare comparative advantage in the Ricardian model with that in the Heckscher-Ohlin model?
5. Prove or disprove the following:
 - a) If a country is opened to foreign trade, a potential Pareto improvement takes place irrespective of whether world prices are different from or the same as the pre-trade domestic prices.
 - b) Consider a two-commodity world. Suppose that there is a tax at the proportional rate t_1 on the consumption of good 1. Then a tariff at the same rate on good 2 will restore Pareto efficiency.
 - c) A country is opened to trade. Suppose that the tariff on a labor-intensive good is so high that the good is not imported. Then the labor income increases above its original level.
 - d) Tariffs are more damaging to consumer welfare than quotas.
 - e) Absolute advantage is a necessary condition for comparative advantage.
6. Consider a two-country world. These countries devote their resources to the production of meat and grain. If country 1 devotes all its resources to meat production, it can produce 64 units. Grain can be produced instead of meat according to the production possibilities frontier $(m_1 - 72)(g_1 - 18) = 144$, where m_1 (g_1) is country 1's output of meat (grain).

The maximum possible output of meat for country 2 is 16 units. Its production possibilities frontier is $(m_2 - 18)(g_2 - 72) = 144$, where m_2 (g_2) is country 2's production of meat (grain).

All inhabitants of the two countries have the same utility function given by $U(g_i^j, m_i^j) = \log g_i^j + \log m_i^j$, where g_i^j (m_i^j) is the consumption of grain (meat) in country i ($i = 1, 2$) by a representative citizen j .

- a) Initially there is no trade. Each has a perfectly competitive equilibrium. Calculate, for each nation separately, the competitive equilibrium values for the output of meat, the price of meat, the output of grain and the price of grain.
 - b) Free trade develops between the two nations. Transport costs are negligible. Calculate at the competitive free-trade equilibrium:
 - i) The price ratio,
 - ii) Country 1's output of meat and grain,
 - iii) Country 2's outputs of meat and grain,
 - iv) Country 1's consumption of meat and grain,
 - v) Country 2's consumption of meat and grain.
7. Robinson Crusoe produces coconuts (C) and fish (F) according to the production functions $C = \sqrt{L_C}$ and $F = \sqrt{L_F}$, where L_C (L_F) is the number of hours spent gathering coconuts (fishing). Crusoe has decided to work for 16 hours, consequently, $L_C + L_F = 16$. His utility function for fish and coconuts is given by $U(F, C) = \log F + \log C$.
- a) If Robinson cannot trade with the rest of the world, how will he choose to allocate his labor? What will the optimal levels of F and C be? What will his utility be?
 - b) Suppose now that trade is opened and Robinson can trade fish and coconuts at a price ratio of $p_F/p_C = 2$. How will he adjust production to take advantage of world prices?
8. Consider an economy with just one technique available for the production of two goods, food and cloth:

<i>Good</i>	<i>Food</i>	<i>Cloth</i>
Labor per unit output	1	1
Land per unit output	2	1

- a) Suppose total labor equals 80 units and total land equals 120 units. Sketch the production possibilities frontier.

- b) Explain why the production possibilities frontier is concave.
 - c) If consumers insist on trading 4 units of food for 5 units of cloth, what is the relative price of food?
9. An economy with fixed terms of trade produces two goods, *A* and *B*. The equilibrium production of each unit of good *A* requires 6 units of labor and 4 units of capital, while good *B* requires 2 units of labor and 2 units of capital. The economy has a factor endowment of 50 units of labor and 40 units of capital.
- a) If equal amounts of goods *A* and *B* are consumed in the home country and its factor endowments are fully employed, will the country export good *A* or *B*?
 - b) Suppose the country was endowed with 60 units of capital and 50 units of labor, what would be the pattern of trade?

22

Market Failures and the Theory of Second-Best

GENERAL DISCUSSION

The efficiency of a perfectly competitive price system depends crucially on the assumptions that form the basis of the competitive model. However, actual markets may differ from competitive markets in many ways. The number of impediments to perfect competition that we might discuss is practically infinite. These sources of market failures, however, can be classified into five general groupings that include most possible cases: (i) non-convexities (ii) non-competitive behavior, (iii) externalities (iv) public goods, and (v) informational externalities. In chapter 16 we indicated that a convexity assumption is necessary for the Second Fundamental Theorem of Welfare Economics to hold. If not, there may exist allocations which are Pareto optimal, but which cannot be achieved by a decentralized market mechanism. These cases are called market failures resulting from non-convexities.

An important assumption underlying the fundamental theorems is that all market participants are price takers, that is, they make their decisions under the assumptions that their actions do not influence market prices.

This is reasonable when the market is populated with numerous buyers and sellers. In many cases and for different reasons there are sufficiently few persons on one side of the market and they know that their actions may alter market prices. Markets that are organized as oligopolies or in which there is a monopsony on the buying side are considered in this category, in addition to the usual monopoly case. The essential effects of all such markets is that the marginal revenue (or marginal expense in the case of monopsony) differs from the market price. Consequently, relative prices no longer reflect relative marginal costs, which in turn means that the price system does not convey the information necessary to ensure efficiency.

Externalities, as we will see, arise when one consumer's utility and/or a firm's production set get affected by the activities of other economic agents. The cause of market failure in the case of a public good is jointness in consumption. Two types of informational externalities that impede the functioning of a perfectly competitive market are adverse selection and moral hazards. Externalities can be positive or negative. If the relations between the concerned agents (say consumers) are beneficial, then we say that the externalities are positive. That is, positive externality arises if one consumer enjoys a benefit from the decision taken by the other consumer. (The opposite holds for a negative externality case.) In either case, market prices do not fully reflect the marginal benefit of the consumption. More precisely, a person making an economic decision on the basis of market prices does not enjoy the full marginal social benefit or bear the full marginal cost of his actions. We elaborate this issue in Chapter 23.

A public good is a good which is jointly consumed by persons. Since it is not possible to exclude persons from consumption of a jointly consumed good, there may not be any incentive for individual to pay for the good. This cause of market failure is known as the free-rider problem (see Chapter 24).

The problem of moral hazards arises due to asymmetry of information. If some persons have information that others do not have, then the former may be able to exploit in the way a monopoly firm exploits its market power. One case of this is referred to as moral hazard. The problem of adverse selection results from improper selection of a sample required for the purpose in mind. We analyze this issue in greater detail in Chapter 25.

It is clear that the five issues discussed above are relevant to any economic system. One may argue that a competitive price system could still be advocated for those sectors of the economy in which these impediments do not exist. By using the price system in these sectors, the Pareto rules for allocational efficiency would be satisfied there; this would be a second best

solution to the problem. But unfortunately unless one corrects all of these distortions/market failures so as to bring about a 'best' Pareto efficient allocation, this approach to find a second-best solution is not correct.

Lipsey and Lancaster (1956) addressed this problem in a general framework of policy optimization and reached an essentially negative conclusion. Their general theory for the 'second-best optimum' states that in general, if one or more of the Pareto conditions are unattainable, then others, still attainable, are not desirable. From this theorem it follows that there is no a priori way to rank different states of the world on the basis of the remaining Pareto optimality conditions. It is not true that having more (but not all) of the efficiency conditions satisfied is necessarily superior to having fewer hold.

The central idea underlying the general theorem of the second-best can be explained graphically. Suppose that a society's production possibilities for the two goods are represented by the frontier FF' in Fig. 22.1 and that preferences are given by the indifference curves. There is a constraint in the economy that makes the actual optimum point B unattainable. This constraint is represented by the line CC' . That is, combination of the goods to the top right of the line CC' cannot be achieved because of the constraint. Society wishes to maximize welfare represented by the indifference curves subject to the constraint CC' . From the figure it is clear that the welfare maximizing point need not be on the frontier FF' . This demonstrates the principle negative result of the theory of the second-best: if all the conditions of Pareto efficiency cannot be fulfilled, it is not true that to satisfy some of them is the best policy. In such a case an optimum situation can be attained only by departing from the remaining Paretian conditions. The optimum situation finally achieved may be regarded as a second best optimum because it is achieved subject to a constraint which, by definition, does not allow attainment of a Paretian optimum.

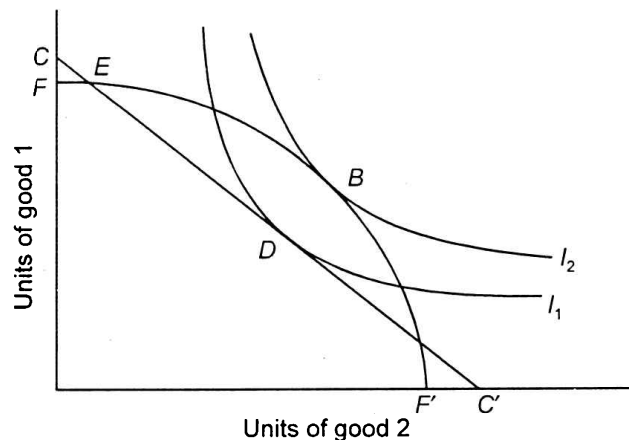


Fig. 22.1 Theory of Second-Best

Appendix 22.1

THEORY OF SECOND BEST

To state the problem technically, suppose we need to maximize an objective function $F(x_1, x_2, \dots, x_n)$ subject to the feasibility constraint $G(x_1, x_2, \dots, x_n) = 0$. Let us consider the Lagrangian

$$\mathcal{L} = F(x_1, x_2, \dots, x_n) + \lambda G(x_1, x_2, \dots, x_n),$$

where λ is the Lagrange multiplier. Then differentiating $\mathcal{L}$ partially with respect to $x_1, x_2, \dots, x_n$ and λ , and setting the partial derivatives to zero, we get the first order conditions

$$\begin{aligned} \frac{\delta \mathcal{L}}{\delta x_1} &= F_1(x_1, x_2, \dots, x_n) + \lambda G_1(x_1, x_2, \dots, x_n) = 0, \\ &\vdots \\ \frac{\delta \mathcal{L}}{\delta x_n} &= F_n(x_1, x_2, \dots, x_n) + \lambda G_n(x_1, x_2, \dots, x_n) = 0, \end{aligned}$$

and
$$\frac{\delta \mathcal{L}}{\delta \lambda} = G(x_1, x_2, \dots, x_n) = 0,$$

where $F_i(x_1, x_2, \dots, x_n)$ [$G_i(x_1, x_2, \dots, x_n)$] is the partial derivative of F (G) with respect to x_i , $i = 1, \dots, n$. We can rewrite the above conditions as:

$$\begin{aligned} \frac{F_1(x_1, x_2, \dots, x_n)}{F_n(x_1, x_2, \dots, x_n)} &= \frac{G_1(x_1, x_2, \dots, x_n)}{G_n(x_1, x_2, \dots, x_n)}, \\ &\vdots \\ \frac{F_{n-1}(x_1, x_2, \dots, x_n)}{F_n(x_1, x_2, \dots, x_n)} &= \frac{G_{n-1}(x_1, x_2, \dots, x_n)}{G_n(x_1, x_2, \dots, x_n)}, \\ G(x_1, x_2, \dots, x_n) &= 0. \end{aligned}$$

We have a second best situation if at least one of the first $(n - 1)$ equalities does not hold. Observe that the equality $G(x_1, x_2, \dots, x_n) = 0$ has to hold because it is given a priori. But the other $(n - 1)$ equalities are derived as optimizing conditions. For instance, if we have $[F_1(x_1, x_2, \dots, x_n)/F_n(x_1, x_2, \dots, x_n)] = k[G_1(x_1, x_2, \dots, x_n)/G_n(x_1, x_2, \dots, x_n)]$, where $k \neq 1$, then a second-best situation arises.

Suppose, for example, that F is the social welfare function in a two-person two-good (1 and 2) community, then F is a function of $U_A(x_A) = U_A(x_A^1, x_A^2)$ and $U_B(x_B) = U_B(x_B^1, x_B^2)$ where U_A is person A 's utility function, x_A^1 is person A 's consumption of good 1, and so on. The constraint function G becomes the transformation function $T(X^1, X^2) = 0$, where X^1 is the total amount of good 1, that is, $X^1 = x_A^1 + x_B^1$. Similarly X^2 will denote the total of good 2. Thus, we are considering the problem of maximizing social welfare.

We have shown in Chapter 18 (Appendix 18.1) that in this case the Pareto equalities or *the first best conditions* are:

$$\frac{\delta U_A(x_A^1, x_A^2)/\partial x_A^1}{\delta U_A(x_A^1, x_A^2)/\partial x_A^2} = \frac{\delta T/\partial X^1}{\delta T/\partial X^2},$$

$$\frac{\delta U_B(x_B^1, x_B^2)/\partial x_B^1}{\delta U_B(x_B^1, x_B^2)/\partial x_B^2} = \frac{\delta T/\partial X^1}{\delta T/\partial X^2},$$

$$T(X^1, X^2) = 0.$$

A second-best situation arises here if at least one of the first two equalities does not hold. There is no doubt that the Lipsey-Lancaster theorem has given a negative implication in the sense that undermines the Pareto efficiency principle stating that if any one of the Pareto efficiency conditions is not met, then satisfying the remaining conditions is not necessarily desirable. It means that the second-best optimality conditions are much more difficult than the simple Pareto optimality conditions. However the second-best framework is essential for treating economic problems in the real world.

EXERCISES

1. What is market failure? What are the sources of this problem?
2. Consider a one-person economy which produces two goods (1 and 2) according to the production functions $x_1 = \sqrt{L_1}$ and $x_2 = \sqrt{L_2}$, where $x_1(x_2)$ is the output of 1(2) and L_1 (L_2) is the number of hours devoted by the person to the production of good 1(2). The person has decided to work for 200 hours during a certain period so that $L_1 + L_2 = 200$. His utility function for the goods is given by $U(x_1, x_2) = \sqrt{x_1 x_2}$. Suppose that the production for the goods is constrained by $x_1 + x_2 = 16$.
 - a) Draw a graph to illustrate a second best solution.
 - b) Is your solution efficient?

23

Externalities

Externalities occur whenever the activities of one economic agent affect the activities of another agent in ways that are not taken into account by the operation of the market. That is, externalities are concerned with the side-effects on the bystanders of the economic activities of others. Some of these side-effects of economic activity are beneficial, and others are harmful. The beneficial externalities are called positive externalities; the harmful ones are the negative externalities.

Externalities can take place between consumers, between firms and consumers, and also between firms. If a consumer's preference is directly affected by goods consumed by another consumer or a firm's production, then we say that consumption externality exists. Many of the impressive and exciting sights of a large city are the skyscrapers built by private firms. This is an example of a positive consumption externality (or external economy in consumption). Another example of an externality of this type can be the benefit that a person gets from a better-lit footpath when his neighbor installs a light outside the gates of the neighbor's house. But suppose your neighbor is a motorcycle enthusiast and does not believe in using a muffler. Then it is probable that his comings and goings at all hours will cause noise for you and his other neighbors. This is an example of negative consumption externality (or external diseconomy in consumption).

When the production of one firm gets affected by the choice of a consumer or the production of another firm, we say that production externalities exist. The pollination of an apple orchard by a honey producer's bees improves productivity of the apple industry. Conversely, since the bees feed on apple blossoms, an increase in apple production will improve productivity of the honey industry. Thus, in this classic case we have mutual positive production externalities, or, external economy in production (Meade, 1952). Next, suppose two firms are located on the same river. Firm *A*, the upstream firm, dumps its wastes into the river, while firm *B*, the downstream firm, uses river water for processing its output. If firm *A* increases its output (hence wastes), firm *B*'s production suffers. To produce the same output with dirtier water firm *B* uses more labor, more chemicals, etc. (More specifically we may assume that firm *B* is a fishery so that as the amount of wastes dumped into the river increases, its output gets adversely affected to a higher extent.) This is an example of negative production externality, or external diseconomy in production—firm *A*'s production negatively affects firm *B*'s production.

In all the above situations the problem is that there are no markets for externalities, whether they are in consumption or in production, whether they are positive or negative. That is, there is no market for noise created by motorcycles or wastes dumped into the river and so on. In the case of negative externalities the decision maker does not know how much his decision really costs. In the case of positive externalities, the decision maker does not know how much his decision really helps.

Let us now recall the assumptions underlying the fundamental theorems of welfare economics, the results that link Pareto optimality and competition. These assumptions are: (a) a consumer's utility depends on his own consumption bundle only, and (b) a firm's production set remains unaffected by other firms' production decisions. In this chapter we relax these assumptions and examine the economic consequence of externalities.

Since externalities are caused by interdependence of people, resources and things, the link between Pareto efficiency and competition is broken when externalities are present. The only way to cope with this problem is to control the rate of externality-producing activity. In this chapter we will see how appropriate corrective mechanisms can once again achieve Pareto efficiency.

PRODUCTION EXTERNALITIES

To illustrate production externalities analytically, let us consider the two-firm example again, where the upstream firm *A* produces some amount of

good 1, x_A^1 , and also produces a certain amount of wastes, s , which it dumps into the river. The downstream firm (B) gets adversely affected by firm A 's wastes.

We denote firm A 's cost function by $C_A(x_A^1, s)$ and firm B 's cost function by $C_B(x_B^2, s)$, where x_B^2 is firm B 's production of its good, 2, and s is the amount of wastes produced and dumped into the river by firm A . (We assume here that the input markets that firms A and B face are perfectly competitive. Recall that if factor markets are perfectly competitive, a firm's cost function can be written as a function of its output only.) Note that s directly affects firm B 's cost of producing good 2. We assume that firm B 's cost of production increases as the wastes dumped into the river increase, $\Delta C_B(x_B^2, s)/\Delta s > 0$, and that increased wastes do not increase firm A 's cost of production; $\Delta C_A(x_A^1, s)/\Delta s \leq 0$.

Firm A 's profit maximization problem is

$$\max_{x_A^1, s} p_1 x_A^1 - C_A(x_A^1, s)$$

and firm B 's profit maximization problem is

$$\max_{x_B^2} p_2 x_B^2 - C_B(x_B^2, s),$$

where p_1 and p_2 are competitive prices of goods 1 and 2 respectively. Since firm B does not have any control over the amount of wastes dumped into the river, it cannot regard s as an optimizing variable.

The profit maximizing conditions for firm A are

$$p_1 = \frac{\Delta C_A(x_A^1, s)}{\Delta x_A^1}, \quad (23.1)$$

$$0 = \frac{\Delta C_A(x_A^1, s)}{\Delta s}. \quad (23.2)$$

For firm B the only profit maximizing condition turns out to be

$$p_2 = \frac{\Delta C_B(x_B^2, s)}{\Delta x_B^2}. \quad (23.3)$$

According to these conditions, at profit maximizing point the price of each good should be equal to its marginal cost. Since the price of wastes dumped into the river is zero for firm A , the profit maximizing condition requires that the firm will dump wastes until the cost of an additional unit is zero.

We can easily recognize the presence of externality in the model formulated above. Firm *B* cares about wastes but has no control over it. When firm *A* increases its output, it increases the cost of production for firm *B*. Because this cost is external to firm *A*, it ignores the cost when making its decision of how much to produce. The increase in cost of production of firm *B* associated with an increase in wastes dumped by firm *A* is a part of the social cost of production of firm *A*'s output. In general, firm *A* will dump too much waste since it ignores the impact of that on firm *B* and good 1 will be over-produced.

We may illustrate the situation graphically. The demand for firm *A*'s output is given by the curve *DD* and *MC* denotes *A*'s marginal cost curve. The curve *MC'* records the social marginal cost of firm *A*. *MC'* differs from *MC* by the amount of extra costs that firm *A* imposes on others (here only firm *B*) in the society. (Ignore the curve *DD'* for the moment.)

The normal workings of the market will cause output level $\bar{x}_A^1$ to be produced, since at this output market price is equal to the private marginal cost *MC* (Eq. [23.1]). Since the other profit maximizing condition says that the marginal cost of wastes dumped is zero, firm *A* ignores the difference between *MC'* and *MC*.

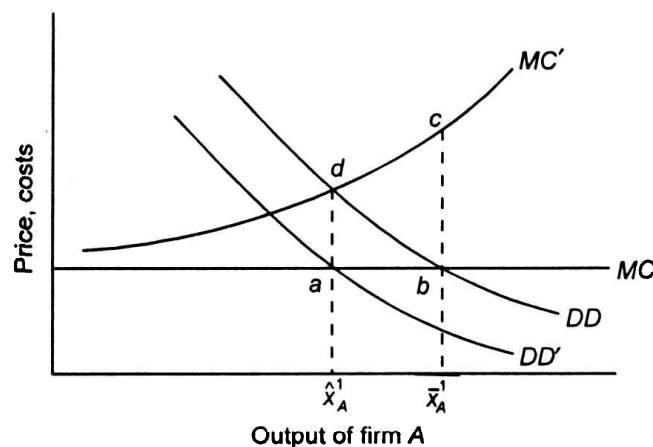


Fig. 23.1 Costs of an Externality

RESOLUTION OF PRODUCTION EXTERNALITIES: TRADITIONAL WAYS

Within the confines of the simple model we have considered above to analyze externalities, there is a number of potential solutions to the allocational problems posed. We will first examine two 'traditional' solutions: internalization, and taxation of costs. Then we will see that in some circumstances

externalities can be accommodated by the normal workings of the market and traditional solutions may be unnecessary.

Merger and Internalization

A traditional cure for the allocational distortions caused by production externalities would be merging of two firms. If there was a single firm operating both the plants of A and B , it would recognize the detrimental effect that firm A has on the production function of firm B . Then there will be no externality! We say that the externality has been internalized as a result of the merger.

Attempts to internalize externalities in productions are quite common. It is often the case that an organization will expand in size with the purpose of encompassing all the spillover effects of its activities. Firms, for example, may merge to capture external benefits. Recreational site developers (ski areas, golf courses, resort hotels) often operate the service industries (hotels, gas stations, shops) near their projects. In this way they are able to internalize the positive benefits that such developments provide to the service industries. Similarly, in the case of the apple orchard and bee keeper example mentioned earlier, it may be possible that more profits could be earned if one or both firms coordinated their activities either by mutual agreement or by sale of one of the firms to the other. Therefore market, more precisely, profit maximization may encourage/signal internalization.

Let us now go back to the production externality problem discussed in the previous section. Before the merger, each firm had the right to produce whatever amount of output it wanted, regardless of what the other firm did. After the merger, the combined firm has the right to control the production of outputs of goods 1 and 2.

The merged firm's profit maximization problem is

$$\max_{x_A^1, x_B^2, s} p_1 x_A^1 + p_2 x_B^2 - C_A(x_A^1, s) - C_B(x_B^2, s).$$

The optimality conditions turn out to be

$$p_1 = \frac{\Delta C_A(x_A^1, s)}{\Delta x_A^1}, \quad (23.4)$$

$$p_2 = \frac{\Delta C_B(x_B^2, s)}{\Delta x_B^2}, \quad (23.5)$$

$$0 = \frac{\Delta C_A(x_A^1, s)}{\Delta s} + \frac{\Delta C_B(x_B^2, s)}{\Delta s}. \quad (23.6)$$

The last condition says that the merged firm will take into account the effect of wastes dumped on the marginal costs of firms *A* and *B*. That is, under merger firm *A* takes the social cost of its production into account. In other words, the merged firm would now pay the full social marginal cost of production of good 1 because it also produces good 2. Consequently, in Fig. 23.1 the marginal cost curve for producing good 1 becomes MC' . This in turn shows that the profit maximizing output of good 1 becomes $\hat{x}_A^1$. This is the optimal output level from a social point of view. At this output level the marginal benefit of good 1's production (what people are willing to pay for the good) is exactly equal to the marginal social cost. Thus, as indicated earlier, under normal workings of the market good 1 will be over-produced ($\hat{x}_A^1 < \bar{x}_A^1$).

We can rewrite Eq. (23.6) as

$$-\frac{\Delta C_A(x_A^1, s)}{\Delta s} = \frac{\Delta C_B(x_B^1, s)}{\Delta s} > 0 \quad (23.7)$$

The right hand side of Eq (23.7), which is denoted by MC_B , measures the (positive) marginal cost to the downstream firm *B* of more wastes. Therefore, the left hand side, which we denote by $-MC_A$, gives the social marginal cost, the marginal cost of firm *A* from dumping more wastes. The upstream firm independently dumps wastes up to the point where its marginal cost from dumping more waste is zero. But under merger, firm *A* dumps wastes up to the point where the effect of marginal increase in wastes is equal to the marginal social cost, which is positive. This positive marginal social cost counts the impact of wastes on the costs of both firms. Hence the merged firm will dump less wastes than the independent firm *A*. In other words, when the true social cost of externality involved in the production of firm *A*'s output is taken into account, the optimal dumping of wastes takes place and this optimal amount is smaller than the amount dumped by independent firm *A*. Since the optimal level of wastes is determined by Eq. (23.6), the intersection of the two marginal costs on the left and right hand sides of Eq. (23.7) also gives the optimal waste level. (See Fig. 23.2.)

As stated earlier, at the efficient level of wastes dumped, the amount that firm *A* is willing to pay for an extra unit of waste should equal the social cost generated by dumped wastes. This is consistent with the efficiency arguments

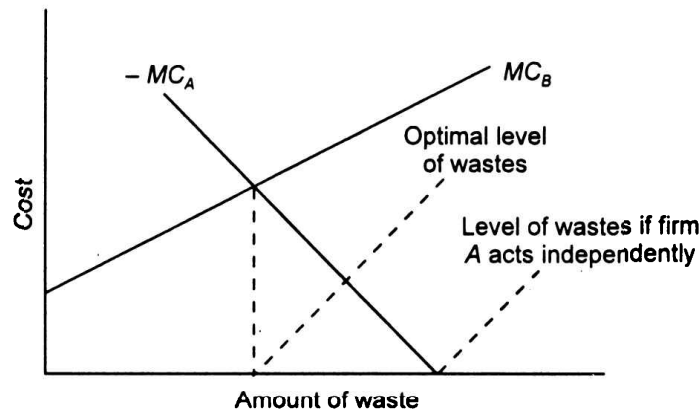


Fig. 23.2 Externality and Private and Social Costs

provided in earlier chapters. In this case under the corrective mechanism, which ensures equality between private and social costs, the free market will determine a Pareto efficient amount of output of each good.

Pigouvian Taxation

The government could impose a suitable tax on the firm generating externality. The imposed tax would cause the output of firm *A*, good 1, to be cut back and would cause inputs to be shifted out of production of good 1. This classic remedy of the externality problem was suggested by Pigou (1920).

The taxation problem can be illustrated in terms of Fig. 23.1. A specific tax of amount t would cause the demand curve DD to shift to DD' . With this new demand curve, the private profit maximizing output will be $\hat{x}_A^1$, the marginal external damage done by producing good 1 is given by the distance ad , which is the amount t . By taxing the good, its effective demand is reduced and resources are shifted out of its production.

To discuss the solution analytically, note that under taxation the profit maximization problem of firm *A* becomes

$$\max_{x_A^1, s} p_1 x_A^1 - C_A(x_A^1, s) - ts,$$

which gives the following optimality conditions

$$p_1 = \frac{\Delta C_A(x_A^1, s)}{\Delta x_A^1}, \quad (23.8)$$

$$t = - \frac{\Delta C_A(x_A^1, s)}{\Delta s}. \quad (23.9)$$

Comparing these two equations to Eq. (23.7), we observe that letting

$$t = \frac{\Delta C_B(x_B^2, s)}{\Delta s} \quad (23.10)$$

will make these conditions the same as the conditions that characterize the optimal level of wastes.

The central problem with Pigouvian taxes is that regulators require empirical information about the optimal level of wastes in order to impose the tax. But once the optimal level of wastes is known, firm A may be asked to dump exactly that much.

Property Rights and the Coase Theory

In order for the decentralized market mechanism to work, firms and individuals should be able to exchange claims on the right to use a factor or consume a good. This requires a well-defined system of property rights that excludes persons/firms from using a good or factor for which they have not paid. Property right establishes the legal owner of a resource and specifies the way in which the resource may be used. It allows agents to exclude themselves from consuming an unwanted commodity or factor without compensation. Two major types of property specifications are common property and private property. Common property is solely owned by the society at large. For such properties it is prohibitively costly or otherwise infeasible to assign and/or enforce property rights to individuals. Private property, on the other hand, is directly owned by individuals who have, within prevailing legal structures, some say over how it is used. Private property may be either exchangeable or non-exchangeable, depending on whether the good in question can be traded or not. (An interesting example of a non-exchangeable property is an individual's vote, a private good provided by the state.)

When we have well-defined property rights in the good involving externality, the concerned parties can trade from their initial endowments to a Pareto efficient state. If property rights are not well-defined, practical problems with externalities arise. For instance, we have difficulties if you have the right to silence but your neighbor believes that he has the right to ride a motorcycle without a muffler. Or, a smoker believes that he has the right to smoke but a non-smoker believes that he has the right to clean air. In our two-firm example, firm A believes that it has the right to dump wastes into

the river but firm B believes that it has the right to clean water. Situations where property rights are not well-defined can lead to inefficient production of externalities, which means that there should be routes for making the concerned persons/firms better off by changing the production of externalities. When property rights are well-defined and mechanisms exist to allow for negotiations between concerned parties, the parties can trade their rights to produce externalities in the same way that they trade rights to consume, to produce goods in the absence of externalities.

For our two-firm externality example, it will be interesting to consider the nature of the property right and negotiation possibilities. Suppose that property rights were defined so as to give 'ownership' of the river to one of the firms but the firms were free to bargain over how the river might be used. It might be thought if the ownership were given to firm A , wastes would be dumped, whereas if the right were given to firm B , the river would remain pure. This might not be the case, because this conclusion disregards the bargaining between two parties. If bargaining is costless, the two parties left on their own will arrive at the efficient output $\hat{x}_A^1$ and this result will be independent of who owns the river. This result is known as the Coase theorem, after the economist who proposed it (Coase, 1960). Formally, the Coase theorem states the following: Given any imperfection in the market that results in an inefficient (non-Pareto efficient) allocation, then, if transaction costs are negligible, negotiations among the concerned agents will lead to the attainment of a Pareto efficient state.

We now demonstrate the Coase theorem for our two-firm example. Suppose, initially, that firm A owns the river. Then it must impute some cost of this ownership into its cost function. This cost of river ownership is determined by the next-best alternative use of the river. In this case only firm B has some alternative use for the river (to keep it clean), and the amount that this firm would be willing to pay to keep the river clean is equal to the external damage done by the wastes dumped. Therefore, if firm A calculates its cost correctly, its marginal cost curve (including the cost of river ownership) will be MC' in Fig. 23.1. Firm A will therefore produce $\hat{x}_A^1$ and sell the remaining rights of river use to firm B for a fee of some amount between abd (A 's loss from producing $\hat{x}_A^1$ rather than $\bar{x}_A^1$) and $abcd$ (the maximum amount that B would be willing to pay to avoid output of good 1 increasing from $\hat{x}_A^1$ to $\bar{x}_A^1$).

A similar result emerges if firm B owns the river. In this case firm A would be willing to pay an amount up to the total profit it earns to achieve the right

to dump wastes into the river. Firm B will accept a payment as long as it exceeds the cost resulting from dumping of wastes. The ultimate point of bargaining will be for firm A to offer a payment to firm B for dumping wastes associated with output level $\hat{x}_A^1$ into the river. Firm B will not sell the rights to dump any more waste because what firm A will pay in exchange falls short of the costs associated with dumped wastes. Therefore, an efficient point is reached through bargaining between the two firms. This completes the proof of the Coase theorem.

It is interesting to note that in both situations some production of good 1 (hence dumping of wastes) takes place. Having no output of good 1 (and no waste dumped) becomes inefficient. Thus, there is some optimal level of pollution and this level is determined thorough bargaining between concerned parties.

We can also demonstrate the Coase theorem analytically. Suppose firm B has the right to clean water and wishes to sell the right to firm A to allow dumping of wastes. Let r be the price per unit of waste dumped and s be the amount of wastes that firm A dumps. (Note that when firm A was acting independently, it used to treat r as zero.) Then the profit maximization problems of firms A and B are

$$\max_{x_A^1, s} p_1 x_A^1 - rs - C_A(x_A^1, s),$$

$$\max_{x_B^2, s} p_2 x_B^2 + rs - C_B(x_B^2, s).$$

Since firm A buys the right to dump s units of wastes into the river, the term rs enters with a negative sign in its profit expression. Because firm B earns a revenue of rs by selling the right, rs appears with a positive sign in B 's expression.

The profit maximizing conditions are

$$p_1 = \frac{\Delta C_A(x_A^1, s)}{\Delta x_A^1}, \quad (23.11)$$

$$r = - \frac{\Delta C_A(x_A^1, s)}{\Delta s}, \quad (23.12)$$

$$p_2 = \frac{\Delta C_B(x_B^2, s)}{\Delta x_B^2}, \quad (23.13)$$

$$r = \frac{\Delta C_B(x_B^2, s)}{\Delta s}. \quad (23.14)$$

The conditions of Eqs. (23.12) and (23.14) show that the marginal cost to firm *A* of reducing waste dumping should be equal to the marginal benefit to firm *B* of that reduction. Equality between these two marginal costs is necessary for production of the optimal pollution level. This is the same condition we had in Eq. (23.7). We can also interpret this equality condition as follows: If the price of pollution is adjusted until the demand for wastes equals the supply of wastes, we will have an efficient level of wastes dumped, just as with any other good.

Conversely, suppose that firm *A* has the right to dump waste and firm *B* will be willing to pay firm *A* for reduction of wastes. Now suppose firm *A* dumps wastes up to certain amount, say, s_m . But firm *B* will pay firm *A* to reduce this. The profit maximization problems now become

$$\max_{x_A^1, s} p_1 x_A^1 + r(s_m - s) - C_A(x_A^1, s),$$

$$\max_{x_B^2, s} p_2 x_B^2 - r(s_m - s) - C_B(x_B^2, s).$$

The optimality conditions turn out to be,

$$p_1 = \frac{\Delta C_A(x_A^1, s)}{\Delta x_A^1}, \quad (23.15)$$

$$r = - \frac{\Delta C_A(x_A^1, s)}{\Delta s}, \quad (23.16)$$

$$p_2 = \frac{\Delta C_B(x_B^2, s)}{\Delta x_B^2}, \quad (23.17)$$

$$r = \frac{\Delta C_B(x_B^2, s)}{\Delta s}, \quad (23.18)$$

which are the same as the conditions of Eqs (23.11) to (23.14). Thus, the optimal amount of dumping is independent of who owns the river.

It should be noted that the Coase theorem depends crucially on the assumption that bargaining is costless. If there are costs associated with

bargains, then we will have to compare those costs to the potential gains resulting from bargaining. If the gains exceed the bargaining costs, then only will the Coasian result hold. Another important point is that the demand for the good causing externality does not depend on the distribution of income. Therefore, the Coase theorem implicitly assumes that there are no income effects.

CONSUMPTION EXTERNALITIES

To illustrate consumption externality let us imagine two roommates A and B , who have preferences over 'watching TV ' (good 1) and 'money' (good 2). We suppose that two types of programs are available on TV —while channel I telecasts only sports programs, the sole program of channel II is music. It is assumed that both consumers like money but that A likes to watch sports programs and dislikes music while B likes to listen to music. Let the utility functions of the two consumers be

$$U_A(x) = x_A^1 x_A^2 - x_B^1, \quad (23.19)$$

$$U_B(x) = x_B^1 x_B^2, \quad (23.20)$$

where $x = (x_A^1, x_A^2; x_B^1, x_B^2)$ with $x_A^1 (x_B^1)$ being the time in a day that $A(B)$ spends for watching TV and $x_A^2 (x_B^2)$ is the amount of money possessed by $A(B)$. Since A dislikes music he gets irritated when B listens to music. This is expressed by showing dependence of A 's utility on x_B^1 , the time that B spends for listening to music, in a negative way. Thus, this is a case of negative consumption externality.

The initial endowments of the two consumers are $w_A = (10, 0)$ and $w_B = (0, 10)$. By $w_A = (10, 0)$, we mean that initially A does not possess any money though he owns the TV and the total time in a day for which TV programs are telecast is 10 hours. In contrast $w_B = (0, 10)$ means that initially B owns the entire amount of money, which is 10. Trade in this economy will mean that A will allow B to watch TV for some time in exchange for money.

To start the analysis let us first solve for the set of Pareto efficient allocations. Since $x_A^1 + x_B^1 = 10$ and $x_A^2 + x_B^2 = 10$, we can rewrite U_A as

$$\begin{aligned}
U_A(x) &= x_A^1 x_A^2 - (10 - x_A^1) \\
&= x_A^1 x_A^2 + x_A^1 - 10
\end{aligned} \tag{23.21}$$

Next, we calculate the marginal rate of substitution of good 2 for good 1 for A:

$$\begin{aligned}
MRS \text{ for } A &= \frac{\text{marginal utility of good 1}}{\text{marginal utility of good 2}} \\
&= \frac{\Delta U_A(x)/\Delta x_A^1}{\Delta U_A(x)/\Delta x_A^2} \\
&= \frac{x_A^2 + 1}{x_A^1},
\end{aligned} \tag{23.22}$$

and for person B:

$$\begin{aligned}
MRS \text{ for } B &= \frac{\text{marginal utility of good 1}}{\text{marginal utility of good 2}} \\
&= \frac{x_B^2}{x_B^1}.
\end{aligned} \tag{23.23}$$

Then, to determine the locus of tangency points of the two individuals' indifference curves, we set MRS for A equal to MRS for B:

$$\frac{x_A^2 + 1}{x_A^1} = \frac{x_B^2}{x_B^1} = \frac{10 - x_A^2}{10 - x_A^1}. \tag{23.24}$$

Solving this in terms of x_A^1 we get

$$x_A^2 = \left(1 + \frac{1}{10}\right) x_A^1 - 1. \tag{23.25}$$

Fig. 23.3 shows the contract curve, the locus of Pareto efficient tangency points, for this two-person economy.

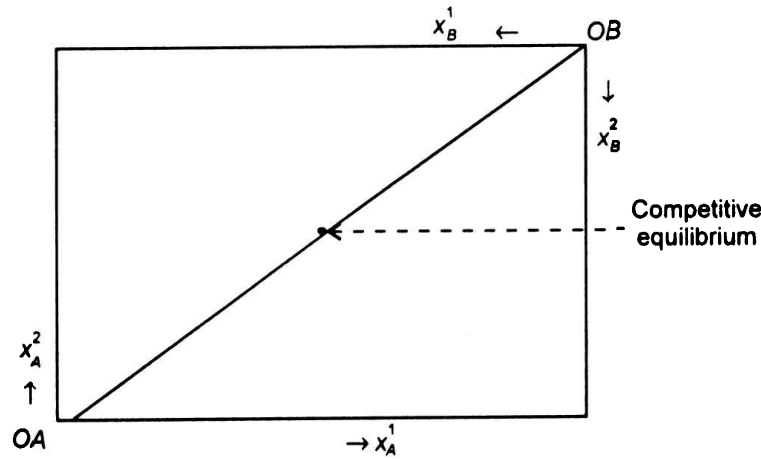


Fig. 23.3 Consumption Externality

The next step is to calculate a competitive equilibrium. In competitive equilibrium persons *A* and *B* will act as price takers. Person *A* wants to maximize his utility $U_A(x) = x_A^1 x_A^2 - x_B^1$ subject to his budget constraint $p_1 x_A^1 + p_2 x_A^2 = 10 p_1 + 0 \cdot p_2 = 10 p_1$, where p_1 and p_2 are the competitive prices of goods 1 and 2 respectively. Since *A*'s income $10 p_1$ already includes his initial endowment $(10, 0)$, in this private optimization case we do not replace x_B^1 by $10 - x_A^1$. Now *A*'s utility maximizing conditions are

$$\frac{\text{marginal utility of good 1}}{\text{marginal utility of good 2}} = \frac{p_1}{p_2} \quad (23.26)$$

$$\text{and} \quad p_1 x_A^1 + p_2 x_A^2 = 10 p_1. \quad (23.27)$$

On substituting the expressions for marginal utilities of goods 1 and 2 for *A*, we can rewrite the above conditions as

$$\frac{x_A^2}{x_A^1} = \frac{p_1}{p_2} \quad (23.28)$$

$$\text{and} \quad p_1 x_A^1 + p_2 x_A^2 = 10 p_1. \quad (23.29)$$

Person *B*, on the other hand, wants to maximize his utility $U_B(x) = x_B^1 x_B^2$ subject to $p_1 x_B^1 + p_2 x_B^1 = 0 \cdot p_1 + 10 p_2 = 10 p_2$. Since the marginal utilities of goods 1 and 2 for *B* are x_B^2 and x_B^1 respectively, the utility maximizing conditions for this person are

$$\frac{x_B^2}{x_B^1} = \frac{p_1}{p_2} \quad (23.30)$$

$$\text{and} \quad p_1 x_B^1 + p_2 x_B^2 = 10 p_2. \quad (23.31)$$

From Eqs. (23.28) and (23.29) we have $x_A^1 = 5$. Next, from Eqs. (23.30) and (23.31) it follows that $x_B^2 = 5$. Since the initial endowments of the two persons are $w_A = (10, 0)$, $w_B = (0, 10)$, it must be the case that $x_B^1 = 5$ and $x_A^2 = 5$. Thus, the unique competitive equilibrium allocation here is $[(x_A^1 = 5, x_A^2 = 5); (x_B^1 = 5, x_B^2 = 5)]$. Note that from Eq (23.28) we have $p_1/p_2 = x_A^2/x_A^1 = 5/5 = 1$. Therefore, the price vector $(p_1 = 1, p_2 = 1)$ can be regarded as the competitive price vector. Clearly, the unique competitive equilibrium allocation $[(5, 5); (5, 5)]$ does not satisfy the tangency condition.

$$x_A^2 = \left(1 + \frac{1}{10}\right)x_A^1 - 1.$$

The competitive equilibrium in fact lies above the contract curve shown in Fig. 23.3. Consequently, it is not Pareto efficient. When externalities are present, a competitive equilibrium allocation need not be Pareto efficient, and the First Fundamental Theorem breaks down.

RESOLUTION OF CONSUMPTION EXTERNALITIES: PIGOUVIAN TAXES

Externalities weaken the link between competition and Pareto efficiency. There is a remedy that is consistent with the decentralized price mechanism. This remedy is the introduction of per unit (or marginal) taxes or subsidies on the consumption of goods that induce externalities. They will depend on the quantities consumed. Furthermore, they should be designed to encourage a person's consumption of a good if that consumption has positive external effects, and to discourage a person's consumption of a good if that consumption generates negative external effects.

Let us now return to the example of the previous section. In that example B 's consumption of good 1 has a negative external effect on person A . But when person B makes his consumption decision in the standard competitive equilibrium framework, he does not take into account the negative impact of this consumption on A 's utility. This suggests that this consumption of good 1 should be taxed.

Let t be a tax paid by person B for each unit of good 1 that he consumes. The connection between competition and efficiency will be reestablished if there exists a t , a price vector $p = (p_1, p_2)$ and cash transfers T_1, T_2 , so that when A 's budget constraint is

$$p_1 x_A^1 + p_2 x_A^2 \leq 10p_1 + T_1 \quad (23.32)$$

and B 's budget constraint is

$$(p_1 + t) x_B^1 + p_2 x_B^2 \leq 10 p_2 + T_2 \quad (22.33)$$

the two will move to a Pareto efficient allocation through the competitive mechanism. In fact, there exist such t, p, T_1 and T_2 and with the appropriate choice of T_1 and T_2 we move to any desired Pareto efficient allocation.

To see this, let $y_A = (5, 4.5)$ and $y_B = (5, 5.5)$. Since the allocation $y = (y_A, y_B)$ satisfies Eq. (23.25), it is Pareto efficient. Now, the marginal rate of substitution for person A at y_A is

$$\frac{4.5}{5} = \frac{p_1}{p_2} \quad (23.34)$$

and his budget constraint is

$$5p_1 + 4.5 p_2 = 10 p_1 + T_1 \quad (23.35)$$

If we let $T_1 = 0$, we have $p_1/p_2 = 4.5/5$. Thus, if we normalize prices by setting $p_2 = 10$, we have $p_1 = 9$.

Next, the marginal rate of substitution for person B at y_B is

$$\frac{5.5}{5} = \frac{p_1 + t}{p_2} \quad (23.36)$$

and his budget constraint is

$$5(p_1 + t) + 5.5 p_2 = 10 p_2 + T_2. \quad (23.37)$$

Given $p_1 = 9, p_2 = 10$, from Eq. (23.36) it follows that $t = 2$. Putting the values of p, p_2, t in Eq. (23.37) we have $T_2 = 10$. Note that $t \cdot y_B^1 = 10 = T_1 + T_2$, so that, the total government cash transfer equals its commodity tax revenue.

Therefore, if $T_1 = 0$, $T_2 = 10$ and $t = 2$, the competitive mechanism, modified by T_1 , T_2 and t , will take the economy to the Pareto efficient allocation (y_A, y_B) . Thus, the introduction of the tax t reestablishes the connection between competition and efficiency.

To explain the intuitive reasoning behind this calculation, note that at y_A the marginal utility to person A of his own consumption of good 1 is 4.5. Now, if person B increases his consumption of good 1 by 1 unit, person A would have to increase his consumption of good 1 by $1/4.5$ units in order to remain as well-off as before. Thus, there is a marginal rate of substitution between A 's consumption of good 1 and B 's consumption of good 1 and this is given by the marginal disutility to person A of person B 's consumption of good 1 divided by the marginal utility to person A of his own consumption of good 1. This is a measure of the damage done to A by B 's consumption of good 1. We call this marginal external damage or cost of person B 's consumption of good 1. Note that this marginal external cost is measured in terms of units of good 1. Therefore, by multiplying this by the price of good 1, we get the marginal external cost in money terms. Since $p_1 = 9$, this amount turns out to be $p_1 \cdot 1/4.5 = 2$. But the required tax was 2. That is, the tax equals the value of the external damage or cost. This is intuitively reasonable: if person B is causing Rs. 2 worth of external damage for every extra unit of good 1 he consumes, then the appropriate way to achieve efficiency is to impose a tax of Rs. 2 per unit on each extra unit he consumes. This way externality is internalized.

We can also apply the Coase theorem to internalize externalities here. Suppose person A has a legal right to silence. It can easily happen that A would prefer to trade some of his right to silence in exchange for some money. That is, A is allowing himself to be bribed to consume some of B 's noise. It is easy to imagine that A and B are trading to a Pareto efficient point.

Conversely, suppose B has a legal right to listen to loud music and (hence to noise) and A would have to bribe B to reduce noise. We can easily imagine the agents trading to a mutually preferred point.

The exact position of the efficient point is determined by the property rights and the precise mechanism which A and B use to trade. This is like the usual Edgeworth box analysis, but described in a somewhat different set-up. Thus, we see that whenever we have a market for noise, a competitive equilibrium becomes Pareto efficient.

THE PROBLEM OF COMMONS

There are many examples of externalities resulting from the existence of common property. This is not a new problem—it dates back to at least Middle Ages. The problem was recognized in the context of use of the ‘commons’. The ‘commons’ is a term that refers to a plot of land that was reserved by the lord of the manor for the common use of his tenants. The fertility of the commons’ soil gradually eroded because of the misuse of the plot by the tenants through overgrazing by their cattle, with maintenance being neglected.

Another classic example of a common property externality is the case of a fishery (Gordon, 1954). When a fisherman increases his fishing, he reduces the equilibrium stock of fish and raises the cost of finding a fish to other fishermen. Since this cost is external to the fisherman, he ignores it when making his decision of how many fish to catch.

The common property here is the inventory of fish in the common place: once a fish is caught, property rights on it are established. The result of this common property externality is overfishing, which reduces the stock of fish in the common place. Additional output can be obtained by using more and more variable factors such as labor and fishing boats. The average and marginal cost will eventually rise since the common ground itself is of a fixed size. Since the place is commonly owned, fishermen treat it as if it were free despite the fact that it is a scarce resource of given size. Free entry into the fishing ground along with free use of the ground implies that the output of the fishing ground will expand up to the point where price equals marginal private cost, which is smaller than the marginal social cost. The difference between these two costs gives us the external marginal cost imposed on other fishermen when one fisherman increases his catch by one unit. Because of the common property nature of the ground, fishermen do not take this into account.

Each fisherman has a negligible impact on the total stock of fish, but the total efforts of all the fishermen together result in serious depletion. Property rights on the stock of fish can be partly restored by imposing fishing quotas. But his corrective device does not work at all if the fish can travel outside national boundaries (for example, in a sea). The problem can also be solved by banning new entries, and increasing the mesh size of nets of the fishermen. If the fishing ground were owned by a firm or the government, and users were charged a price equal to the marginal external cost, then common property resource problems would not exist, and the fishing ground would be used optimally.

Applied Example 23.1 Air Pollution and Externality

The West Bengal Pollution Control Board ordered the closure of the kitchen and generator units of two hotels at Digha, in Midnapore district on January 3, 1996 (*The Statesman*, Calcutta, January 4, 1996). These hotels were charged with polluting the sea water with effluents from the generator units and kitchen. The same directive instructed a third hotel to set up a chlorination unit and a fourth hotel to ensure normal discharge of water according to the board's specifications. Several hotels and restaurants in Calcutta were also asked to install pollution control measures immediately.

These hotels and restaurants were found to produce too much pollution and that is why they had to face punitive action or were given statutory warnings. One approach to controlling the amount of pollution is to require that the hotels meet certain standards in the amounts of pollution they generate.

Following Varian (1993), we can develop a simple model to determine the optimal emission of pollution. Suppose there are n hotels and hotel i 's emission quota is s_i , $i = 1, 2, \dots, n$. The cost of achieving an emission quota s_i is $C_i(s_i)$. We wish to minimize the total costs of achieving the emissions quota, subject to the aggregate constraint. That is, we need to

$$\text{minimize } C_1(s_1) + C_2(s_2) + \dots + C_n(s_n) \quad (23.38)$$

$$\text{subject to } s_1 + s_2 + \dots + s_n = S. \quad (23.39)$$

The solution requires that the marginal cost of emission control must be equal across hotels. If not, we can decrease the total cost of emission control by reducing the quota of a hotel with a higher marginal cost and increasing the quota of a hotel with a lower marginal cost.

Suppose there is a market for pollution. If a hotel exactly meets its pollution quota it faces no action or fines. But if it reduces its pollution by more than its quota, it can sell the extra 'right to pollute' on the open market. Hotels that find it easy to reduce pollution will sell credit to hotels that find it costly to reduce pollution. In equilibrium, the market price for the right to emit one unit of pollution will be equal to the marginal cost of reducing pollution by one unit. This requirement turned out to be the condition for the optimal pattern of pollution. Thus, the market for pollution produces the efficient pattern of pollution automatically.

EXERCISES

Multiple Choice

1. If a firm has been creating pollution and if it is asked to find a method of waste disposal; which is not harmful to the society, the result will probably be
 - a) A lower price for firm's product.
 - b) A higher price and a higher equilibrium quantity.
 - c) A lower price and a higher equilibrium quantity.
 - d) A higher price and a lower equilibrium quantity.
2. If a product generates external benefits, then the demand curve reflecting social benefits
 - a) Will coincide with the demand curve that shows private benefits.
 - b) Will be upward sloping.
 - c) Will be to the left of the demand curve that shows private benefits.
 - d) Will be to the right of the demand curve that shows private benefits.
3. Which of the following statements is not true?
 - a) If social costs are higher than private costs, external costs are present.
 - b) If a product generates external costs, fewer resources should be devoted to its production.
 - c) Consumers will buy too much of the product that creates external benefits.
 - d) In the presence of externalities, resources are not used optimally.
4. A market economy is likely to underproduce those products which generate
 - a) Social benefits.
 - b) Social costs.
 - c) External benefits.
 - d) External costs.
5. Which of the following is likely to produce external costs?
 - a) A cigarette.
 - b) A flower garden.
 - c) A highway.
 - d) A lighthouse.

Quantitative Problems and Questions

1. Prove or disprove the following
 - a) Externalities affect efficient allocation of resources, through divergence between private and social costs.
 - b) There are mechanisms in the market system that can bring social costs into producers' private cost considerations.
 - c) When property rights are well-defined, it may be possible to eliminate externalities.
 - d) The Coase theorem holds under the assumption that there should be income effects.
 - e) The problem of pollution can be solved from the point of view of supply and demand.

2. Consider a two-person exchange economy with the following characteristics

$$U_A(x) = x_A^1 x_A^2 - x_B^1 x_B^2,$$

$$U_B(x) = x_B^1 x_B^2.$$

Suppose that the initial endowments of the two consumers are $w_A = (1, 0)$ and $w_B = (0, 1)$.

- a) How do you interpret the existence of the term $x_B^1 x_B^2$ with a negative sign in person A's utility function?
 - b) Discuss the optimality of a competitive equilibrium allocation.
 - c) Calculate the taxes and transfers which will generate an efficient allocation.
3. Suppose that both chemical and eyeglass industries in a city are perfectly competitive. Chemical production raises the marginal cost of eyeglass in the sense that smoke from the chemical industry leaves particles on equipment used for making eyeglass, and these particles are required to be removed.
 - a) Demonstrate graphically how market forces misallocate resources.
 - b) Explain how reallocation might be brought about through a tax on chemical production or through costless bargaining.
 - c) Is internalization through merger possible?
4. Replace the utility function $U_A(x)$ in Eq. (23.19) by $U_A(x) = x_A^1 x_A^2 + x_B^2$. Assume that the utility function of person B is given by Eq. (23.20). The

initial endowments are the same as before. Calculate the subsidy that will establish a connection between efficiency and competition.

5. Consider two firms A and B which use labor L as the only factor of production to produce a particular good. The production functions of the two firms are $x_A = \sqrt{L}$ and $x_B = \sqrt{L} - \frac{x_A}{2}$. The output price and the wage rate are both equal to 1.
 - a) Determine the competitive equilibrium production plan.
 - b) Is the competitive equilibrium efficient? If not, calculate the amount of tax that should be imposed on firm 1 to make the modified equilibrium efficient.
6. What is meant by negative and positive externalities? Give some examples.
7. What is meant by private cost and social cost? Show how divergences between the two occur and how this results in misallocation of resources.
8. How does taxation internalize external costs?
9. How can externalities be resolved? What is meant by the expression 'to internalize externalities'?
10. What is the Coase theorem? Discuss the assumptions on which this theorem is based.

24

Public Goods

In the previous chapter we analyzed certain kinds of externalities, and discussed the mechanisms for correcting externality problems. However, our discussion has been concerned solely with private goods. (A good is said to be private if it conveys benefits only to the purchaser. For instance, a domestic car is a private good.) We will now examine the theory of production and consumption of goods whose character is public, rather than private.

How do we define a public good? Two important properties that characterize public goods are non-exclusivity and non-rivalry. A good is exclusive if it is easy to exclude individuals from benefitting from the good. For most private goods such exclusion is possible: I can easily be excluded from consuming a glass of milk if I do not pay for it. A good is non-exclusive if it is impossible to exclude individuals from benefitting from the good. A good is non-rival if one person's consumption does not decrease the amount available to others. Equivalently, we can say that a non-rival good is one for which additional units can be consumed at zero marginal social cost. A good is rival if one person's consumption reduces the amounts available to others. Consumption of additional units of rival goods involves some marginal costs of production. Consumption of one more glass of milk by someone, for example, requires that some resources be devoted to its production. Private goods are both exclusive and rival.

Many goods that are non-exclusive are also non-rival. National defence and mosquito control are two examples. Once a defence system is established, everyone in a country benefits from it. Similar comments apply to goods such as mosquito control or inoculation against diseases. In these cases, once the programs are implemented, nobody in the community can be excluded from those benefits. Similarly consumption of these goods by one person does not reduce the amount available to others. Thus, additional consumption of these goods takes place at zero marginal cost. Goods that are non-exclusive and non-rival are called public goods. Other examples of public goods are *TV* and radio broadcasts, highways, lighthouses, police and fire protection and so on.

Some goods may possess one property, but not the other. For instance, the use of a bridge during off-peak hours may be non-rival, but it is possible to exclude potential users by imposing toll booths. Goods of this type are called club goods (Buchanan, [1965]). A cable *TV* broadcast is another example of this type of good. We may also have privately produced club goods such as a swimming pool or a golf course. In this case a private producer makes available a good which is consumed jointly by members.

We may have some goods that are not excludable, but are rival. That is, given the quantity of the good, allowing an additional individual to use the good does not prevent the previous users from using it, but it does reduce benefits attained by previous users. Therefore there is some opportunity cost involved in admitting additional users. This phenomenon of partial rivalness is called congestion in economics literature. The opportunity cost of allowing more users of the public good, or the reduction in benefits to those already consuming it, is called congestion cost. An example of a public good of this type is a crowded street. Any body can use it, but one person's use reduces the amount of space available to others. Sometimes goods of this type and club goods are referred to as impure public goods. On the other hand, non-rival, non-exclusive goods are called pure public goods.

There are certain goods which are private, but sometimes they are treated as public goods. For instance, education and health care are private goods. But in many countries education (upto some level) and health care (for low income groups) are provided publicly.

There exist public goods on the production side also. In this case a factor of production is used simultaneously by a number of firms. A good of this type is called a public intermediate good. The fruits of new discoveries, public information and weather reports might be examples.

What are the efficiency implications of public goods? Like externalities, public goods possess certain properties which render the market mechanism

an inefficiency device for allocating resources to them. What then should be done? How should the public goods be financed? That is, are private markets a good mechanism for allocation of public goods? Or do we need other mechanisms, like voting? What are the optimal quantities for public goods? We shall explore these and similar questions in this chapter.

EFFICIENT SUPPLY OF DISCRETE PUBLIC GOOD

In this section we consider a two-agent, two-good model. One good is a private good and the other good is a public good. We write g for the quantity of the public good. Here we assume that the public good is available only in a discrete amount. That is, either the good is available in that amount or it is not available at all. A concrete example is having one streetlight or not having it all. A second example can arise when two persons who share an apartment buy a TV and put it in the living room. Since both of them can watch the TV, it becomes a public good.

We write w_1 and w_2 to denote the initial wealths of persons 1 and 2 respectively. Suppose person i decides to contribute e_i to the public good ($i = 1, 2$). Letting $y_1(y_2)$ be person 1(2)'s money left over to spend on the consumption of the private good, we can write the budget constraints of the two individuals as

$$\begin{aligned} e_1 + y_1 &= w_1, \\ e_2 + y_2 &= w_2. \end{aligned} \tag{24.1}$$

Suppose it costs c amount of money to have one unit of the public good g . Therefore, one unit of the public good will be provided if the two persons jointly pay the cost c . In other words, the technology to provide the public good is given by

$$\begin{aligned} g &= 1 \text{ if } e_1 + e_2 \geq c, \\ &= 0 \text{ if } e_1 + e_2 < c. \end{aligned} \tag{24.2}$$

The utility function of person 1 depends on the amount of money y_1 he spends on the consumption of the private good and the public good g , where g will either be zero, indicating no public good, or one, indicating that one unit of the public good will be there. Therefore, we write person 1's utility function as $U_1(y_1, g)$. U_1 is increasing, that is, it is increasing in consumption of both the public and the private good. Similarly person 2's increasing utility function is $U_2(y_2, g)$. Observe that we use the subscript 1(2) to indicate person 1(2)'s consumption of the private good. Since there is jointness in the consumption of the public good, it has no subscript.

Now, when should the public good be provided? The public good should be provided if both people become better off having the good and paying their share than not having it. That is, providing the public good should Pareto dominate not providing it. Equivalently, having the public good is a Pareto improvement over not having it.

It will be Pareto improvement to provide the public good if there is some pattern of consumption (e_1, e_2) such that $e_1 + e_2 \geq c$ and

$$\begin{aligned} U_1(w_1, 0) &< U_1(w_1 - e_1, 1), \\ U_2(w_2, 0) &< U_2(w_2 - e_2, 1). \end{aligned} \quad (24.3)$$

Observe that in the context of provision of the public good, two kinds of allocations are of interest. One allocation is that the good is not provided. In this case each person spends his entire wealth on his private consumption. That is, person i 's utility in this case is given by $U_i(w_i, 0)$, $i = 1, 2$. The other allocation is the one where the public good is provided. Then using the budget constraints we see that in this second situation person i 's utility function takes the form $U_i(w_i - e_i, 1)$. Therefore, having the public good is a Pareto improvement over not having it if Eq. (24.3) holds.

Now, let t_1 be the reservation price of person 1. This price is the maximum amount that person 1 would be willing to give to get one unit of the public good. That is, t_1 is that price such that person 1 is indifferent between paying t_1 for having the public good and not having the public good at all. Using person 1's utility function we can define t_1 implicitly as

$$U_1(w - t_1, 1) = U_1(w_1, 0). \quad (24.4)$$

If person 1 pays t_1 for the public good he will have $w_1 - t_1$ for private consumption. Therefore, his utility is $U_1(w_1 - t_1, 1)$ when he decides to pay t_1 for the public good. In case he decides not to have the public good, the entire wealth w_1 will be available for private consumption which gives him a utility of $U_1(w_1, 0)$. We determine t_1 by setting equality between these two utilities. Similarly, person 2's reservation price t_2 is defined implicitly by

$$U_2(w_2 - t_2, 1) = U_2(w_2, 0). \quad (24.5)$$

From the first inequality of Eq.(24.3) and from (24.4) we have

$$U_1(w_1 - t_1, 1) = U_1(w_1, 0) < U_1(w_1 - e_1, 1). \quad (24.6)$$

Similarly, the second inequality of Eq. (24.3) along with Eq. (24.5) gives us

$$U_2(w_2 - t_2, 1) = U_2(w_2, 0) < U_2(w_2 - e_2, 1). \quad (24.7)$$

Since utility is increasing in private consumption, from Eqs. (24.6) and (24.7) we have

$$\begin{aligned} w_1 - t_1 &< w_1 - e_1 \\ \text{and} \quad w_2 - t_2 &< w_2 - e_2. \end{aligned} \quad (24.8)$$

from which it follows that

$$\begin{aligned} t_1 &> e_1 \\ \text{and} \quad t_2 &> e_2. \end{aligned} \quad (24.9)$$

Adding the two inequalities in Eq. (24.9) we see that

$$t_1 + t_2 > e_1 + e_2 \geq c. \quad (24.10)$$

The two inequalities in Eq. (24.9) state that for each person the maximum willingness to pay is greater than his contribution to the public good. From our derivation it is clear that if it is Pareto improvement to provide the public good, then the condition of Eq. (24.9) and hence Eq. (24.10) must hold. In other words, the condition of Eq. (24.9) and hence Eq. (24.10) is a necessary condition for acquisition of the public good to be a Pareto improvement. This is evident, since if a consumer can acquire the good by paying less than the maximum he is willing to pay, then acquisition will benefit him.

According to the condition of Eq. (24.10) the sum of the willingness to pay for the public good exceeds the cost of providing it. We can show that this condition turns out to be a sufficient condition for Pareto improvement under acquisition of the public good. To see this, we choose $e_1(e_2)$ slightly less than $t_1(t_2)$ so that $e_1 + e_2 \geq c$ and

$$\begin{aligned} U_1(w_1 - e_1, 1) &> U(w_1, 0), \\ U_2(w_2 - e_2, 1) &> U(w_2, 0), \end{aligned} \quad (24.11)$$

hold. From Eq. (24.11) it follows that when $t_1 + t_2 > c$ holds, it is Pareto improving to have the public good.

We note that the condition describing the Pareto improving provision of the public good depends on each person's willingness to pay, which in turn depends on the initial wealth distribution (w_1, w_2) , and the cost of provision c . We may compare this condition with the efficiency condition for providing a private good, in the case of a private good, if a person is willing to pay the cost of producing the good, then it is efficient to provide it. But here we

require the weaker condition that the sum of the willingness to pay is higher than the cost of provision.

PRIVATE FINANCING OF DISCRETE PUBLIC GOOD AND THE FREE RIDER PROBLEM

In this section we wish to examine how effective a private market is for providing public goods. Suppose that for our 2 person 2 good example the reservation price of both the persons is Rs 200 and cost of providing the public good is Rs 300. Therefore, the sum of the willingness to pay is higher than the cost of provision, which means that it is Pareto improving to provide the public good.

To start the analysis, we suppose that person 1 takes the lead, and makes the first purchase of the public good. Thus, he gets Rs 200 worth of benefits (since his reservation price is Rs 200) but has to pay Rs 300 for these benefits.

Now, if person 1 has purchased the public good, what does person 2 do? He enjoys the benefits from 1's purchase without paying for those benefits. The public good being a jointly-consumed commodity, person 2 cannot be excluded from its consumption though he is not paying anything for it. This problem is known as the free rider problem.

Free riding is quite similar to the prisoner's dilemma that we have examined in the chapter on game theory. To construct the pay-off matrix for the free riding game, note that if person 1 buys the public good, then his net benefit is -Rs 100. (This follows from the fact that his reservation price is Rs. 200, but the cost of the public good is Rs 300.) When person 1 buys the public good, person 2 takes a free ride and enjoys a benefit of Rs 200. When they decide independently to buy the good, each person enjoys a benefit of -Rs. 100. If nobody decides to buy the public good, then each person enjoys a benefit of zero. We summarize these observations in the following pay-off matrix.

Table 24.1 Pay-off Matrix for Free Riding Game

		<i>Person 2</i>	
		Contribute	Don't Contribute
Person 1	Contribute	(-100, -100)	(-100, 200)
	Don't contribute	(200, -100)	(0, 0)

The dominant strategy equilibrium for this game is (don't contribute, don't contribute). If person 1 decides to contribute to the cost of the public good, person 2 will take a free ride. If person 1 decides not to contribute, then person 2's interest will also be not to contribute. Hence (don't contribute, don't contribute) is the dominant strategy equilibrium. That is, the net result is that the good will not be provided at all, though it is efficient to do so.

If person 1 contributes to the cost of the good, and both persons use it, then a Pareto improvement occurs if person 2 makes a side payment to person 1. Since person 1's net benefits are –Rs 100, any side payment between Rs 100 and Rs 200 will result in a Pareto improvement. In fact this kind of cost sharing would probably happen in practice. But in many cases it is difficult to avoid the free riding problem. When there are more than two persons, there are many people on whom one may wish to free ride. This becomes Pareto inefficient from society's viewpoint.

Therefore, purely independent decisions may not result in an efficient provision of public good. It will be necessary to use more complicated ways of financing a public good.

EFFICIENT SUPPLY OF CONTINUOUS PUBLIC GOOD

Now we assume that the public good can be provided in any continuous amount. An example is radio or TV broadcasting time. We continue to consider the 2 person 2 good model. As before we write g for the amount of the public good. Let $C(g)$ denote the cost of providing the public good. Assume also that y_1 and y_2 are private consumptions of persons 1 and 2 respectively and e_1 and e_2 are their contributions to the cost of the public good.

The total amount that the two persons spend on their private and public consumption must add up to their initial wealth:

$$y_1 + y_2 + C(g) = w_1 + w_2 \quad (24.12)$$

The definition of Pareto efficiency requires that given any arbitrarily fixed level of utility for person 2, person 1's utility should be as large as possible. Let us fix the utility of person 2 at U_2^0 . Then we can write the problem as

$$\max_{y_1, y_2, g} U(y_1, g) \quad (24.13)$$

such that

$$U_2(y_2, g) = U_2^0,$$

$$y_1 + y_2 + C(g) = w_1 + w_2. \quad (24.14)$$

We demonstrate in Appendix 24.1 that the appropriate optimality condition for this problem is that the sum of the marginal rates of substitution between the private good and the public good for the two persons should equal the marginal cost of providing the public good:

$$MRS^1 + MRS^2 = MC(g). \quad (24.15)$$

Named after Paul A. Samuelson, this condition is called the Samuelson (1955) Optimality Condition for Public Goods. Intuitively it makes a great deal of sense, for it says that at the optimum, the marginal cost of supplying the last unit of the public good in terms of the private good foregone just equals the sum of the marginal benefits that all users of the increment of the public good obtain in terms of the private good.

To illustrate this condition, suppose that a flood control program was under consideration and assume that 100,000 individuals would benefit from the plan. Assume also that each individual would be willing to trade away the benefits that the flood control program promises in exchange for one unit of a composite private good. Clearly, then, the social marginal rate of substitution of the flood control plan for the composite private good is 1 to 100,000. In the aggregate, the individuals would be willing to give up 100,000 units of the composite private good in exchange for the benefits of flood prevention.

In the case of a discrete public good the optimality condition requires that the sum of the willingness to pay should be higher than the cost. In the case when the public good can be supplied in any continuous amount, the optimality condition is that the sum of the marginal willingness to pay should equal the marginal cost of providing the good. If the sum of the marginal willingness to pay for the public good becomes higher (smaller) than the marginal cost, it is appropriate to supply more (less) of the public good.

The problems raised by the non exclusive nature of public goods can also be examined in a partial equilibrium framework by considering the demand curve associated with such goods. In the case of a private good, the market demand curve is determined by summing individuals' demands horizontally. At any price, the quantity demanded by each individual is added to calculate the total quantity demanded in the market. The market demand curve shows the marginal evaluation that individuals place on an additional unit of output.

To determine how society values some level of public good provision, we must ask how each individual values this level of output and then sum these

valuations. Thus, for a public good (which is supplied in a fixed quantity to everyone), individual *MRS* curves must be added vertically. This idea is pictured conceptually in Fig. 24.1. Here the total *MRS* curve for the public good (*M*) is the vertical sum of the individual *MRS* curves. Each point on the *M*-curve represents social marginal evaluation of the particular level of public goods' expenditure. Producing one more unit of the public good would benefit everyone. To evaluate this benefit it is necessary to sum each individual's personal evaluation of the good.

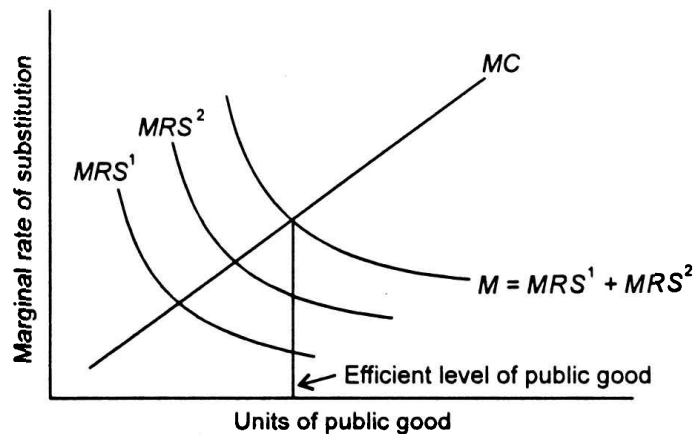


Fig. 24.1 Efficient Provision of Public Good

It may be worthwhile to compare the efficiency condition for a public good with the efficiency condition for a private good. For any two private goods, the marginal rate of substitution must be identical for all individuals and this common *MRS* must equal the rate at which good can be technically transformed. There is no ambiguity if we say that the social *MRS* must equal the social rate of product transformation. In the case of a public good, it is possible to speak of the rate of product transformation of this good for some private good. Exactly the same kind of resource reallocation will take place between private and public producers as takes place between two private producers when the composition of output changes. In producing more for national defense, some trucks will have to be given up, just as they would be if we decided to produce more vehicles. An important difference arises in the context of defining the social marginal rate of substitution as the sum of individual marginal rates of substitution between a public good and a private good. For a private good marginal evaluation (in terms of substitution for another private good) is the same for all persons, though different persons can consume different amounts of the good. But for a public good, marginal evaluations (in terms of substitution for a private good) are different, though each person must consume the same amount of the good.

PRIVATE FINANCING OF CONTINUOUS PUBLIC GOOD AND THE FREE RIDER PROBLEM

Let us assume here that the private market mechanism is used to supply the public good. To simplify the analysis we suppose that a consumer's utility function represents quasilinear preferences. That is, utility is the sum of the quantity of the private good consumed plus a well-behaved function of the quantity of the public good produced and available to all. Such a utility function is said to be separable between private and public consumption. Therefore $U_1(y_1, g)$ can be written as $y_1 + V_1(g)$, where V_1 is strictly concave. Similarly, $U_2(y_2, g) = y_2 + V_2(g)$, where V_2 is strictly concave.

We suppose for simplicity that $C(g) = g$, which implies that the marginal cost of providing a unit of the public good is constant at 1 {that is, $(C(g + \Delta g) - C(g))/\Delta g = 1$ }. Therefore, the total money amount of the public good provided is g . This also means that the ratio at which the private good can be transformed into a public good is one to one. The feasibility constraint $y_1 + y_2 + C(g) = w_1 + w_2$ now becomes $y_1 + y_2 + g = w_1 + w_2$.

Suppose that person 1 makes the first purchase of the public good. He wants to choose a g to maximize his utility

$$U_1(y_1, g) = y_1 + V_1(g) \quad (24.16)$$

subject to his budget constraint

$$y_1 + g = w_1. \quad (24.17)$$

Substituting for y_1 , person 1 wants to maximize $w_1 + V_1(g) - g$. This is equivalent to maximizing $V_1(g) - g$, the vertical distance between the function $V_1(g)$ and the straight line $l_1 = g$ that passes through the origin and has a slope equal to 1. The vertical distance is highest at the point Q , where V_1 is tangent to the line l_2 , which is parallel to l_1 .

At the point of tangency the slope of the V_1 function is equal to the slope of the line l_2 , which is 1. Now, for sufficiently small Δ 's, the marginal utility from the public good $[V_1(g + \Delta g) - V_1(g)]/\Delta g$ coincides with the slope of the V_1 function. If we denote the marginal utility function by V'_1 , then the utility maximizing condition requires that

$$V'_1(g) = 1. \quad (24.18)$$

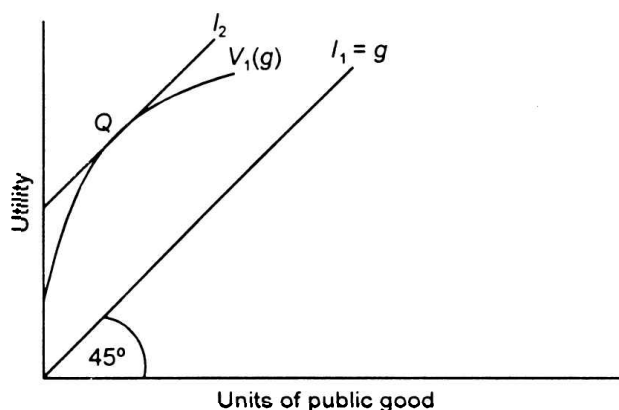


Fig. 24.2 Private Provision of Continuous Public Good

It may be interesting to note that V'_1 can also be regarded as the marginal rate of substitution between the public good and the private good for person 1. To see this, recall that MRS^1 is equal to $[(U_1(y_1, g + \Delta g) - U_1(y_1, g))/\Delta g]$ divided by $[(U_1(y_1 + \Delta y_1, g) - U_1(y_1, g))/\Delta y_1]$. Given that $U_1(y_1, g) = y_1 + V_1(g)$, the denominator of MRS^1 becomes 1. Since the numerator coincides with $V'_1(g)$ for small Δ 's, $V'_1(g)$ becomes MRS^1 . That is, we can rewrite Eq. (24.18) as $MRS^1 = 1$. (It may be worthwhile to mention that the condition of Eq. [24.18] will determine a unique level of the public good. For instance, if $V_1(g) = \log g$, then $V'_1(g) = 1/g$, which on substitution in Eq. [24.18] gives $g = 1$.)

Similarly, if person 2 consumes both goods and maximizes his utility independently then the optimizing condition is $MRS^2 = 1$. But it may happen that person 2 will observe that the amount contributed by person 1 is sufficient and it is unnecessary for him to contribute toward the public good at all.

Given that person 1 has already made a contribution, person 2 can only decide whether or not he wants to increase the amount of the public good. If we denote the solution to Eq. (24.18) by $\bar{g}$, then we have two different cases to analyze, namely, $V'_2(\bar{g}) < 1$ and $V'_2(\bar{g}) \geq 1$. If $V'_2(\bar{g}) < 1$, then marginal utility to person 2 from the public good is less than the marginal cost of the public good. In this case if person 2 was to curtail his consumption of the private good by one unit and increase the amount of the public good by one unit, then he would be worse off. So he will not do that. Therefore, he will take a free ride on person 1's consumption. If $V'_2(\bar{g}) > 1$, then person 2 would benefit if he reduces his private consumption and purchases more units of the public good. He would want the quantity of the public good increased until $V'_2(g) = 1$. A similar situation will arise for person 1 to decide if person 2 makes the first purchase.

Therefore, the final equilibrium in the private market for public and private goods will have the following properties: (i) For at least one person i , we have $V'_i(g) = 1$, (ii) for all i , $V'_i(g) \leq 1$, where $i = 1, 2$, (iii) The details of who pays how much will be determined by who makes the first purchase and who the second. But if $V'_1(g)$ or $V'_2(g) = 1$, and if $V'_2(g)$ or, $V'_1(g) > 0$, then $V'_1(g) + V'_2(g) > 1$, which means that the Samuelson optimality condition is violated. There will not be enough of the public good produced.

This problem becomes clearer if we assume that the functions V_1 and V_2 are the same, say V . Now, if 1 takes the first step and purchases $\bar{g}$ amount of the public good, then person 2 will take a free ride. This follows from the fact that the utility maximizing amount of public good for person 2 is also $\bar{g}$. Then $V'_1(\bar{g}) + V'_2(\bar{g}) = 2V'(\bar{g}) = 2 > 1$. Therefore, there is a difference between $\bar{g}$ and the optimal quantity of the public good.

We illustrate the free riding issue below graphically. Person 1 takes the first step. Person 2 takes a free ride and spends the entire wealth on private consumption. Since person 2 cannot decrease the amount of the public good, his budget line takes the shape shown in the figure.

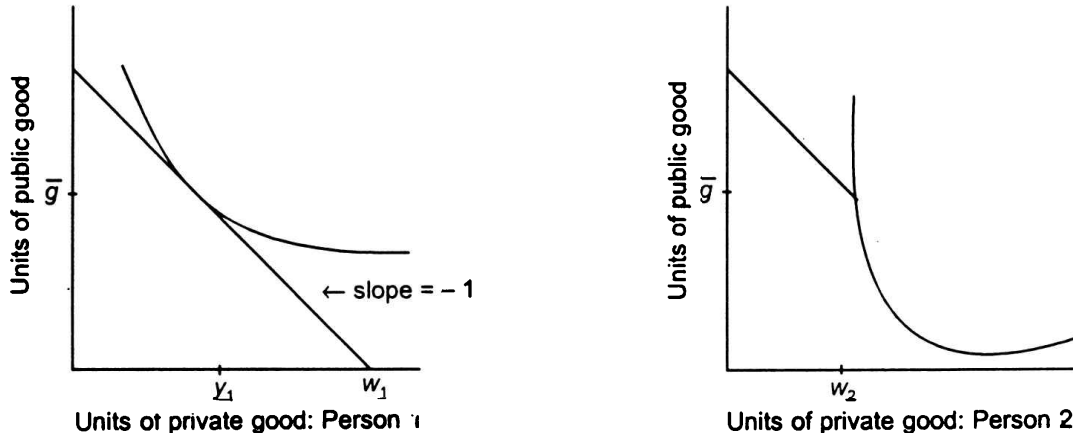


Fig. 24.3 The Free Riding Problem

Therefore, the main problem with the private provision of public goods is the problem of free riding. Those who enjoy the public good without paying for it will never signal their desire. As a result, sufficiently enough quantity of the public good may not be provided.

From the above discussion it is clear that public goods, in some way, should be publicly financed. Before we discuss alternative financing principles for

a public good, let us make a brief comparison between public and private goods. This will show why we need different types of provision principles for a public good. Purchasers of exclusive private goods can appropriate the benefits of those goods entirely for themselves. A cake bought by person *A* yields no benefit to person *B*. The resources that were used to produce the cake can be seen as contributing to *A*'s utility only and he is willing to pay whatever this is worth to him. The resources cost of a private good purchased by a person can be attributed to him. For a public good this cannot be the case. In purchasing a public good a person will not be able to appropriate all the benefits of the good. Since others cannot be excluded from benefitting from the good, society's utility obtainable from the resources employed for producing the public good will exceed the utility of the single person who pays for it. The resource cost cannot be attributed solely to the purchaser. But the potential purchasers do not incorporate in their expenditure decisions the benefits that their purchases have on others. Therefore, private markets may under-allocate resources to public goods. This means that alternative mechanisms are required to determine the provision of public goods. We discuss some such mechanisms below.

LINDAHL'S TAX SCHEME

A number of economists have examined how public goods might be provided publicly and financed through taxation. One approach relies on the assumption that individuals would agree to be taxed in exchange for the benefits that the public good provides. A very clear statement about how such an equilibrium might arise was provided by Lindahl (1919).

We assume again for simplicity that $C(g) = g$, so that the marginal cost of the public good is constant at 1. We shall let T_1 stand for person 1's total tax payment. When g units of the public good are provided, $T_1 = r_1 g$, where r_1 stands for person 1's tax share. Similarly, for person 2, $T_2 = r_2 g$.

Person 1's maximization problem takes the form

$$\max_{y_1, g} U_1(y_1, g) \quad (24.19)$$

such that

$$y_1 + r_1 g = w_1. \quad (24.20)$$

The optimization condition for this problem is that the marginal rate of substitution between the public and the private goods is r_1 , the tax share of person 1 (see Appendix 24.1 for a derivation). Formally,

$$MRS^1 = r_1. \quad (24.21)$$

The optimal amount of g that we obtain from Eq. (24.21) as function of r_1 and w_1 is consumer 1's demand function for the public good.

Similarly, person 2 maximizes $U_2(y_2, g)$ with respect to y_2 and g subject to his budget constraint $y_2 + r_2 g = w_2$, where r_2 is his tax share. (Clearly, $r_2 = 1 - r_1$.) In this case the optimizing condition becomes

$$MRS^2 = r_2. \quad (24.22)$$

Person 2's demand for the public good will depend on his tax share r_2 and initial wealth w_2 .

A Lindahl equilibrium is defined to be a vector of tax shares (r_1, r_2) and a level of output $\bar{g}$ for the public good, such that for each person i , when his tax share is r_i , his desired level of public good equals $\bar{g}$, where $i = 1, 2$. That is, $\bar{g}$ maximizes the utilities $U_1(y_1, g)$ and $U_2(y_2, g)$ of persons 1 and 2 subject to their respective budget constraints $y_1 + r_1 g = w_1$ and $y_2 + r_2 g = w_2$.

It is clear from the definition that a Lindahl equilibrium equates tax shares and marginal benefits. That is, what a person pays is connected to what he gets. Adding the left hand side (right hand side) of Eq. (24.21) with the corresponding side of Eq. (24.22), we have

$$MRS^1 + MRS^2 = r_1 + r_2 = 1 \quad (24.23)$$

since r_1 and r_2 are the tax shares of consumers 1 and 2. Given that the marginal cost of producing the public good is 1, Eq. (24.23) shows that Lindahl equilibrium output for a public good satisfies the Samuelson optimality condition.

Deriving a Lindahl equilibrium requires knowledge of the optimal tax shares of different persons. The government might ask individuals how much they are willing to pay for the public good. In responding to this question, individuals may feel that they should understate their true preferences fearing that they will have to pay what the goods is worth to them in the form of taxes. Therefore, the proper strategy for an individual is to understate true preferences with the expectation that others will bear the burden of paying for the good. Since no one can be excluded from enjoying the benefits of a public good once it is provided, the best strategy is to become a free rider. Therefore, the problem with the Lindahl tax mechanism is the problem of demand revelation. People will not reveal their true feelings about the public good.

VOTING

Since private provision of public good and the Lindahl tax scheme do not work very nicely, we must look elsewhere for an ideal public finance scheme. One mechanism can be voting. Voting is used as a social decision process in many institutions.

Since voting is not very interesting for a two-person case, we will assume that there are n persons, where n is odd. The voting principle that we are going to discuss here as a social decision criterion yields a positive result when n is even also. A discussion on this is beyond the scope of the book. The interested readers may look at Black's book (Black, 1958). Suppose that these n persons are voting on the amount of a public good, more precisely, on the magnitude of expenditure on some public good.

Each person has preferences over alternative levels of expenditure. He has a most-preferred level of expenditure and his valuation about other expenditure levels will depend on their closeness to the most preferred level. Decisions will be made by majority rule, that is, we will assume that policies will be adopted if they are favored by the next whole integer above $n/2$ voters or more. (Thus, if there are $n = 5$ persons, then a policy will be adopted if it is voted for by 3 persons or more.)

M. de Condorcet, a French social theorist, observed an important peculiarity of the majority voting systems in the 1780s. The Condorcet observation says that a majority voting may not yield an equilibrium but instead may cycle among alternative options. To illustrate the Condorcet paradox, which is popularly referred to as the paradox of voting, suppose that there are three voters (A , B and C) choosing among three policy options representing three levels of spending on a particular public good ($[l]$ low, $[m]$ medium, or $[h]$ high). Preferences of A , B and C among the three policy options are indicated by inequality signs as shown in Table 24.2. That is, the signs indicate that A prefers option l to option m , B prefers option m to option l and so on.

Table 24.2 Paradox of Voting

Voter	Spending on the Public Good			
	l	m	h	l
A	>	>	<	
B	<	>	>	
C	>	<	>	

Now, consider a vote between policy options l and m . Here l would win, since it is favored by A and C and opposed by only B . In a vote between m and h , m would win, again by 2 votes to 1. But if we consider a vote between l and h , then h would win and consequently social choices would cycle. (The social preference relation does not satisfy the transitivity condition.)

The paradox of voting is quite disturbing. One may therefore ask what restrictions on preferences will rule it out. A fundamental result established by Black (1958) showed that equilibrium voting outcomes can always occur in cases where voters' preferences are 'single-peaked'. To understand single-peakedness, we first illustrate in Fig. 24.4 the preferences of the three voters considered above by assigning hypothetical utility levels to options l , m and h . The utility levels are assumed to be consistent with the preferences recorded in Table 24.2. For person A , utility decreased consistently as the expenditure on the public good increased.

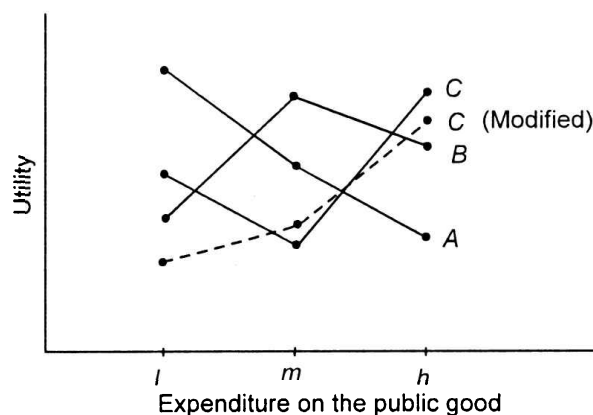


Fig. 24.4 Single-Peaked Preference and Median Voter

To explain this, note that each person cares about the level of the public good and also about the amount he has to contribute to it. A higher level of expenditure means a higher amount of the public good but individuals also have to pay higher taxes to finance the good. Since A 's utility decreases uniformly over expenditure levels, it seems he is more concerned about expenditures on the public good than the benefits he can derive from it. For person B , the utility of expenditure on the public good rises at first due to benefits of the public good but then it falls gradually due to the cost of providing it. Clearly, person C is more concerned about the benefits than the costs.

For persons A and B preferences are single-peaked. With single-peaked preferences, the utility of different expenditure levels should be either

monotone (uniformly decreasing or increasing) or it should rise until the most preferred point and then fall. Single-peakedness thus requires that utility should not go up, down and then up again. Person *C*'s preferences fail to satisfy this property and hence they are not single-peaked.

If each individual has single-peaked preferences, then social preferences revealed by majority voting will not exhibit intransitivity (Black, 1958). To see this suppose that *C* had preferences represented by the broken line in Fig. 24.4. In that case, option *m* would win since it would defeat both *l* and *h* by 2 votes to 1. That is, the choice of the 'median' voter (*B*) whose preferences are between the preferences of *A* and the (modified) preferences of *C* wins. The choice of the median expenditure means that (excluding the median person) one-half of the population is willing to spend more and one-half wants to spend less. This result is intuitively reasonable. In a vote between the median expenditure and any amount smaller than that, the former wins, since the median person and persons richer than him will vote in its favor. Similarly, if there is a vote between the median expenditure and any amount higher than that, then the median person and everyone poorer than him will favor the median expenditure and hence it wins.

There are certain advantages of the median expenditure system of public finance. First, it is relatively simple. It can be easily understood by the people reporting their desired expenditures on the public good and voting on these expenditures.

But there are certain disadvantages also. The median outcome means that $(n - M)/2$ persons want more and $(n - M)/2$ persons want less, where *M* is the median person. It does not state explicitly how much more of the public good is wanted. Since the Samuelson optimality condition takes this kind of information into account, voting may not lead to an efficient level of the public good. Furthermore, given certain agendas, people may vote against a preferred expenditure level at one stage in order to achieve a better outcome at a later stage. That is, people may have incentives to misrepresent their true preferences. As a consequence, the final outcome may be manipulated.

THE DEMAND-REVEALING TAX SCHEME

We have seen that private provision of a public good generally results in less than an optimal amount of the public good. Voting also may not produce an efficient level of output. People may not reveal their exact demands for a public good under the Lindahl tax scheme. We therefore return to our search

for an ideal tax scheme. This scheme should: (i) lead to the optimal output for the public good; and (ii) provide no incentive for the individuals to misrepresent their demands for the public good. Because of this latter property, this public finance rule is called the demand-revealing scheme. It was proposed independently by Clarke (1971) and Groves (1976).

We begin our discussion by assuming that individual utility functions are of quasilinear type. That is, person i has a utility function of the type $U_i(y_i, g) = y_i + V_i(g)$, where $i = 1, 2, \dots, n$. We do not impose any restriction on n , that is, n may be odd or even. Let us think of an association (may be a municipality, corporation or a public good board) that is considering supply of a public good, which has constant marginal cost of 1. (We assume that the public good cost function $C(g)$ equals g .) The association asks each person about his feeling for the public good. Let us suppose that these feelings are represented by V_i functions. We will soon establish that nobody will be able to gain by lying, so the true V_i functions will be reported. But for now when i reports his V_i function, it should be viewed with suspicion.

The association then employs the Samuelson optimality condition to solve for an amount of the public good. That is, assuming that Δg is sufficiently small, it solves for a value of g that satisfies the equation.

$$\sum_{i=1}^n \frac{V_i(g + \Delta g) - V_i(g)}{\Delta g} = 1. \quad (24.24)$$

This exercise is equivalent to the maximization of aggregate net benefit from the public good $\sum_{i=1}^n V_i(g) - g$. That is, the association maximizes the expression

$$\sum_{i=1}^n V_i(g) - g. \quad (24.25)$$

If V_i functions are reported correctly, the value of g that maximizes Eq. (24.25) is the optimal level for the public good. Let us denote the value of g that solves (24.24) (or, maximizes Eq. [24.25]) by $\bar{g}$.

The association sets total tax payment T_i 's of different individuals. Assume that it does so according to the following rule

$$T_i = \bar{g} - \sum_{j \neq i}^n V_j(\bar{g}). \quad (24.26)$$

That is, person i 's tax T_i equals the total cost of the public good less the aggregate utility the persons other than i enjoy from the public good.

We now investigate whether under this tax scheme the individuals have incentives to reveal their V_i functions truly. Given that the association wishes to produce $\bar{g}$ amount of the public good, person i maximizes $V_i(\bar{g}) + y_i$ subject to $y_i + T_i = w_i$. That is, he wants to maximize $V_i(\bar{g}) + w_i - T_i$, or $V_i(\bar{g}) - T_i$. Substituting the value of T_i from Eq. (24.26) we see that person i wants to maximize the following expression for his after-tax utility

$$V_i(\bar{g}) - [\bar{g} - \sum_{j \neq i}^n V_j(\bar{g})]. \quad (24.27)$$

Person i knows that if he reports his V_i function correctly then the association will choose $\bar{g}$ to maximize

$$\sum_{i=1}^n V_i(g) - g = \sum_{j \neq i}^n V_j(g) + V_i(g) - g = V_i(g) - [g - \sum_{j \neq i}^n V_j(g)]. \quad (24.28)$$

It thus turns out that if person i reports his V_i function truly, the association will attempt to maximize what i wants to maximize. That is, the association's goal coincides with that of i and hence i does not gain anything by lying. Observe that this holds true irrespective of what other persons ($j \neq i$) might do—person i maximizes his utility by reporting V_i correctly; whether V_j 's ($j \neq i$) are true or false is irrelevant. Telling the truth is the optimal strategy for each person no matter what the other persons do. That is, truth is a dominant strategy here. Further, as we have stated earlier, since the demand-revealing tax scheme satisfies the Samuelson optimality rule it generates the optimal output level.

However, the procedure we have described so far for collecting taxes has a serious disadvantage: tax collections may not cover the proposed expenditure $\bar{g}$. To see this, suppose that in a four-person society for a bridge to be constructed we need to incur an expenditure of $\bar{g} = \text{Rs. } 1,500$. Assume that $V_1(\bar{g}) = 0$, $V_2(\bar{g}) = 700$, $V_3(\bar{g}) = 300$ and $V_4(\bar{g}) = 600$. Then $T_1 = -100$, $T_2 = 600$, $T_3 = 200$ and $T_4 = 500$. Negativity of T_1 means that person 1 will get a subsidy rather than paying a tax. We note that the total of taxes equals only Rs. 1,200, which is less than the total expenditure of Rs. 1,500.

It is therefore necessary to ensure that the total tax proceeds will at least cover the expenditure on the public good. Formally we require that

$$\sum_{i=1}^n T_i \geq \bar{g}. \quad (24.29)$$

It is evident that if we add a constant term that does not depend on the evaluation function V_i or $\bar{g}$ to person i 's tax function, then person i 's incentive to tell the truth remains unaltered. Let us consider the expression which is independent of V_i and $\bar{g}$:

$$m_i = \max_g \sum_{j \neq i}^n \left[V_j(g) - \frac{g}{n} \right].$$

The term $[V_j(g) - g/n]$ is person j 's net benefit from the public good amount g when he is required to pay the equal $(1/n)$ th share of the expenditure on g . Then m_i is the maximal value of the sum of net benefits of all persons except i . This maximal value of the aggregate net benefits is calculated when i has not expressed his demand for the public good.

In contrast the term

$$a_i = \sum_{j \neq i}^n \left[V_j(\bar{g}) - \frac{\bar{g}}{n} \right] \quad (24.30)$$

can be regarded as the aggregate net benefit for all persons except i when i expresses his demand for the public good. Hence we can interpret $m_i - a_i$ as the loss person i imposes on others by revealing his demand for the public good.

Now, $m_i - a_i + \bar{g}/n$ equals

$$\bar{g} - \sum_{j \neq i}^n V_j(\bar{g}) + m_i = T_i + m_i = \bar{T}_i \text{ (say)}. \quad (24.31)$$

Thus $\bar{T}_i$ can be regarded as the new tax for person i which is obtained by adding m_i to his old tax T_i . In view of the fact

$$\begin{aligned} m_i &= \max_g \sum_{j \neq i}^n \left[V_j(g) - \frac{g}{n} \right] \\ &\geq \sum_{j \neq i}^n \left[V_j(\bar{g}) - \frac{\bar{g}}{n} \right] \end{aligned}$$

$$= \alpha_i, \quad (24.32)$$

we have $\bar{T}_i = m_i - \alpha_i + \bar{g}/n \geq \bar{g}/n$. Therefore $\sum_{i=1}^n \bar{T}_i \geq \bar{g}$, that is, $\bar{T}_i$'s fulfil the requirement of Eq. (24.29).

Therefore we have developed a taxation scheme for financing the public good under consideration with the following properties: (i) everyone reports his valuation about the public good correctly, (ii) the optimal quantity $\bar{g}$ of the public good is produced, and (iii) the total tax revenues are sufficient to pay for $\bar{g}$.

Despite the nice properties of the taxation scheme developed above, it has some disadvantages. There is always the possibility of a budget surplus, that is, we may have $\sum_{i=1}^n \bar{T}_i > \bar{g}$. In such a case we have $\sum_{i=1}^n y_i + \bar{g} < \sum_{i=1}^n w_i$. Thus, some real wealth remains unused. Therefore, the distribution $(\bar{g}, y_1, \dots, y_n)$ cannot be Pareto optimal, even though $\bar{g}$ satisfies the Samuelson optimality principle. Several economists argue that it might be sensible to donate the surplus to another country. Another disadvantage of this taxation scheme is that it only works with quasilinear preferences.

Example 24.1

Suppose out of 20 persons in a society each of 10 persons is ready to pay Rs. 10 for each extra unit of a mosquito control program. On the other hand, each of the remaining 10 persons is willing to pay Rs. 20 for every additional unit of the program. The cost function of the project is $C(g) = g^3$. We want to determine the optimal level of mosquito control.

The sum of the marginal rates of substitution between the public good and the private good across individuals is given by $10 \times 10 + 20 \times 10 = 300$. The marginal cost of the public good is $3g^2$. Solving the equation $3g^2 = 300$ we get $g = 10$, which is the optimal level of mosquito control. (Note that the equation $3g^2 = 300$ yields two roots $g = 10, -10$. We ignore the negative root because the desired quantity is positive.)

Applied Example 24.1 'Vidyasagar Setu' as a Public Good

The 'Vidyasagar Setu' (Vidyasagar Bridge) on the Hooghly river near Calcutta is a public good. Because of the non-exclusive nature of the bridge, no charge can be made for its use. The question then arises: how can one best finance current expenditure and capital costs? Economic efficiency dictates that the

price charged should equal marginal cost, and since the marginal cost of supplying this service is zero, then no price should be charged. However, this navigational path is more than just a safety device free to all; it aims at providing a service that will allow vehicles to travel by the shortest possible route.

The State Government of West Bengal has decided to impose some duty on the vehicles passing over the bridge. This toll is small for two-wheelers, private cars and taxis, and bigger vehicles are required to pay higher amounts. Since any vehicle that uses the bridge is required to pay some charge, there is no scope of free riding. Given that these payments allow the vehicles to use the bridge, it can be argued that the payment reflect willingness-to-pay of the vehicle owners. Otherwise, they would not use the bridge.

Appendix 24.1

THE SAMUELSON OPTIMALITY CONDITION (EQ. [24.15])

We need to solve the problem

$$\max_{y_1, y_2, g} U_1(y_1, g)$$

such that $U_2(y_2, g) = U_2^0$, $y_1 + y_2 + C(g) = w_1 + w_2$.

Consider the Lagrangian:

$$\mathcal{L} = U_1(y_1, g) - \lambda_1[U_2(y_2, g) - U_2^0] - \lambda_2[y_1 + y_2 + C(g) - w_1 - w_2].$$

Differentiating the Lagrangian partially with respect to y_1 , y_2 , g , λ_1 and λ_2 and setting the derivatives to zero, we get

$$\frac{\delta \mathcal{L}}{\delta y_1} = \frac{\delta U_1(y_1, g)}{\delta y_1} - \lambda_2 = 0,$$

$$\frac{\delta \mathcal{L}}{\delta y_2} = -\lambda_1 \frac{\delta U_2(y_2, g)}{\delta y_2} - \lambda_2 = 0,$$

$$\frac{\delta \mathcal{L}}{\delta g} = \frac{\delta U_1(y_1, g)}{\delta g} - \lambda_1 \frac{\delta U_2(y_2, g)}{\delta g} - \lambda_2 \frac{dC(g)}{dg} = 0,$$

$$\frac{\delta \mathcal{L}}{\delta \lambda_1} = U_2(x_2, g) - U_2^o = 0,$$

$$\frac{\delta \mathcal{L}}{\delta \lambda_2} = y_1 + y_2 + C(g) - w_1 - w_2 = 0.$$

We rewrite the third equation as

$$\frac{1}{\lambda_2} \frac{\delta U_1(y_1, g)}{\delta g} - \frac{\lambda_1}{\lambda_2} \frac{\delta U_2(y_2, g)}{\delta g} = \frac{dC(g)}{dg}. \quad (24.33)$$

From the first and second equations it follows that

$$\lambda_2 = \frac{\delta U_1(y_1, g)}{\delta y_1},$$

$$\frac{\lambda_2}{\lambda_1} = - \frac{\delta U_2(y_2, g)}{\delta y_2}.$$

Substituting the above expressions for λ_2 and λ_2/λ_1 in Eq. (24.33) we get

$$\frac{\delta U_1(y_1, g)/\delta g}{\delta U_1(y_1, g)/\delta y_1} + \frac{\delta U_2(y_2, g)/\delta g}{\delta U_2(y_2, g)/\delta y_2} = \frac{dC(g)}{dg}$$

which is just

$$MRS^1 + MRS^2 = MC(g).$$

THE LINDAHL TAX SCHEME

We need to solve the problem

$$\max_{y_1, g} U(y_1, g)$$

subject to $y_1 + r_1 g = w_1$.

Consider the Lagrangian:

$$\mathcal{L} = U_1(y_1, g) - \lambda(y_1 + r_1 g - w_1)$$

where λ is the Lagrange multiplier. Differentiating $\overline{\mathcal{L}}$ partially with respect to y_1 , g and λ and equating the derivatives to zero, we get

$$\frac{\delta \mathcal{L}}{\delta y_1} = \frac{\delta U_1(y_1, g)}{\delta y_1} - \lambda = 0,$$

$$\frac{\delta \mathcal{L}}{\delta g} = \frac{\delta U_1(y_1, g)}{\delta g} - \lambda r_1 = 0,$$

$$\frac{\delta \mathcal{L}}{\delta \lambda} = y_1 + r_1 g - w_1 = 0.$$

From the first equation we have

$$\lambda = \frac{\delta U_1(y_1, g)}{\delta y_1}$$

which when substituted in the second equation gives

$$\frac{\delta U_1(y_1, g)}{\delta g} = r_1 \frac{\delta U_1(y_1, g)}{\delta y_1}.$$

From the above equation we have

$$\frac{\delta U_1(y_1, g)}{\delta g} \bigg/ \frac{\delta U_1(y_1, g)}{\delta y_1} = r_1,$$

or, $MRS^1 = r_1,$

which is the desired result.

EXERCISES

Multiple Choice

1. Which of the following is an example of a public good?
 - a) A cigarette.
 - b) A bus.
 - c) A lighthouse.
 - d) An automobile.
2. Public goods can lead to market failure because
 - a) They create external costs.
 - b) They create social costs.
 - c) They cannot be sold easily in markets.

- d) They cannot be paid for through taxes.
3. Which of the following best describes the free rider problem?
- a) Some private goods create external benefits for those who have not paid.
 - b) Some people think that the environment is a free resource and therefore abuse it.
 - c) Some goods cannot be sold in markets because the benefits they confer are available to all, whether they have paid or not.
 - d) Your neighbor gets pleasure from observing your flower garden.
4. Which of the following statements is not correct about the demand-revealing tax scheme for a public good?
- a) Honesty is a dominant strategy, so everybody tells the truth.
 - b) The optimal quantity of the public good is produced.
 - c) The tax revenues are always sufficient to pay for the optimal quantity of the public good.
 - d) The outcome showing the optimal quantity of the public good and quantities of the private good demanded by different individuals is Pareto efficient.
5. Which of the following is not an example of a public good?
- a) Education.
 - b) A park.
 - c) A storm warning system.
 - d) A bus.

Quantitative Problems and Questions

1. Suppose that there are only two individuals in a society. The demand curve of person 1 for mosquito control is given by $d_1 = 100 - p$. For person 2 the demand curve for mosquito control is $d_2 = 200 - p$.
- a) Suppose mosquito control is a public good. What would be the optimal level of mosquito control if it could be produced at a constant marginal cost of Rs. 50 per unit?
 - b) If mosquito control were a private good, how much might be produced?
 - c) Can you explain the reason for the two answers being different?

2. Suppose that the production possibilities frontier for an economy that produces one public good (g) and a private good (X) is given by $X^2 + 100g^2 = 5,000$. This economy consists of 100 individuals. The i th individual's utility function is given by $\sqrt{gx_i}$, where x_i is individual i 's share of private good production ($x_i = X/100$).
 - a) If the market for X and g were perfectly competitive, what levels of those goods would be produced? What would be the typical individual's utility in this situation?
 - b) What are the optimal production levels of g and X ? What would be the typical individual's utility here? How should individuals be taxed to finance the production of g ?
3. Explain how the free rider problem leads to market failure.
4. Does the demand-revealing, that is, the Clarke-Groves tax scheme result in a Pareto efficient allocation? Does this tax scheme produce an optimal amount of output?
5. Prove or disprove the following.
 - a) The optimal level of a public good is given by the intersection of the collective demand curve with the marginal cost curve, where the collective demand curve is the vertical sum of individual demand curves.
 - b) There is no difference between the construction of the market demand curve for a private and a public good.
 - c) Because of the free rider problem, public goods are likely to be underproduced.
 - d) Consumption of additional units of a public good involves positive marginal social costs of production.
 - e) The majority voting principle generates the optimal quantity of a public good.
6. An economy is populated with n individuals, each with a utility function of the form $U(g, x_i) = \log g + x_i$, where g is the quantity of the public good and x_i is the level of the competitive private good demanded by person i . Each person is obliged to pay an equal share $1/n$ of the public good.
 - a) What is the optimal amount of the public good?
 - b) What is the amount of the public good produced when we adopt the majority voting principle? In this case you may assume that n is odd.

7. Suppose a discrete public good is to be provided to a society consisting of two persons. Each person i states a bid, a_i . If the sum of the bids covers the cost of providing the good, then the good will be provided and each person pays his bid amount. Otherwise the good is not provided and nobody pays anything. What is the efficient equilibrium? Can there be any other equilibrium?
8. Suppose there are three persons in a society who vote on whether the government should undertake a specific project. Let the benefits of a particular project be Rs. 150, Rs. 140, Rs. 50 for persons I, II and III respectively.
 - a) If the project costs Rs. 300 and these costs are to be shared equally, would a majority vote to undertake the project? What would be the net benefits to each person under such a scheme? Would total net benefits be positive?
 - b) Suppose that the project costs Rs. 375 and again the costs are to be shared equally. Now would a majority vote for the project and would total net benefits be positive?
9. Why is a storm warning system a public good? Why does it fall into this category?
10. There is a bridge connecting two sides of a river. The demand for trips from one side of the bridge to the other is given by $t = 20 - 0.001x$, where x is the trips per day and t is the time per trip in hours. The cost per trip is $2 + 0.001x$. There are no other travel costs and the value of time is Re. 1 per hour for all trips.
 - a) Determine the optimal number of trips.
 - b) Determine the amount of tax that a person should pay so that optimal utilization of the bridge is ensured.
11. Robinson Crusoe and Friday are required to decide as to how many hours each of them should work to build a fort to protect themselves from cannibals. Show how the optimum amount of defence is produced under truthful revelation about their working hours.

25

Economics of Information

The information structure of a market situation gives us a precise meaning to the concept of imperfect or incomplete or imprecise information. Let us consider the case of uninformed buyers of a commodity examined in Stigler's 1961 classic article. The potential buyers of the commodity wish to know which shop in the town sells it at the lowest price. In this case information would be complete if each potential buyer knew which shop charges which price. Then the cheapest shop would be known with certainty. However, given that the buyers do not know which shop charges the lowest price, we say that the information structure is asymmetric, in that complete information prevails on one side of the market (e.g. on the selling side) while incomplete information characterizes the other side (viz. the buying side). Analogously, some agents may have information which other agents on the same side of the market do not have.

Asymmetry of information may also arise in the context of the quality of items being sold in the market. For instance, it may be difficult for a firm to know the productivities of its employees, though an employee may be having some concrete idea about his work efficiency. A buyer of a used car may not be knowing whether the car is good or not. In contrast the seller of the car may have a good idea about the quality of the car.

There are some important distinctions between 'economics of uncertainty' and 'economics of information', though both study the implications of incomplete information and of the resulting uncertainty. In their survey on the two fields Hirschleifer and Riley (1979) suggest that decision making under uncertainty is concerned with 'event uncertainty', while the topic of the present chapter is 'market uncertainty'. The former focuses on its implications for the decision making of an individual consumer or firm. It may go further to derive consequences for the market as a whole, but always under the assumption that uncertainty is the same for all market participants. It is therefore natural for theories of insurance decisions or investment decisions to claim competitive markets in which there is no strategic interplay between the agents. In contrast when the focus is on the economics of information, strategic interactions become essential and the application naturally concerns markets characterized by monopolistic competition or oligopoly.

In the case of market uncertainty agents are unsure about the supply-demand offers of other economic agents. As stated earlier, buyers may be unsure about the prices charged or the qualities offered by the seller. Or sellers are unsure about the prices buyers are ready to pay. Sellers may be producers of consumer goods, but they may also be job applicants who wonder what salary to accept, or insurance companies wondering about the risk they are covering, or bankers trying to ascertain the solvency of potential money borrowers. Or sellers may be oligopolists who are not well-informed of the prices or the production rates or the costs of production of their competitors. Therefore, informational asymmetry characterizes market uncertainty as opposed to event uncertainty.

LIMITED INFORMATION ABOUT QUALITY

In many markets consumers do not have any information about how quality varies across brands. Examples of markets where consumers are often not sure about quality include those for professional services (doctors, electricians, lawyers etc.), processed foods, electronic products, used goods. In these markets asymmetry of information exists: sellers know the quality of the good, but buyers do not. If consumers have limited information about a product's quality, one of two undesirable results occur: either equilibrium does not exist, or if equilibrium exists, the quality produced is different (usually lower) than in a world of perfect information. For example, often only the lowest-quality products are produced. Thus, resources are used inefficiently compared to markets with perfect information.

THE MARKET FOR LEMONS

We now turn to a situation where buyers are imperfectly informed about the quality of the seller's product. The product is a desirable good such that buyers will discover its quality only by using it. Probably the best known study of this kind is Akerlof's (1970) classic case of the market for lemons (bad cars). Akerlof shows that where sellers have perfect information and consumers have extremely limited information, a market may not exist, or only the lower quality product may be sold.

Consider the market for cars. Neither the seller nor the buyer of a new car knows if it is a good car (a plum) or a bad one (a lemon) that will break down repeatedly. An owner learns of the quality of a car after using it for a while.

Now consider the market for used cars. In this market there is asymmetry of information between buyers and sellers. The sellers, who are the current owners, know the quality of the cars, but the buyers do not. At most, a potential buyer knows the probability of getting a good car. As a result, both good and bad used cars must sell for the same price, since buyers cannot separate them.

Bad cars are overvalued and good cars are undervalued in this market. For example, suppose consumers believe that half the used cars in the market are lemons worth Rs. 20,000 and the other half are plums worth Rs. 40,000. Suppose also that consumers are risk neutral, that is, they are indifferent between having Re. 1 and having something that has a 50% probability of being worth nothing and a 50% probability of being worth Rs. 2. Then the value to a typical customer of a randomly selected car will be Rs. 30,000 ($= \text{Rs. } 20,000 \times 0.5 + \text{Rs. } 40,000 \times 0.5$). Therefore, a buyer who thinks that a car might turn out to be good desires to pay more than the value of a lemon ($\text{Rs. } 20,000 < \text{Rs. } 30,000$), but at the same time he thinks that it might be a lemon and hence does not want to pay the full value of a plum ($\text{Rs. } 30,000 < \text{Rs. } 40,000$).

As a consequence, owners of a good used car do not have an interest in selling it because they cannot get a good price for it. Most used cars traded must be lemons. Because only lemons are sold, buyers know that they are getting lemons and will pay only the price of a lemon. There is no market for plums. There is thus an analogy with Gresham's Law, which states that 'bad money drives out good money'. (This law was formulated by Sir Thomas Gresham (1519–79).) According to this law, if two monetary measures circulate at values different from their true value, then only the lower-value measure survives. For example, if gold and silver coins are used, and their face values are lower than their metal values, then the more undervalued coins are melted down first.)

This example can be easily extended to many qualities of cars. But the result remains the same. By applying the same sort of reasoning we can see that the lowest-quality cars eventually drive all other cars out of the market. The source of this market failure is an externality problem generated by the selling side. When a seller decides to sell a lemon he affects the buyer's perceptions of the quality of the average car on the market. This reduces the price that the buyers are willing to pay for an average car and hence plums are undervalued. That is, the owners of the plums will find the market price offered by ignorant buyers insufficient to induce them to sell their cars, whereas owners of lemons will accept such offers readily. Thus the sellers cannot fully capture the value of producing higher-quality products. The market failure is a consequence of this externality.

ADVERSE SELECTION

The type of informational externality discussed in the previous section that impedes the functioning of the used car market is referred to as adverse selection. This sort of problem also arises in markets for insurance and for home repair.

In an insurance market, insurers are able to observe only the average probability that the policy holders will make a claim and not the probability of a particular policy holder doing so. Healthy senior persons often have problems buying medical insurance. The people who want to buy insurance the most are the ones who are likely to need it the most and therefore why can't insurance companies merely adjust the price of a policy to match the risk? The reason appears to be the same kind of adverse selection as in the used car example: as the price of a policy rises only the worst risks are interested in buying the policy. Price based on average probability in the group of worst risks is a misleading indication. The insurance company will not get an unbiased selection of customers: rather they will get an adverse selection. As the price level rises (for example, with age), only those who believe that they have increasingly high risks buy the insurance. If the individual can better determine his health condition than can the insurance company, the insurance company sells a disproportionate number of policies to the least healthy persons because for them it is a good buy. Individuals with small risks will not buy because it is too expensive given what they know. Since only high risk individuals buy it, costs are high. Adverse selection arises here because individuals are in a better position to assess the risks they face than an insurance provider can. A sample of people considered by Akerlof reports that although 63 percent of people aged

between 45 and 54 had hospital insurance, only 31 percent over 65 had such an insurance.

Similarly, suppose some good carpenters use high-quality materials and bad carpenters use low-quality materials. If homeowners do not know the honesty of a carpenter for many years (for instance, bad materials break down in 6 years and good materials break down in 11 years) and pay both types the same amount, then bad carpenters drive out the good ones, whose costs are higher.

Thus, adverse selection arises in situations where one side of the market cannot observe the type or quality of the goods on the other side of the market. That is why it is also called a hidden information problem.

Although not in all markets with asymmetric information only the lowest-quality product is sold, there is always an inefficiency in these markets compared to a world with perfect information: quality levels are very low. These inefficiencies generate from a discrepancy between private and social returns. Unfortunately, these inefficiencies cannot be corrected by government intervention, since providing perfect information is often prohibitively expensive.

In an equilibrium with imperfect information, the marginal firm faces a market price that equals its opportunity cost, such as the value of the used car to the firm. That is, the marginal seller is indifferent between selling and not selling. Since the opportunity cost of better-quality units is higher than the market price, the quality of the marginal seller's product is the highest quality observed in the market.

In a market of this type since the buyers cannot distinguish good products from bad products, in equilibrium the private value (the amount that a buyer is willing to pay) of the highest quality product is equal to the average value. That is, if consumers are able to recognize the difference in quality, they are willing to pay more for a better-quality product. The high-quality seller must split the benefits of his product with a low-quality product by raising the average price. If firms cannot capture the value of producing higher-quality goods (an externality), then they have an incentive to produce goods of low quality.

MORAL HAZARD

Another informational externality that arises in the insurance industry is moral hazard. The moral hazard problem arises when one can influence the occurrence of an event against which he is insured, but the seller of the insurance policy cannot monitor or detect this action. In other words, moral hazard occurs when insurance coverage against some event may actually change the expected value

of loss. It is well known that coverage of certain types of medical procedures such as psychiatric or dental care increases potential usage substantially. Once a person is covered by insurance, his cost for such services becomes zero, so that it is not surprising that demand increases. In order to control this, many insurance policies contain cost-sharing provisions so that patients directly share some percentage (usually small) of the expenses their treatment entails.

As a second example let us consider the case of a household desiring to purchase insurance against the loss arising from theft. The price at which the insurance can be offered depends on the probability of the loss occurring. If the insured household is able to influence the probability of loss, say by varying expenses on preventive activity, then the moral hazard problem arises. After all, if loss takes place then the household gets reimbursement from the insurance company. In the case of complete-reimbursement the household has no incentive to take any preventive activity at all. This lack of incentive to take protective measures or care is called moral hazard. Analogously, fires of suspicious origin often take place in heavily insured buildings, possibly because of the desire to claim insurance proceeds. If it is possible to observe the amounts of care, then the insurance company can base its rates on these amounts. In practice, insurance companies give rates to households and business depending on the preventive measures they take against theft/fire. We note the implicit trade-off involved in all these cases: too much insurance means too little care and too small insurance means adequate care.

The moral hazard problem also arises in the context of guarantees or warranties provided by sellers of different products like consumer durables (e.g. refrigerators, vacuum cleaners, cars, TV sets, radios etc.). In this case the terms of warranty may create an incentive for the buyer to give the product insufficient maintenance or to handle it carelessly. This is a first moral hazard problem. A similar problem arises on the seller side, when the actual quality is not observable by buyers. A seller may not maintain high quality. There is thus a double moral hazard. It is double, because information is imperfect on both sides: buyers are imperfectly informed about product quality, but the seller are imperfectly informed about the way in which buyers will subsequently use their product. If the buyer of the product has to pay a part of the costs of repair, then it is possible to limit abuses on the buyer's side. A seller may provide information in the form of standards and certification to demonstrate quality. Standards are established for a good by defining a scale for evaluating that product. Certification means that a particular product has been found to meet a standard. For example, the mark of the Indian Standard Institution on a consumer good in India certifies that it is a good quality product.

Thus, moral hazard arises in situations when one side of the market cannot observe the actions of the other. That is why it is sometimes referred to as a hidden action problem.

SOLVING THE PROBLEM: CASES OF EQUAL INFORMATION

The problem of chasing out high-quality items by low-quality items is a consequence of asymmetry of information. Under symmetry of information, markets may exist. The symmetry of information can be of two types: either both sides know the quality or neither knows.

If both buyers and sellers know the quality of used cars, then price will reflect quality. There will be no external effect. We then have competitive markets with perfect information.

If sellers and buyers are equally ignorant about the quality (as in the case of new cars), then high and low-quality products are sold at a price that reflects an average of two qualities. Therefore, with symmetric but imperfect information markets may exist. If information can be obtained at relatively low costs, consumers may obtain the information and the market works smoothly. If costs are rather high, information may not be collected and inefficiency will arise.

Information may be obtained in a number of ways. First, as we have stated earlier, sellers of high-quality goods may supply information by providing credible guarantees and warranties. Second, a manufacturer may use his reputation for signalling that his products are of high quality. A seller supplying high-quality products may expect repeat purchase by a consumer and therefore may not provide low-quality products. Third, an expert may be able to supply reliable information to a consumer. A car mechanic will be able to tell the quality of a used car. We have already stated that an alternative way of providing information about quality is through certification.

Observe that information is a public good; a good, if it is supplied to anyone, can be supplied to others at no extra cost. It is possible that no firm can profitably provide it because the supplier may not capture all the benefits. Since information is a public good, a firm may not provide any information with the hope that some other firm will bear the burden of providing information. That is, there may be a free rider problem. Therefore, in some cases information may not be available at all. Information may be costly. A consumer will not collect it beyond the point where the marginal benefit from its collection equals the marginal costs of collecting it. Employing a mechanic to know the quality of a car may be costly. Information may not be reliable. Consumers may not

have sufficient knowledge to make use of available information for judging quality. For instance, many intelligent persons may not be competent enough to judge the quality of various computers.

But this is not to say that information, if available at relatively low cost, will not be useful. For instance, as we have stated, accurate information about the quality of used cars ensures that there will be no inefficiency and higher prices are charged for high quality cars. Since information may take the form of a signal; in the next section we examine whether signalling enables smooth functioning of a market.

SIGNALLING

We saw that in the model of the used car market asymmetric information about the quality of cars generates problems in the market. Since bad used cars drive out good used cars, the owners of good used cars may try to convey the fact that they have good cars to sell. That is, they would like to convince the potential buyer about the quality of the cars through some signal. One sensible signal here can be the provision of a credible warranty by the seller to the buyer of a car. This can be a promise to pay the buyer a certain amount if the car turns out to be bad. By providing consumers with equal information, such a seller is able to charge a higher price that reflects the higher quality of his product.

The purpose of this section is to explain how such signalling can be used as a *self-selection mechanism* that generates information endogenously. The endogenous nature of this self-selection mechanism results from the circumstances that the other side of the market (which receives the signal) uses this mechanism for its own advantage in order to make the appropriate choice. To illustrate this, let us consider some examples of signals. Producers signal through advertising, job applicants invest in education in order to be able to show diplomas, borrowers signal their low risk by refusing to pay high interest rates and so on. Such signals can be used to choose high-quality products, the more productive employee, the low-risk borrower etc. (These examples also explain why a signal can be regarded as a self-selection mechanism: advertising shows that the product is worth promoting, because advertising is not advantageous for bad-quality products. Even if a diploma says nothing about quality, it shows at least that the person worked hard to obtain it. The willingness to pay low interest rates may generate from the borrower's goodwill that he is not a defaulter, he follows the terms and conditions imposed by the lender.) If both sides of the market are satisfied with the working of the mechanism, some kind of equilibrium emerges.

For simplicity we will analyze a labor market model examined by Spence (1974). Spence defines signals as manipulable attributes or activities that convey information. For job applicants signals are diplomas, previous work experience, and so on, which the employer has to interpret. There are other attributes that cannot be altered, such as race or sex. These are not signals; they may be called *indices*. Individuals also have different abilities which are innate. Some people are more intelligent, more business oriented, or whatever, than others. The basic assumption is that applicants know their abilities and use signals to inform employers about this in a credible way.

Suppose there are two types of workers: Group I has a marginal product of m_1 and group II has a marginal product of m_2 , where $m_2 > m_1$ because of differences in innate abilities. The employers know (from past experience, say) that the proportion of people who have a marginal product of m_1 is q , so that the remainder $(1 - q)$ has a marginal product of m_2 .

The labor market is assumed to be perfectly competitive. The employer's production function is linear, so that the total output produced is $m_1 Lq + m_2 L(1 - q)$, where L is the total number of workers.

If m_1 and m_2 were known, the employer would pay each person his marginal product. But in the absence of signalling the employer is unable to distinguish between the two groups and will thus pay everybody the expected marginal product $\bar{w} = qm_1 + (1 - q)m_2$. But then the members of group I receive a wage that is higher than their marginal product ($\bar{w} > m_1$) and members of group II are hurt. The gain to the former is $qm_1 + (1 - q)m_2 - m_1 = (1 - q)(m_2 - m_1)$, which increases as their proportion in the population falls. The loss to the latter is $m_2 - qm_1 - (1 - q)m_2 = q(m_2 - m_1)$ which decreases as their proportion in the population increases. The able members of group II have an interest to convince the employer that they deserve more. However, since the unable workers could also pretend to belong to group II and the employer has no way of checking this, more information is necessary to distinguish between them.

The message from group II workers would be convincing if it were translated into their having better education than group I workers. The employer pays attention to such a signal since it allows him to compete for talented workers and to avoid the costs associated with the training of workers. To model this situation, Spence assumes that greater ability implies lower cost of education. 'Education' is measured by a composite index of years and performance in education and is denoted by e . Let $e_1(e_2)$ be the amount of education earned by group I (II) workers. The cost of education of group I workers is $c_1 e_1$ and that for group II workers is $c_2 e_2$. Education becomes a signal of capability because it is costly and because this cost is negatively correlated with capability.

Faced with this situation the employer will associate an expected productivity to each education level and offer a corresponding wage. How will the potential employees react to the announcement of such a wage schedule? Equilibrium in this model depends on a worker's decision of how much education to acquire and an employer's decision of how much to pay workers with different education levels.

To determine equilibrium, suppose an employer who faces the two groups believes that if the level of education e is smaller than $\bar{e}$, then productivity is m_1 with probability 1 and that if e is greater than $\bar{e}$, then productivity is m_2 with probability 1, where $\bar{e}$ is just some value of the composite index for education. The wage schedule offered is: wage rate $w = m_1$ for $e < \bar{e}$ and $w = m_2$ for $e \geq \bar{e}$, if the market is competitive.

Will this wage schedule, given the costs of education, provide an incentive for members of group II to acquire at least education level $\bar{e}$, and for members of group I to be satisfied with an education smaller than $\bar{e}$, so that the two groups are effectively separated and innate capabilities are correctly revealed? To answer this we must analyze the choice of education of the two groups.

The optimal choice of e for any member of either group is to choose that value of e for which the difference between the wage increase $(m_2 - m_1)$ and the cost of education is maximum. This is so, because a person is acquiring education with the expectation that education will increase wage rate. Note that now since the employer can distinguish between workers with different capabilities, he will pay either m_1 or m_2 depending on marginal productivity. Since m_2 and m_1 are fixed, we are simply minimizing the cost of education. Note that for members of group I, the difference $(m_2 - m_1) - c_1 e$, when $e < \bar{e}$, is maximized if $e = 0$. But a member of group I will choose a zero educational level only if his benefit is less than the cost of attaining the education level $\bar{e}$, that is, $(m_2 - m_1)$ is less than $c_1 \bar{e}$, or, $(m_2 - m_1)/c_1 < \bar{e}$. Therefore, his decision rule is as follows

$$\text{Choose } e = 0 \text{ if } \frac{(m_2 - m_1)}{c_1} < \bar{e}. \quad (25.1)$$

The members of group II maximize $(m_2 - m_1) - c_2 e$ subject to $e \geq \bar{e}$. The maximum is achieved at $e = \bar{e}$. In this case the condition for benefit to exceed cost is $(m_2 - m_1) > c_2 \bar{e}$, or $(m_2 - m_1)/c_2 > \bar{e}$. Therefore, the decision rule of a member of group II is

$$\text{Choose } e = \bar{e} \text{ if } \frac{(m_2 - m_1)}{c_2} > \bar{e}. \quad (25.2)$$

From the conditions of Eqs. (25.1) and (25.2) it turns out that the groups are separated and each group chooses the appropriate education or signal if

$$\frac{(m_2 - m_1)}{c_1} < \bar{e} < \frac{(m_2 - m_1)}{c_2}. \quad (25.3)$$

Given our assumption that $m_1 < m_2$ and that $c_2 < c_1$, there must exist an $\bar{e}$ satisfying the boundary condition of Eq. (25.3).

The critical education level $\bar{e}$ satisfying condition in Eq. (25.3) depicts a signalling equilibrium which separates each type of worker making a choice from the other type. That is why this type of signalling equilibrium is also called a separating equilibrium. The employer's beliefs are confirmed here by market experience. If the employer would have put $\bar{e}$ above $(m_2 - m_1)/c_2$, group I would have continued to prefer $e = 0$ (because $m_2 - m_1 < c_1 \bar{e}$) but group II would also have decided not to attain education at all (because $m_2 - m_1 < c_2 \bar{e}$). If the employer would have put $\bar{e}$ below $(m_2 - m_1)/c_1$, the two groups would have chosen $\bar{e}$. In both cases, education cannot work as a signal for separating the more capable from the less capable. The consequence is a pooling or non-separating equilibrium. Here all workers are paid a wage based on their average ability. In contrast, by choosing a critical education level in the correct range given by Eq. (25.3) the entrepreneur has realized a separating signalling equilibrium. In this case no worker has any incentive to change his behavior.

We can now take a closer look at the signalling equilibrium. Since any value of $\bar{e}$ satisfying condition in Eq. (25.3) will do, there can be an infinite number of possible equilibrium values for $\bar{e}$. Of course, any of those values enables the employer to separate the two groups of workers.

Next, these equilibria are not equivalent from the viewpoint of welfare. An increase in the critical level education $\bar{e}$ obligates members of group II to spend more on education without affecting members of group I (they stay at $e = 0$). Everybody may be worse off with signalling than without signalling, because the wage $\bar{w}$ (without signalling) may be higher than the net return $m_2 - c_2 \bar{e}$ to group II. (Recall that $\bar{w} > m_1$.) In general, a signalling equilibrium with a lower $\bar{e}$ Pareto dominates equilibria with higher values of $\bar{e}$, because a smaller value of $\bar{e}$ implies spending less on education. This property follows from the fact that $\bar{e}$ is used only as a screening device. Therefore, signalling may lead to inefficiencies.

Note also that members of group II are willing to attain education at their own interest only. They want to send a signal to the employer because that will

distinguish them from less capable workers which in turn enables them to earn more. The amount of output produced under signalling is the same as it would be if there were no signal at all (Since education does not increase the productivity).

But it should be mentioned that signalling always does not lead to inefficiency. For instance, in the case of used cars warranties may help trade facilitation. Signalling equilibrium will Pareto dominate non-signalling equilibrium here. Therefore, each case needs to be investigated on its own merit.

THE PRINCIPAL-AGENT PROBLEM

The purpose of this section is the study of the principal-agent problem, or simply the agency problem. Our discussion will show that the topic involves asymmetric information in a very natural way.

The standard version of this problem is the following: one person (the principal) contracts with another person (his agent) to perform some work for him. Standard examples are client-lawyer, patient-doctor, insurer-insured person and manager-worker relationships. Another interesting example is the case of a landowner who is hiring a farmer to do the farming for him. One possible contract here can be sharecropping where the worker and the landowner each gets some fixed percentage of the output. Therefore the payment to the worker depends in part on the observed output, but the worker and the principal share the risk of output fluctuations.

The pay-off to the principal depends partially on random events over which the agent has no control and partially on the behavior of the agent. In the absence of randomness, the principal can infer the actions of the agent from the outcome achieved and an optimal contract in such a case would be one that paid the agent some particular amount if the optimal outcome occurred and nothing otherwise. In case the principal can monitor the actions of the agent, then he will give a detailed list of instructions to the agent, and the agent will be paid a fixed sum if the instructions are followed and nothing otherwise. The problem arises if randomness characterizes the outcome and it is not possible to monitor the actions of the agent. How does the principal-agent relationship work when the agent knows that he is not going to be monitored?

An optimal contracting arrangement in the principal-agent framework can be defined as follows: A contract is optimal if it maximizes the expected utility of the principal for any expected utility of the agent, subject to the condition that the agent finds it worthwhile to participate in the contract. Now, under a

fixed price contract, that is under a contract which compels the agent to supply the required services at a fixed price per unit, the agent has incentives to keep costs under control. But if the work is too risky then a risk-averse agent will require a large premium over expected cost to undertake the work. If the contract is of cost-plus type, a contract under which the agent is paid the entire cost he incurs on the work plus a fixed sum of money, then there is no strong incentive for the agent to keep costs under control because he receives the same money irrespective of costs. Thus, there are two important aspects of the contracting problem: the agent must be given some incentive that will encourage him to act in the interest of the principal and the allocation of risk involved in the work efficiently. These two issues are often in conflict with one another and it is this fact that makes the problem complicated.

To illustrate the problem analytically, let us suppose that a person the (principal) owns a plot of land but he is unable to work on the land himself. Therefore, he employs another person (an agent) to do the farming for him. To make the analysis easy, let us suppose that there is only a finite number of output levels $(x_1, x_2, \dots, x_n)$. Let a and b be two actions that can be chosen by the agent out of some set of feasible actions. These actions influence the probability of occurrence of different output levels. We denote the probability that the output level x_i will occur if the agent chooses action $a(b)$ by $\pi_{ia}(\pi_{ib})$. Let $q_i = q(x_i)$ be the amount that the principal pays the agent if output level x_i is observed. Then the expected profit of the principal if the agent chooses action a , say, is

$$\sum_{i=1}^n (x_i - q_i) \pi_{ia}. \quad (25.4)$$

Next, suppose the agent is risk-averse and maximizes the expected utility from the payment. We assume that the agent finds actions costly and let $C(a)$ be the cost of action a . This cost enters into his utility function linearly. Therefore, the agent's expected utility less cost, if he chooses action a , is given by $\sum_{i=1}^n U(q_i) \pi_{ia} - C(a)$, where U stands for the Von Neumann-Morgenstern utility function of the agent. Since the agent is a utility maximizer he will choose action a if

$$\sum_{i=1}^n U(q_i) \pi_{ia} - C(a) \geq \sum_{i=1}^n U(q_i) \pi_{ib} - C(b) \quad (25.5)$$

and will choose action b otherwise. This constraint is referred to as the *incentive compatibility constraint*. It says that the expected utility to the agent from choosing action a is at least as large as the expected utility from action b .

The agent may have other alternatives available that give him some utility U_0 (say). This utility may come from working elsewhere or from leisure. The principal's work will provide an incentive to the agent if the utility from the work is at least as large as U_0 . That is,

$$\sum_{i=1}^n U(q_i)\pi_{ia} - C(a) \geq U_0. \quad (25.6)$$

This is the *participation constraint*.

The principal wishes to maximize his expected profit of Eq. (25.4) subject to the constraint Eqs. (25.5) and (25.6). The maximization will take place over a and q_i . The agent will choose an action a that he finds most profitable for him given the incentive system (q_i) set up by the principal. Since the principal is aware of this, he will offer the q_i that is best for him. That is, the principal is taking the action of the agent as a constraint while offering q_i . This means that he will offer q_i so that the agent's desired action becomes the principal's desired action.

If the payment is based on action rather than on output, then the principal is to determine the expected profit from each action by the agent and then induce the action that maximizes the principal's expected profit.

We have stated earlier that an optimal incentive scheme should involve some degree of risk sharing between the agent and the principal. Such a case can arise when the actions of the agent are hidden so that only outputs are the basis of incentive payments. In this case the payments to the agents are random. This is the most interesting case of the principal agent problem. Since the payments are based on outputs, the principal will pay the agent less when less output is produced. But at the same time the principal has to ensure that the agent will do the work for him (the participation constraint has to be fulfilled). This is the trade-off in designing an optimal incentive scheme.

Attempts have been made to develop results along this line. To illustrate one such result suppose that there is no incentive problem, and that the only issue is risk sharing. In such a case the principal's optimization problem is maximizing the expected profit

$$\sum_{i=1}^n (x_i - q_i)\pi_{ia}. \quad (25.7)$$

subject to

$$\sum_{i=1}^n U(q_i)\pi_{ia} - C(a) \geq U_0 \quad (25.8)$$

where maximization takes place over q_i . The solution to this optimization problem is $q_i = \text{constant}$ (see Appendix 25.1 for a technical derivation). That is, the principal insures the agent against all risks. This is desirable since the agent is risk averse and the principal is risk neutral. (The principal is risk-neutral because his expected profit is linear in x_i .) Under the risk-neutrality assumption the principal is indifferent to risk and thus there is no need to trade off incentives for risk sharing.

A great deal remains to be done in the context of the principal-agent problem. One disturbing feature of the problem is that the principal knows the agent's utility function and incorporates this into the contract with agent. But it is not clear how the principal can get this information. Acquiring this information may be harder than monitoring the agent's actions. Discussions on the principal-agent relationships under imperfect information about an agent's utility function and also under alternative assumptions about the individuals' attitudes towards risk are important topics of current research in the economics of information.

Applied Example 25.1 Village Panchayats and Information

Grameen banks (village banks) in rural West Bengal, a state in India in which village panchayats work very efficiently, sanction loans at a very minimal interest rate to rural inhabitants to commence their investment activities. In terms of hierarchy a village panchayat is the lowest level administrative organization in rural areas. Grameen banks take the help of panchayats to grant loans under different development schemes. The bank depends on the panchayats because the panchayats have better access to information about who are good and bad credit risks than the bank officials and the panchayats are in a better position to monitor the progress of the loan repayments. The first one is an adverse selection problem and the second one is a problem of moral hazard.

The Grameen banks address these two problems with the help of panchayats. Since persons can get loans under rural development schemes only if they get approval from panchayats, they have strong incentives to convince panchayats that they are not bad credit risks. A panchayat being very much a local organization has enough information about loan demanders. Therefore, the banks are able to solve the adverse selection problem by seeking information from panchayats. Next, the moral hazard problem of monitoring the progress of repayments can also be done by the panchayat, again because it is a regional organization and has enough information about the activities of the loan seekers.

Appendix 25.1

VALUE OF INFORMATION

In the absence of perfect information a decision maker may not be able to know what the consequence of a particular action will be. Better information can reduce uncertainty and can therefore lead to better decisions that generate increased utility levels. Information is valuable because it may allow an individual to revise his estimates of probabilities of choosing options and to take advantage of these revisions. Even small revisions of probability may be quite valuable.

To illustrate the value of information suppose that information can be measured by the number of messages (s) received. The decision maker will revise his probabilities on the basis of s . Let us consider a framework in which a decision maker forms opinions about the subjective probabilities of two states of the world: good time and bad time. His consumption of contingent commodities in these periods are denoted by x_g and x_b respectively. Let $\pi_g(\pi_b)$ be the probability of good (bad) time. These probabilities are functions of s . The individual's goal is to maximize

$$\text{expected utility} = \pi_g U(x_g) + \pi_b U(x_b)$$

subject to

$$m = p_g x_g + p_b x_b + p_s s$$

where $U(x_g)$ [$U(x_b)$] is the utility in good (bad) time, $p_g(p_b)$ is the price of the contingent commodities in good (bad) time, p_s is the per unit cost of information messages (that is, the cost of a phone call to gather price information, and so on) and m is the individual's income.

Setting up the Lagrangian for this problem

$$\mathcal{L} = \pi_g U(x_g) + \pi_b U(x_b) + \lambda(m - p_g x_g - p_b x_b - p_s s)$$

generates the following first order conditions for a maximum:

$$\frac{\partial \mathcal{L}}{\partial x_g} = \pi_g U'(x_g) - \lambda p_g = 0,$$

$$\frac{\partial \mathcal{L}}{\partial x_b} = \pi_b U'(x_b) - \lambda p_b = 0,$$

$$\begin{aligned}\frac{\delta \mathcal{L}}{\delta s} &= \pi_g U'(x_g) \frac{dx_g}{ds} + \pi_b U'(x_b) \frac{dx_b}{ds} + U(x_g) \frac{d\pi_g}{ds} + \\ &\quad U(x_b) \frac{d\pi_b}{ds} - \lambda p_g \frac{dx_g}{ds} - \lambda p_b \frac{dx_b}{ds} - \lambda p_s = 0, \\ \frac{\delta \mathcal{L}}{\delta \lambda} &= -p_g x_g - p_b x_b - p_s s + m = 0,\end{aligned}$$

where U' is the derivative of U .

The first two of these equations imply that at optimal position the ratio of expected marginal utilities should be equal to the price ratio p_g/p_b . According to the third equation the individual should purchase additional information up to the point at which the expected marginal gain from an additional message equals the cost of the message and the budgetary reallocation promoted by the message. It may be important to observe that the additional messages increase utility in two ways: (a) they allow the individual to revise his choices of x_g and x_b for getting higher expected utility at the revised probability estimates, and (b) they may increase the expected utility associated with a particular choice of x_g and x_b by revising the probabilities of the states. The first of these seems to be a very important reason for acquiring information.

MORAL HAZARD

Assume that there is a possibility that a risk-averse individual with initial wealth W_0 will incur a loss L . The probability of loss is given by π . This probability can be reduced if the individual adopts some preventive measures for which he has to spend M . Let U be the individual's utility function.

In the absence of insurance coverage the wealths in the two situations (no loss and loss) are $W_1 = W_0 - M$ (no loss), $W_2 = W_0 - M - L$ (loss). The individual will choose M to maximize

$$E = \text{Expected Utility} = (1 - \pi)U(W_1) + \pi U(W_2).$$

The first order condition for a maximum is

$$\begin{aligned}\frac{\delta E}{\delta M} &= -U(W_1) \frac{d\pi}{dM} - (1 - \pi) U'(W_1) + \\ &\quad U(W_2) \frac{d\pi}{dM} - \pi U'(W_2) = 0\end{aligned}$$

where U' is the derivative of U . We can rewrite this condition as

$$\pi U'(W_2) + (1 - \pi) U'(W_1) = [U(W_2) - U(W_1)] \frac{d\pi}{dM}.$$

This condition says that the individual should take preventive measures up to the point where the expected marginal utility when a preventive measure is taken by spending one more unit of money is equal to the reduction ($d\pi/dM < 0$) in the expected value of utility that might be encountered in the case of wealth loss.

Now suppose that the individual can purchase insurance coverage that pays him X if a loss occurs and for this he has to pay a premium p (which depends on X). Wealths in the two situations are now given by

$$\begin{aligned} W_1 &= W_0 - M - p, \\ W_2 &= W_0 - M - p - L + X. \end{aligned}$$

The individual wants M and X to take those values which will maximize his expected utility. Assume that the insurance provider can monitor precautionary activities and knows the probability of loss, and therefore charges a fair premium $p = \pi X$. Under this assumption $W_1 = W_0 - M - \pi X$ and $W_2 = W_0 - M - L + (1 - \pi) X$. With full insurance coverage ($X = L$), the first order condition for a utility maximizing choice of M is

$$\begin{aligned} \frac{\delta E}{\delta M} &= -(1 - \pi) U'(W_1) \left(1 + L \frac{d\pi}{dM} \right) - U(W_1) \frac{d\pi}{dM} - \\ &\quad \pi U'(W_2) \left(1 + L \frac{d\pi}{dM} \right) + U(W_2) \frac{d\pi}{dM} = 0. \end{aligned}$$

Under the assumption that $W_1 = W_2$,

$$-L \frac{d\pi}{dM} = 1.$$

This condition says that at the utility maximizing choice the marginal cost of an extra unit of prevention (which is one here) should equal the marginal reduction in the expected loss provided by that extra expenditure. Thus, with full insurance and a fair premium, precautionary measures are taken at socially optimal levels.

Our analysis has assumed that insurance providers know the probability that an individual will incur a loss and can therefore charge a fair premium. This assumption is quite doubtful. In the simplest case the insurance provider may set a premium based on the average probability of loss experienced by some group of persons. However in this case each individual has an incentive to adopt minimum precautionary measures. This can be seen easily for $X = L$ (full coverage) case. In this case $W_1 = W_2$ irrespective of p and M . Since p does not depend on M , utility is maximized if $M = 0$. This distorting effect of moral hazards arises from the information asymmetry resulting from insurers' inability to monitor preventive measures taken by insured persons.

PRINCIPLE-AGENT PROBLEM

To maximize the principal's expected profit

$$\sum_{i=1}^n (x_i - q_i) \pi_{ia}$$

with respect to q_i subject to the participation constraint

$$\sum_{i=1}^n U(q_i) \pi_{ia} - C(a) \geq U_0.$$

Let us suppose that the participation constraint is satisfied with equality. We then write the Lagrangian for this optimization problem as

$$\mathcal{L} = \sum_{i=1}^n (x_i - q_i) \pi_{ia} - \lambda \left[\sum_{i=1}^n U(q_i) \pi_{ia} - C(a) - U_0 \right],$$

where λ is the Lagrange multiplier. Differentiating $\mathcal{L}$ partially with respect to q_i and λ , and setting the derivatives to zero, we get the first order conditions

$$-\pi_{ia} - \lambda U'(q_i) \pi_{ia} = 0,$$

$$\sum_{i=1}^n U(q_i) \pi_{ia} - C(a) - U_0 = 0.$$

From the first of the above conditions we have

$$U'(q_i) = 1/\lambda, \text{ a constant.}$$

This means that q_i must be independent of i . That is, q_i is a constant.

EXERCISES**Multiple Choice**

1. If consumers have limited information about a product, which of the following is true?
 - a) Equilibrium may not exist.
 - b) If equilibrium exists then the quality produced is lower than that produced in a world of perfect information.
 - c) In equilibrium, if it exists, resources are used inefficiently compared to markets with perfect information.
 - d) All of the above.
2. Healthy senior persons often have problems buying medical insurance because of
 - a) Adverse selection.
 - b) Moral hazards.
 - c) (a) and (b).
 - d) None of the above.
3. If one can influence the occurrence of an event against which he is insured, but the seller cannot detect this action, then this is a problem of
 - a) Moral hazard.
 - b) Adverse selection.
 - c) (a) and (b).
 - d) None of the above.
4. Consider two groups of workers, where the first group has a constant marginal product of m_1 and the second group has a constant marginal product of m_2 , $m_1 < m_2$. The constant marginal costs of acquiring education by these groups are C_1 and C_2 ($C_1 > C_2$). Then a critical education level $\bar{e}$ that depicts a signalling equilibrium must satisfy
 - a) $\frac{m_2 - m_1}{C_1} < \bar{e} < \frac{m_2 - m_1}{C_2}$.
 - b) $C_2(m_2 - m_1) < \bar{e} < C_1(m_2 - m_1)$.

$$c) \frac{C_2}{(m_2 - m_1)} < \bar{e} < \frac{C_1}{(m_2 - m_1)}.$$

$$d) C_1 m_1 = \bar{e} = C_2 m_2.$$

5. In the cost-plus type principal-agent relationship,
- The principal bears the total cost.
 - The agent bears the total cost.
 - The principal and the agent share the cost equally.

Quantitative Problems and Questions

- Distinguish between the economics of uncertainty and the economics of information.
- What is Gresham's law? Discuss the role of this law in the market for lemons.
- There are two types of workers: high-ability and low-ability workers. High-ability workers earn Rs. 5,000, low-ability workers earn Rs. 3,000. Firms are not able to measure workers' abilities but they can observe whether a worker has some special degree. Workers' utility depends on the difference between their wages and the costs they incur for obtaining the special degree.
 - If the costs of obtaining the special degree is the same for high-ability and low-ability workers, can there exist a separating equilibrium in which high-ability workers get high-wage jobs and low-ability workers get low-wage jobs?
 - What is the maximum amount that a high-ability worker would pay to obtain a special degree?
- Consider an individual with current wealth of Rs. 100,000 who faces a 25% chance of losing his car worth Rs. 20,000. If he installs an antitheft device that costs Rs. 1,950, then the probability of theft reduces from 0.25 to 0.15. The person's utility function is of the type $U(W) = \log W$. Do you think that the person will take this precaution?

Suppose now that the person can buy full insurance coverage for a premium of Rs. 5,200. The insurance company does not make any effort to monitor installation of antitheft devices. Do you think that the individual will buy the insurance, or take precaution against theft?

Next, assume that the insurance providers can monitor installation of antitheft devices. Suppose that it would cost Rs. 10 to determine whether an owner had installed such a device. In this case the insurance premium for a person with an antitheft device would be Rs. 3,210. Will the person buy the insurance?

5. A person spends his income (2) on the consumption of two goods 1 and 2. The price of good 1 is 0.25. The price of good 2 which is sold in two shops is either 0.75 or 1.25. The person does not know which shop charges which price. He can choose the shop either randomly by flipping a fair coin or by receiving information through a phone call by spending 0.23. The person's indirect utility function is $V = 0.5m \left(\sqrt{p_1 p_2} \right)^{-1}$, where m is the person's income and p_1 and p_2 are the prices of goods 1 and 2. Do you think that it will be better for the person to receive information through a phone call? Why?
6. Consider a market of used cars. There are two types of cars: bad cars worth Rs. 5,000 each to their present owners and good cars worth Rs. 15,000 each to their owners. There are 100 cars of each type. The demand for good-quality used cars is $X = 180 - 0.005p$ and that for bad-quality cars is $X = 150 - 0.005p$. If buyers can differentiate between two types of cars, bad cars will sell at Rs. 10,000 and good cars will sell at Rs. 16,000. If there is inability of demanders to differentiate among the cars, why are both demanders and suppliers of good cars made worse off? What might they do for mutually beneficial trade to take place? What is the maximum amount of consumer surplus created by trade in market equilibrium?
7. Refer to question no. 4 above. If insurers know which owners had installed antitheft devices, they could formulate price policies accordingly. Owners without the device would face a probability of loss of 0.25 and would pay Rs. 5,000, as premium for insurance coverage. Owners with devices have a probability of loss of 0.15 and face a premium of Rs. 3,000. As before, the individual utility function is of the type $U(W) = \log W$. If only low-risk owners can buy a certificate indicating installation of an antitheft device, how much would they pay for it? How much would forged certificates have to cost to prevent high-risk owners from using them?
8. A football league association is trying to decide how much price to charge for season tickets for its matches. The association knows that the demand function for tickets is of the form $x = a - bp$ where x is the number of season tickets demanded, and p is the price per season ticket. They know that a

= 150,000 but there is uncertainty about the value of b . It is assumed that $b = 500$ with probability $1/3$, that $b = 1000$ with probability $1/3$ and that $b = 2000$ with probability $1/3$. The association will choose season tickets for maximizing expected revenue as a monopolist.

- a) What value of p will be chosen? What is the expected revenue?
 - b) Suppose that the value of b is known with certainty. What is the value of such information to the association?
9. Explain why insurance is not available in the following risky situations:
- a) A firm is engaged in research and development to invent a medicine for cancer. The firm would like to insure against the possibility that it will not be able to invent such a medicine.
 - b) A student wishes to insure the loss in lifetime earnings if he fails to earn some degree.
10. An agent can produce x units of outputs for the principal using the cost function $C(x) = x^2/2$. The principal pays the worker a constant wage per unit of output plus a lumpsum. The worker can achieve a utility level zero if he works elsewhere. What is the optimal incentive scheme for the agent? How much profit would the principal earn if the agent's utility from alternative employment is 1?
11. Prove or disprove the following:
- a) The source of market failure in a market of many quality cars is an externality problem generated by the selling side.
 - b) In an equilibrium with symmetric imperfect information the private value of the highest-quality product is equal to its average value.
 - c) Information may be regarded as a public good.
 - d) Signalling never leads to inefficiencies.
 - e) Moral hazard is an informational externality problem
12. 'Adverse selection is a hidden information problem, whereas moral hazard is a hidden action problem'. Discuss this statement.
13. In what way can a signal be regarded as a self-selection mechanism?

26

Law and Economics

Economic analysis has developed paradigms that seem to provide a powerful analytical framework for the study of law. The economic models of self-interest, whose objectives are efficient allocation of resources, appear to offer a neat view of the world. Such a view may be quite comforting to a lawyer who generally deals with complex problems created by the diversity of values among people.

In fact, there are many areas in which the interests of economists often dovetail with those of lawyers. For instance, the antitrust field, criminal justice, liability rules, pollution, poverty, racial discrimination, land use planning, accident prevention, landlord-tenant relations and urban government, among others, are active areas of cooperation.

This chapter is concerned with several topics in law and economics. The areas covered are crime and punishment, liability law, contract law and antitrust law. Clearly, given many active areas of cooperation between economists and lawyers, these topics constitute only a small sample. The analyzed models will give some idea about the interrelationships between law and economics.

CRIME AND PUNISHMENT

Many crimes are motivated by economic considerations. For instance, a bank robber robs banks because he needs money. Similarly, a person gets involved in pick pocketing because of his demand for money. In this section we look at the economic choice calculus that underlies criminal activity. Economists have developed a framework that explains how rational persons trade off returns from legal and illegal activities. Gary S. Becker is a pioneering contributor to this literature.

An economic choice framework suggested by Becker (1968) regards crime, excepting crimes of passion, as an economic activity with rational participants. A person commits a crime if his expected utility exceeds the level of utility he could derive from alternative (legal) activities. Therefore, a person commits a criminal offense, not because his motivation differs from that of other persons, but because his options and the evaluation of their benefits and costs are different.

In the framework considered by Becker, the number of crimes committed by an individual depends on his probability of convictions, the expected severity of his punishment, and variables reflecting his legal-income earning potentiality, environment and tastes. This can be represented as

$$n_j = c_j(p_j, f_j, u_j) \quad (26.1)$$

where n_j is the number of offenses committed by person j in a given period, p_j is his probability of conviction per offense, f_j is his expected punishment per offense, and u_j is a composite variable representing all other influences. Both p_j and f_j may be assumed to depend on the judge, jury, prosecutor etc. that person j receives.

Increasing the probability of convictions or the severity of punishment increases the incentive not to commit crimes. That is, n_j is decreasing in p_j and f_j . This becomes more evident if we look at the expected utility of person j from an offense. This expected utility is given by

$$p_j U_j(y_j - f_j) + (1 - p_j) U_j(y_j) \quad (26.2)$$

where y_j is his income from an offense. For instance, if j is a shoplifter then y_j is the value of the item lifted. U_j is person j 's utility function and f_j is to be interpreted as the monetary equivalent of the punishment. Note that we can rewrite Eq. (26.2) as $U_j(y_j) - p_j[U_j(y_j) - U_j(y_j - f_j)]$. Therefore, given increasingness of the utility function in income, an increase in f_j decreases Eq. (26.2). The above expression is clearly decreasing in p_j also. The effect of changes in some

components of u_j could also be anticipated. For instance, a rise in legal income or level of education is expected to reduce n_j . A shift in the form of punishment, say, from a fine to imprisonment, would tend to reduce the number of offenses, since they cannot be committed while in prison. This example shows how environment can change n_j .

The total number of offenses is the sum of all n_j 's and would depend on the set of p_j , f_j and u_j 's. Although the variables p_j , f_j and u_j are different for each individual j because of differences in such factors as intelligence, education, wealth and family upbringing, Becker hypothesizes that the total number of offenses n_0 can be expressed as the function

$$n_0 = c(p, f, u) \quad (26.3)$$

where p , f , u represent the average values of these variables for all persons in the community. (We can define $p = \sum_{j=1}^k n_j p_j / \sum_{j=1}^k n_j$, where k is the number of persons in the society. Similar definitions hold for f and u .) It is assumed that n_0 is decreasingly related to p and f .

The criminal through his crime imposes some harm on other people. The amount of harm would tend to increase with the activity level. We represent this as

$$H_j = h_j(n_j) \quad (26.4)$$

where h_j is increasing in n_j . The concept of harm is quite close to that of externality and the function relating the amount of harm to the offense level is similar to a function through which activities cause externalities.

The social value of the gain to offenders also tends to increase with the number of offenses. That is, the gain G_j where

$$G_j = g_j(n_j) \quad (26.5)$$

is also increasing in n_j . The net damage done to the society due to person j 's activity is simply the difference between the harm and gain and can be written as

$$\begin{aligned} D_j &= d(n_j) = h_j(n_j) - g_j(n_j) \\ &= H_j - G_j. \end{aligned} \quad (26.6)$$

For example, the net damage of a murder can be measured by the loss in earnings of the victim. But other factors like the value placed by the society on the life itself could also be considered.

We write D , the total net damage done to the society, as a function of n_0 and assume that

$$\begin{aligned}
 D &= d(n_0) = h(n_0) - g(n_0) \\
 &= H - G
 \end{aligned}
 \tag{26.7}$$

where $H = h(n_0)$ is the total harm imposed on the society and $G = g(n_0)$ is the total value of the gain to offenders. D , H and G are increasing in n_0 and hence monotone in p and f .

Next, it is clear that the more that is spent on policemen, court personnel and specialized equipment, the easier it is to discover offenses and convict offenders. One can postulate a relation between the output of police and court activity, and the total number of offenses and average probability of conviction and write it as $A = l(n_0, p)$, where l is a production function. The cost C of apprehension and conviction depends on A and we write it as $C = K(A) = K[l(n_0, p)] = q(n_0, p)$, say. That is, the relationship between the total cost C and the number of offenses and conviction probability is expressed through the functional form q . It is extremely likely that A is increasing in n_0 and p , and since increased activity would be more costly, C is also increasing in n_0 and p .

The cost of different punishments to an offender can be made comparable by converting them into their monetary equivalent or worth. For example, the cost of an imprisonment is the discounted sum of the earnings foregone and the value placed on the restrictions on freedom.

Punishments affect not only offenders but also other members of the society. For instance, imprisonment requires expenditure on guards, supervisors, building, food etc. However, fines paid by offenders are received as revenues by others. The total social cost of punishments is the cost to offenders plus the cost or minus the gain to others. It is convenient to express social costs in terms of offender cost. We write the social cost as bf , where the coefficient of multiplication $b > 0$ varies over punishments. For instance, b is very small for fines, while $b > 1$ for torture, probation, parole, imprisonment etc.

We now consider the social goal. If the aim simply were deterrence, the probability of conviction p could be made very close to 1, and punishments, f , could be made to exceed the gain. By doing this, the number of offenses, n_0 , could be reduced substantially. But an increase in p , increases the social cost of offenses through its effect on the cost of punishments bf . Similarly, if the goal were to make 'the punishment fit the crime', p could be set close to 1 and f could be equated to the harm imposed on the rest of the community. But here again we have very high social costs resulting from increases in p and f .

Therefore what is required is a criterion that will give due weight to the damages from offenses, the cost of apprehension and conviction, and the social

cost of punishment. We assume the existence of a function that measures social loss from offenses and write it as

$$L = F(D, C, bf, n_0) \quad (26.8)$$

where L is increasing in D , C and bf . It may be convenient to work with a simpler formulation and we therefore suppose that L is additive across components. More precisely,

$$L = D(n_0) + C(p, n_0) + bfpn_0. \quad (26.9)$$

Since bf is the loss per offense, $bfpn_0$ is the total social loss from punishment. (Note that pn_0 is the number of offenses punished.)

Assume that b is a constant. Then the decision variables with which we can work are p and f . (Since we can change C through p , we prefer to work with p .) Then the objective is to minimize the social loss function of Eq. (26.9) with respect to p and f . The optimizing conditions turn out to be

$$\frac{\Delta D}{\Delta n_0} + \frac{\Delta C}{\Delta n_0} = -bpf \left(1 - \frac{1}{e_f} \right) \quad (26.10)$$

$$\text{and} \quad \frac{\Delta D}{\Delta n_0} + \frac{\Delta C}{\Delta n_0} + \frac{\Delta D / \Delta p}{\Delta n_0 / \Delta f} = -bpf \left(1 - \frac{1}{e_p} \right) \quad (26.11)$$

where
$$e_f = -\frac{f}{n_0} \frac{\Delta n_0}{\Delta f}.$$

and
$$e_p = -\frac{p}{n_0} \frac{\Delta n_0}{\Delta p}. \quad (26.12)$$

(See Appendix 26.1 for a derivation.) Since n_0 is decreasing in f and p , both e_f and e_p are positive.

The term on the left hand side of Eq. (26.10) ([26.11]) is the marginal cost of increasing the number of offenses n_0 through a reduction in $f(p)$. A reduction in p partly reduces the cost of combating offenses, and hence the marginal cost of increasing n_0 must be less when p rather than when f is reduced. This also follows technically in view of the fact that the third term on the left hand side of Eq. (26.11) is negative (since $\Delta n_0 / \Delta f$ is negative). The left hand side of Eq. (26.10) is clearly positive. Now, the average revenue given by $-bpf$ is negative, but the marginal revenue given on the right hand side of Eqs. (26.10) and (26.11) is not necessarily negative and would be positive if the elasticities e_f and e_p

were less than unity. Since the loss is minimized when marginal cost equals marginal revenue the optimal value of e_f must be less than unity. Now the optimal value of e_p could exceed unity only when the third term on the left hand side of Eq. (26.11), which is negative, outweighs the first two terms which are positive. But $e_p > 1$ is a reversal of the usual equilibrium condition for a profit-maximizing firm, which is that the absolute value of the elasticity of demand is greater than unity. (Recall that in the usual case the average revenue is positive, but in this case the average revenue is negative.) Therefore, e_p must be less than unity.

Since the marginal cost of changing n_0 through a change in p is less than that of changing n_0 through f , the marginal revenue from p is also less than that from f . From Eqs. (26.10) and (26.11) it follows that this is possible if and only if $e_p > e_f$. This latter inequality means that the elasticity of response of offenses with respect to a change in p exceeds its response to a change in f . (This is certainly the case if carrying out the punishment is costly, as it is with probation, imprisonment, or parole.) Thus, minimization of social loss (via minimal supply of offenses) requires a high value of p . That is, offenders' activities are subject to high risk. More explicitly, the loss from offenses is minimized if p and f are selected from those regions where offenders, on balance, are risk-preferers. Now a relatively high value of p means that the number of offenses also reduces. This in turn means that the real income of offenders would be less. Hence 'illegal activities would not pay' in the sense that real income received would be less than what could be received in less risky legal activities. Therefore, the conclusion that 'crime does not pay' is an optimality condition and not an implication about the efficiency of the police and judicial systems.

If offenders were risk-neutral, a reduction in p compensated by an equal increase in f would leave pf , n_0 , $D(n_0)$, $bfpn_0$ unchanged but would reduce loss, because the cost of apprehension and conviction would be lowered by reduction in p . The loss therefore gets minimized if p is lowered arbitrarily close to zero and f is raised sufficiently so that the product pf would induce the optimal number of offenses. Thus, since $e_p = e_f$, and if n_0 depends only on pf , and $C = 0$ if $p = 0$, the two equilibrium conditions of Eqs. (26.10) and (26.11) reduce to the single equation

$$\frac{\Delta D}{\Delta n_0} = - bpf \left(1 - \frac{1}{e_f} \right). \quad (26.13)$$

From this equation and the condition that n_0 is a function of pf , the equilibrium values of pf and n_0 can be determined.

If offenders were risk-avoiders, then the loss would be minimized by setting p arbitrarily close to zero, since a reduction in p reduces not only C but also n_0 and thus D and $bpf n_0$.

LIABILITY LAW AND ITS ECONOMIC CONSIDERATIONS

Though tort law or liability law has a long history, it is today one of the most dynamic branches of law. Loosely speaking, a tort is a civil (seldom a criminal) wrong. Such a wrong takes place when one party, usually unintentionally, destroys another party's initial entitlement by imposing a negative externality. The courts can then provide a remedy in the form of damages.

By what rule should we determine whether a person, having destroyed an initial entitlement, is liable and therefore should pay for the tortious act? Brown (1973) has identified eight liability rules. Here we discuss these rules very briefly.

- a) No Liability: The victim is liable under all circumstances.
- b) Strict Liability: The injurer is liable under all circumstances.
- c) Negligence Rule: The tort of negligence is perhaps the most common tort. An injured individual can seek compensation from his injurer for damages under this rule if the injurer's actions were negligent, that is, the injurer's action fell below the standard of those of a 'reasonable' man and thereby caused injury. An injury without proof of negligence and of injurer's deficient conduct is not compensatory.
- d) Strict Liability with Contributory Negligence Rule: The injurer is liable if the victim is not found negligent.
- e) Negligence Rule with Contributory Negligence: The injurer is liable if he is negligent and the victim is not; the victim is liable otherwise.
- f) Strict Liability with Dual Contributory Negligence: The victim is liable if he is negligent and the injurer is not; the injurer is liable otherwise.
- g) Relative Negligence: The injurer is liable under this rule if the money that could be spent by the injurer on avoidance is more cost-effective than that which could be spent by the victim.
- h) Comparative Negligence Rule: Under this rule liability is shared by the injurer and the injured according to the relative liability of the two parties. That is, in this case the injurer must pay a fraction of the extent of damage done.

To analyze the behavior of the victim and the injurer under different liability rules, let us consider a model developed by Shavell (1987). For expositional ease we assume that the negative externality imposed by the injurer is an accident. Then the injurer can be regarded as a person driving down the road and the victim as one crossing the road. Suppose x stands for the effort of the injurer to prevent accident. In other words, x is the amount of care taken by the injurer. Similarly, y represents the effort of the victim to avoid accident. That is, y is the amount of care taken by the victim. Now, exercising more care will be costly to the injurer and the victim. We write $C_i(x)[C_v(y)]$ for the cost to the injurer (victim) for taking the level of care $x(y)$. We suppose that the expected loss of the accident depends on the care taken by both parties and write it as $L(x, y)$. It is natural to assume that L is decreasing in both x and y .

The appropriate social objective is the minimization of total costs:

$$\min_{x,y} [C_i(x) + C_v(y) + L(x, y)]. \quad (26.14)$$

There are two conditions for determining the optimal levels of care. The first one is that the marginal cost of increasing care by the injurer equals the marginal benefit to the victim of this more careful behavior. The second condition says that the marginal cost borne by the victim from increasing his level of care should equal the marginal benefit he achieves from the reduction in expected accident loss. (These conditions are derived rigorously in Appendix 26.1.) We will refer to the levels of care taken by the victim and the injurer that minimize the social loss in Eq. (26.14) as socially optimal levels of care and denote them by $\bar{y}$ and $\bar{x}$.

Now, under the no liability rule the injurer will choose a zero level of care and the victim will choose that level of care that minimizes his total costs. Since the injurer has no incentive to take care here, this rule is not socially optimal.

In the case of strict liability the victim has no incentive to take care and the optimal choice for him is $y = 0$. The injurer will choose that level of care which is optimal for him. This rule also is not socially optimal.

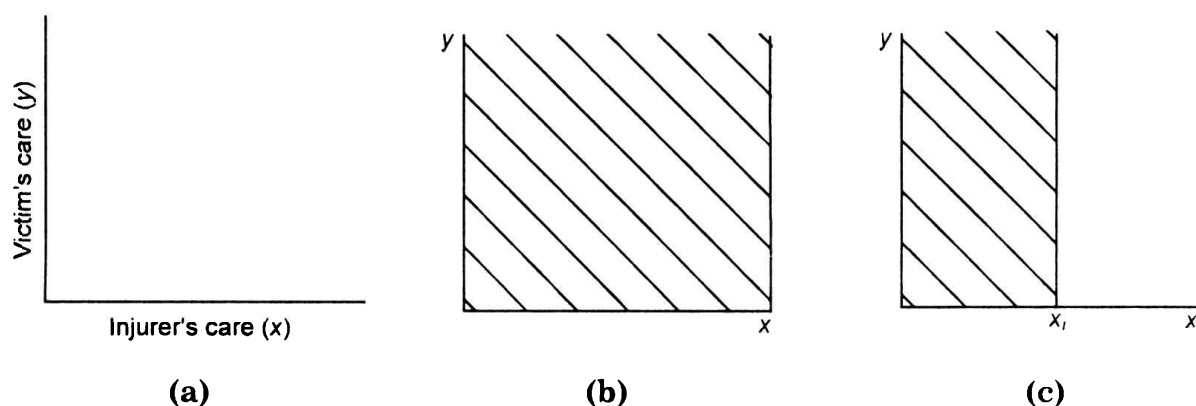
Under the negligence rule, the injurer will be liable for accident losses only if the level of care taken by him is less than an amount determined by the court. Let x_l be the legal standard of negligence for the injurer. This amount is also referred to as 'due care' or 'reasonable care'. For $x = x_l$, the injurer will not be liable. Now, suppose the victim chooses the socially optimal care $\bar{y}$. Let us then examine the possibility $\bar{x} = x_l$. Would the injurer choose $x > \bar{x}$? No, because

if $x = \bar{x}$, the injurer will not have any liability and it is expensive to undertake excessive care. Next, $x < \bar{x}$ is also ruled out, since this level of care is less than the legal standard and the injurer faces the total expected costs of the damage in such a case. When the injurer chooses the due care standard $x = \bar{x}$, the victim faces $C_v(y) + L(\bar{x}, y)$. Hence the victim will minimize this expression with respect to y and the solution is clearly $\bar{y}$. Therefore, each party choosing the socially optimal level of care is a Nash equilibrium under the negligence rule. Under standard assumptions this equilibrium is unique.

Under the strict liability with contributory negligence rule the injurer is liable unless the victim's level of care is the legal standard y_l . Clearly, if $y_l = \bar{y}$, then both parties will choose socially optimal levels of care.

Finally let us consider the case of the comparative negligence rule. Under this rule the injurer pays a fraction α of the loss $L(x, y)$. Therefore, he will minimize $C_i(x) + \alpha L(x, y)$ with respect to x . The solution will be lower than the socially optimal level of care $\bar{x}$. The victim will choose y to minimize $C_v(y) + (1 - \alpha)L(x, y)$ which will generate a level of care lower than $\bar{y}$.

From the above analysis it appears that if the court can set an appropriate level of legal standards for either party, then some version of the negligence rule will result in socially optimal decisions by both parties. To ensure that both parties exercise the socially optimal levels of care, the court can impose a cost of an accident on both the victim and the injurer. If the injurer is fined an amount equal to the cost he imposes on the victim, then he will have appropriate incentive to minimize the cost. Since the victim also bears a cost, which is equal to the cost imposed by the injurer, he will also have an incentive to avoid the accident. Observe that in this case the victim is not compensated by the injurer. Each of the parties must face the full cost of an accident for the scheme to work.



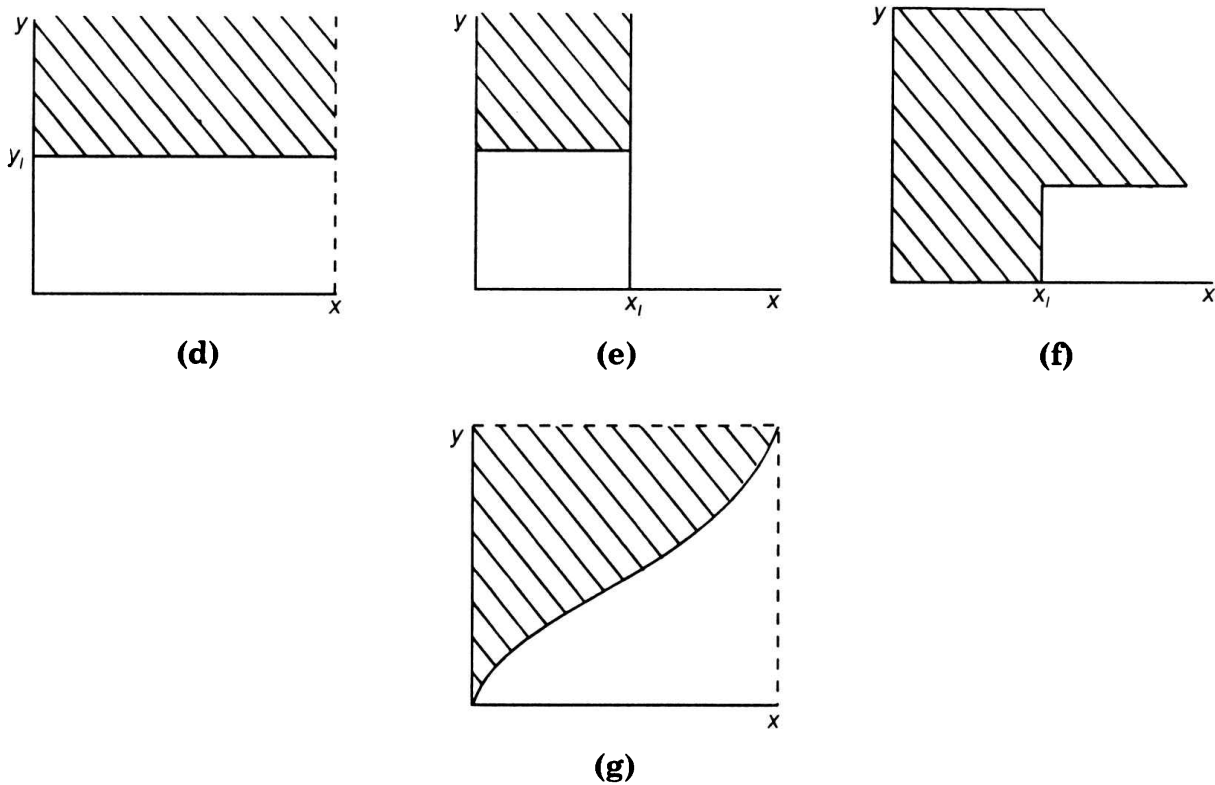


Fig. 26.1 Liability rules: (a) No Liability; (b) Strict Liability; (c) Negligence Rule; (d) Strict Liability with Contributory Negligence; (e) Negligence Rule with Contributory Negligence; (f) Strict Liability with Dual Contributory Negligence; (g) Relative Negligence.

We now present graphically the different liability rules, except that of comparative negligence. The injurer is liable in the scratched area. Thus in panel (a) (no liability), there is no scratched area since the injurer is never liable, but in panel (b) (strict liability), the entire area is scratched since the injurer is liable under all circumstances.

An important point that needs to be mentioned here is that the formulation we have considered in Eq. (26.14) is a bilateral formulation—both the injurer and the victim are involved. We can also consider a unilateral framework where the expected loss L to the victim from an accident depends on the level of care exercised by the injurer only. In such a case the social objective is to minimize the total loss imposed on both parties: $\min_x [C_i(x) + L(x)]$. It is easy to see that the strict liability rule and the negligence rule generate the socially optimal levels of care. However, under the strict liability rule the victim retrieves the total cost of an accident from the injurer, while under the negligence rule the

victim gets compensation only for actions in which the injurer did not exercise care. But it is rare that an accident occurs due to one party's fault only. Therefore, the formulation of (26.14) is more general than the unilateral formulation.

ANTITRUST VIOLATIONS AND MULTIPLE DAMAGES

Loosely speaking, antitrust laws are antimonopoly laws. These laws are the major policy vehicles that governments use to influence the ways in which firms compete with each other. They deal mainly with monopolistic practices, unfair competition, price fixing, merger of firms, and similar activities. In the US the Clayton Act, 1914 limits undertakings that tend to lessen competition substantially or create a monopoly. Mention may also be made of the Sherman Act. Section 2 of the Sherman Act, 1890 condemns monopoly that does not result from normal industrial development; it does not condemn monopoly that is achieved from competition on the merits. For the UK, Japan and India the corresponding legislations are the Monopolies and Restrictive Trade Practices (Inquiry and Control) Act, 1948; the Antimonopoly Law, 1947; and Monopolies and Restrictive Trade Practices (MRTP) Act, 1969. The MRTP Act was introduced to check the concentration of economic power, control the growth of monopolies and prevent restrictive and unfair trade practices. Restrictions were imposed on the expansion of enterprises, amalgamations and takeover of business enterprises including acquisition of shares falling within the purview of this Act.

We will now analyze a simple model of antitrust violation, developed by Salant (1987). Consider the case of a price-fixing conspiracy among sellers of a good produced at constant marginal cost c . That is, sellers collude to form a cartel. In the absence of antitrust enforcement, buyers will take as given the price they are charged and will select the amount they wish to purchase by equating the value of the marginal product of the good to its per unit price. Anticipating this reaction, the colluding sellers will charge a price that maximizes their profits. The resulting equilibrium is the solution to the following optimization problem

$$\max_p (p - c)q(p) \quad (26.15)$$

where $q(p)$, as a function of p , is the consumer demand function that the cartel faces. We denote the solution to this monopoly profit-maximization problem by (q_m, p_m) .

Now, suppose a buyer can sue the cartel and recover multiple damages if he succeeds in proving that the sellers violated antitrust laws by conspiring to fix

prices. Let t denote the damage multiple. That is, if suit is won, the buyers will receive a damage of t times the profit of the cartel. (In the US at present $t = 3$.) We denote the probability of winning the suit by π . Thus the expected damage is

$$d(q) = \pi t(p - c)q. \quad (26.16)$$

The cartel's optimization problem is now given by

$$\max_p (p - c)q(p) - d[q(p)]. \quad (26.17)$$

Substituting $d[q(p)]$ from Eq. (26.16) into the cartel's objective function in Eq. (26.17) we get the revised optimization problem

$$\max_p [1 - t\pi](p - c)q(p). \quad (26.18)$$

Clearly, for positivity of the cartel's profits we require $t\pi < 1$. Note that in the revised optimization problem the cartel has an incentive to set the price above the marginal cost.

Since $t\pi < 1$ is a constant, the solution to the above problem is also (q_m, p_m) . That is there is no change in the behavior of the cartel. The antitrust penalty that the cartel pays to the buyers is simply $t\pi$ times its profits. We can therefore interpret this penalty as a profit tax on the cartel.

If $t\pi > 1$, then for any $p > c$ the cartel would expect to pay in fines more than it receives in profits. It would then be optimal to set a price equal to the marginal cost. Therefore, the firms will not have the incentive to collude if the amount of damage and the probability of detection are high.

So far we have assumed that consumer demand is independent of the level of damages. But the buyers may have increased incentive to get damaged for collecting the multiple-damage provision of the antitrust law.

To see this, suppose that a consumer spends his entire income m on the consumption of the good produced by the cartel and a composite good. Assume without loss of generality that the price of the composite good is 1. Suppose also that the consumer has a quasilinear utility. Then the consumer's optimization problem is to maximize

$$U(q, y) = V(q) + y \quad (26.19)$$

$$\text{subject to} \quad pq + y = m \quad (26.20)$$

where y is the quantity of the composite good and the function V is increasing and strictly concave. Substituting y from Eq. (26.20) into Eq. (26.19) we rewrite the consumer's optimization problem as

$$\max_q U(q, y) = \max_q V(q) + m - pq. \quad (26.21)$$

When the consumer can bring suit against the cartel, he has an expected earning of $d(q)$ from the suit. Therefore the utility maximization problem now becomes

$$\max_q V(q) + m - pq + t\pi(p - c)q \quad (26.22)$$

which we can rewrite as

$$\max_q V(q) + m - [p - \pi t(p - c)]q. \quad (26.23)$$

The expression in the third bracket in Eq. (26.23) looks like a price in the consumer's maximization problem. The consumer pays p for each unit of the good he buys, but there is also an expected return of $\pi t(p - c)$ for each unit bought.

In view of the above we now write the cartel's optimization problem as

$$\max_p [1 - t\pi](p - c)q(\bar{p}) \quad (26.24)$$

where $\bar{p} = p - t\pi(p - c)$. Observing that $\bar{p} - c = p - t\pi(p - c) - c = (1 - t\pi)(p - c)$, we can rewrite Eq. (26.24) as

$$\max_{\bar{p}} (\bar{p} - c)q(\bar{p}) \quad (26.25)$$

Thus, the consumer and the cartel face the same effective price $\bar{p}$. Note that the optimization problem in Eq. (26.25) is the same as the monopoly profit maximization problem. Hence the value of $\bar{p}$ that maximizes Eq. (26.25) must be p_m . If we write p_1 for the price that the cartel actually charges, then

$$p_m = p_1 - t\pi(p_1 - c) \quad (26.26)$$

which demonstrates that $p_m < p_1$. Therefore, the cartel charges a price higher than the monopoly price, since it is required to pay some amount of money in the form of penalty. However, the effective price is exactly the same to the cartel and the consumer. The reason for having a higher price actually charged by the cartel here is that in this case we have incorporated the antitrust violation penalty in the consumer's choice problem.

CONTRACT LAW

A contract is an agreement implying a promise for a future exchange. The purpose of the law of contracts is to facilitate the process of exchange. If the

parties involved agree, the law makes the agreement a contract and for some types of agreements society provides legal enforcement remedies if one of the parties breaches the contract.

Following Hirsch (1979) we will now consider a simple framework which focuses on the activities and costs associated with the formation and performance of a contract. In this model there is a seller-creditor A who decides to sell all of his k goods to a buyer-borrower B . There are three stages in the transaction. At the initial stage t_0 , A has k goods and B has money or earning capacity but no good. In the contract formation stage t_1 , both A and B incur certain transaction costs. The two main components of these costs are the costs of negotiating the contract and the costs of preparing and signing the contract form.

Next, at the contract performance stage t_2 , both A and B incur some transaction costs in connection with performance. If the seller-creditor and the buyer-borrower perform as agreed in the contract, then these transaction costs can be bookkeeping costs and check-writing (collection) costs. In the absence of complete performance by both parties, additional costs can occur to A in each of three successive contract cases: (i) default is threatened and A decides to take self-help measures; (ii) default occurs and A takes self-help measures; and (iii) default occurs and A initiates judicial coercive action under due process. Likewise default-related transaction costs can occur for party B .

In connection with case (i); self-help can take the form of the seller-creditor threatening to take back the commodity or to initiate legal action. The associated cost can range from efforts at persuasion to lawyers' fees. For the buyer-borrower, the related transaction cost is virtually nil in most instances.

Once the buyer has defaulted (case [ii]), the seller will take self-help. For instance, suppose after due notice is given, the seller wishes to repossess the good, say, a car. If he does so, he will have to incur some transaction costs associated with the repossession, such as hiring a man for the job. Moreover, he may recover a value that is smaller than the outstanding debt. The value of the repossessed car will depend on the rate of depreciation and the down payment. This repossession deprives the buyer of use and resale of the item bought on credit. The repossessed car of a travelling salesman may make his living quite difficult.

For case (iii), both A and B incur court fees, attorney fees, time costs etc. The court may call for such execution technique as repossession or garnishment. Normally, the costs associated with judicial coercive collection are higher for

the buyer-borrower than for the seller-creditor, particularly in consumer transactions. The borrower may get his reputation damaged because of the court case. The costs of repossession or garnishment may be very high for him.

After performance is completed at the end of t_2 , A gets back k goods along with an interest payment I . The length of the period t_2 , which can vary from a day or two to a number of years, is important for various reasons. One reason relates to the risk of price level changes. Usually the longer t_2 is, the greater is the price-level risk. This risk goes hand-in-hand with the risk of default.

We can identify some strategies that can be adopted during the contract formation stage for reducing the transaction costs of a party. For instance, the seller-creditor, if the buyer-borrower defaults, can take certain steps that will increase the latter's costs, thus reducing the chances of default. These steps can be: (i) the buyer is required to pay his installments to a finance company which has bought the right to collect payments; (ii) the seller may insist on larger cash down-payments, (iii) the buyer-borrower can borrow items against some security deposit, default on a single payment may permit the seller to become owner of a part or whole of the deposit.

A major goal of contract law is expectation protection, that is, providing protection for agreed-upon expectations to be fulfilled. Another goal is the maintenance of incentives; parties should be encouraged to honor their promises. However, courts do not explicitly recognize this latter goal. The goal of expectation protection is quite clear—the offended party should be given a sum of money in damages to place him in as good a position as he would have enjoyed had the contract been honored. For instance, suppose according to a contract a party sells 100 pieces of an item at price p_1 on some day and on that day the market price is p_2 which is greater than p_1 . The buyer will want to enforce the contract whereas the seller will want to avoid enforcement. If the seller breaches, the buyer will purchase the pieces on the open market at price p_2 . The difference $(p_2 - p_1)$ would be the amount that would compensate the buyer for not obtaining the good at price p_1 .

But when will a contract be enforced by both parties? If parties with equal bargaining power negotiate about risks that both understand equally, then damages can be automatically ascertained in an organized market and contract will be enforced. That is, parties, with equal bargaining power have complete knowledge of each other's utilities. They also have substantial and similar knowledge of the risk involved in the transaction, as a whole. Then any bargained for allocation of risks incorporated in the contract should be enforceable. Intuitively the result is quite clear. No party has any market power

(bargaining powers are the same) and complete information exists. Consequently, under assumption of independent utilities we should have an optimal contract.

Applied Example 26.1 Probability of Conviction and Severity of Punishment

During the 18th and 19th centuries there was tendency in many Anglo-Saxon countries to punish those convicted of criminal offenses rather severely. But at the same time the probability of conviction and capture was set at very low values (see Radzinowicz, 1948). This tendency can be explained as follows. A high probability of conviction absorbs more public and private resources in the form of policemen, judges and juries etc. Therefore a decrease in this probability clearly reduces the expenditure on combating crime. Due to severity of punishment, more precisely, given that expected punishment remains unchanged, there is no offsetting increase in the amount of damages or the cost of punishments. The system of setting a small probability of conviction could, in fact, be a consequence of continuous political pressure to keep expenditure on police relatively low.

Appendix 26.1

MINIMIZATION OF SOCIAL LOSS FROM CRIME

We are to minimize

$$L = D(n_0) + C(p, n_0) + bfpn_0$$

with respect to f and p . Differentiating L partially with respect to f and p and setting the partial derivatives to zero, we get

$$\frac{\delta L}{\delta f} = \frac{\delta D}{\delta n_0} \frac{\delta n_0}{\delta f} + \frac{\delta C}{\delta n_0} \frac{\delta n_0}{\delta f} + bpf \frac{\delta n_0}{\delta f} + bpn_0 = 0$$

and

$$\frac{\delta L}{\delta p} = \frac{\delta D}{\delta n_0} \frac{\delta n_0}{\delta p} + \frac{\delta C}{\delta n_0} \frac{\delta n_0}{\delta p} + \frac{\delta C}{\delta p} + bpf \frac{\delta n_0}{\delta p} + bfn_0 = 0.$$

We rewrite the above equations as

$$\frac{\delta L}{\delta f} = D' n_{of} + C' n_{of} + bpf n_{of} + bpn_0 = 0$$

and
$$\frac{\delta L}{\delta p} = D' n_{op} + C' n_{op} + C'_p + bpf n_{op} + bfn_0 = 0.$$

In n_{op} and n_{of} are not zero, we can divide these expressions by them and recombine terms to get

$$D' + C' = -bpf \left(1 + \frac{n_0}{fn_{of}} \right)$$

and
$$D' + C' + C'_p \frac{1}{n_{op}} = -bpf \left(1 + \frac{n_0}{pn_{op}} \right).$$

These two equations, on slight rearrangement, yield

$$D' + C' = -bpf \left[1 - \frac{1}{(-f/n_0)n_{of}} \right],$$

$$D' + C' + C'_p \frac{1}{n_{op}} = -bpf \left[1 - \frac{1}{(-p/n_0)n_{op}} \right]$$

which are the required conditions.

In order to express these conditions in the forms given by Eqs. (26.10) and (26.11) we rewrite the derivatives in terms of Δ 's. That is, replace $D' = \frac{\delta D}{\delta n_0}$ by $[D(n_0 + \Delta n_0) - D(n_0)]/\Delta n_0 = \Delta D/\Delta n_0$ and so on. Then from the last two equations we get

$$\frac{\Delta D}{\Delta n_0} + \frac{\Delta C}{\Delta n_0} = -bpf \left(1 - \frac{1}{(-f/n_0) \frac{\Delta n_0}{\Delta f}} \right), \quad (26.27)$$

$$\frac{\Delta D}{\Delta n_0} + \frac{\Delta C}{\Delta n_0} + \frac{\Delta C/\Delta p}{\Delta n_0/\Delta p} = -bpf \left(1 - \frac{1}{(-p/n_0) \frac{\Delta n_0}{\Delta p}} \right)$$

which are the conditions given by Eqs. (26.10) and (26.11).

MINIMIZATION OF SOCIAL LOSS FROM ACCIDENT

We are to minimize

$$S = C_i(x) + C_v(y) + L(x, y)$$

with respect to x and y .

Differentiating S partially with respect to x and y and setting the derivatives to zero, we get

$$\frac{\delta S}{\delta x} = \frac{dC_i(x)}{dx} + \frac{\delta L(x, y)}{\delta x} = 0,$$

$$\frac{\delta S}{\delta y} = \frac{dC_v(y)}{dy} + \frac{\delta L(x, y)}{\delta y} = 0.$$

We rewrite the above equations as

$$\frac{dC_i(x)}{dx} = - \frac{\delta L(x, y)}{\delta x},$$

$$\frac{dC_v(y)}{dy} = - \frac{\delta L(x, y)}{\delta y}.$$

Now $\frac{dC_i(x)}{dx}$ is the marginal cost to the injurer when he increases care by one unit. Since $-L(x, y)$ can be regarded as a benefit, $-\delta L(x, y)/\delta x$ is the marginal benefit enjoyed by the victim due to the decrease in the expected cost of an accident. We can interpret the second equation analogously.

EXERCISES

Multiple Choice

1. In the case of a unilateral accident law, under the negligence rule, the injurer's level of care is
 - a) Less than the optimal level.
 - b) Greater than the optimal level.
 - c) Equal to the optimal level.
 - d) Any one of the above is possible.
2. For minimization of social loss from offenses we require the elasticity of offences with respect to the probability of conviction to be
 - a) Less than unity.
 - b) Equal to unity.
 - c) Greater than unity.
 - d) Any one of the above is possible.
3. The total social cost of punishments to offenders can be approximated by the cost to offenders
 - a) Plus the cost to others.
 - b) Minus the cost to others.
 - c) Minus the gain to others.
 - d) (a) or (c).
4. The expected utility of an offender is
 - a) An increasing function of the probability of conviction.
 - b) A decreasing function of the probability of conviction.
 - c) Not monotonic in the probability of conviction.
 - d) Nothing can be concluded unambiguously.
5. Under strict liability in the unilateral accident law framework, the injurer
 - a) Bears the full cost of an accident.
 - b) Does not bear the cost of accident at all.
 - c) Bears only an insignificant portion of the cost of an accident.
 - d) Bears a major portion of the cost of an accident.

Quantitative Problems and Questions

1. Give some examples to show affinity between law and economics.
2. What strategies can be taken for reducing the transaction costs of a seller-creditor during a contract formation stage?
3. Suppose that the constant marginal costs of taking care by the victim and the injurer are $1/4$ and $1/9$ respectively. If the expected loss from an accident is $L(x, y) = 1/x + 1/y$, where $x(y)$ is the level of care taken by the injurer (victim), calculate the optimal values of x and y .
4. Prove or disprove the following:
 - a) In the case of a bilateral accident law, under the negligence rule, each party choosing the socially optimal level of care is a Nash equilibrium.
 - b) The elasticity of the number of offenses (n_0) with respect to the probability of conviction (p) is greater than the elasticity of n_0 with respect to the expected punishment from offense (f) if and only if the marginal revenue from p is less than that from f .
 - c) In the case of a bilateral accident law under the comparative negligence rule, optimality requires that each party pay fifty percent of the loss of an accident.
 - d) If parties with equal bargaining power negotiate about risks that both understand equally, then a contract can be enforced.
 - e) If a cartel is found guilty of price fixing then consumers will be able to collect more damages the more they consume.
5. 'Crime does not pay' may be regarded as an optimality condition. Use Becker's 'crime and punishment' framework to analyze this statement. Explain the role of attitudes towards risk in this context.
6. Give some examples to illustrate how crimes are motivated by economic considerations.
7. Use graphs to explain alternative liability rules.
8. Clearly derive the social loss from a criminal offense as a function of the net damage done to the society, the cost of conviction, the punishment cost and the number of offenses.
9. State and explain some conditions that are sufficient to ensure that Nash equilibrium of cares adopted under the negligence rule (bilateral accident law) is unique.

10. Show that if consumer demand does not change with the level of damages, the antitrust penalty that a cartel faces is equivalent to a profit tax. How does your answer change if consumers behave strategically, that is, if they attempt to get damaged in order to collect penalty?
11. Suppose that the constant marginal costs of taking care by the victim and the injurer are $1/4$ and $1/9$ respectively. The expected loss from an accident is $L(x, y) = 1/x + 1/y$. The two parties share the resulting loss from an accident. If the optimal care taken by the victim is 1, determine the fraction of costs he is required to pay. What is the injurer's optimal care?
12. Suppose that a cartel faces the market demand curve $X = 53 - p$ and the cartel's constant marginal cost of production is 5. Assume that a consumer's probability of winning a suit against the cartel for conspiring to fix prices is $1/6$ and if the suit is won the consumer receives damages three times the profits of the cartel. What is the cartel's actual price?
13. Harms from criminal activities can be regarded as externalities. Discuss this statement very briefly.

27

Mathematics

The purpose of this chapter is to provide a brief review of some of the mathematical concepts that are used in the text. Since the discussion aims only at clarifying the concepts, it will not be very rigorous.

FUNCTIONS

By a function or transformation we mean a rule that gives a relationship between numbers. Examples of functions are $y = \log x$, $y = x^2$. In these examples we say that ‘ y is a function of x ’. In each case, corresponding to each number x we get a unique number y according to the rule of association specified in the form of the function.

If we denote the rule of association by f , then we write $y = f(x)$ to mean that y depends on x according to the rule f . In the above examples, $f(x) = \log x$ and $f(x) = x^2$.

Given a function $y = f(x)$, the variable x is called, an independent variable and y is called a dependent variable. These terms are based on the notion that x can vary independently but the value of y depends on the value of x . The variable x is also referred to as the argument of the function $y = f(x)$.

A function may not be defined for all values of x . For instance, $f(x) = \log x$ cannot be evaluated for $x \leq 0$. However, the function $f(x) = |x|$, where $|x|$ is the modulus of x , that is, $|x| = x$ if $x \geq 0$ and $|x| = -x$ if $x < 0$, is defined for all values of x . The set of numbers for which a function is defined is called the domain of the function. The set of values of a function (corresponding to the x values) in the domain is called the range of the function. Thus, for the function $f(x) = \log x$, the domain is the set of all positive numbers and the range is the set of all possible numbers (positive, negative and zero).

The relation $y = f(x)$ is an explicit function, since y is expressed explicitly in terms of x . If the functional relationship between y and x is denoted by $g(y, x) = 0$, y is an implicit function of x . Specification of a particular value of x implicitly defines a value of y such that the expression on the left hand side becomes zero when the appropriate values of x and y are substituted. Thus, the expression $y - x^2 = 0$ is an implicit function, while $y = x^2$ is an explicit function.

By equating a function to some particular number, we get an equation. For instance if we equate the function $f(x) = x^2$ to some number θ , say 5, then we get the equation $x^2 = 5$. A value of x that satisfies the equation is called a solution to the equation. Thus, for the equation considered here there are two solutions $x = \sqrt{5}, -\sqrt{5}$.

An identity is a relationship between variables that holds for all values of the variables. For instance, $x + y \equiv y + x$ is an identity. The symbol $\equiv$ means that the left hand side and the right hand side are equal for all values of the two variables. Similarly, $(x + 1)^2 \equiv x^2 + 2x + 1$ is an identity, since it is true for all values of x .

PROPERTIES OF FUNCTIONS

A function is said to be increasing if its value increases when the independent variable increases. More precisely, if the value of x increases from x_1 to x_2 , then increasingness of f requires that $f(x_2) > f(x_1)$. Similarly, f is decreasing if $f(x_2) < f(x_1)$, where $x_2 > x_1$. Thus the function $f(x) = -x$ is decreasing for all values of x . (In some texts increasingness and decreasingness are referred to as monotonic increasingness and monotonic decreasingness respectively. It is then said that a function is monotonic or monotone if it is monotonically increasing or monotonically decreasing.) Clearly, for an increasing or a decreasing function there will be a unique value of x for each value of y . A function that relates x to y in this fashion is called an inverse

function. Thus, an inverse function expresses y as a function of x . For instance, the inverse of the function $y = f(x) = \log x$ is the exponential function $x = e^y$, where e is the base of natural logarithms (which is equal to 2.7183 . . .). For all values y_1, y_2 of y , the function e^y satisfies $e^{y_1+y_2} = e^{y_1} \cdot e^{y_2}$. An example of a function for which an inverse does not exist is $y = x^2$. (in this case $\sqrt{y} = x$ and $-\sqrt{y} = x$ have the property that y becomes equal to x^2 .)

A function f is convex over its domain if

$$f[\theta x_1 + (1 - \theta)x_2] \leq \theta f(x_1) + (1 - \theta) f(x_2) \quad (27.1)$$

for all x_1, x_2 in the domain of f , and all $0 \leq \theta \leq 1$. It is strictly concave if the inequality in Eq (27.1) becomes strict for all $x_1 \neq x_2$, $0 < \theta < 1$. A function f is concave over its domain if

$$f[\theta x_1 + (1 - \theta)x_2] \geq \theta f(x_1) + (1 - \theta) f(x_2) \quad (27.2)$$

for all x_1, x_2 in the domain of f , and all $0 \leq \theta \leq 1$. It is strictly concave if the inequality in Eq (27.2) becomes strict for all $x_1 \neq x_2$, and all $0 < \theta < 1$.

We can illustrate the above concepts graphically. The graph of a function is a pictorial representation of the function. The dependent variable is usually plotted on the vertical axis and the independent variable on the horizontal axis. The graph then shows the relationship between the dependent and independent variables.

Let x_1 and x_2 , where $x_1 < x_2$, be two values of the independent variable. Then $\bar{x} = \theta x_1 + (1 - \theta)x_2$ is a weighted average of the values of x_1 and x_2 , where the weights are given by θ and $(1 - \theta)$ respectively. This weighted average is popularly known as a convex combination of x_1 and x_2 . Clearly, $x_1 < \bar{x} < x_2$. We denote the functional values $f(x_1)$, $f(x_2)$ and $f(\bar{x})$ by y_1 , y_2 and $\bar{y}$ respectively. If we denote the weighted average $\theta f(x_1) + (1 - \theta) f(x_2)$ by $\hat{y}$, then strict convexity of f requires that $\bar{y} < \hat{y}$. In Fig. 27.1 we have joined the points $A = (x_1, y_1)$ and $B = (x_2, y_2)$ by a straight line. The point C on the graph has the coordinates $(\bar{x}, \bar{y})$. Moving vertically up from C we hit the line AB at D whose coordinates are $(\bar{x}, \hat{y})$.

From Fig. 27.1 we see that the vertical coordinate of D lies above that of C , that is, $\bar{y} < \hat{y}$, which is demanded by strict convexity. Since we want this condition to hold for all $0 < \theta < 1$, the entire line AB will lie above the graph. This condition holds for all values of the dependent variable. We can therefore say that a function is strictly convex if the straight line joining any two points on the graph of the function lies entirely above the graph. Examples of strictly convex functions are $f(x) = x^2$ and $f(x) = e^x$.

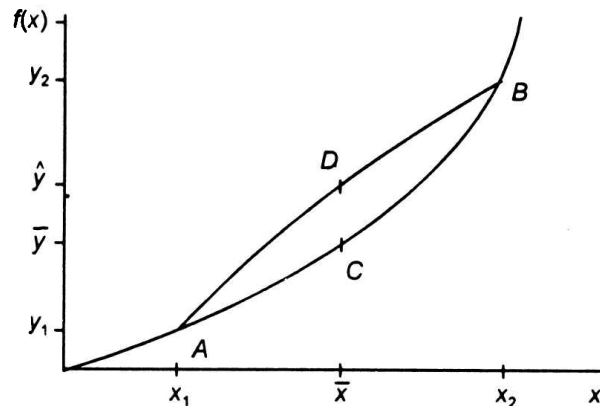


Fig. 27.1 Strictly Convex Function

Similarly, strict concavity of a function means that the straight line joining any two points on the graph of the function lies entirely below the curve. Examples of strictly concave functions are $f(x) = x^{1/2}$ and $f(x) = \log x$.

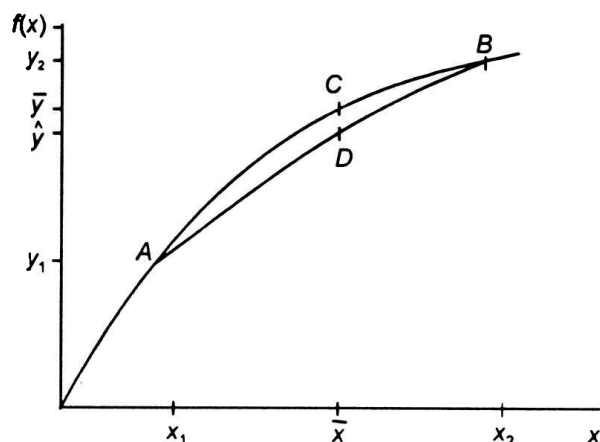


Fig. 27.2 Strictly Concave Function

Functions that are both convex and concave, that is, functions satisfying the inequalities of Eqs. (27.1) and (27.2), are called linear. Linear functions are neither strictly convex nor strictly concave. Examples of linear functions are $y = 5x + 3$ and $y = 20x$. In general, any function of the form $y = ax + b$, where a and b are constants, is linear. If the constant a is positive, then the function $y = ax + b$ is also called an affine function. Note that the constant b is the vertical intercept of the linear function $y = ax + b$. (This intercept is the value of the function when $x = 0$.) On the other hand, the horizontal intercept, which is the value of x when $y = 0$, is given by $-b/a$.

A function f is said to quasiconcave over its domain if

$$f[\theta x_1 + (1 - \theta)x_2] \geq \min[f(x_1), f(x_2)] \quad (27.3)$$

for all x_1, x_2 in the domain of f , and all $0 \leq \theta \leq 1$. It is strictly quasiconcave if the inequality in Eq. (27.3) becomes strict for all $x_1 \neq x_2$ and all $0 < \theta < 1$.

To illustrate quasiconcavity graphically let us consider the graphical representation of the linear function $y = f(x) = x$. Consider two values x_1 and x_2 of the independent variable x . Then the weighted average $\bar{x} = \theta x_1 + (1 - \theta)x_2$ lies between x_1 and x_2 . We write y_1, y_2 and $\bar{y}$ for the functional values $f(x_1), f(x_2)$ and $f(\bar{x})$ respectively. Clearly, $\min(y_1, y_2) = y_1$. From Fig. 27.3 it follows that $\bar{y} > y_1$, which is demanded by strict quasiconcavity. Therefore, the function $f(x) = x$ is strictly quasiconcave. This example shows that a quasiconcave function need not be concave. Given that for any $0 \leq \theta \leq 1$, $\theta f(x_1) + (1 - \theta)f(x_2) \geq \min[f(x_1), f(x_2)]$, comparing Eqs.(27.2) and (27.3) it turns out that a concave function is always quasiconcave.

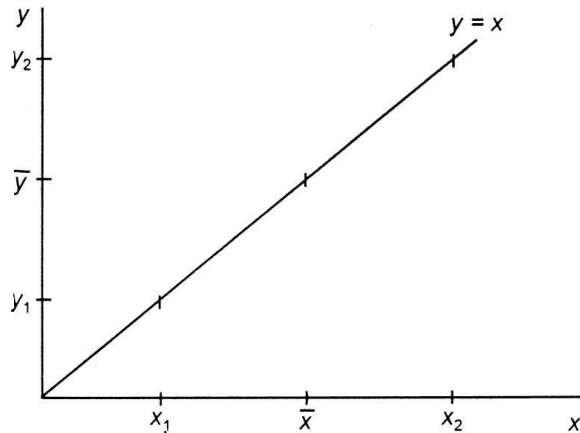


Fig. 27.3 Strictly Quasiconcave Function

An advantage of quasiconcavity is that it remains preserved under any increasing transformation. To illustrate this consider the strictly quasiconcave function $y = f(x) = x$. It is easy to see that the increasing transformation $y^4 = [f(x)]^4 = x^4$ of $y = x$ is also strictly quasiconcave. (Observe that the transformation y^4 is in fact a composite mapping in the sense that we can write it as $h(y) = h[f(x)]$, where $h(y) = y^4$. That is, we write y^4 as a composition of functions f and h .) However, a concave function does not possess this preservation property. To see this let us choose the strictly concave function $y = f(x) = x^{1/2}$. Then the increasing transformation $y^4 = [f(x)]^4 = x^2$ of $y = x^{1/2}$ becomes strictly convex rather than being strictly concave.

CONTINUITY AND DERIVATIVES

A function is said to be continuous if its graph does not contain any jump. Thus, the graph of a continuous function can be drawn without lifting pencil from paper. For instance, the function $f(x) = x$ is continuous.

Assume that the function $y = f(x)$ is continuous. Suppose that Δx denotes a marginal change in the independent variable x . That is, Δx is a small change in x . Then the value of the function will change by the quantity Δy . Hence $y + \Delta y = f(x + \Delta x)$. We can write the change in the value of the function as

$$\Delta y = f(x + \Delta x) - f(x). \quad (27.4)$$

Dividing both sides of Eq (27.4) by Δx , we get the rate of change of y per unit of change of x :

$$\frac{\Delta y}{\Delta x} = \frac{f(x + \Delta x) - f(x)}{\Delta x}. \quad (27.5)$$

This rate of change measures how y changes when x changes by 1 unit.

The derivative of $y = f(x)$, denoted by dy/dx , or $f'(x)$, is defined as the rate of change of $f(x)$ as Δx approaches zero:

$$\frac{dy}{dx} = f'(x) = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}. \quad (27.6)$$

That is, the derivative is the limit of the rate of change of y with respect to x as the change in x goes to zero. If the graph of $f(x)$ is drawn, the derivative calculated at $x = a$ is the slope of the curve representing $f(x)$ at $x = a$. For the function whose graph is shown in Fig. 27.4, the rate of change defined in Eq (27.5) is the ratio BC/AC . The derivative is the slope of the tangent DAE to the curve at a given point $x = a$. (A tangent to a function at some point $x = a$ is a linear function that touches the curve at $x = a$.) That is, the slope of the tangent DAE to the curve at $x = a$ is the limiting value of the ratio BC/AC .

If a function has a positive derivative at some point, then it is increasing around that point. In contrast a function possessing a negative derivative at some point will be decreasing around that point. If the derivative is zero at some point, then the function is a constant near that point.

The second derivative of a function is the derivative of the derivative of that function. If $y = f(x)$, the second derivative of $f(x)$ with respect to x , denoted by d^2y/dx^2 , or $f''(x)$, is

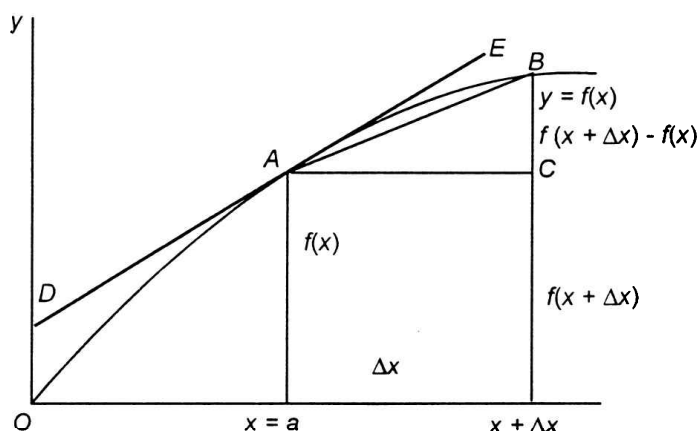


Fig. 27.4 Derivative of a Function

$$\frac{d^2y}{dx^2} = f''(x) = \lim_{\Delta x \rightarrow 0} \frac{f'(x + \Delta x) - f'(x)}{\Delta x}. \quad (27.7)$$

The second derivative, which is the rate of change of the first derivative, measures the curvature of a function. A function with a negative second derivative at some point is strictly concave around that point; its slope is decreasing. A function with a positive second derivative at some point is strictly convex around that point; its slope is increasing. A function with a zero second derivative at some point is flat near that point.

RULES FOR FINDING DERIVATIVES

The repeated application of the definition of a derivative can be tedious. Therefore several elementary derivatives are presented below. Proofs of these results can be found in any elementary book on calculus.

1. If c is a constant, then for the function $y = c$,

$$\frac{dy}{dx} = \frac{dc}{dx} = 0. \quad (27.8)$$

2. If a is a constant, then for $y = x^a$,

$$\frac{dy}{dx} = \frac{dx^a}{dx} = ax^{a-1} \quad (27.9)$$

3. If $y = \log x$, then

$$\frac{dy}{dx} = \frac{1}{x}. \quad (27.10)$$

4. If $y = a^x$, where $a > 0$ is a constant, then

$$\frac{dy}{dx} = a^x \log a. \quad (27.11)$$

A particular case of this rule is $de^x/dx = e^x$.

Suppose that $f(x)$ and $g(x)$ are two functions of x . Then

$$5. \frac{d[f(x) + g(x)]}{dx} = f'(x) + g'(x).$$

Similarly $\frac{d}{dx} [f(x) - g(x)] = f'(x) - g'(x)$.

$$6. \frac{d[f(x)g(x)]}{dx} = f'(x)g(x) + f(x)g'(x).$$

Rule 6 is the product rule of differentiation.

$$7. \frac{d}{dx} \left[\frac{f(x)}{g(x)} \right] = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}$$

provided $g(x) \neq 0$.

Finally, if $y = g(z)$, $z = h(x)$, so that $y = g[h(x)] = f(x)$ (say), then

$$8. \frac{dy}{dx} = \frac{dy}{dz} \frac{dz}{dx} = \frac{dg(z)}{dz} \cdot \frac{dh(x)}{dx}.$$

This result is called the chain rule, which shows how one variable (x) affects another variable (y) through its influence on some intermediate variable (z). For instance, let $y = g(z) = \log z$ and $z = h(x) = x^2$. Then the composite function is $y = f(x) = g(h(x)) = \log x^2$. By the chain rule, $df(x)/dx = \frac{1}{z} \cdot 2x = \frac{2x}{x^2} = 2/x$.

PARTIAL DERIVATIVES AND TOTAL DIFFERENTIAL

Often a variable y depends on more than one variable. For instance, a firm's production function depends on the quantity of labor, capital and land devoted to the production process. In these circumstances this dependence of the variable y on a series of variables $x_1, x_2, \dots, x_n$ is denoted by $y = f(x_1, x_2,$

$\dots, x_n$). Then the partial derivative of $f(x_1, \dots, x_n)$ with respect to x_1 is just the derivative of the function f holding $x_2, x_3, \dots, x_n$ fixed. More precisely, the partial derivative with respect to x_1 is

$$\frac{\delta y}{\delta x_1} = \frac{\delta f(x_1, \dots, x_n)}{\delta x_1} = \lim_{\Delta x_1 \rightarrow 0} \frac{f(x_1 + \Delta x_1, \dots, x_n) - f(x_1, x_2, x_3)}{\Delta x_1} \quad (27.12)$$

This rule holds for the other variables $x_2, x_3, \dots, x_n$ also. Therefore, in general the partial derivative of $f(x_1, x_2, \dots, x_n)$ with respect $x_i, i = 1, \dots, n$, is

$$\begin{aligned} \frac{\delta y}{\delta x_i} &= \frac{\delta f(x_1, \dots, x_i, \dots, x_n)}{\delta x_i} \\ &= \lim_{\Delta x_i \rightarrow 0} \frac{f(x_1, \dots, x_{i-1}, x_i + \Delta x_i, x_{i+1}, \dots, x_n) - f(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n)}{\Delta x_i} \end{aligned} \quad (27.13)$$

As an illustrative example suppose $y = f(x_1, x_2) = ax_1^2 + bx_1x_2 + cx_2^2$. Then $\delta y/\delta x_1 = 2ax_1 + bx_2$, $\delta y/\delta x_2 = bx_1 + 2cx_2$.

As a first step for defining the total differential of a function that depends on several variables, let us observe from Eq. (27.6) that for a single variable function $y = f(x)$ we can write dy as

$$dy = f'(x)dx. \quad (27.14)$$

The identity of Eq. (27.14) records the fact that the change in $y(dy)$ is equal to the change in $x(dx)$ times the slope of the function $[f'(x)]$.

This analogy applies to a several variable function $y = f(x_1, \dots, x_n)$ as well. Suppose we are changing only one of the x_i 's, say x_1 by dx_1 , while holding all the others constant. The change in y (that is, dy) that would result from this change in x_1 is given by

$$dy = \frac{\delta f(x_1, \dots, x_n)}{\delta x_1} dx_1 = f_1 dx_1 \quad (27.15)$$

where f_1 is the partial derivative of $f(x_1, x_2, \dots, x_n)$ with respect to x_1 . This says that the change in y is equal to the change in x_1 times the slope measured in the x_1 direction.

If all the x_i 's are varied by small amounts, the total effect on y will be the sum of the effects such as shown in Eq. (27.15). Therefore, the total change in y is defined to be

$$dy = f_1 dx_1 + f_2 dx_2 + \dots + f_n dx_n \quad (27.16)$$

where f_i is the partial derivative of $f(x_1, x_2, \dots, x_n)$ with respect to x_i and dx_i is the change in x_i , $i = 1, 2, \dots, n$. The expression given by Eq. (27.16) is called the total differential of y (or f) and is analogous to the expression for the single variable case given by Eq. (27.14). This equation is intuitively reasonable: The total change in y is the sum of changes brought about by varying each of x_i 's.

The total differential in Eq. (27.16) can be used to demonstrate the chain rule for a several variable function. Suppose $y = f(x_1, x_2)$ and $x_1 = g(z)$ and $x_2 = h(z)$ so that $y = f[g(z), h(z)]$.

The total differential of $y = f(x_1, x_2)$ is

$$dy = f_1 dx_1 + f_2 dx_2 \quad (27.17)$$

where f_i is the partial derivative of $f(x_1, x_2)$ with respect to x_i , $i = 1, 2$. Dividing this equation by dz we get

$$\begin{aligned} \frac{dy}{dz} &= f_1 \frac{dx_1}{dz} + f_2 \frac{dx_2}{dz} \\ &= f_1 \frac{dg(z)}{dz} + f_2 \frac{dh(z)}{dz}. \end{aligned} \quad (27.18)$$

Hence calculation of the effect of z on y requires examination of how z affects the determinants x_1 and x_2 of y .

HOMOGENEOUS FUNCTIONS

A function f depending on n variables, $x_1, x_2, \dots, x_n$, is said to be homogeneous of degree k if for all $(x_1, x_2, \dots, x_n)$ and for all $t > 0$, $f(tx_1, tx_2, \dots, tx_n) = t^k f(x_1, x_2, \dots, x_n)$. The most common cases of homogeneous functions are $k = 0$ and $k = 1$. If f is homogeneous of degree zero, doubling all its arguments leaves f unchanged. If f is homogeneous of degree one, doubling all its arguments will double the value of f . Sometimes a homogeneous function of degree one is also called a linear homogeneous function. We encounter functions of homogeneous of degree one in the chapter on the theory of production. In our discussion on consumer behavior, we show that all demand functions are homogeneous of degree zero.

If a function f is homogeneous of degree k , that is, if $f(tx_1, tx_2, \dots, tx_n) = t^k f(x_1, \dots, x_n)$ for all $(x_1, \dots, x_n)$ and for all $t > 0$, then

$$x_1 \frac{\delta f(x_1, \dots, x_n)}{\delta x_1} + x_2 \frac{\delta f(x_1, \dots, x_n)}{\delta x_2} + \dots + x_n \frac{\delta f(x_1, \dots, x_n)}{\delta x_n} = k f(x_1, \dots, x_n). \quad (27.19)$$

This is known as Euler's theorem for homogeneous functions. In fact, the converse is also true, that is, if Eq. (27.19) holds for all $(x_1, \dots, x_n)$, then f must be homogeneous of degree k .

OPTIMIZATION

If $y = f(x)$, then f achieves a maximum at a , if $f(a) \geq f(x)$ for all x . If $f(x)$ achieves its maximum at $x = a$, then $df'(a)/dx = 0$ and $d^2f(a)/dx < 0$. These conditions are referred to as the first order and second order conditions for a maximum. According to the first order condition the tangent to the graph of f at a must be horizontal. Therefore, the function f must be flat at $x = a$. Such points where $f'(a) = 0$ are called stationary points of f . The second order condition means that the function is strictly concave around a . Therefore, for a function $y = f(x)$ which is strictly concave over its domain, $x = a$ is maximum of f if $f'(a) = 0$.

A function f is said to achieve its minimum at $x = a$ if $f(a) \leq f(x)$ for all x . The first and second order conditions for $f(x)$ to achieve its minimum at $x = a$, are $df(a)/dx = 0$ and $d^2f(a)/dx > 0$. Here also the first order condition means that the tangent to the graph of f at a is horizontal. The second order condition says that f is strictly convex around a . Thus we claim that for a function $y = f(x)$, which is strictly convex over its domain, $x = a$ is a minimum of f if $f'(a) = 0$.

By an argument similar to that given for the case of one variable, the first order conditions for $y = f(x_1, x_2)$ to achieve the maximum at some point (a, b) are

$$\frac{\delta f(a, b)}{\delta x_1} = 0, \quad (27.20)$$

$$\frac{\delta f(a, b)}{\delta x_2} = 0, \quad (27.21)$$

Second order conditions are required to ensure that the point (a, b) is a true maximum. Given the complex nature of these conditions, we do not wish to discuss them here. It may be mentioned that most functions we encounter

in this book obey these restrictions and we are therefore interested only in applying the first order conditions.

So far attention has been centered on finding the maximum value of the function f without restricting the choice of the independent variable(s). In most economic problems all values of the independent variable(s) are not permissible. For instance, in choosing items to consume, an individual is not able to choose any quantities desired. Rather, choices are constrained by the consumer's income. This kind of problem is referred to as the constrained optimization problem.

The general way of writing such an optimization problem is

$$\max_{x_1, x_2, \dots, x_n} f(x_1, x_2, \dots, x_n)$$

such that $g(x_1, x_2, \dots, x_n) = C$. The function $f(x_1, x_2, \dots, x_n)$ is called the objective function and $g(x_1, x_2, \dots, x_n) = C$ is called the constraint. A solution to the above problem is $(a_1, a_2, \dots, a_n)$ such that $f(a_1, a_2, \dots, a_n) \geq f(x_1, x_2, \dots, x_n)$ for all values of $x_1, x_2, \dots, x_n$ that meet the condition $g(x_1, x_2, \dots, x_n) = C$. Methods for solving the constrained maximization problems using the Lagrangian multiplier method, which involves a mathematical technique that has a useful economic interpretation, are discussed in the appendices to several chapters, for instance Chapters 3 and 9.

CLOSED AND OPEN SETS

A set is said to be closed if it contains its boundaries. For instance, the set $A = [0, 1] = \{x : 0 \leq x \leq 1\}$ is closed, since it contains its lower and upper boundaries 0 and 1. On the other hand, a set is open if it does not contain its boundaries. Thus the set $B = (0, 1) = \{x : 0 < x < 1\}$ is open, since it does not contain its boundaries.

References

1. Akerlof, G. (1970). The Market for Lemons: Quality Uncertainty and The Market Mechanism, *Quarterly Journal of Economics*, 84.
2. Arrow, K.J. (1951). An Extension of the Basic Theorems of Classical Welfare Economics, in J. Neyman (ed.) Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, *University of California Press*, Berkeley.
3. Arrow, K.J. (1951a). Social Choice and Individual Values, *Wiley*, New York.
4. Arrow, K.J. (1971). Essays in The Theory of Risk Bearing, *North-Holland*, Amsterdam.
5. Arrow, K.J., H.B. Chenery, B.S. Minhas and R.M. Solow (1961). Capital Labour Substitution and Economic Efficiency, *Review of Economics and Statistics*, 43.
6. Bain, J.S. (1951). Relation of Profit Rate of Industrial Concentration: American Manufacturing 1396-40, *Quarterly Journal of Economics*, 65.
7. Baumol, W.J. (1967). Business Behaviour, Value and Growth, Revised Edition, *Harcourt*, New York.
8. Baumol, W.J., J.C. Panzar and R.D. Willig (1982). Contestable Markets and The Theory of Industry, *Harcourt*, San Diego.
9. Baye, M.R., M.E. Deily and D.W. Jansen (1991). Marginal and Total Production Cost Indices: Theory and Applications, *Journal of Productivity Analysis*, 2.
10. Becker, G.S. (1964). Human Capital, *Columbia University Press*, New York.
11. Becker, G.S. (1968). Crime and Punishment: An Economic Approach, *Journal of Political Economy*, 76.

12. Bergson, A. (1938). A Reformulation of Certain Aspects of Welfare Economics, *Quarterly Journal of Economics*, 52.
13. Bertrand, J. (1883). Review of Cournot, *Journal des Savants*.
14. Black, D. (1958). The Theory of Committees and Elections, *Cambridge University Press*, London.
15. Blackorby, C., D. Donaldson and J.A. Weymark (1984). Social Choice Theory with Interpersonal Utility Comparisons: A Diagrammatic Introduction, *International Economic Review*, 25.
16. Bliss, C. (1987). Distribution Theory, in J. Eatwell, M. Milgate and P. Neuman (eds.) The New Palgrave: A Dictionary of Economics, Vol. I, *Macmillan*, London.
17. Boadway, R. and N. Bruce (1984). Welfare Economics, *Blackwell*, Oxford.
18. Brown, J.P. (1973). Toward an Economic Theory of Liability, *Journal of Legal Studies*, 2
19. Buchanan, J.M. (1965). An Economic Theory of Clubs, *Economica*, 32.
20. Carlton, D.W. and J.M. Perloff (1990). Modern Industrial Organization, *Scott-Foresman*, Illinois.
21. Chakravarty, S.R. (1990). Ethical Social Index Numbers, *Springer-Verlag*, New York.
22. Chakravarty, S.R. (1996). Issues in Industrial Economics, Second Print, *Avebury*, Aldershot.
23. Chakravarty, S.R. and N. Chattopadhyay (1994). On Indices of Marginal and Total Production Cost, *Statistical Papers*, 35.
24. Chamberlin, E. (1956). The Theory of Monopolistic Competition, *Harvard University Press*, Cambridge.
25. Chattopadhyay, M., C. Neogi and S. Maiti (1993). Growth and Instability in Crop Production in Eastern India, *Asian Economic Review*, 35.
26. Clarke, E.H (1971). Multipart Pricing of Public Goods, *Public Choice*, 11.
27. Coase, R. (1960). The Problem of Social Cost, *Journal of Law and Economics*, 3.
28. Cournot, A. (1838). Researches into the Mathematical Principles of the Theory of Wealth, *Macmillan*, New York.
29. Deaton, A. and J. Muellbauer (1980). Economics and Consumer Behaviour, *Cambridge University Press*, London.
30. Debreu, G. (1959). Theory of Value, *Wiley*, New York.
31. Dorfman, R., P.A. Samuelson and R.M. Solow (1958). Linear Programming and Economic Analysis, *McGraw Hill*, London.
32. Edgeworth, F.Y. (1881). Mathematical Psychics, *Routledge*, London.
33. Fishburn, P.C. (1980). Continua of Stochastic Dominance Relations for Unbounded Probability Distributions, *Journal of Mathematical Economics*, 7.

34. Fisher, I. (1930). *The Theory of Interest*, Macmillan, New York.
35. Foley, D. (1967). Resource Allocation in the Public Sector, *Yale Economic Essays*, 7.
36. Friedman, M. and L.J. Savage (1948). The Utility Analysis of Choices Involving Risk, *Journal of Political Economy*, 56.
37. Frisch, R. (1960). *Planning for India: Selected Explorations in Methodology*, Asia Publishing House, New Delhi.
38. Gale, D. (1960). *Theory of Linear Economic Models*, McGraw Hill, London.
39. Gordon, H.S. (1954). The Economic Theory of a Common Property Resource: The Fishery, *Journal of Political Economy*, 62.
40. Groves, T. (1976). Information, Incentives and Internalization of Production Externalities, in S.A.Y. Lin (ed.) *Theory and Measurement of Economic Externalities*, Academic Press, New York.
41. Harsanyi, J.C. (1967). Games of Incomplete Information Played by Bayesian Players, *Management Science*, 14.
42. Harsanyi, J.C. (1973). Games with Randomly Distributed Payoffs; A New Rationale for Mixed Strategy Equilibrium Points, *International Journal of Game Theory*, 2.
43. Hawkins, D. and H.A. Simon (1949). Some Conditions of Macroeconomic Stability, *Econometrica*, 17.
44. Hay, D.A. and D.J. Morris (1991). *Industrial Economics and Organization*, Oxford University Press, London.
45. Heckscher, E.F (1919). *Economist Tidskrift* (in Swedish), 21; an English translation appeared under the title 'The Effects of Foreign Trade on the Distribution of Income', in H.S. Ellis and L.A. Motzler (eds.) *Readings in the Theory of International Trade*, Blackiston, Philadelphia, 1949.
46. Henderson, J.M. and R.E. Quandt (1971). *Microeconomic Theory: A Mathematical Approach*, McGraw Hill Kogakusa, Tokyo.
47. Hicks, J.R. (1939). *Value and Capital*, Clarendon Press, Oxford.
48. Hicks, J.R. (1939a). Foundations of Welfare Economics, *Economic Journal*, 49.
49. Hirsch, W.Z. (1979). *Law and Economics: An Introductory Analysis*, Academic Press, New York.
50. Hirschleifer, J. and J.G. Riley (1979). The Analysis of Uncertainty and Information—An Expository Survey, *Journal of Economic Literature*, 17.
51. Hotelling, H. (1929). Stability in Competition, *Economic Journal*, 39.
52. Houthakker, H. (1950). Revealed Preferences and The Utility Function, *Economica*, 17.
53. Intriligator, M.D. (1980). *Econometric Models, Techniques and Applications*, Prentice-Hall, London.

54. Jacquemin, A. (1987). *The New Industrial Organization*, Clarendon, Oxford.
55. Kaldor, N. (1939). Welfare Propositions in Economics, *Economic Journal*, 49.
56. Kaldor, N. (1955-56). Alternative Theories of Distribution, *Review of Economic Studies*, 23.
57. Koopmans, T.C. (1960). Stationary Ordinal Utility and Impatience, *Econometrica*, 28.
58. Kreps, D.A. (1990). *A Course in Microeconomic Theory*, Princeton University Press, New Jersey.
59. Leontief, W.W. (1951). *The Structure of American Economy, 1919-1939*, Second Edition, Oxford University Press, New York.
60. Leontief, W.W. (1953). Domestic Production and Foreign Trade; the American Position Re-examined, *Proceedings of the American Philosophical Society*, 37.
61. Leontief, W.W. (1964). An International Comparison of Factor Costs and Factor Use: A Review Article, *American Economic Review*, 54.
62. Leontief, W.W. and Others (1953). *Studies in the Structure of the American Economy*, Oxford University Press, New York.
63. Lerner, A.P. (1934). The Concept of Monopoly and The Measurement of Monopoly Power, *Review of Economic Studies*, 1.
64. Lindahl, E. (1919). *Just Taxation—A Positive Solution*, Translated and Reprinted in R.A. Musgrave and A.J. Peacock (eds.) *Choices in The Theory of Public Finance*, Macmillan, New York.
65. Lipsey, R.G. and K.J. Lancaster (1956). The General Theory of Second Best, *Review of Economic Studies*, 24.
66. Luce, R. and H. Raiffa (1957). *Games and Decisions*, Wiley, New York.
67. Madden, P. (1986). Concavity and Optimization in Microeconomics, *Blackwell*, Oxford.
68. Majumder, A. (1986). Consumer, Expenditure Pattern in India: A Comparison of the Almost Ideal Demand System and the Linear Expenditure System, *Sankhya, The Indian Journal of Statistics*, 48, Series B.
69. Marshall, A. (1961). *Principles of Economics*, Ninth Edition, Macmillan, London.
70. McKenzie, L.W. (1959). On The Existence of a General Equilibrium for a Competitive Market, *Econometrica*, 27.
71. Meade, J.E. (1952). External Economies and Diseconomies in a Competitive Situation, *Economic Journal*, 62.
72. Minhas, B.S. (1962). The Homohypallagic Production Function, Factor Intensity Reversals, and the Heckscher-Ohlin Theorem, *Journal of Political Economy*, 60.
73. Muth, J. (1961). Rational Expectations and The Theory of Price Movements, *Econometrica*, 29.

74. Narayana, N.S.S., K.S. Parikh and T.N. Srinivasan (1991). Agriculture, Growth and Redistribution of Income: Policy Analysis with a General Equilibrium Model of India, *North-Holland*, Amsterdam and Allied, New Delhi.
75. Nash, J.F. (1951). Noncooperative Games, *Annals of Mathematics*, 45.
76. Ohlin, B. (1933). Interregional and International Trade, *Harvard University Press*, Cambridge, Massachusetts.
77. Pigou, A.C. (1920). The Economics of Welfare, *Macmillan*, New York.
78. Pratt, J.W. (1964). Risk Aversion in Small and Large, *Econometrica*, 32.
79. Quirk, J.P. (1986). Intermediate Microeconomics, *Science Research Associates*, Chicago.
80. Radzinowicz, L. (1948). A History of English Criminal Law and Its Applications from 1950, Vol.1, *Stevens and Sons*, London.
81. Rawls, J. (1971). A Theory of Justice, *Harvard University Press*, Cambridge, Massachusetts.
82. Ricardo, D. (1817). On the Principles of Political Economy and Taxation, *John Murray*, London; in P. Sraffa (ed.) The Works and Correspondence of David Ricardo, Vol.1, *Cambridge University Press*, Cambridge, 1951.
83. Saini, G.R. (1969). Resource—Use Efficiency in Agriculture, *Indian Journal of Agricultural Economics*, 24.
84. Salant, S.W. (1987). Treble Damage Awards in Private Lawsuits for Price Fixing, *Journal of Political Economy*, 95.
85. Samuelson, P.A. (1947). Foundations of Economic Analysis, *Harvard University Press*, Cambridge Massachusetts.
86. Samuelson, P.A. (1949). Economics, *McGraw Hill*, New York.
87. Samuelson, P.A. (1950). Evaluation of Real National Income, *Oxford Economic Papers*, 2.
88. Samuelson, P.A. (1951). Abstract of a Theorem Concerning Substitutability in an Open Leontief Model, in T. Koopmans (ed.) Activity Analysis of Production and Consumption, *Wiley*, New York.
89. Samuelson, P.A. (1953). Diagrammatic Exposition of a Theory of Public Expenditure, *Review of Economics and Statistics*, 37.
90. Samuelson, P.A. (1956). Social Indifference Curves, *Quarterly Journal of Economics*, 70.
91. Schmalensee, R. (1989). Inter-Industry Studies of Structure and Performance, in R. Schmalensee and R.D. Willig (eds.) Handbook of Industrial Organization, *North Holland*, Amsterdam.
92. Schumpeter, J. (1950). Capitalism, Socialism and Democracy, *Harper*, New York.
93. Scitovsky, T. (1941). A Note on Welfare Propositions in Economics, *Review of Economic Studies*, 9.

94. Selten, R. (1975). Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games, *International Journal of Game Theory*, 4.
95. Sen, A.K. (1976). Real National Income, *Review of Economic Studies*, 48.
96. Sen, A.K. (1977). On Weights and Measures: Informational Constraints in social Welfare Analysis, *Econometrica*, 45.
97. Shavell, S. (1987). Economic Analysis of Accident Law, *Harvard University Press* Cambridge, Massachusetts.
98. Smith, A. (1776). An Inquiry into The Nature and Causes of The Wealth of Nations, E. Cannon (ed.) *Menthuen*, London 1904.
99. Spence, M. (1974). Market Singaling, *Harvard University Press*, Cambridge, Massachusetts.
100. Stackelberg, von, H. (1934). Market Form Und Gleichgewicht, *Julius Springer*, Vienna.
101. Stigler, G.J. (1954). The Early History of Empirical Studies of Consumer Behaviour, *Journal of Political Economy*.
102. Stigler, G.J. (1961). The Economics of Information, *Journal of Political Economy*, 69.
103. Stolper, W. and P.A. Samuleson (1941). Protection and Real Wages, *Review of Economic Studies*, 9.
104. Stone, J.R.N. (1954). Linear Expenditure Systems and Demand Analysis:An Application to the Pattern of British Demand, *Economic Journal*, 64.
105. Tobin, J. (1958). Liquidity Preference as Behaviour Towards Risk, *Review Economic Studies*, 26.
106. Varian, H.R. (1974). Equity, Envy and Efficiency, *Journal of Economic Theory*, 9.
107. Varian, H.R. (1993). Intermediate Microeconomics: A Modern Approach, Third Edition, *Norton*, New York.
108. Von Neumann, J. and O. Morgenstern (1944). The Theory of Games and Economic Behaviour, *Princeton University Press*, New Jersey.
109. Wald, A. (1951). On Some Systems of Equations in Mathematical Economics, *Econometrica*, 19.
110. Walras, L. (1874). Elements of Pure Economics, Trans. W. Jaffee, Ricad D. Irwin, *Homewood*, 1954.

Index

A		average cost	267
absolute advantage	616	average cost pricing	366
ad valorem tax	57	fixed cost	261
adding-up controversy	602	total cost	261
adjustment process	317-320	variable cost	261
adverse selection	708	product	230
advertising	373, 408	revenue	288-289
affine function/transformation	14, 203, 752	revenue product	459
aggregate/ total demand	500	Averch-Johnson effect	368
aggregate excess demand	501	B	
Air India	413	backward bending labor	
allocation	493	supply curve	467
efficient	495	backward induction	442
fair	546	bad	28
feasible	493	barriers to entry	
allowable costs	366	artificial	362
Anglo-Saxon	743	natural, technical	362
annuities	177	barriers to trade	633
antimonopoly law	738	barter trade	621
antitrust law	738	basic feasible solution	584
argument of a function	749	solution	583-584
Arrow's impossibility theorem	547	theorem of linear programming	584
	-549	battle of sexes	433
asymmetric information	705	Bayesian game	446
auctioneer	404, 499	Benthamite social welfare	
autarky	618	function	551
		Bergson-Samuelson social	

welfare function	554	choice under uncertainty	202-203
Bernoulli-Nash social welfare function	552	Clarke-Groves tax	695
Bertrand equilibrium	404	classical utilitarian	551
game	437	Clayton Act	738
Bihar	216	closed set	760
bilateral formulation	737	club goods	679
bilateral monopoly	483	coalition	498
Bliss' model	603	Coase theorem	664
bliss point	29	Cobb-Douglas	
bond	178	production function	242
borrower	166	utility function	34
Brazil	413	cobweb diagram	328
budget		collusion	396, 438
constraint	53, 166	commodity bundle	17
line	53	common property	663
set	53	commons, problem of	673
C		community indifference curve	625
Calcutta	9, 674, 698	comparability requirement	550
Cambridge	599	comparative advantage	528-529, 617-619
capacity, excess	373	theorem	623
capital goods	182, 226	region	622-623
human	184-185	comparative statics	7, 80, 308
intensity	226	compensated demand curve	100
services of	182, 226	compensating variation	118
cardinal utility	8-9	compensation criterion	543
cartel	395-396	Hicks	544
certainty equivalent	206	Kaldor	543
CES production function	242	Samuelson	557
ceteris paribus assumption	3	Scitovsky	544-545
chain rule	756	competitive equilibrium	500, 504
change in demand	96	industry	6,286-288
change in quantity demanded	97	market	6,286-288
charitable donation	43	market and Pareto efficiency	505

Index			769
complements		continuous compounding	189
gross	103	continuous function	754
net	104	contract	
perfect	26, 65	cost plus	717
complete preferences	18	curve	496
composite function	753	fixed price	717
composite good	55	incentive	717
concave preferences	30	law	740
utility function	12, 205	convex	
concentration	402	combination	751
concentration ratio	402	function	751
conditional factor demand	460	indifference curves	21
Condorcet paradox	692	isoquant	233
conduct by firms	410	preferences	18, 19
conjectural variation	414	set	20
congestion	679	cooperative game	426
cost	679	cooperative of consumers	208
Consolidated coffee	410	core	497-498, 505-506
consols	177	corn laws	326
constant cost industry	318	corner solution	61-62
constant elasticity demand	146	corporation	224
constant returns to scale	238	cost	
constant sum game	427	average	261, 267
constrained maximization	760	average fixed	261
consumer		average total	261
impatient	171	average variable	261
patient	171	fixed	256
consumer preferences	17	long run	254-255, 266
consumer price index	123	long run average	267-269
consumer's surplus	116	long run marginal	267-269
contestable market	320	marginal external	672
contingent commodity/		opportunity	254
consumption	199, 720	short run	256
continuity of aggregate		short run marginal	261-262
excess demand	503	social marginal	659, 661

sunk	443-444	Digha	674
transactions	741	diminishing marginal productivity	229
variable	256	marginal rate of substitution	24
costly information	711	marginal rate of technical	
Cournot equilibrium	398	substitution	235
game	437	marginal utility	11
crime and punishment	729	discounted present value	175
D		discriminating monopolist	353
damage, marginal external	672	diseconomies of scale	269
decreasing cost industry	319	disequilibrium	6, 499
decreasing returns to scale	239	distortionary tax	510
deductive reasoning	5	diversification	408
Delhi	9	dividend	224
demand		divisibility	21
compensated	100	domain	750
constant elasticity	146	dominant firm	391
elasticity	142	dominant strategy	429
excess	501	equilibrium	430
factor, conditional	460	dominated strategy	430
factor, derived	460	double mark-up	485
factor, profit maximizing	460	downstream monopolist	483
function	62	dual problem	590
market	139	due care	735
ordinary	95	duopoly	390
revealing tax	694	durable good	185
dependent variable	749	E	
derivative	754	economic rent	452
partial	756-757	economic survey	124
rules for calculating	755	economics	2
second	754	economies of scale	268
derived factor demand	460	economy, pure exchange	493
dictator	549	Edgeworth box	493-494
differential	756	Edgeworth-Bowley box	520
differentiated products	369	efficiency	7, 494, 505, 524, 541

elasticity of conjecture	424	existence of a competitive	
elasticity of demand	142	equilibrium	502-504
and revenue	144	exit	287
of offense	733	exogenous variable	150
of output	244, 602	expansion path	260
of substitution	241	expected utility	202
of supply	300, 311-313	expected value	198
endogenous variable	150	expenditure elasticity	160
Engel curve	84	function	68-69
Engel's law	149	extensive form game	428
generalized	161	extensive margin of production	231
entry deterrence	443- 444	external cost, marginal	672
envelope curve	271	damage, marginal	672
theorem	128	diseconomies	319, 657
envy	546	economies	319, 656
equation	750	externalities	
equilibrium	6, 301, 314, 321, 500	consumption	656
and excess demand	501	production	657
subgame perfect	442	<i>F</i>	
equilibrium of output markets		factor demand	
long run	314	conditional	258, 460
short run	301-302	derived	258, 460
equity	175, 509	profit maximizing	460
equivalent variation	119	factor intensity	628
Euler's theorem	759	factor intensity reversal	628-629
Europe	398	factors of production	227
events		factor-price equalization theorem	631-
collectively exclusive	198		632
mutually exclusive	198	fair allocation	546
excess demand	305 , 501	rate of return	367
exchange economy	493	fairness	546
exclusive good	678	feasibility	321
exhaustible resource	191	feasible allocation	493
		region	575

solution	583-589	normal form	428
firm	6, 223	tree	440
firm conduct	410	variable sum	427
first-degree price discrimination	354	zero sum	427
First Fundamental Theorem		general equilibrium	7, 492-493
exchange	505	generalized Bernoulli-Nash social	
production	528	welfare function	552
first mover advantage	443-444	utilitarian	552
first order conditions	759	Giffen good	92
fixed costs	256	Gini index	559
fixed proportions	237	social welfare function	559
folk theorem	443	good	
football club	486	Giffen	92
free entry and exit	320	impartial	82
free rider	652, 683, 687	inferior	82
free trade	623	luxury	149
Friedman-Savage utility function	222	neutral	28
function	749	normal	82
concave	751	private	678
continuous	754	public	679
convex	751	Grameen Bank	719
decreasing	750	Gresham's Law	707
explicit	750	gross complements and substitutes	103
homogeneous	758	H	
implicit	750	Hawkins-Simon conditions	572-574
increasing	750	Heckscher-Ohlin Theorem	630
monotone	750	Herfindahl-Hirschman measure	402
G		Hicks criterion	544
gains from trade	496, 619-620, 624	Hicks demand curve	100
game		Hicks substitution effect	93, 105, 107, 111
constant sum	427	hidden action	711
cooperative	426	hidden information	709
extensive form	428	Hindustan Aluminium Company	409
noncooperative	426		

homogenous function	758	indifference	17, 18
homothetic preferences	86	indifference curves	20, 21
homothetic production function	240	indifference map	20
Hooghly river	698	indirect revealed preference	108
Hotelling's model of location	373	indirect utility function	67
human capital	184-185, 633	individualistic welfare function	554
I		Indonesia	413
IATA	413	inductive reasoning	5
identity	750	industrial organization theory	407
impartial good	82	industry	6
impatience	170	constant cost	318
implicit function	750	decreasing cost	319
import quota	634	increasing cost	319
incentive compatability constraint	717	industry equilibrium	
preservation	510	long run	314
income consumption curve	83-85	short run	301-302
effect	89-95	industry performance	410
expansion path	83	industry supply curve	
offer curve	83	long run	315
tax	71-72	short run	299
increasing cost industry	319	inequality	509
function/transformation	32, 750	inferior good	82
returns to scale	238	inflation rate	167
independence of irrelevant		information externalities	652
alternatives	549	information, value of	720
independent variable	749	initial endowment	493
index numbers	112-115	input	225
India	190, 216, 332, 335, 375, 409,	input-output analysis	567
413, 593, 617-620, 637, 738		input-output coefficient	571
Indian cotton textile industry	275	insurance	207
Indian postal system	375	integration	
Indian Railways	375	conglomerate	408
Indian Standard Institution	710	horizontal	408
Indian Statistical Institute	593	vertical	408

intensive margin of production	230	L	
interest rate	164-165	labor intensity	226
interior maximum	61	labor market	464-467
internal rate of return	195	labor supply	464
internalization of externalities	660	labor theory of value	621
International Business Machines	391	Laffer curve	473
international economics		effect	474
micro aspects	615	Lagrange multiplier	73-74
pre-Ricardian views	616-617	Lagrangian	74
intertemporal		Laspeyres	
budget constraint	166-167	price index	113
budget line	167	quantity index	115
choice	170-171	Law of demand	95
intramarginal	367	diminishing marginal rate of	
inverse		substitution	24
demand curve	16, 97	diminishing marginal rate	
function	750	of technical substitution	235
supply curve	297	diminishing marginal utility	11
investment	182	diminishing marginal	
isocost line	257	productivity	229
isoprofit line	526	variable proportions	182, 232
iso quant	232	leadership	
isorevenue line	582	price	393
isowelfare curves	552	quantity	391
Italy	396, 398	learning by doing	391
J		leisure	464
Japan	617-620, 637, 738	Leontief paradox	632
joint production	528-529	Leontief production function	237
K		Lerner's market power index	352, 401
Kaldor criterion	543	lexicographic ordering	37
k-firm concentration ratio	402	L'Hospital's rule	247
kinked demand curves	405	liability laws	734
Kodak	391	limit pricing	445
Kuwait	616	Lindahl equilibrium	691

linear expenditure system	67	market failure	651
linear function	752	market power	352, 400
linear homogenous	239, 758	market supply	299, 315-316
linear programming	574, 579	market structure	407
long run	227	mark-up pricing	352
average cost	267-269	Marshallian demand curve	95
cost function	254 - 255, 266	stability	305-308
equilibrium	314	Maruti car	159
marginal cost	267-269	matching pennies	431
supply curve	311-313	material balance equation	574
loss minimization	293, 350-351	matrix	585
lumpsum tax	57	maturity date	178
luxury good	149	maximax social welfare function	564
M		maximin social welfare function	552
macroeconomics	5	maximin strategy	434
majority voting, paradox of	692	maximization of social welfare	553
Malaysia	413	maximum	759
Malthusian population theory	606	mean	208
marginal change	12, 23	mean-variance analysis	211
cost	261-262, 267-269	median	694
cost pricing	365	mercantilism	616
cost, social	659, 661	mercury cartel	396, 398
external cost	672	merger	408, 660
external damage	672	microeconomics	5
product	228	Midnapore	674
rate of technical substitution	234	Ministry of Finance, Government	
rate of time preference	169	of India	124
rate of transformation	522	mixed strategy	431-432
revenue	145, 289	modern utility theory	16
revenue product	459	modulus	750
utility	9, 38	money illusion	63
willingness to pay	24, 97	monopolistic competition	368
Massachusetts	599	monopoly	343
market demand	138-139	benefits of	368

deadweight loss	360	<i>O</i>	
natural	363	objective function	581, 760
regulation	365	offer curves	626
rent	350	Oil and Natural Gas Commission	375
and unions	482	oligopoly	389-390
monopsonistic competition	478	oligopsony	475
exploitation	479	OPEC	332, 397
monopsony	475, 477	open set	760
monotone function	750	opportunity cost	165, 254, 617
monotonicity	18, 19	optimal choice	60-61
moral hazard	709-710	optimal contract	716
MRTTP Act	738	optimal strategy	696
multiple equilibria	307	ordinal utility	31-32
Mumbai Stock Exchange	322	Orissa	216
<i>N</i>		output effect	461
Nash equilibrium	428	<i>P</i>	
natural monopoly	363	Paasche	
necessary condition	25	price index	114
necessities	149	quantity index	115
net complements and substitutes	104	panchayat	719
neutral good	28	paradox of voting	692
New Bank	408	Pareto consistency	549
nominal rate of interest	165	dominated	541
non-credible threat	442	efficiency	7, 494-495, 505, 541
non-dictatorship	549	improvement	360, 681
non-exclusive goods	678	non-comparable	541
non-linear pricing	355	superiority	541
non-rival goods	678	partial derivative	756-757
normal equations	152	partial equilibrium	492
normal form game	428	participation constraint	718
normal good	82	partnership	223
normative analysis	3-4	payoff	426-427
numeraire price	502	pepper cartel	413

perfect competition	6, 286-288	principle of revealed preference	108
perfect complements	26, 65, 237	prisoner's dilemma	434-435
perfect divisibility	21	private good	678
perfect price discrimination	354	probability	198
perfect substitutes	25, 237	cumulative	209
perfectly elastic	143, 301	producer's surplus	298
perfectly inelastic	143, 301	production	223
perpetuities	177	production function	225
Pigouvian tax	662, 670	CES	242
pivot	586-588	Cobb-Douglas	239, 242
plant	227	fixed proportion/Leontief	237
pooling equilibrium	715	homothetic	240
portfolio	212	production isoquant	232
positive analysis	3	production possibilities frontier	520, 619-620
predatory pricing	445	set	520-521
preference	17	production process	226
complete	18	efficient	226, 524
concave	30	inefficient	226
convex	18, 19	production set	225
monotone	18	vector	519
reflexive	18	profit	223, 290, 526
transitive	18	economic	290
present value	175	excess	317
price discrimination		long run	317
first-degree	353	normal	290
second-degree	353	programming, linear	579
third-degree	353	property rights	663
price follower	390, 393-394	proprietorship	223
leader	390, 393-394	public good	651-652, 679
offer curve	97, 506	impure	679
rigidity	405-406	intermediate	679
setter	343	pure	679
taker	287, 394	public good board	695
primal	590	Punjab	560
principal-agent problem	716		

Punjab National Bank	408	rental equivalent price	188
purchasing power	56	rental rate	255
pure strategy	432	Renusagar Power Supply Company	409
Q		repeated game	437
quality	707	reservation price	99, 115, 304, 681
quantity		residual demand function	394
follower	390-393	returns to scale	
leader	390-393	decreasing	239
tax	57	constant	238
quasiconcave	40, 41	increasing	238
quasilinear	87	revealed preference	107-108
quasirent	454	strong axiom	112
quotas	633	weak axiom	109
R		revenue	144, 289, 526
range	750	Ricardian model of distribution	605-608
rate of interest	164	Ricardian trade theory	617-623
rate of return	164	ridge lines	233
rate of return regulation	367	risk	
rational expectations	329	averse	204-205
rationing	57	lover	205
Rawlsian social welfare function	552	neutral	205
reaction function	392	risk aversion	
rectangular hyperbola	35	absolute	216
reflexivity	18	relative	220
regulation of monopolies	365	rival good	678
relative abundance	630	Robinson Crusoe economy	530
relative price	56	S	
Reliance Industries	409	sales maximization	243
Reliance Petrochemical	409	Samuelson criterion	557
rent		Samuelson substitution theorem	577-578
economic	454	Samuelson optimality condition	685
land	455	satiation	29
Ricardian	456		
seeking	454		

saturation point	9	Bernoulli-Nash	552
saving	174	generalized Bernoulli-Nash	552
saving-investment equilibrium	184	generalized utilitarian	552
Scitovsky criterion	544-545	individualistic	554
second best theory	652-653	maximax	564
second-degree price discrimination	353	maximin	552
second fundamental theorem	509	Rawlsian	552
exchange	509	utilitarian	552
production	528	South Korea	646
second order condition	759	Spain	396, 398
security	742	stability	
self selection mechanism	712	Marshallian	305- 308
separable utility	132	Walrasian	305- 308
separating equilibrium	715	Stackelberg	
sequential games	440	equilibrium	393
shadow prices	592	game	440
Shephard's lemma	128	standard deviation	209
Sherman Act	738	Statesman	674
short run	227	stationary point	759
shutdown decision	293	stochastic dominance	210, 218
signalling	712	stock market	224
simplex method	584	Stolper-Samuelson theorem	634-635
single-peaked preferences	693	Stone-Geary utility function	67
slack variable	583	strategic complements and	
slope	754	substitutes	451
Slutsky		strategy	426
demand curve	107	mixed	432
equation	90	pure	432
substitution effect	93, 105-107	strategy space	427
social indifference curve	552	strictly concave	
social marginal cost	659, 661	function	751
social preference	547-548	preference	30
social welfare function	551	strictly convex	
Bergson-Samuelson	554	function	751

preference	18	lumpsum	57
Strong Axiom of Revealed Preference		quantity	57
Preference	112	value	310
structural variables	583	tax incidence	309
structure, market	407	technical efficiency	227, 238
subadditivity	363	inefficiency	226
subgame	441	terms of trade	640
subsidy		third-degree price discrimination	353
ad valorem	57	tie-in-sales	355
lumpsum	57	tit-for-tat	437
quantity	57	Tokyo round	637
substitutes		tort law	734
gross	102-103	total demand	500
net	104	differential	757
substitution effect		product curve	229
Hicks	93, 105, 107, 111	trade triangle	626
Slutsky	93, 105-107	transaction costs	741
sunk cost	443-444	transformation	14, 223
supply curve		transitive	18, 19
labor	464-465	trigger strategy	439
long run	311-312	true cost of living index	134
market/industry	299, 315-316	two-part tariff	355
short run	297	two-tier pricing	367
surplus	457	<i>U</i>	
sustainability	321	UK	738
symmetric oligopoly	401	unanimity principle	541
synergy	408	uncertainty	
T		choice under	202
tariff	633	event	706
Tata Tea	409	market	706
tatonnement process	305	undocumented workers	468
tax		union	482
ad valorem	57	unrestricted domain	548
income	71	upstream monopolist	483

Index			781
USA	398, 632, 637, 738	voting	692
usage fee	355	W	
user cost	188	Walras' law	501
util	16	Walrasian equilibrium	500
utility function		Walrasian stability	305-308
cardinal	8	Weak Axiom of Cost Minimization	259
ordinal	31-32	Profit Maximization	294, 474
utility possibility curve	555	Revealed Preference	109
frontier	556	welfare maximization	553
Uttar Pradesh	244	welfare triangle	361
V		West Bengal	216, 719
value judgement	548	Pollution Control Board	674
variable cost	256	State Government	699
factors	227	willingness to pay	682
variable-sum game	427	Z	
variance	209, 211	zero-sum game	427
Vidyasagar Setu	698	zero profit	304
von Neumann Morgenstern utility	202		

The Book

An analytical and geometric methodology has been adopted in presenting the text.

Many chapters are supplemented by mathematical appendices, even though the main text is comparatively free of mathematics.

Applied examples related to the theoretical concepts have been included in each chapter and different types of exercises like multiple choice, prove-disprove questions and quantitative problems and questions are also included.

In addition to covering traditional aspects of microeconomics, the book also discusses some recent developments like game theory, externalities, public good, information and law.

Topics such as linear models, theory of distribution and international trade have also been described, which are normally not found in texts on microeconomics.

This book should be useful as class material for undergraduate and graduate students of economics and business.

The Author

Obtained his bachelor of statistics (Hons.) as well as his Masters from the Indian Statistical Institute, Calcutta and also his doctorate from the same institute.

Since 1997 he has been associated in various research and teaching positions with the Indian statistical Institute. He has been a visiting professor in universities in Canada, Germany, Israel, Japan, France, Hongkong and Italy.

So far, he has published over 58 research articles in national and international journals and authored two other books-*Ethical Social Index Numbers* and *Issues in Industrial Economics*.

Dr. Satya R. Chakravarty was also awarded the Mahalanobics Memorial Prize in 1994 and is listed in various international sources.

ISBN 81-7764-348-7

ALLIED PUBLISHERS PVT. LIMITED