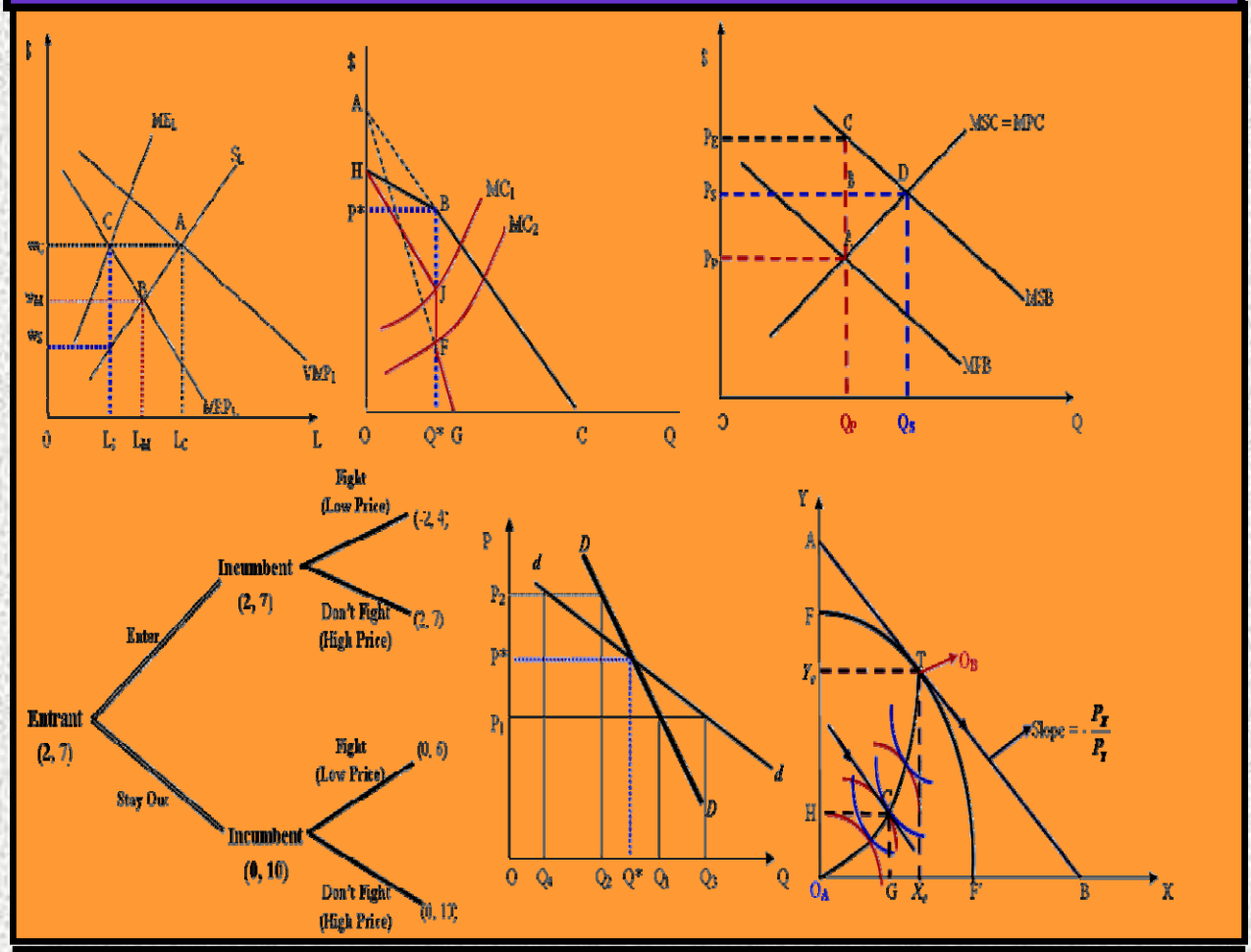
MICROECONOMICS II

A TEACHING MATERIAL

BY HASSEN ABDA



JIMMA UNIVERSITY
FACULTY OF BUSINESS AND ECONOMICS
DEPARTMENT OF ECONOMICS

2009
JIMMA

PREFACE

The study of microeconomics occupies the central place in the study of both theoretical and applied courses in economics, regardless of whether the issues are microeconomic or macroeconomic in nature. There is no course in the department of economics, if not for the technical ones, for which microeconomics is not a basic requirement. Thus to tackle the bottlenecks surrounding this basic course should be the first step in solving problems in the larger domain – in the department.

As to the curriculum of the university (and all universities in the nation), the subject matter microeconomics is taught divided into two parts: microeconomics I and microeconomics II – each with three credits hours. This teaching material is for the second part of the course.

Microeconomics II begins with the discussion of the interaction between producers and consumers in the product markets along the lines of monopolistic competition (chapter one) and oligopoly (chapter two). Some recent development in explaining the behavior of oligopolistic firms is the theory of games, and an introduction to this tool of analysis is illuminated in the third chapter. We then turn our face away from the product market to the factor market, and consider factor pricing in different market structures and the related issue of income distribution in the fourth chapter. The fifth chapter puts aside the partial equilibrium analysis of individual markets in isolation, and sheds light on the general equilibrium analysis that deals with the interrelations among different markets and economic agents. Along with this general equilibrium analysis, welfare economics is also discussed in the same chapter. The final chapter, chapter six, deals with the causes of market failure that lay ground for government intervention.

In addition to the numerical problems solved as in-chapter illustrative examples, each chapter is accompanied by a set of review questions. The answers to some selected end-of-chapter questions are provided following the final chapter.

The various references used in writing this teaching material are disclosed following the section of answers to selected questions.

I am sure that students will find this teaching material very helpful; particularly, the ease with which it is presented, the illustrative local examples, and the exercises included in it. Besides, I expect it to be of much help in keeping uniformity of the course coverage across instructors. As a positive externality, it would also initiate similar efforts in helping students out of difficulties.

Finally, I believe that suggestions and comments of readers are very useful and I am very glad to listen to your opinions regarding the content (coverage), elaborations of points raised, or any aspect of this material.



Hassen Abda

June 2009

Jimma

TABLE OF CONTENTS

PAGE

Preface i
Table of Contents iii

CHAPTER ONE: MONOPOLISTIC COMPETITION..... 1
1.1 Monopolistic Competition Defined 2
1.2 Product Differentiation, the Demand Curves and Costs of the Firm 3
1.3 The Concept of Industry and Product Group 8
1.4 Equilibrium of the Firm 9
1.5 Excess Capacity and Welfare Loss 16
1.6 Criticisms of Chamberlin's Theory of Monopolistic Competition..... 22
1.7 Summary 23
1.8 Review Questions 25

CHAPTER TWO: OLIGOPOLY..... 30
2.1 Oligopoly Defined..... 31
2.2 Non-Collusive Oligopoly 34
2.2.1 Cournot's Duopoly Model 34
2.2.2 Bertrand's Duopoly Model 41
2.2.3 Stackelberg's Duopoly Model..... 46
2.2.4 The 'Kinked-Demand Curve' Model..... 50
2.3 Collusive Oligopoly 52
2.3.1 Cartels..... 53
2.3.1.1 Cartels Aiming at Joint-Profit Maximization (Centralized Cartels) 53
2.3.1.2 Market-Sharing Cartels 57
2.3.2 Price Leadership..... 59
2.3.2.1 The Low-Cost Price Leadership 59
2.3.2.2 The Dominant-Firm Price Leadership..... 66
2.3.2.3 Barometric Price Leadership..... 69
2.4 Contestable Markets 70
2.5 Long Run Efficiency Implications and Welfare Effects of Oligopoly 71
2.6 Summary 74
2.7 Review Questions..... 75

CHAPTER THREE: INTRODUCTION TO GAME THEORY..... 82
3.1 Some Basic Concepts 83
3.2 Dominant Strategy and Nash Equilibrium..... 85
3.3 Mixed Strategies..... 89
3.4 The Prisoners' Dilemma and Cartel Cheating 92
3.5 Repeated Games and Enforcing a Cartel 96
3.6 Sequential Games and a Game of Entry Deterrence 97
3.7 Summary 104
3.8 Review Questions..... 105

CHAPTER FOUR: FACTOR PRICING AND INCOME DISTRIBUTION 109
4.1 Factor Pricing under Perfect Competition 110
4.1.1 The Demand for Factors of Production 111
4.1.1.1 The Demand of a Firm for One Variable Productive Factor 111
4.1.1.2 The Demand of a Firm for a Variable Factor when There Are Several Variable Inputs 115
4.1.1.3 The Market Demand Curve for an Input 119
4.1.2 Factor Supply 121
4.1.2.1 The Supply of Labor 121
4.1.2.2 The Supply of Other Factors 128
4.1.3 Factor Pricing 129
4.2 Factor Pricing in Imperfectly Competitive Markets 130
4.2.1 Monopolistic Power in the Product Market 130
4.2.1.1 Demand of the Firm for a Single Variable Input 131
4.2.1.2 Demand of a Firm for a Variable Factor when There Are Several Variable Factors 132
4.2.1.3 Market Demand and Supply of Factors 133
4.2.1.4 Factor Pricing 134
4.2.2 Monopsonistic Power in the Factor Market 135
4.2.2.1 A Monopsonist Using a Single Variable Factor 136
4.2.2.2 A Monopsonist Using Several Variable Factors 140
4.2.3 Bilateral Monopoly 141
4.3 Elasticity of Factor Substitution, Technological Progress and Income Distribution 144
4.3.1 Elasticity of Factor Substitution and the Shares of Factors of Production 146
4.3.2 Technological Progress and Income Distribution 149
4.4 Summary 152
4.5 Review Questions 155

CHAPTER FIVE: GENERAL EQUILIBRIUM ANALYSIS AND WELFARE ECONOMICS 161
5.1 General Equilibrium Analysis 162
5.1.1 Partial versus General Equilibrium Analysis 162
5.1.2 General Equilibrium in a Two-Factor, Two-Commodity, Two-Consumer (2x2x2) Economy 165
5.1.2.1 Equilibrium of Production (Efficiency in Factor Allocation) 166
5.1.2.2 Equilibrium of Consumption (Efficiency in Distribution of Commodities) 172
5.1.2.3 Simultaneous Equilibrium of Production and Consumption (Efficiency in Product-Mix) 173
5.2 Welfare Economics 177
5.2.1 Definition of Welfare Economics 178
5.2.2 The Utility Possibilities Frontier (UPF) and the Grand Utility Possibilities Frontier (GUPF) 179
5.2.3 Social Decision Mechanisms 182
5.2.3.1 Majority Voting Approach (MVA) 183
5.2.3.2 Rank-Order Voting Approach (ROVA) 184

5.2.4 Social Welfare Function *187*
5.2.5 Fair Allocations, Envy and Equity *193*
5.3 Summary *197*
5.4 Review Questions *199*

CHAPTER SIX: MARKET FAILURE ***203***
6.1 Externalities *204*
6.2 Public Goods *209*
6.3 Market Power *211*
6.4 Asymmetric Information *213*
6.4.1 Adverse Selection (The Hidden Information Problem) *214*
6.4.2 Market Signaling *216*
6.4.3 Moral Hazard (The Hidden Action Problem) *218*
6.4.4 Information Asymmetry, Monitoring Costs and Microfinance *220*
6.5 Summary *223*
6.6 Review Questions *225*

ANSWERS TO SELECTED REVIEW QUESTIONS ***230***
REFERENCES…….. ..... ***231***

CHAPTER ONE
MONOPOLISTIC COMPETITION

Introduction

In an introductory course of microeconomics (Microeconomics I in our case), the perfectly competitive market structure is used to illuminate the interaction between firms and consumers (or the market mechanism). This is done as perfect competition is the easiest model to handle. However, the assumptions of perfect competition are rarely (if at all) met in the real world markets.

The best effort usually made to account for the existence of non-competitive markets is to discuss the behaviors of pure monopolists. Though pure monopoly may exist, it is not a typical (representative) market structure.

In general, both perfect competition and pure monopoly are useful tools of analysis; but neither of the two depicts the common phenomena of the real world. These real world phenomena are best analyzed using the market structures that fall between the two extremes, namely, monopolistic competition and oligopoly.

The oligopoly market structure will be considered in the second chapter, and we now turn our face to monopolistic competition.

Chapter Objectives

After completing this chapter, students would be able to:

- Describe the monopolistically competitive market structure, and compare and contrast it with perfect competition and pure monopoly.
- Explain the short run and long run equilibrium of a firm in monopolistic competition.

- Discuss the costs and benefits of advertising.
- Discuss the existence of excess capacity and the social welfare aspects of monopolistic competition.
- Comment on Chamberlin's model of monopolistic competition and appreciate its contributions.

1.1 Monopolistic Competition Defined

As the name implies, monopolistic competition is a blend of competition and monopoly. Put differently, monopolistic competition possesses some elements or characteristics of perfect competition and draws some elements from pure monopoly. The competitive element arises because there are many sellers, each of which is too small to significantly affect the other sellers. In addition, firms can enter and leave a monopolistically competitive industry. Thus, the major elements of perfect competition found in monopolistic competition are:

- the existence of many small firms which are unable to significantly affect each other (when seen individually), and
- the possibility of entry and exit.

The monopolistic element arises from product differentiation. That is, since the product of each seller is similar but not identical, each seller has a monopoly power over the specific product it sells. This monopoly power, however, is severely limited by the existence of close substitutes. In other words, even if the product of a monopolistic competitor is unique and this grants the firm some power of price-making (just like a pure monopolist), the product has close substitutes (unlike that of a pure monopolist) which limits the power of a monopolistic competitor.

Examples of monopolistic competition include the markets of the numerous brands of soap (B-29, Wabel, 777 …), toothpaste (Aquafresh, Colgate, CloseUp …), stationeries (SinarLine, MAMCO …), cigarettes (Rothmans, Marlboro, Nyala, Dunhill …), etc.

Monopolistic competition is most common in retail and service sectors of many economies.

Notice that one could list a lot of real world examples for monopolistic competition more comfortably than he/she could do for perfect competition or pure monopoly. The relative dominance of monopolistic competition over perfect competition and pure monopoly led to the evolution of some new analytical tools and ideas. As such, in response to the rising concern in the 1920s and 1930s that models of perfect competition and pure monopoly were too extreme to serve as analytical models of business firms and market behavior, Edward Chamberlin (1933) and Joan Robinson (1933) contributed the pioneering works on monopolistic competition. Though both works have similar implications, we adopt the theory of Chamberlin here following the trend in the field.

1.2 Product Differentiation, the Demand Curves and Costs of the Firm

This section explains the types and sources of price discrimination; the nature and kinds of the demand curves monopolistically competitive firms face; as well as the categories and behaviors of costs these firms incur.

Product Differentiation:

Product differentiation is generally intended to distinguish the product of one producer from that of the others in the 'industry'. A firm whose product has close substitutes (strong competitors) is busily engaged in trying to make its product *better* or *look better* than the substitute products. That is, monopolistically competitive firms have to secure some market share by making their products "different".

This product differentiation might be attributed to:

- The subjective judgment of the consumer,
- The quality (for instance, the durability) of the product,

- The characteristics of the product such as taste, color ---,
- Sales promotion, packaging, trademarks, etc.

Product differentiation can be real or fancied.

Real product differentiation: exists when there are differences in the specification of the products (chemical composition/ingredients), or differences in the factor inputs, or the location of the firm that determines the convenience with which the product is accessible to the consumer, or the services offered by the producer during times of sale. For more clarity, two brands of soap (say, B-29 and Peacock) are said to be really differentiated either:

- if they are made of different ingredients, or
- if the producer (or retailer) of one is more conveniently accessible to consumers than that of the other, or else
- if there is a difference only in the quality of services say the whole distributors of the two brands offer to their buyers.

As long as there is a difference in what the consumer or the buyer actually gets, whether it is a difference of chemical composition or a difference of the accompanying service, the product differentiation is real.

Fancied (spurious/imaginary) product differentiation: exists when the products are basically the same but the consumer is persuaded, via advertising or other selling activities, that the products are different. It is established by advertising, difference in packaging, difference in design, or simply by brand name. As a hypothetical example, even if Fine and Abyssinia pocket tissue papers might be of the same quality for an individual in all aspect, the individual may still tilt towards witnessing the superiority of one only because of the difference in color of packaging. This is just one example of spurious product differentiation; think of such differences between/among close substitutes in your locality.

Whatever the case, the aim of product differentiation is to make the product unique in the mind of the consumer.

The Demand Curves:

One major implication of product differentiation is that the producer has some power in the determination of its product's price. The firm is not a price-taker but a price-maker even though it faces the keen competition of close substitutes offered by other firms. Hence, product differentiation gives rise to a negatively sloping demand curve. This demand curve is less than perfectly elastic as the firm can alter price, but more elastic than the demand curve a pure monopolist faces because of the availability of close substitutes in monopolistic competition. Unlike a wheat-farmer who cannot individually affect the market price of wheat (nearest example of perfect competitor), the producer of a cigarette brand (say, Marlboro) can set the price for its product. However, as opposed to the Ethiopian Telecommunications Corporation (ETC) which possesses an absolute power to set its price (fare), the extent to which the cigar producer can choose (and set) price is very limited. This is because while there is no competitor to ETC that threatens its market share, the cigar producer fears loss of customers (smokers) to producers of other brands.

We distinguish between two types of demand curves in the theory of monopolistic competition – the planned sales curve and the actual sales curve.

i. **The planned sales curve** – this is a demand curve that shows how much the firm will sell if it varies its price from the ongoing level under the assumption that other firms maintain their existing prices. If the firm reduces its price and other firms maintain their prices, the firm can expect a considerable increase in sales since it will be able to attract buyers away from other firms in the group ('industry') and increase sales to existing customers. If the firm increases its price and other firms maintain their prices, the firm can expect a

considerable decrease in sales since it will lose business to other firms in the group.

For example, if Bedele Brewery notably lowers the price of a bottle of beer and other brewery companies do not follow the price reduction, then Bedele can expect a considerable rise in its sales both as the currently existing consumers consume more at the lower price and as some consumers will shift away from other brands to Bedele. If, to the contrary, Bedele raises its price and the others do not join the price rise, some loyal consumers will consume less at the higher price and some others will shift away to other brands.

Thus, assuming that each firm expects its actions to go unfelt by the others, each believes its demand curve to be quite elastic. This demand curve (the planned sales curve) is shown as ***dd*** in Figure 1.1 below. If the firm under consideration charges P_2 per unit, it would sell Q_4 units of its output; similarly, Q* and Q_3 units would be sold at prices P* and P_1, respectively. Under the assumption we made regarding what this firm thinks about its competitors, the decision of this firm to change prices implies a movement along the ***dd*** curve.

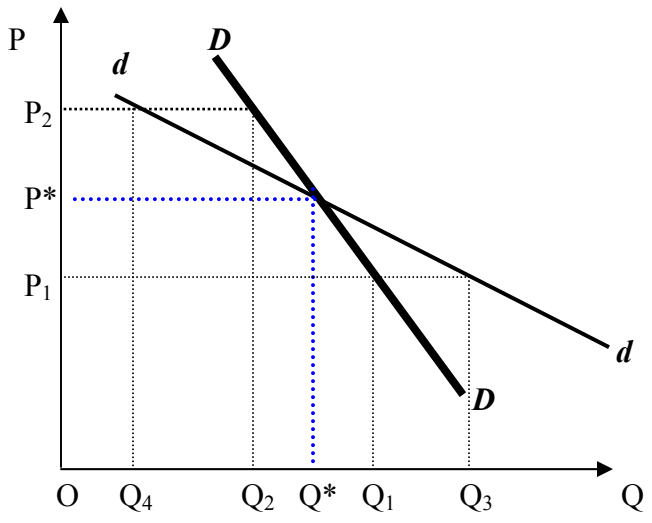


Figure 1.1: Two Kinds of Demand Curves

ii. **The actual sales curve** – this is the demand curve based on the supposition that all firms raise or lower their prices by the same amount as this firm under consideration. This demand curve, also known as the share-of-the-market curve, is shown as ***DD*** in Figure 1.1. If this firm reduces its price from P* to P_1 (on the expectation of raising sales to Q_3) and all other firms reduce their prices to P_1 as well, this firm will sell only Q_1 units of output. Similarly, if this firm increases its price to P_2 and all other firms increase their prices to P_2 as well, this firm will sell Q_2 units of output. Thus, if firms follow both price-rises and price-cuts of a firm, no shift of customers is expected to result from a price change. The only change that takes place is the change in the quantity consumed of each firm's product in a direction opposite the direction of price change. If this is the case, then the position of the ***DD*** curve will be affected only by the entry into or exit out of the particular market where the market share of each firm is affected.

The actual sales curve (***DD***) is less elastic than the planned sales curve (***dd***), since price reductions by this firm will expand its sales by a greater amount if other firms do not meet the price reduction, and price rises by this firm will decrease its sales by a greater amount if other firms do not meet the price increase (than if they do). For example, if other firms do not follow the decision of a firm to raise price from P* to P_2, the firm's sales will fall from Q* to Q_4, rather than from Q* to Q_2.

Costs:

From the theory of costs (you covered in Microeconomics I), recall that the average variable cost (AVC), average total cost (ATC) and marginal cost (MC) are all U-shaped. Chamberlin adopted this shape of costs of the traditional theory of the firm. The average variable cost (AVC), marginal cost (MC) and average total cost (ATC) curves are all U-shaped implying that there is only a single level of output which correspond to optimal production.

Nevertheless, there are some new concepts that Chamberlin added to the traditional theory of costs. He introduced the selling costs to the theory of the firm for the first time. The recognition of product differentiation provides the rationale for the selling expenses incurred by a firm. He also argued that the selling-costs curve is U-shaped, i.e., there are economies and diseconomies of scale of advertising as output changes. The cost of advertising per unit of output initially declines and then eventually rises as output expands. Thus, there is a level of output where the ATC of advertising is minimized.

The U-shaped selling costs, added to the U-shaped production costs, yield U-shaped average, average variable, and marginal cost curves. Hence, whenever you come across the costs (or cost curves) of a monopolistically competitive firm, remember that they include both production and selling costs.

1.3 The Concept of Industry and Product Group

To enhance the comparison of the concept of industry in monopolistic competition to the same concept in perfect competition and pure monopoly, recall how industry was defined in the latter two cases. Under perfect competition, an industry encompasses all firms producing exactly identical (homogenous) products. Under pure monopoly, since there are no two firms which produce the same product, a given firm is an industry in itself.

When products are differentiated, one cannot define an industry as a collection of firms producing identical (homogeneous) products. In this narrow sense, each firm having a distinct product is an industry in itself, exactly as a pure monopolist is. However, unlike the pure monopolist which produces a completely unique product, in monopolistic competition, there are firms which produce closely related (or very similar) products that satisfy the same need. Despite some slight differences among the various brands of soap, soap producers are in keen competition with each other and their behaviors are very similar (more or less the same). Thus, it helps if we study the behavior of soap producers in general rather than studying the behavior of the producer of each brand of soap separately.

However, product differentiation creates difficulties in the analytical treatment of the industry. Heterogeneous products cannot be added to form the market demand and supply schedules as in the case of homogeneous products. Furthermore, there isn't a single equilibrium price for the differentiated products, but a cluster of prices. That is, although the producers of the various brands of soap are to be studied together, the equilibrium price of one brand (say, Wabel) is likely to be different from the equilibrium price of another (say, 777). As such we will have a cluster of equilibrium prices for soap taken as a single commodity.

Chamberlin was (perhaps) the first economist who tried to define industry in monopolistic competition by lumping together firms producing very closely related commodities (in terms of their use). He referred to such group of firms as a 'product group'. Thus, whenever the term industry is used in reference to monopolistic competition, it is to mean a product group.

1.4 Equilibrium of the Firm

A perfectly competitive firm is a price-taker. It can choose the level of output only. This means that quantity of output is the only choice variable for a perfect competitor. On the other hand, a pure monopolist can choose either the price it charges or the quantity of output it produces, but cannot simultaneously choose both. What about a monopolistically competitive firm? The choice-related variables for a monopolistically competitive firm are: (i) price (or quantity), (ii) product variation, and (iii) selling expenses.

In this section, we examine how a monopolistically competitive firm determines its best level of output and price in the short run and long run on the assumption that the firm has already decided on the characteristics of the product to produce and on the selling expenses to incur. Once the firm has decided on product variation and selling expenses, it can choose only the price it charges (which also implies the level of output chosen). Note

that the choice of price automatically determines the optimal level of output and thus we do not treat price and quantity as two distinct variables of choice.

As for a firm under any type of market structure, the best level of output for a monopolistically competitive firm is determined by the equality of marginal revenue and marginal cost (i.e., MR = MC). In the short run, the firm maximizes its profit at a point where MR = MC provided that price is at least as high as the average variable cost (i.e., as long as $P \geq AVC$). There are three possible situations (profit levels) for a monopolistically competitive firm in a short run equilibrium: production at positive profit, production at normal (zero) profit, and production at loss.

If the price the firm charges per unit of output (P) exceeds the average cost (ATC) at the equilibrium level of output, then the firm enjoys positive economic profit (excess/abnormal profit). This is because if P > ATC, then the firm receives more money from the sale of each unit of output than what it incurs to produce and sell a unit on average.

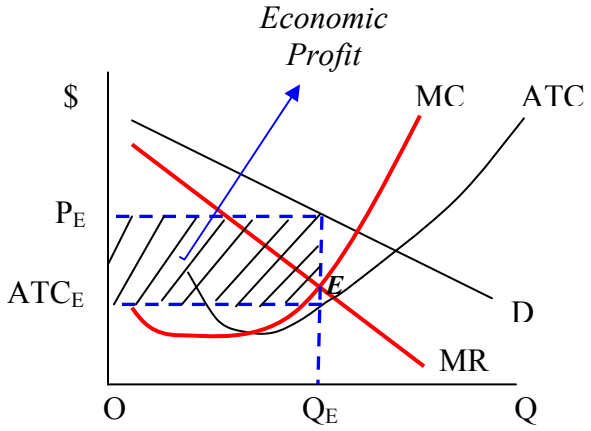
The firm gets only normal or zero profit if P = ATC (if the demand curve is tangent to the ATC curve) at a point corresponding to the equilibrium level of output. In this case, the firm incurs a per unit cost (implicit plus explicit costs) that is exactly equal to what the sale of each unit of output generates.

The third possibility – production at loss – materializes if P < ATC at equilibrium[1]. Even if the firm incurs a loss, it will continue to be in business as long as the total revenue generated from the sale of the product, at least, covers the total variable cost (i.e., as long as TR ≥ TVC or P ≥ AVC).

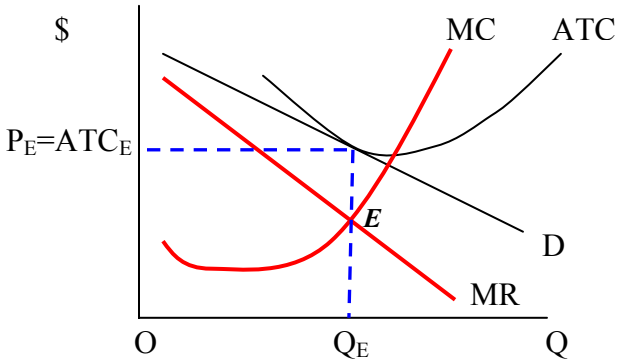
[1] To easily understand how the relationship between P and ATC could be translated to the level of profit, multiply both P and ATC by the level of output (Q). By doing so you would notice that if P > ATC, then P.Q > ATC.Q. Now recall that P.Q is total revenue (TR) and that ATC.Q is the total cost (TC). This means that if P > ATC, then TR > TC, which more clearly indicates that profit (TR – TC) is positive. The other two cases (where P = ATC and P < ATC) can be analyzed in the same way.

As can be clear from the third case above, the total fixed cost does not enter the short run decision of a firm to either shutdown or continue operation as it could not be recovered (avoided) by closing down. Fixed costs are incurred whether the firm chooses to produce or to cease operation. Hence, to decide whether to shutdown or to continue operation, the firm examines whether the revenue from producing and selling its product can cover its total variable cost or not.

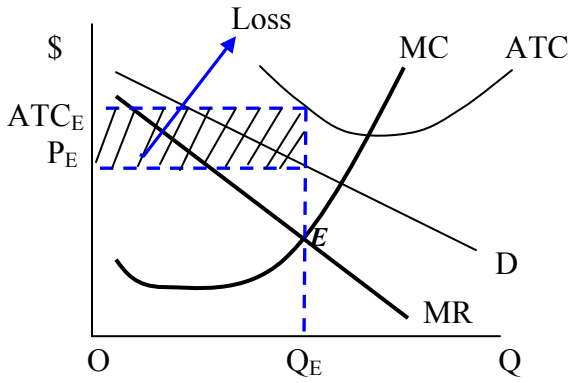
Figure 1.2 below depicts the three possible short run equilibria of a firm in a monopolistically competitive market structure.



(a) Short Run Equilibrium with Excess Profit



(b) Short Run Equilibrium with Normal Profit



(c) Short Run Equilibrium at Loss

Figure 1.2: Short Run Equilibrium of a Firm in Monopolistic Competition

In the long run, however, the monopolistically competitive firm breaks even. In other words, for a monopolistic competitor, production at positive profit or production at loss is unrealistic and doesn't exist in the long run. This is one similarity between perfectly competitive and monopolistically competitive firms. However, there is a difference between how the two categories of firms reach this long run equilibrium. While perfectly competitive firms reach the long run (zero profit) equilibrium as result of free entry and exit, monopolistically competitive firms reach the long run (zero profit) equilibrium because of two forces:

i. Free entry into and exit out of the product group, and
ii. Price competition of the existing firms.

How this long run equilibrium is achieved through the price adjustments of the existing firms and through the entry and exit of firms into and out of the product group, is discussed in the following paragraphs.

Recall the two types of demand curves monopolistically competitive firms face: the planned sales curve and the actual sales curve. One conclusion from that discussion which is very important for our present discussion is the following.

*Price adjustments of the existing firms are shown by shift in the planned sales (**dd**) curve, while entry (or exit) cause shifts in the actual sales (**DD**) curve.*

In order to be able to analyze the equilibrium of the firm and that of the industry on the same diagram, Chamberlin made two 'heroic assumptions', namely:

i. that firms have identical costs, and
ii. that consumers' preferences are evenly distributed among the different products.

That is, although the products are differentiated, all firms are assumed to have identical demand and cost curves.

Under these assumptions, the price in the market will be unique. If all firms have the same cost and demand curves, then the equilibrium of one of the firms is also the equilibrium of every other firm. Hence, what we state for a firm (in our forthcoming discussion) applies to every other firm. This implies that our knowledge of the situation of a single firm will enable us tell the situation of the product group at large.

Moreover, Chamberlin didn't use marginal revenue and marginal cost curves, though they are implicit in his analysis. To keep the diagram as easy as possible, we do not impose the two curves (the MR and MC curves) on our graphs; we make them implicit just as Chamberlin did.

A stable equilibrium is achieved when the planned sales curve is tangent to the LAC curve and the planned (expected) sales are equal to the actual sales. That is, equilibrium attains if the actual sales (*DD*) curve cuts the planned sales (*dd*) curve at the point of tangency of the later (*dd*) with the LAC curve (See Figure 1.3 below).

As a starting point, suppose that our representative firm is in short run equilibrium at point $\boldsymbol{e_0}$, where d_0d_0 intersects D_0D_0. At this point, with the right plant size, the firm enjoys positive (abnormal) profit since P > ATC.

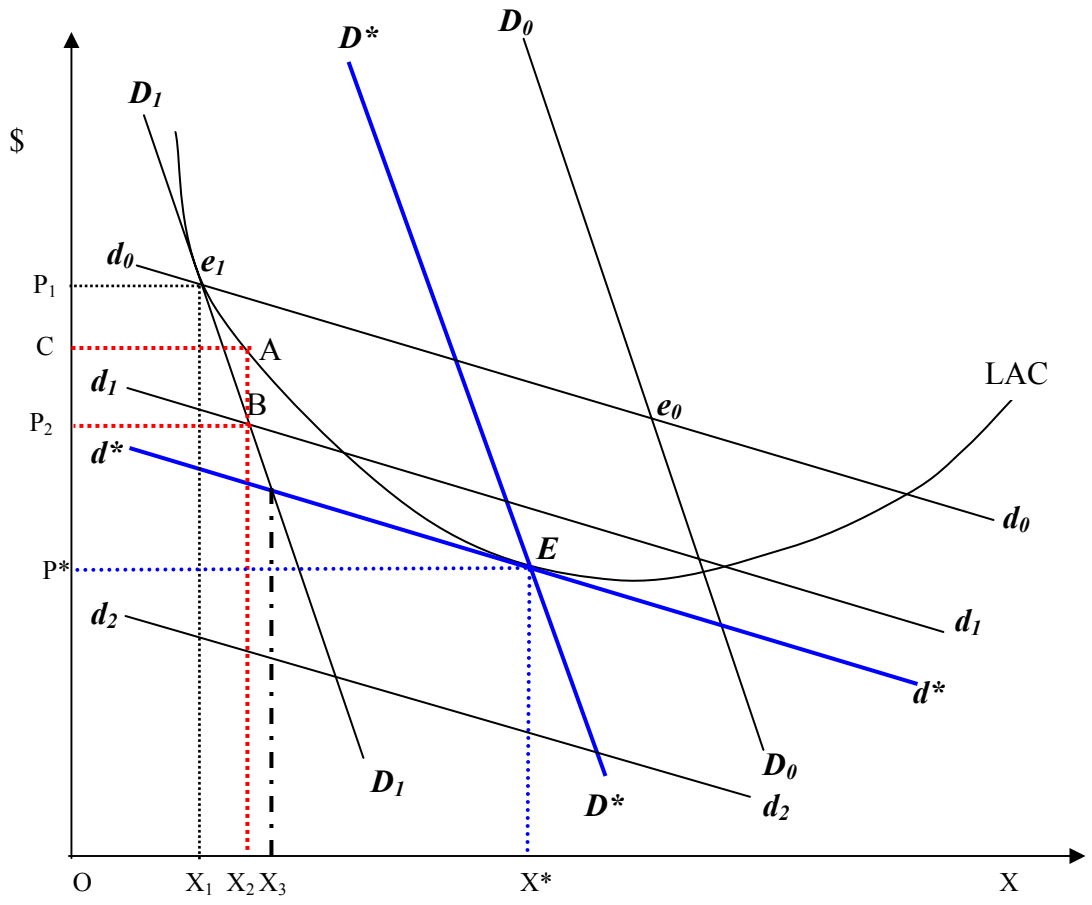


Figure 1.3: Long Run Equilibrium in Monopolistic Competition

In the long run, this excessive profit attracts new firms to the industry. Remember that the entry of new firms reduces the market shares of the existing firms, shifting the actual sales curve to the left. The entry of new firms continues until the actual sales curve shifts from D_0D_0 to D_1D_1. One might think that $\boldsymbol{e_1}$ is the long run equilibrium (with price P_1 and output X_1) since only normal profits are earned at this point. However, this is not the case. This is because each firm takes d_0d_0 as its demand curve; and, consequently believes that if it reduced price, its sales would expand along d_0d_0 and profits would increase.

But each firm has exactly the same incentive and all firms reduce their price. As all firms reduce price, the planned sales curve slides down D_1D_1 (from the initial position d_0d_0)

and each firm realizes a loss instead of a positive profit. For example, when the planned sales curve is at position d_1d_1, the representative firm has reduced its price to P_2 but, as all firms act similarly, the firm produces X_2 units at a total loss equal to the area ABP_2C. However, the firm acts on the 'myopia' curve d_1d_1 and so long as this lies above the LAC, it believes that it can obtain positive profits by cutting its price. The loss increases still further as the planned sales curve slides further down (to d^*d^*) along D_1D_1. One might think that the process would stop when the planned sales curve becomes tangent to the LAC curve (at E). This would be the case if the firm could produce X*. However, there are too many firms in the industry and the share of the firm is only X_3 (on D_1D_1). The firm, still on the 'myopia assumption', believes that it can reach X* if it reduces price to P*. However, all firms do the same and the planned sales curve falls (from d^*d^* to d_2d_2). Now the planned sales curve (d_2d_2) lies below the LAC curve at all levels of output implying loss for whatever level of production.

As a result, the financially weakest firms will eventually begin to leave the industry. As the number of firms declines because of the exit of these weak firms, any surviving firm will have a larger share: the actual sales curve moves (from D_1D_1) to the right. As the actual sales curve moves up, the firm begins to revise its expectations about the demand for its product and thus its planned sales curve also starts to shift up (up from d_2d_2). Exit will continue until:

- the planned sales curve becomes d^*d^*, tangent to the LAC curve, and
- the actual sales curve becomes D*D*, and as a result
- D*D*cuts d*d* at the point of tangency of d*d* and LAC (point E).

Equilibrium is then stable at point E with normal profits earned by all firms and no entry or exit taking place. The equilibrium price P* is unique and each firm has a share equal to OX*.

To conclude, the interplay of the two forces – the free entry and exit of firms which shift the actual sales (*DD*) curve and the price competition of the existing firms which shifts the planned sales (*dd*) curve – will ensure the long run equilibrium of normal profit.

1.5 Excess Capacity and Welfare Loss

As shown in the previous section, the long run equilibrium of a monopolistically competitive firm is defined by the point of tangency of the demand (*dd*) curve and the LAC curve. At this point, MR = MC and LAC = P, but P > MC. As a result, price will be higher and output will be lower as compared to the long run equilibrium of a perfectly competitive firm. There are also too many firms in the product group, each producing an output less than the socially optimal level (at a point on the falling part of the LAC). The difference between the level of output indicated by the lowest point on the LAC curve and the monopolistic competitor's output when in long run equilibrium measures *excess capacity*. Put differently, excess capacity is the difference between the ideal level of output corresponding to the minimum LAC (Q_I in Figure 1.4) and the output actually attained in the long run equilibrium (Q_A in the same figure).

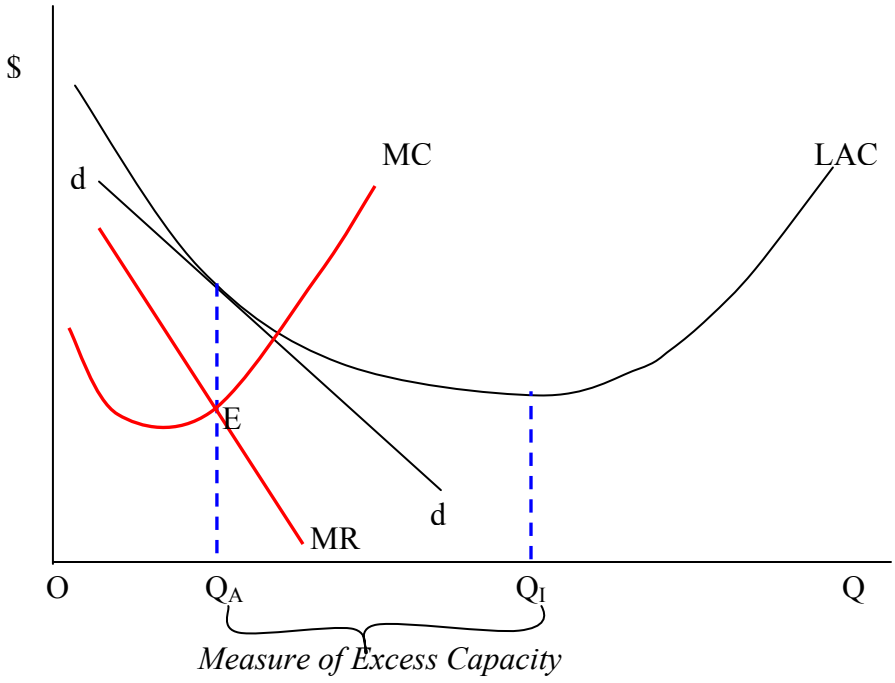


Figure 1.4: Excess Capacity

Chamberlin doesn't agree with this measure of excess capacity however. According to Chamberlin, long run equilibrium under monopolistic competition does not give rise to

excess capacity so long as the market is characterized by active price competition. In his view, excess capacity arises when free entry is coupled with the *absence* of active price competition. Under the conditions of free entry and active price competition, Chamberlin argues that, the equilibrium output will be close to the minimum cost output, because firms will be competing along their individual demand (*dd*) curves that are very elastic. And any difference between the equilibrium level of output reached under such conditions and the level of output corresponding to the minimum point on LAC curve is justifiable. That is, consumers do desire variety of products; or, product differentiation reflects the desires of consumers who are willing to pay the higher price in order to have choice among differentiated products. The higher cost, resulting from producing to the left of the minimum LAC, is thus socially acceptable. Consequently, the difference between minimum cost output Q_I and actual output Q_A (in Figure 1.4) is not a measure of excess capacity but rather a measure of the 'social cost' of "differentness" which is considered as a desirable quality.

Chamberlin takes Q_A (in Figure 1.4 above) as the potential or ideal level of output and any deviation from it, which results from absence of price competition, as a measure of excess capacity. If firms avoid price competition and instead enter into non-price competition, there will be excess capacity (restriction of output) and higher price in each firm and for the industry. In this event, the firm ignores its planned sales curve (*dd* in Figure 1.5 below) and concerns itself only with its market share (share-of-the-market curve or the actual sales curve). In other words, *DD* becomes the relevant demand curve.

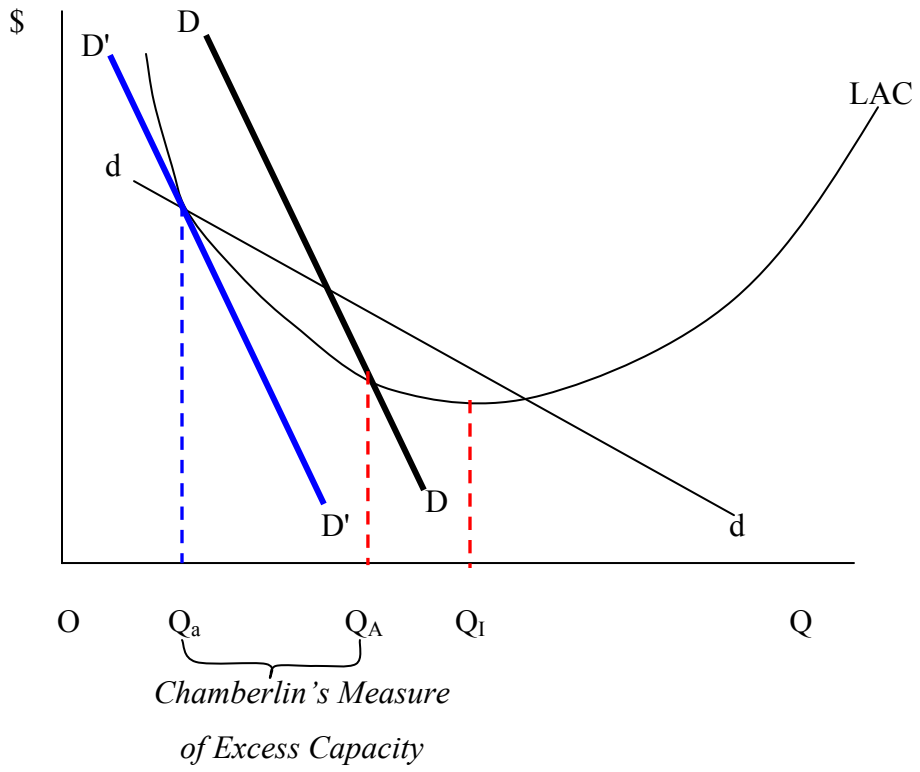


Figure 1.5: Excess Capacity According to Chamberlin

The long run equilibrium (in this case) is reached only after free entry has shifted the DD curve to a position of tangency with the LAC curve (D'D'). Each firm will produce Q_a units of output, as opposed to Q_A (what could have been the case under active price competition). The difference between Q_A and Q_a gives what Chamberlin regards as excess capacity.

Consider a hypothetical example from the soap industry. According to Chamberlin, there won't be excess capacity if the existing soap producers are engaged in active price competition and if any positive profit is eroded away by free entry into (or any prolonged loss leads to exit out of the product group. If, however, our soap producers do not compete on price basis, there will be excess (unutilized) capacity even though free entry and exit ensures zero profit in the long run.

In general, the social welfare implications of monopolistic competition are ambiguous.

On the one hand, each firm produces less than the socially optimal output (as compared to prefect competition) and charges P > MC. That is, the price that consumers are willing to pay for the last unit is greater than the extra cost incurred to produce and sell that unit.

On the other hand, if each firm were somehow forced to produce this level of output at marginal cost price, private enterprise would no longer represent a viable economic system because this requires production at loss, which is impossible in the long run. Even by selling its product at a price above marginal cost, a monopolistically competitive firm just breaks even. If it were to charge P = MC, it would incur loss in the long run. This implies that a monopolistically competitive firm, if forced to charge P = MC, will leave the market. Or, in short, monopolistic competition will be eliminated. Elimination of monopolistic competition itself represents loss of social welfare as freedom of choice to consumers (the quality of 'differentness') will be lost. This implies that the existence of monopolistic competition has a positive contribution to social welfare.

Thus, to say a word on the social welfare impact of monopolistic competition, one has to compare these two opposing implications of its existence and take the net. In addition to the above two points, the welfare effects of monopolistic competition depend on the costs and benefits of advertising (or selling activities in general).

From the viewpoint of a monopolistically competitive firm, advertising can induce consumers to attach importance to its product's distinguishing features. Hence, increases in advertising expenditure reduce the product's price elasticity of demand (or increase the monopoly power of the firm).

From the society's point of view, advertising has a variety of social effects, some positive, and some negative. The following are the well-established disadvantages and advantages of advertising in the literature.

Disadvantages:

i. Some advertising misleads and creates false needs (*advertising is manipulative*). In the face of misleading advertising, consumers could make purchasing decisions that are far from optimal from their own point of view (but fine from the advertiser's perspective).

ii. If advertising makes consumers recognize certain brands and if it encourages them to be loyal to these brands, sellers may have more power to raise prices without losing sales to competitors. If for instance, a firm producing one of the soap brands (say, brand "X") becomes successful at convincing consumers about the superiority of its product/brand to others, it could charge a price higher than that of its competitors. That is, advertising could grant a strong monopoly power to this producer.

Advantages:

i. It is costly and time consuming for consumers to find the best price for a product they want to buy. If advertising (of various stores' prices) enables consumers to identify and get to the lowest-price seller of an item at a lower (additional) cost, the identification of this lowest-price store is worthwhile – and the consumers will be better off as a result of this advertising (*advertising is informative*). For example, it might cost you an hour (with an opportunity cost of 40 Birr) to find the lowest-price seller of a particular brand of shirt in Jimma. Suppose you can access the price catalogues of all shirt sellers from a weekly commercial paper for which you have to pay 5 Birr. Then you would be better off by paying 5 Birr and buying the shirt within few minutes than devoting an hour of your time to searching for the lowest-price seller yourself (sacrificing about 40 Birr).

ii. In contrast to the view that advertising promotes monopoly power, some economists argue that advertising tends to increase the price elasticity of demand for products. According to these economists, the greater the advertising

effort, the more price-elastic the demand for a good becomes. And the more elastic the demand for a good, the more competitive is the market for that good – and frequently the lower is its price. Instead of granting a strong monopoly power to the producer of soap brand X, an intense advertising in the soap industry might make this producer feel that its own customers have got a large number of alternative soap brands in the market to shift to.

iii. If advertising is successful in enlarging the market for a firm's product, and if this product's average cost of production falls as more of it is produced, then the saving in production costs may more than offset the advertising cost. Significant economies of scale can be realized by achieving a relatively larger sales volume than if sellers are constrained by bans/restrictions on advertising. If a producer of a particular soft drink could enlarge the market for its product via advertising, the unit cost of production might decline as the firm produces the product in a larger quantity. This may reduce the total average cost if the reduction in the average cost of production (because of the economies of scale) more than offsets any rise in the average advertising cost.

Because advertising is of many kinds, it really is impossible to generalize about whether or not it is socially beneficial. The answer depends on the nature of the advertising and the circumstances under which it takes place. Questions like: "Is the advertising informative or manipulative?" and "Does it enhance or restrict competition?" etc. should be answered before any assertive words are forwarded on the net benefit of advertising.

Thus, in general, the social welfare impacts of the existence of monopolistic competition are inconclusive. The net effect depends on:

- ☞ whether the harm from the fact that $P > MC$ is less than, greater than, or equal to the benefit from having variety of products for choice; and
- ☞ the overall effects of advertising (and other selling activities).

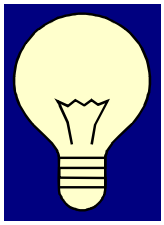
1.6 Criticisms of Chamberlin's Theory of Monopolistic Competition

The effort of Chamberlin to model the behavior of real world firms remains highly significant in economics. However, a number of important criticisms have also been made of his theory. The following are among the criticisms or the reasons for which economists have grown somewhat dissatisfied with the theory.

1) The definition of the set of firms to be included in the product group is extremely ambiguous. Even though we might agree to include firms producing commodities that satisfy similar wants into the same product group, it is not easy to tell how closely related two wants (or commodities) are. Shall we include producers of B-29, 777, Wabel, and OMO together, or put producers of OMO into a different category (product group)? It is not easy!
2) It is unrealistic to assume that firms selling differentiated products have the same demand and cost curves.
3) In markets where there are many small sellers, product differentiation has been found to be slight. As a result, the demand curve facing the monopolistic competitor is close to being horizontal. Under these circumstances, the perfectly competitive model provides a good approximation to the monopolistically competitive solution, and it is also much easier to use for analysis. That is, the easy-to-handle model of perfect competition is a good enough approximation to study the behavior of monopolistically competitive markets where product differentiation is only a slight one. This renders the model of monopolistic competition unnecessary.
4) In many markets where there are strong brand preferences, it usually turns out that there are only a few producers, so that the market is oligopolistic rather than monopolistically competitive. Those monopolistically competitive markets which are not approximated by perfect competition are rather oligopolistic; and not really monopolistically competitive.
5) Even in a market where there are many small sellers of a product, a change in price by one seller may have little impact or no effect on most other sellers of the

product located far away from it, but the price change will have a significant impact on competitors in the immediate vicinity. These nearby firms are, therefore, likely to react to a reduction in price or to an increased promotional effort on the part of the nearby competitor. The nearby competitor is also likely to be aware of this fact and consider it in deciding to change its price or in undertaking a new promotional effort. In cases such as this, the oligopoly model is more appropriate than the model of monopolistic competition.

Despite these serious criticisms, the monopolistically competitive model does provide some important insights such as the emphasis on product differentiation and selling expenses, which are also applicable to oligopolistic markets.



1.7 Summary

1. *Monopolistic competition is the form of market organization in which there is a large number of sellers of a differentiated product, and entry into and exit from the 'industry' (product group) are rather easy in the long run. It has elements both from perfect competition and pure monopoly. Product differentiation (real or fancied) gives each seller some monopoly power, which is, however, severely limited by the availability of close substitutes. The choice-related variables for a monopolistically competitive firm are product variation, selling expenses, and price.*
2. *In the theory of monopolistic competition, there are two kinds of demand curves: the planned sales curve which shifts because of price competition of firms and the actual sales curve (or the share-of-the-market curve) which is shifted by entry of new firms and exit of the already existing ones.*
3. *The short run equilibrium price and output of a monopolistic competitor occur when marginal revenue, MR (based on the planned sales (dd) curve) equals*

marginal cost, MC. The actual sales (DD) curve of the firm is steeper than the planned sales (dd) curve and passes through the point on the dd curve corresponding to the point of intersection of MC and MR. It is possible for the firm to earn a positive profit, normal (zero) profit or incur loss (as long as P≥ AVC).

4. *The long run equilibrium position of a monopolistically competitive firm is the point at which the planned and actual sales curves intersect each other and the planned sales curve is tangent to the LAC curve. In the long run, firms in monopolistic competition, as in perfect competition, get normal (zero) profit. The adjustment to this long run normal profit is brought about by: (1) free entry into and exit from the industry, and (2) active price competition of the existing firms.*
5. *Firms in monopolistic competition realize the long run normal profit at a point on the falling part of the LAC curve. The difference between the ideal level of output corresponding to the minimum point of LAC curve and the actual level of output gives the measure of excess capacity according to some economists. However, Chamberlin argues that this divergence measures the quality of 'differentness' (differentiated product) and no excess capacity exists in monopolistic competition so long as free entry and exit is accompanied by active price competition.*
6. *The social welfare aspects of monopolistic competition are generally ambiguous. On the one hand, each unit is produced at a cost below the price that consumers are to pay (P > MC). On the other hand, the existence of product variety (freedom of choice) – the quality of 'differentness' – in monopolistic competition contributes positively to social welfare. The net of these two effects should be taken, and the impacts (advantages and disadvantages) of advertising (and other selling activities) associated with product differentiation must be added to this net. The social welfare impacts of a specific monopolistically competitive market could be commented on, whereas that of monopolistic competition in general is inconclusive.*

7. *Whether or not advertising is socially beneficial depends on whether it is more of informative or misleading, whether it enhances or restricts competition, and whether it helps in realizing significant economies of scale or not.*

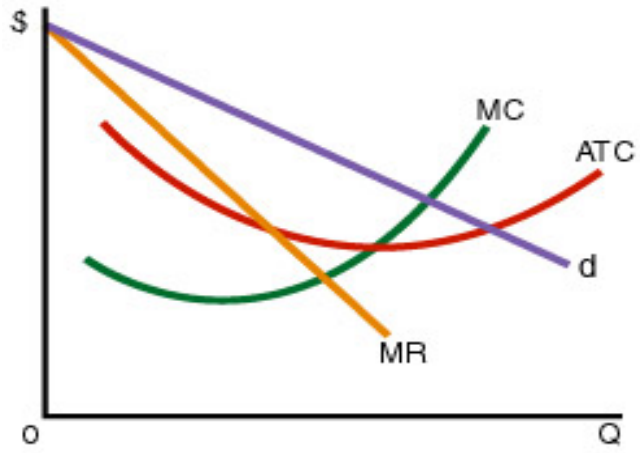


1.8 Review Questions

I. Multiple-Choice Questions

1. What are the characteristics of monopolistic competition?
   A. Many firms, homogeneous product, and artificial barriers to entry.
   B. Many firms, differentiated product, and artificial barriers to entry.
   C. Many firms, differentiated product, and no artificial barriers to entry.
   D. Few firms, differentiated product, and no artificial barriers to entry.
   E. Few firms, homogeneous product, and artificial barriers to entry
2. Which of the following is the preferred strategy of the monopolistic competitor to achieve market power?
   A. Expand plant size in order to achieve economies of scale.
   B. Try to charge the lowest price possible.
   C. Try to minimize the cost of production.
   D. Try to produce a unique product or establish a good reputation.
   E. Try to reproduce the output of other firms as accurately as possible.
3. The demand curve faced by a monopolistic competitor is likely to be _____ elastic than the demand curve faced by a perfectly competitive firm, and _____ elastic than the demand curve faced by a pure monopolist.
   A. less, more
   B. more, less
   C. less, less
   D. more, more

4. To maximize profit, a monopolistic competitor:
   A. Follows the rule of profit maximization unique to perfectly competitive firms.
   B. Follows the same rule used by purely monopolistic firms, since they earn higher profits than competitive firms.
   C. Follows the rule of profit maximization, which applies to any firm regardless of market structure.
   D. Follows an entirely different rule from the one used by a competitive firm or by a pure monopolist.
   E. Sets price equal to marginal cost.
5. The graph below shows the revenue and cost structures of a monopolistically competitive firm. How is this graph similar to, or different from, the graph that depicts a pure monopoly?



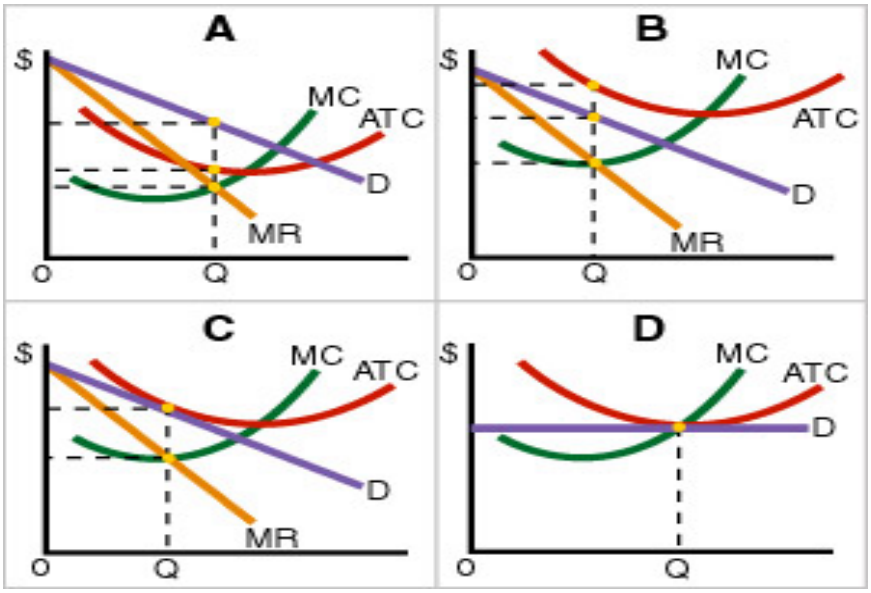
   A. There is no difference. Both graphs are identical.
   B. Marginal cost and average cost look the same as the pure monopolist's, but they are not the same concepts.
   C. The demand curve is downward sloping, but it is not the same concept as the pure monopoly's demand curve.
   D. Both graphs are entirely different. There is no similarity at all.
6. As new firms enter a monopolistically competitive market (over the long run), each of the existing firms will produce _____ output and charge a _____ price.
   A. more, higher
   B. more, lower

C. less, higher

D. less, lower

E. less, the same

7. Which of the graphs below (A, B, C, or D) describes the situation of a monopolistic competitor in the long run?



A. A

B. B

C. C

D. D

E. Both A and B

8. An inefficiency associated with the monopolistic competitor is that the firm does not realize all of the economies of scale available. Which condition below best describes this outcome?

A. Price is greater than marginal cost.

B. Average total cost is not minimized.

C. Marginal revenue equals marginal cost.

D. Price is greater than marginal revenue.

E. Marginal cost and average cost are equal when average total cost is minimum.

9. Identify the correct statement.
   A. The competitive element in monopolistic competition arises from homogeneity of products.
   B. Each firm in monopolistic competition has some monopoly power, which is limited by the existence of close substitutes.
   C. The adjustment of a firm to the long run normal profit position in monopolistic competition is exactly the same as that of a firm in perfect competition.
   D. The condition of entry puts monopolistic competition nearer to pure monopoly than perfect competition.
   E. None of the above.
10. The product of firm A is identical to that of firm B in terms of ingredients, taste, color, accessibility to consumers, etc. But, the services offered by the two firms during time of sales differ. This type of product differentiation is:
   A. Real
   B. Fancied
   C. Imaginary
   D. Spurious
   E. All except A.
11. Which of the following is not a new concept introduced with the evolution of models of monopolistic competition?
   A. U-shaped AVC, ATC and MC of production
   B. U-shaped selling costs
   C. Product differentiation
   D. Product group
   E. A and B

II. True or False Questions

12. In long-run equilibrium, the demand curve of a monopolistic competitor is tangent to its average total cost curve.

13. Excess capacity is the difference between the ideal and the actual levels of output in the long run.
14. For Chamberlin, excess capacity doesn't exist in monopolistic competition as long as there are free entry and exit, and active price competition.
15. Free entry and exit cause shifts in the planned sales curve whereas price competition of the existing firms cause shifts in the share of the market curve of a monopolistic competitor.
16. Firms in monopolistic competition produce less than the socially optimal level and charge a higher price as compared to perfect competition, but the social welfare implication is indeterminate.
17. It is not easy to tell whether advertising is good or bad for the society.

III. Discussion Questions

18. Define monopolistic competition and give some examples. Identify the competitive and monopoly elements in monopolistic competition.
19. Distinguish between real and imaginary product differentiations.
20. Distinguish between the planned and actual sales curves and explain how the position of each is affected.
21. Why is it difficult or impossible to define industry under monopolistic competition?
22. Assuming that all firms are faced with identical cost and demand structures, compare the long run equilibrium results of a monopolistically competitive market with that of a perfectly competitive one.
23. Discuss the concepts of excess capacity and the social cost of producing and offering consumers a great variety of output.
24. Is advertising a good or a bad thing from the society's point of view? Discuss!
25. Comment on Chamberlin's model of monopolistic competition. Explain both his contributions to the theory of pricing and the criticisms.

CHAPTER TWO
OLIGOPOLY

Introduction

In the introduction to the previous chapter, we mentioned two imperfectly competitive markets which represent the real world more accurately than the two extreme cases of perfect competition and pure monopoly. At that point, we chose to deal with monopolistic competition and postponed the discussion of oligopoly to the subsequent chapter.

Now, we are at the juncture to deal with the oligopolistic market structure. Specifically, the chapter discusses the defining characteristics of oligopoly, the types of oligopoly, the various models of oligopoly behavior, as well as the welfare aspects of the oligopolistic market structure.

Chapter Objectives

After an attentive completion of this chapter, students would be able to:

- Define the oligopoly market structure, and explain its sources and distinguishing characteristics.
- Understand the difficulty to have a single widely accepted model to describe oligopoly.
- Determine the equilibrium price and output under the various models of non-collusive and collusive oligopoly.
- Appreciate the possibility of effective competition even with a few firms (contestable markets).
- Comment on the long run efficiency implications and welfare effects of oligopoly.

2.1 Oligopoly Defined

Oligopoly is the form of market structure in which there are a few sellers of either a homogeneous or a differentiated product. With regard to the type of output produced, the oligopoly market could be similar to perfect competition (producing standardized or identical products) or to monopolistic competition (producing closely related but not identical products). The number of sellers is neither as large as that under perfect competition (or monopolistic competition) nor just one as under pure monopoly. The number of firms is taken to be a few where "how few a few is" has never been defined in numbers. However, the limiting number of firms in an oligopolistic market is two. Under monopolistic competition, we saw that firms do compete on non-price basis, say through spending on selling activities like advertising. In oligopoly, such competition is generally more severe: firms do fight to reduce the market share of their competitors to the extent of incurring loss (themselves) so as derive rivals out of market. They do also exercise various practices to block the entry of potential firms.

If oligopoly could be like perfect competition in some aspects, or similar to monopolistic competition in some other aspects, or else like pure monopoly in some others, then how do we differentiate it? The *distinguishing characteristic* of oligopoly is the interdependence or rivalry among firms (and the implied uncertainty) in the industry. In a particular oligopoly market regardless of the nature of the product produced, the exact number of producers, and the variable of choice, the decision of one firm (oligopolist) affects the behavior of other firms. Consequently, each firm has to make its decisions under some assumptions about the current position of its rival(s) and about how each rival would respond to its action.

There is no single widely accepted theory on oligopoly. That is, we don't have a single model (or theory) to analyze the behavior of oligopolistic firms. This is in contrast to all the previous cases. We used just a single model to analyze the behavior of firms under perfect competition. So did we for the cases of pure monopoly or monopolistic competition.

This lack of single model is attributed to the following characteristics of the market structure:

- ☞ Oligopoly involves uncertainty: no firm is sure about the future action plans or the response strategies of its rivals;
- ☞ The products that firms in an oligopoly market produce may be either homogeneous or heterogeneous; and
- ☞ There are various ways of reaction among competitors.

As a result, there are different oligopoly models each of which is applicable to some specific situation(s). In much of the remaining part of this chapter, we will see some models of oligopoly under some specific assumptions made to characterize the market.

The condition of entry puts oligopoly very close to pure monopoly: there are barriers to entry. The sources of oligopoly are generally the same as those of (pure) monopoly but these sources are somewhat weaker in the case of oligopoly. These sources (or barriers to entry) include artificial and natural barriers.

1. Artificial barriers comprise of factors like:
   - ☞ Control over essential raw materials (inputs), which limits the number of firms from growing large;
   - ☞ Possessing patents and/or copyrights, a condition which enables the existing firms to be the only ones to produce a particular product or to use a particular technology (of production);
   - ☞ Government licensing, which restricts the number of producers to 'a few'; and
   - ☞ Limit-pricing, the practice of charging a price low enough to discourage entry into the industry and voluntarily sacrificing short run profits to secure a larger market-share for maximizing profits in the long run.

2. Natural barriers include such factors as:
   - ☞ Economies of scale operating over a sufficiently large range of output so as to leave only a few firms supplying the entire market; and

- ☞ The requirements of huge capital investments and specialized inputs to enter the market.

Note that limit-pricing is a practice designed to exploit barriers that do exist (e.g., economies of scale) more than an artificial barrier to entry. More likely, firms that practice limit-pricing in the expectation of larger profits in the long run are firms which are planning to expand their production to exploit potential economies of scale (declining unit costs).

The limiting case of oligopoly is duopoly – the market in which there are two firms competing with each other. Recall that oligopolists produce and/or sell either identical or differentiated products. If the product supplied by a particular oligopoly is homogeneous, then we have a *pure oligopoly*. If, on the other hand, the product is differentiated (heterogeneous), then we have a *differentiated oligopoly*.

Generally, firms in oligopolistic markets may make decisions separately (may behave independently) even though the decision of one actually affects the other. Such an oligopoly is termed a *non-collusive oligopoly*. Alternatively, oligopolists may enter into tacit (implicit or unspoken) or overt (open or explicit) agreements regarding their decision(s). This second type of oligopoly is referred to as a *collusive oligopoly*.

The models designed to analyze the behavior of firms under each of these two categories will be discussed in the next two sections: models of non-collusive oligopoly in Section 2.2 and those of collusive oligopoly in Section 2.3. In most of the forthcoming discussions, for the purpose of analytical convenience, it is often assumed that the products within the oligopoly market are identical and that there are just two firms – *pure duopoly*.

2.2 Non-Collusive Oligopoly

Economists, in their efforts to analyze firms' behaviors, have developed a number of models which describe the non-collusive oligopoly. Here, we will look at four popular models of the non-collusive oligopoly: Cournot's duopoly model, Bertrand's duopoly model, Stackelberg's duopoly model and the 'Kinked-Demand Curve' model.

2.2.1 Cournot's Duopoly Model

Cournot's model is the earliest duopoly model, developed in 1838. The original version of the model makes the following "heroic" assumptions:

- The duopolists produce a homogeneous product and have identical costs;
- The marginal (additional) cost of production is zero for both firms;
- Both firms have full knowledge of the linear market demand curve they face; and
- Each firm acts and decides on its own to determine its profit-maximizing level of output on the assumption that the competitor will not change its level of output.

Let's begin our discussion of the model proper by assuming that, of the two firms we have, A and B, firm A is the first to start producing and selling mineral water. Note that this assumption is just for a pedagogical reason; otherwise, the firms are making their decisions simultaneously. Refer to Figure 2.1 below for the subsequent discussion.

Given the demand and marginal revenue curves and the assumption of zero marginal cost (MC), firm A will produce a quantity of output equal to A and sells it at a unit price equal to P_1. This holds as MR = MC (= 0) at point A, which is the mid-way between O and D'.

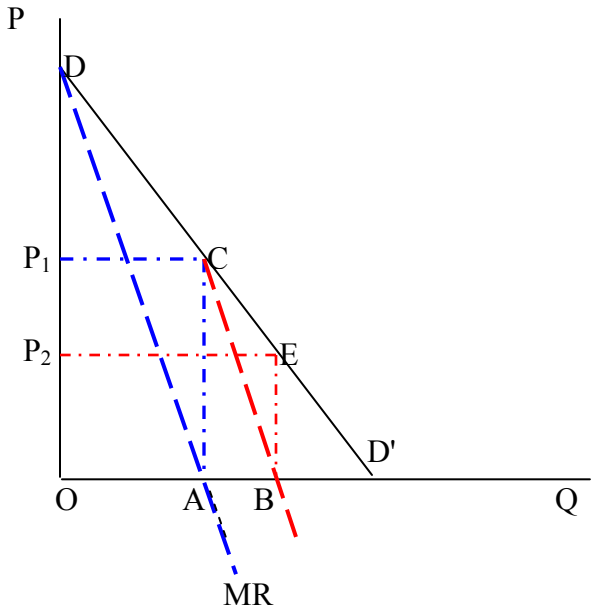


Figure 2.1: Cournot's Duopoly Model

Now, firm B – assuming that firm A will keep its output fixed at OA – considers that its own demand curve is the remainder: CD'. For your convenience, take point A as the origin for firm B. Then, firm B is at the optimal level of production when it produces the amount equal to AB, $(=\frac{1}{2}AD')$, and charges a unit price equal to P_2. Note that firm B supplies half of the unsupplied half or one-fourth of the total market : $(\frac{1}{2}AD'=\frac{1}{2}[\frac{1}{2}]OD'=\frac{1}{4}OD')$.

Now, take the position of firm A again. Firm A, observing that its rival (B) is supplying one-fourth of the total market and assuming that B will hold its quantity constant, decides based on the remaining three-fourth of the market. Hence, A will produce half of the market that is not supplied by B, i.e., $\frac{1}{2}[1-\frac{1}{4}]OD'=\frac{1}{2}[\frac{3}{4}]OD'=\frac{3}{8}OD')$.

From the viewpoint of firm B, observing that A is supplying $\frac{3}{8}$ of the total market and assuming that A will hold its quantity constant, B reacts and produces half of the unsupplied market, i.e., $\frac{1}{2}[1-\frac{3}{8}]OD' = \frac{1}{2}[\frac{5}{8}]OD' = \frac{5}{16}OD'$).

When will this action-reaction come to an end? This action-reaction pattern continues until a stable point (an equilibrium point in economic terms) is reached. At equilibrium, each firm produces and supplies one-third ($\frac{1}{3}$) of the total market (a total of $\frac{2}{3}$); one-third of the market remains unsupplied.

The process below summarizes the action-reaction pattern that takes place between the two firms discussed above.

Instant/Period	***Firm A***	***Firm B***
1	$\frac{1}{2}$	$\frac{1}{2}(1-\frac{1}{2}) = \frac{1}{4}$
2	$\frac{1}{2}(1-\frac{1}{4}) = \frac{1}{2}-\frac{1}{8} = \frac{3}{8}$	$\frac{1}{2}(1-\frac{3}{8}) = \frac{1}{2}[1-(\frac{1}{2}-\frac{1}{8})-\frac{1}{8}] = \frac{1}{4}+\frac{1}{16} = \frac{5}{16}$
3	$\frac{1}{2}(1-\frac{5}{16}) = \frac{1}{2}-\frac{1}{8}-\frac{1}{32} = \frac{11}{32}$	$\frac{1}{2}(1-\frac{11}{32}) = \frac{1}{4}+\frac{1}{16}+\frac{1}{64} = \frac{21}{64}$
. . .	. . .	. . .

At some instant/period *t* (after a large number of rounds of the action-reaction shown above), firm A produces an amount $Q_A = \frac{1}{2}-\frac{1}{8}-\frac{1}{32}-\frac{1}{128}-... = \frac{1}{2}-(\frac{1}{8}+\frac{1}{32}+\frac{1}{128}+...)$.

The expression in braces is a geometric progression[2] with the first term $G_1 = \frac{1}{8}$ and a common ratio $r = \frac{1}{4}$, and thus $Q_A = \frac{1}{2} - (\frac{1/8}{1 - 1/4}) = \frac{1}{2} - \frac{1}{6} = \frac{1}{3}$. Similarly, the optimal quantity of firm B at period t is $Q_B = \frac{1}{4} + \frac{1}{16} + \frac{1}{64} + \frac{1}{256} + ... = \frac{1/4}{1 - 1/4} = (\frac{1}{4})(\frac{4}{3}) = \frac{1}{3}$.

Clear from the discussion so far is that the optimal level of output a firm produces depends on what the rival (opponent) firm produces. For each level of output firm A produces, firm B has some optimal level of production, and vice versa. Let us now relax the assumption of zero marginal cost and see the Cournot's equilibrium of a duopoly market based on the reaction-curves approach.

A reaction curve/function is a curve/function that shows the relationship between a firm's profit maximizing level of output and the amount it thinks its competitor will produce. For instance, for two firms (A and B), firm A's reaction curve/function shows how much output it (firm A) must produce in order to maximize its own profit for every specific level of output its rival (B) produces. The reaction functions of firms under Cournot's behavioral assumption are downward sloping in a quantity set of axes. This will be clear after a moment.

The following numerical example helps you understand the concept of reaction curve/function and master solving the Cournot's equilibrium of duopolists.

[2] The sum (S) of a geometric series with the first term G_1 and the common ratio (the ratio of a term to the term preceding it) *r*, where $0 < r < 1$, is given by: $S = \frac{G_1}{1 - r}$.

Numerical example:

If the market demand for a product supplied by two firms and the costs of the duopolists are: $P = 100 - 0.5X$ where $X = X_1 + X_2$, and, $C_1 = 5X_1$ and $C_2 = 0.5X_2^2$;

- Find the reaction functions of the two firms.
- Find Cournot's equilibrium (the profit maximizing levels of output for the two firms, the market price they charge, and the profit of each firm).

Solution:

Step 1: Define the profit functions of the two firms.

Firm 1's profit is given by $\pi_1 = PX_1 - TC_1$

$$= [100 - 0.5(X_1 + X_2)]X_1 - 5X_1$$

$$= 100X_1 - 0.5X_1^2 - 0.5X_1X_2 - 5X_1$$

$$\boldsymbol{\pi_1 = 95X_1 - 0.5X_1^2 - 0.5X_1X_2}$$

Firm 2's profit is given by $\pi_2 = PX_2 - TC_2$

$$= [100 - 0.5(X_1 + X_2)]X_2 - 0.5X_2^2$$

$$= 100X_2 - 0.5X_1X_2 - 0.5X_2^2 - 0.5X_2^2$$

$$\boldsymbol{\pi_2 = 100X_2 - 0.5X_1X_2 - X_2^2}$$

Step 2: Maximize the profit function of each firm with respect to its own output (find the reaction function of each firm).

The first order condition for profit maximization: $\Rightarrow \frac{\partial \pi_1}{\partial X_1} = 0$ and $\frac{\partial \pi_2}{\partial X_2} = 0$.

$\frac{\partial \pi_1}{\partial X_1} = 95 - X_1 - 0.5X_2 = 0 \quad \Rightarrow \boldsymbol{95 - 0.5X_2 = X_1}$ …….. [1].

This is firm 1's reaction function. It gives us the firm's profit maximizing level of output (X_1) for every level of production by its rival (X_2).

$$\frac{\partial \pi_2}{\partial X_2} = 100 - 2X_2 - 0.5X_1 = 0 \Rightarrow \boldsymbol{X_2 = 50 - 0.25X_1} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots [2].$$

This is firm 2's reaction function. It gives us the firm's profit maximizing level of output (X_2) for every level of production by its rival (X_1).

Step 3: Solve equations 1and 2 simultaneously to find equilibrium quantities of the two firms.

$$\begin{cases} X_1 = 95 - 0.5X_2 \\ X_2 = 50 - 0.25X_1 \end{cases}$$

$$\Rightarrow X_1 = 95 - 0.5(50 - 0.25X_1)$$
$$\Rightarrow X_1 = 95 - 25 + 0.125X_1$$
$$\Rightarrow X_1 - 0.125X_1 = 70$$
$$\Rightarrow 0.875X_1 = 70$$
$$\Rightarrow \boldsymbol{X_1} = {}^{70}\!/_{0.875} = \boldsymbol{80}$$

And, $X_2 = 50 - 0.25X_1 = 50 - 0.25(80)$

$$\Rightarrow \boldsymbol{X_2} = 50 - 20 = \boldsymbol{30}$$

The reaction function of the two firms and the equilibrium quantities are depicted in the figure below.

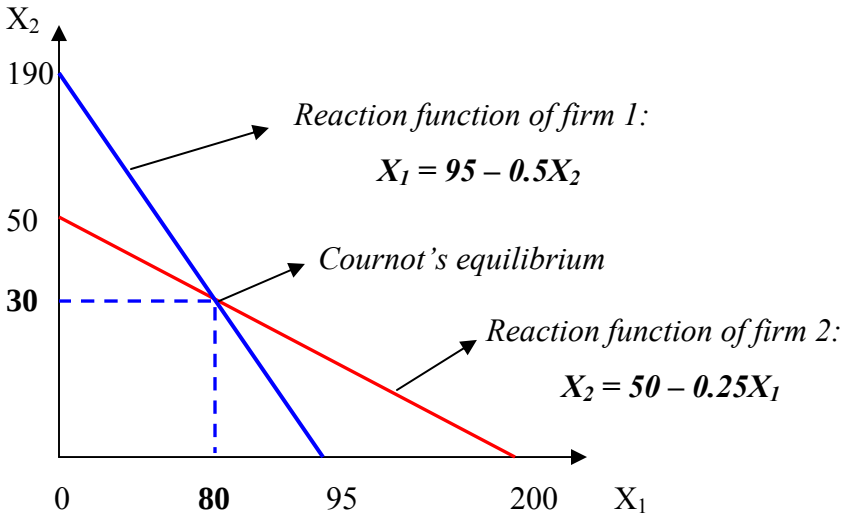


Figure 2.2: Reaction Functions and Duopoly Equilibrium under Cournot's Model

Once we find the equilibrium quantities of the two firms by solving their reactions functions simultaneously, the other values are easy to find.

- ☞ The total output in the market is **X** = X_1 + X_2 = 80 + 30 = **110**. Then, the equilibrium price is determined by substituting this value (X = 110) into the market demand function. Thus, **P*** = 100 – 0.5X = 100 – 0.5(110) = **45.**
- ☞ The profit of each firm at equilibrium is found by substituting the appropriate values into the profit function of the firm (Π_1 or Π_2) from step 1 above. As an exercise, find the profit of each firm by substituting the values obtained into the profit functions! *(Also check the second order conditions by yourselves!).*

Note that whether the quantities supplied by the firms are equal or not depends on whether they face the same cost structure or not. Under the initial assumption of identical cost structure (zero marginal cost for each firm), we found that each firm supplies one-third of the total market. But now, where the costs are different ($C_1 = 5X_1$ and $C_2 = 0.5X_2^2$), the two firms supply different levels of output (i.e., $\mathbf{X_1^* = 80}$ and $\mathbf{X_2^* = 30}$).

Criticisms of the Cournot's Model

Despite its effort to systematically analyze the behavior of duopolists (two rival firms), the Cournot's model is criticized on some grounds. The major points of criticism are the following:

1. The model assumes that firms do not learn from past miscalculations of competitors' reactions. That is, firms are assumed to be so naïve that each assumes that its rival will keep its output constant from time to time though its assumption doesn't materialize. In the context of the earlier discussion, why should firm A assume that its rival (firm B) will keep its output fixed at 5/16 in the third period while firm B had already made some change between periods 1 and 2? If these firms are firms in the real world, they do learn from their past mistakes. If firm A, for its decision in period 2, assumes that its rival keeps on supplying ¼ of the market, but observes that the rival firm produces 5/16 at the end of period 2, then

firm A will never repeat such a mistake. Thus, this assumption of the model is an unrealistic (a heroic) one.

2. It is a closed model in the sense that entry is blocked. The number of firms assumed in the first period (two) remains the same throughout the adjustment process. Only two firms are adjusting to one-another's actions; and no possibility of new firms entering the market (or at least threatening the two firms by trying to enter into the market) is acknowledged.
3. Even if it does not impair (take away) the validity of the model, the assumption of zero production costs is unrealistic. Cournot assumed that the two firms produce at zero marginal costs of production. While this could simplify his analysis, it is not realistic to assume zero production cost. However, as in the numerical example above, the model still holds where positive costs of production are taken.
4. Variables of competition other than quantity are not included in the model. In the real world, firms compete not only in terms of the quantity of output they supply to the market but also in terms of the prices they charge and possibly selling activities. (The latter one holds if we assume a differentiated oligopoly).

2.2.2 Bertrand's Duopoly Model

Bertrand's duopoly model (which was developed in 1883) derives from Cournot's duopoly model we saw above. It differs from Cournot's in that this one assumes that *each firm expects that the rival will keep its price constant*, irrespective of its own decision about pricing. Each firm is faced by the same market demand, and aims at the maximization of its own profit on the assumption that the price of the competitor will remain constant. Like the Cournot's model, it applies to firms that produce the same (homogeneous) good and make their decisions at the same time. In this case, however, the firms choose prices instead of quantities. In short, the difference between Cournot's and Bertrand's models is the difference in the variable of competition – quantity for Cournot and price for Bertrand.

Here begins the arguments of Bertrand. Because the good is homogeneous, consumers purchase only from the lowest-price seller. Thus, if the two firms charge different prices, the lower-price firm will supply the entire market and the higher-price firm will sell nothing. This is a realistic assumption if information is perfect; if two firms (A and B) produce and supply the same commodity to a market and if one of the firms (A) charges Birr 10 per unit while the other (B) charges Birr 6 per unit, then no one with this information would buy from firm A. If, on the other hand, both firms charge the same price, consumers will be indifferent as to which firm they buy from. And, the model assumes that each firm will supply half the market.

Numerical example:

Consider the following example where the market demand for a good is $P = 30 - Q$ (where $Q = Q_1 + Q_2$) and both firms have a marginal cost of 3 (i.e., $MC_1 = MC_2 = 3$). What will be the Bertrand's equilibrium of this market?

Solution:

If the two firms charge the same price, say, Birr 5 per unit, each could get a per unit profit of Birr 2 (assuming zero fixed costs). Each unit of output is being produced at an additional cost of Birr 3 and is being sold for Birr 5. However, under such situation, at least one firm will have the incentive to cut price and undersell the other. That is, if one of the two firms reduces its price to Birr 4 per unit but the other firm keeps on charging Birr 5, the lower-price firm will sell each unit at a marginal profit of Birr 1 (= Birr 4 – Birr 3). By reducing its price, this firm would be able attract all consumers away from its rival firm, an action which possibly raises its profit. This means that P = 5 will not be a stable (equilibrium) price.

In general, as long as price is above the marginal cost (P > MC), there will be an incentive to reduce price, and thus, the equilibrium will be the competitive outcome – where price equals marginal cost.

Hence,

- ☞ the equilibrium price for the good is: P* = MC = **3**.
- ☞ the total equilibrium quantity Q* is obtained by substituting P = 3 into the market demand function: $P = 30 - Q \Rightarrow 3 = 30 - Q \Rightarrow Q^* = \mathbf{27}$.
- ☞ under the assumption of equal market share, each firm supplies $\frac{27}{2} = 13.5$ units (i.e., $Q_1 = Q_2 = \mathbf{13.5}$).

Bertrand's model is criticized on similar grounds as that of Cournot's (criticisms 1, 2 and 4 which we saw under *Criticisms of the Cournot's Model*). That is, the model takes naïve firms which do not learn from past mistakes; only two firms are assumed from the start to the end of the adjustment process; and only one variable of competition (this time price) is recognized. An additional limitation (point of criticism) of Bertrand's duopoly model points to the assumption of *the equal market share* of total sales by the firms.

Recall how firms in perfect competition that produce exactly identical goods compete. Each firm (perfect competitor) recognizes that it cannot choose the price to charge but takes the price as given and looks for a level of production that yields the maximum profit. In general, it sounds more natural to compete by setting quantities rather than prices when firms produce a homogenized good. Price competition would be more natural when products have some degree of differentiation than when products are identical. That is, Bertrand's model could have been better if it considered firms producing a differentiated product. Though Bertrand's model didn't go so far to deal with product differentiation, let us extend this model to the case of price competition with differentiated products.

Numerical example:

To illustrate the extension we proposed to Bertrand's model, assume that the two duopolists have:

- ✍ fixed costs of 10 and 15 (i.e., $TFC_1 = 10$ and $TFC_2 = 15$),

- marginal costs of 3 and 2 (i.e., $MC_1 = 3$ and $MC_2 = 2$), and
- the demand functions given by:
  - $Q_1 = 12 - 2P_1 + P_2$ (firm 1's demand function)
  - $Q_2 = 24 - 4P_2 + 4P_1$ (firm 2's demand function)

Solution:

First note that we are given marginal and fixed costs while the definition of profit requires total cost. Thus, we should obtain the total cost functions of the two firms.

- From $MC_1 = 3$, it follows that $TVC_1 = 3Q_1$, and[3]
- Similarly, from $MC_2 = 2$, we will have $TVC_2 = 2Q_2$.
- Subsequently, $TC_1 = 3Q_1 + 10$ and $TC_2 = 2Q_2 + 15$.

Then, the equilibrium of the two firms setting their prices at the same time (under Bertrand's behavioral assumption) will be determined as follows:

Step 1: Define the reaction functions of the firms by first defining their profit functions.

Firm 1's profit:

$$\pi_1 = TR_1 - TC_1 = P_1Q_1 - (10 + 3Q_1)$$
$$\Rightarrow \pi_1 = P_1(12 - 2P_1 + P_2) - [10 + 3(12 - 2P_1 + P_2)]$$
$$\Rightarrow \pi_1 = 12P_1 - 2P_1^2 + P_1P_2 - 10 - 3(12 - 2P_1 + P_2)$$
$$\Rightarrow \pi_1 = 12P_1 - 2P_1^2 + P_1P_2 - 10 - 36 + 6P_1 - 3P_2$$
$$\Rightarrow \boldsymbol{\pi_1 = 18P_1 - 2P_1^2 + P_1P_2 - 3P_2 - 46}$$

Π_1 is maximized when: $\frac{\partial \pi_1}{\partial P_1} = 0$. $\Rightarrow 18 - 4P_1 + P_2 = 0$

$$\Rightarrow \boldsymbol{P_1 = 4.5 + 0.25P_2} \quad \text{..... (1).}$$

This is firm 1's reaction function.

[3] Note that $\int MC$ (the integral of MC) over the relevant range of output gives us the total cost, i.e., $\int MC = TVC + k$, where k represents the fixed cost for the functions of the type given in our question.

Firm 2's profit:

$$\pi_2 = TR_2 - TC_2 = P_2Q_2 - (15 + 2Q_2)$$
$$\Rightarrow \pi_2 = P_2(24 - 4P_2 + 4P_1) - [15 + 2(24 - 4P_2 + 4P_1)]$$
$$\Rightarrow \pi_2 = 24P_2 - 4P_2^2 + 4P_1P_2 - 15 - 2(24 - 4P_2 + 4P_1)$$
$$\Rightarrow \pi_2 = 24P_2 - 4P_2^2 + 4P_1P_2 - 15 - 48 + 8P_2 - 8P_1$$
$$\Rightarrow \boldsymbol{\pi_2 = 32P_2 - 4P_2^2 + 4P_1P_2 - 8P_1 - 63}$$

Π2 is maximized when: $\frac{\partial \pi_2}{\partial P_2} = 0$. $\Rightarrow \pi_2 = 32 - 8P_2 + 4P_1 = 0$

$$\Rightarrow \boldsymbol{P_2 = 4 + 0.5P_1} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(2).}$$

This is firm 2's reaction function.

Step 2: Solve the two reaction functions simultaneously.

$$\begin{cases} P_1 = 4.5 + 0.25P_2 \\ P_2 = 4 + 0.5P_1 \end{cases}$$

$$\Rightarrow P_1 = 4.5 + 0.25(4 + 0.5P_1)$$
$$\Rightarrow P_1 = 4.5 + 1 + 0.125P_1$$
$$\Rightarrow P_1 - 0.125P_1 = 5.5$$
$$\Rightarrow 0.875P_1 = 5.5$$
$$\Rightarrow \boldsymbol{P_1 = 5.5/0.875 = 44/7 \approx 6.29}$$

$$\Rightarrow P_2 = 4 + 0.5P_1 = 4 + 0.5(44/7)$$
$$\Rightarrow \boldsymbol{P_2 = 50/7 \approx 7.14}$$

Step 3: Substituting $P_1 = \frac{44}{7}$ and $P_2 = \frac{50}{7}$ into the demand functions gives us:

$\Rightarrow Q_1 = 12 - 2(\frac{44}{7}) + \frac{50}{7} = \frac{46}{7}$ and

$\Rightarrow Q_2 = 24 - 4(\frac{50}{7}) + 4(\frac{44}{7}) = \frac{144}{7}$ as the equilibrium quantities.

The following figure summarizes what we found so far for the numerical example.

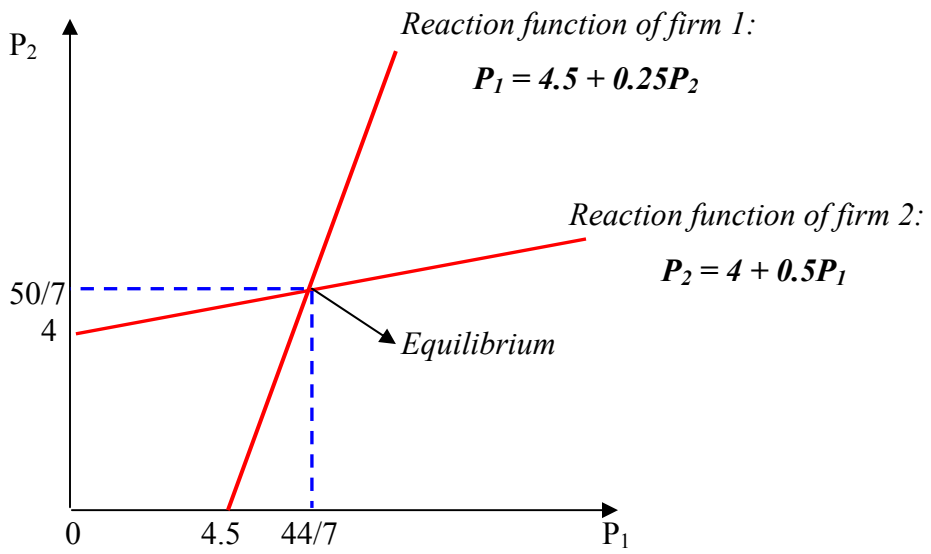


Figure 2.3: Reaction Functions (and Equilibrium) of Firms Producing a Differentiated Product and Competing over Price

As an exercise, find the profit of each firm in the numerical example!

2.2.3 Stackelberg's Duopoly Model

This model was developed by Heinrich von Stackelberg and is an extension of Cournot's model. Like the Cournot's model, this model also assumes two firms producing a homogeneous product and competing over the quantity of output. The difference between Cournot's and Stackelberg's models comes in on whether firms recognize their rivalry or not. Under Cournot's model, the firms are so naïve that they do not recognize their rivalry and make their (output) decisions at the same time. To the contrary, the Stackelberg's model assumes that at least one of the two firms is sufficiently sophisticated to recognize that its competitor is naïve and thus sets its output level ahead of the other.

To visualize how the Stackelberg's model goes and to appreciate how it differs from Cournot's model, consider the numerical example below (which is the same example we used to illustrate the Cournot's model).

Numerical example:

Assume a duopoly market with the demand function $P = 100 - 0.5X$, where $X = X_1 + X_2$, and the costs of the duopolists $C_1 = 5X_1$ and $C_2 = 0.5X_2^2$. Find the Stackelberg's equilibrium:

a. if firm 1 is the Stackelberg's sophisticated leader, and
b. if firm 2 is the Stackelberg's sophisticated leader.

Solution:

(a) Stackelberg's solution with firm 1 considered the sophisticated leader:

By assuming that firm 1 is the leader, we mean that firm 1 knows how firm 2 will respond to its (firm 1's) actions/decisions. But, firm 2 is still assumed to be naïve and just decides on the residual demand (the unsupplied portion of the market) it observes. Thus, firm 1 (the leader):

- calculates the reaction function of firm 2;
- substitutes this reaction function into its own profit function; and then,
- maximizes its profit as if it were a monopolist.

Step 1: Reaction function of firm 2:

$$\pi_2 = PX_2 - TC_2$$

$$= [100 - 0.5(X_1 + X_2)]X_2 - 0.5X_2^2$$

$$= 100X_2 - 0.5X_1X_2 - 0.5X_2^2 - 0.5X_2^2$$

$$\boldsymbol{\pi_2 = 100X_2 - 0.5X_1X_2 - X_2^2}$$

$$\frac{\partial \pi_2}{\partial X_2} = 100 - 2X_2 - 0.5X_1 = 0$$

$$\Rightarrow \boldsymbol{X_2 = 50 - 0.25X_1} \quad \text{.......................................} \quad (1).$$

This is firm 2's reaction function.

Step 2: Substitute this reaction function into the leader's (firm 1's) profit function and maximize.

$$\begin{aligned}\pi_1 &= PX_1 - TC_1 \\ &= [100 - 0.5(X_1 + X_2)]X_1 - 5X_1 \\ &= 100X_1 - 0.5X_1^2 - 0.5X_1X_2 - 5X_1 \\ &= 95X_1 - 0.5X_1^2 - 0.5X_1X_2 \\ \Rightarrow \pi_1 &= 95X_1 - 0.5X_1^2 - 0.5X_1(50 - 0.25X_1) \\ &= 95X_1 - 0.5X_1^2 - 25X_1 + 0.125X_1^2 \\ &= 70X_1 - 0.375X_1^2\end{aligned}$$

$$\Rightarrow \frac{\partial \pi_1}{\partial X_1} = 70 - 0.75X_1 = 0$$

$$\Rightarrow \boldsymbol{X_1 = \frac{70}{0.75} = \frac{280}{3}}.$$ (Check the second order condition)

Step 3: Given the assumption that firm 2 is naïve, it would assume that firm 1 keeps its output at $X_1 = \frac{280}{3}$ units, and thus substitutes this amount into its reaction function and produces:

$$X_2 = 50 - 0.25X_1 = 50 - 0.25(\frac{280}{3}) = \frac{80}{3} \text{ units.}^4$$

Now, it would be easier to find the price charged and the profit of each firm. To find the price charged per unit of output, substitute the values of X_1 and X_2 into the market demand function:

$$P = 100 - 0.5X = 100 - 0.5(\frac{280}{3} + \frac{80}{3}) = 100 - 0.5(\frac{360}{3}) = \frac{120}{3} = 40.$$

Finally, the profit levels of both firms (the leader and the follower) could be found by substituting the values of X_1 and X_2 into their respective profit functions.

[4] Alternatively, firm 2 maximizes $\Pi_2 = PX_2 - TC_2 = [100 - 0.5X_1 - 0.5X_2]X_2 - 0.5X_2^2 = [100 - 0.5(280/3) - 0.5X_2]X_2 - 0.5X_2^2 = (160/3)X_2 - X_2^2$, which yields the same level of output: $X_2 = 80/3$ units.

$$\Rightarrow \pi_1 = 70X_1 - 0.375X_1^2 = 70(\frac{280}{3}) - 0.375(\frac{280}{3})^2 \approx 3267.$$

$$\Rightarrow \pi_2 = 100X_2 - 0.5X_1X_2 - X_2^2 = 100(\frac{80}{3}) - 0.5(\frac{280}{3})(\frac{80}{3}) - (\frac{80}{3})^2 \approx 711.$$

(b) If firm 2 is assumed to be the Stackelberg's sophisticated leader:

As an exercise, follow the steps in (a) above but interchange the positions of the two firms to arrive at the results: X_2 = **35**, X_1 = **77.5**, Π_2 = **918.75**, $\Pi_1 \approx$ **3003**. *What is the price charged per unit of output?*

In the above cases, we assumed that either firm 1 or firm 2 knows the reaction function of its rival, but not both. What happens if each firm knows the reaction function of the competitor? In this case, each firm (duopolist) estimates the maximum profit it would earn (i) if it acted as a leader, (ii) if it acted as a follower, and chooses the behavior which yields the largest maximum profit. Generally, there are three possibilities that could arise: one of the firms may want to be leader and the other follower; each firm might want to be leader; or each firm might prefer to be follower.

- If one of the two firms wants to be leader and the other wants to be follower, then there will be a determinate equilibrium. This will yield the same result as the cases seen in the numerical example above; the follower could be taken as if it didn't know the reaction function of its rival.
- If both firms desire to be followers, then their expectations do not materialize. If one of the rivals doesn't alter its behavior and act as a leader, then the Cournot's equilibrium will be reached. That is, if each duopolist wants to be a follower and recognizes that its rival also wants to be a follower, it first observes what its rival supplies and decides on the basis of the unsupplied portion of the market. The market will then generate Cournot-type equilibrium.
- If both firms want to be leaders, disequilibrium arises. The final outcome of such disequilibrium, according to Stackelberg, is economic warfare or fighting. Where each firm attempts to produce a larger output acting as a leader instead of acting

as a follower, the market might be oversupplied (at a point in time) exerting a downward pressure on price. As a result, profits won't be maximized as planned, and thus the firms start to fight. Equilibrium will be reached either by collusion (some kind of agreement between the firms), or after the 'weaker' firm (the one which cannot withstand the fighting) is eliminated or surrenders/succumbs to the leadership of the other.

2.2.4 The 'Kinked-Demand Curve' Model

Paul Sweezy introduced the kinked-demand model in 1939. This model attempts to explain the price rigidity that is often observed in some oligopolistic markets. Sweezy postulated that if an oligopolist raised its price, it would lose most of its customers because the other firms in the industry would not match the price increase. On the other hand, an oligopolist could not increase its share of the market by lowering its price, since its competitors would immediately match the price reduction. As a result, oligopolists face a demand curve that is highly elastic (sensitive) for price increases and less elastic (insensitive) for price reductions. That is, the demand curve faced by oligopolists has a kink at the established price. Consequently, oligopolists tend to keep prices constant even in the face of changing costs and demand conditions. Hence, interdependence is recognized in Sweezy's model.

The arguments of the model are explained below with the help of Figure 2.4.

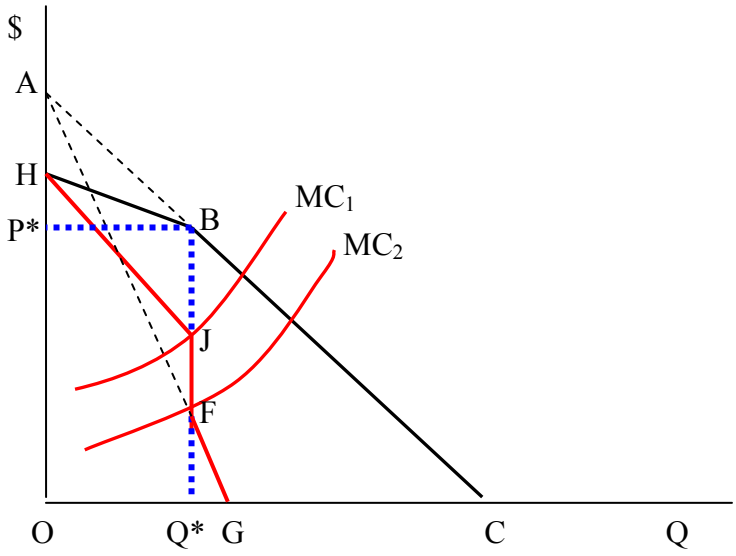


Figure 2.4: The 'Kinked – Demand Curve' Model

The demand curve facing the oligopolist is HBC with a kink occurring at a point corresponding to the prevailing price P* and quantity Q*. This demand curve is more elastic above the kink than below it. In other words, the firm is very likely to lose significantly for raising its price above P* but very unlikely to gain by cutting its price. The oligopolist's marginal revenue curve is given by HJFG (HJ corresponding to the demand curve above the kink – HB, and FG corresponding to the part of the demand curve below the kink – BC). It is discontinuous between J and F. This implies that the oligopolist's output and price remain unchanged even in the face of changing costs. Particularly, for the movement of the firm's marginal cost curve between points J and F, say from MC_1 to MC_2, the equilibrium price and quantity will not be affected.

There is only one case in which a rise in cost will most certainly induce the firm to increase its price, despite the fact that the higher costs pass through the discontinuity of the marginal revenue curve. This occurs when the rise in costs is general (for example, imposition of a sales tax) and affects all firms equally. Under these circumstances, the firm will increase its price with the certainty that the others in the industry will follow, since their costs are similarly affected.

The major weakness of this model is that it cannot explain at what price the kink occurs and how prices are determined. It takes the price as given and explains the rigidity of that given price.

So far we have seen four models of non-collusive oligopoly. To sum up, the first two models – the Cournot's model and the Bertrand's model – assume naïve firms which do not recognize their interdependence in the sense that each firm assumes that its rival will keep its variable of competition (quantity or price) constant. In the third model – the Stackelberg's model – at least one of the two firms recognizes the interdependence and incorporates the response pattern (reaction function) of its rival into its decision-making. The 'kinked-demand' model is another model of non-collusive oligopoly where firms recognize their interdependence and abstain from changing prices even in the face of changing cost conditions.

Despite the fact that firms recognize their interdependence in the last two models seen above, the firms are still assumed to make decisions separately. Nevertheless, as pointed out in the Stackelberg's model, disequilibrium might arise if each firm wants to be a leader and this might lead to some kind of agreement/collusion between the firms. The discussion of models of collusive oligopoly follows.

2.3 Collusive Oligopoly

One way of avoiding the insecurity (uncertainty) arising from the interdependence in oligopoly markets is to enter into collusive agreements. Thus, related to rivalry, incentives to collude represent another distinguishing characteristic of oligopoly. While firms in perfectly and monopolistically competitive markets have the incentive to compete, the incentives of firms in oligopoly may be either to compete or to collude.

Generally speaking, there are two main types of collusion: cartels and price leadership. Each of these forms may represent (imply) either a tacit (secret) agreement or an open collusion. However, it is usually the case that cartels are overt (open) and formal while

price leadership is tacit. Both of these categories of collusive oligopoly – cartel and price leadership – are further divided into sub-categories. We will look at two types of cartels (joint-profit maximizing, and market-sharing cartels) and three types of price leadership (the low cost firm, the dominant firm, and the barometric price leadership models).

2.3.1 Cartels

Cartel is an organization of suppliers of a commodity aimed at restricting competition and/or increasing profits. There are two typical forms of cartels: cartels aiming at joint-profit maximization, and cartels aiming at sharing of the market. These two types of cartel will be discussed one after another.

2.3.1.1 Cartels Aiming at Joint-Profit Maximization (Centralized Cartels)

In this form of cartel, the aim is to maximize the joint (industry) profit. The situation of a centralized cartel is identical with that of a multi-plant monopolist. Just as the multi-plant monopolist looks for the allocation of production among its plants that maximizes its total profit, the centralized cartel also seeks the maximization of its total profit via allocating production among the cartel members (producers). Very often, centralized cartels are pure oligopoly – an oligopoly where all firms produce a homogeneous product.

The firms in a centralized cartel appoint a *central agency* to which they delegate the authority to decide on:

- The *total quantity* to be produced by the industry,
- The *price* at which the output is sold,
- The *allocation of production* among members, and
- The *distribution of the maximum joint-profit* among the members.

To do this, the central agency has to assess the cost structure of the industry (and that of each member firm) and the market demand for the product. For the simple case of two

firms (duopoly) that formed a centralized cartel, a cartel maximizing the joint-profit, the equilibrium is determined as follows:

Given the market demand function for the product, $P = f(X)$ where $X = X_1 + X_2$, and the cost functions of the two firms, $C_1 = f(X_1)$ and $C_2 = f(X_2)$, the cartel maximizes the joint-profit at a point where the marginal cost of production of each firm equals the common marginal revenue. That is, maximizing the joint-profit requires that:

1. $MR = MC_1 = MC_2$, the First Order Condition **(F.O.C)**, and
2. $\frac{\partial^2 \Pi_1}{\partial X_1^2} < 0$ and $\frac{\partial^2 \Pi_2}{\partial X_2^2} < 0$, the Second Order Condition (**S.O.C)**.

Note that these conditions are the same as the conditions for a profit-maximizing multi-plant monopolist.

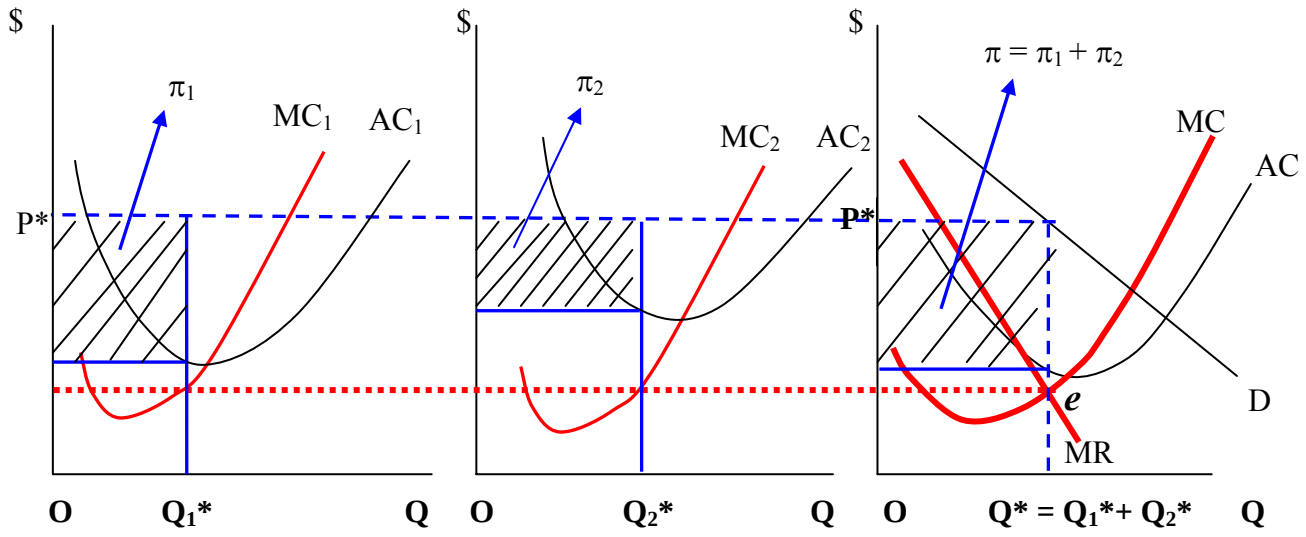


Figure 2.5: Equilibrium Determination under a Centralized Cartel

Note that the profit generated from the production at each firm (π_1 and π_2) does not necessarily go to the generating firm; the total profit π ($= \pi_1 + \pi_2$) could be shared between the two firms on a criterion other than the cost structure of the firms. This is determined by the central agency based on the opening agreements of the firms.

Numerical example:

Consider the following situation in a hypothetical centralized cartel of two firms.

- ☞ The market demand for the product is: $P = 100 - 0.5X$ where $X = X_1 + X_2$;
- ☞ The costs of the colluding firms are: $C_1 = 5X_1$ and $C_2 = 0.5X_2^2$.

Calculate:

a. the level of production that should be allocated to each firm in order to maximize the joint-profit;
b. the per unit price that should be charged; and
c. the maximum profit.

Solution:

a. The objective of the cartel is to maximize joint profit:

$\max_{X_1, X_2} \pi = \pi_1 + \pi_2 = TR_1 + TR_2 - TC_1 - TC_2$.

$$\pi = PX - C_1 - C_2 = P(X_1 + X_2) - C_1 - C_2$$
$$= [100 - 0.5(X_1 + X_2)][X_1 + X_2] - 5X_1 - 0.5X_2^2$$
$$= [100 - 0.5X_1 - 0.5X_2][X_1 + X_2] - 5X_1 - 0.5X_2^2$$
$$= 100X_1 - 0.5X_1^2 - 0.5X_1X_2 + 100X_2 - 0.5X_1X_2 - 0.5X_2^2 - 5X_1 - 0.5X_2^2$$

$\Rightarrow \boldsymbol{\pi = 95X_1 + 100X_2 - 0.5X_1^2 - X_1X_2 - X_2^2}$

Profit maximization requires that: $\frac{\partial \pi}{\partial X_1} = 0$ and $\frac{\partial \pi}{\partial X_2} = 0$. Or, one may use the condition: $MC_1 = MR$ and $MC_2 = MR$.

$\frac{\partial \pi}{\partial X_1} = 95 - X_1 - X_2 = 0$. $\Rightarrow \boldsymbol{95 = X_1 + X_2}$ **…….…….. (1)**

$\frac{\partial \pi}{\partial X_2} = 100 - X_1 - 2X_2 = 0$. $\Rightarrow \boldsymbol{X_1 + 2X_2 = 100}$ **…….……..…. (2)**

Solving equations (1) and (2) simultaneously gives: $\boldsymbol{X_1 = \underline{90}}$ units and $\mathbf{X_2 = \underline{5}}$ units

The total output of the cartel is thus $X = X_1 + X_2 = 90 + 5 = \underline{95}$ units.

b. The market (equilibrium) price is found by substituting X = 95 into the market demand function: P = 100 – 0.5X = 100 – 0.5(95) = **52.5** units**.**

c. The maximum joint profit is found by substituting the values of X_1 and X_2 into the profit equation in (a) above:

$\Pi = 95X_1 + 100X_2 - 0.5X_1^2 - X_1X_2 - X_2^2$

$= 95(90) + 100(5) - 0.5(90)^2 - (90)(5) - (5)^2 =$ ***4525*** units.

Show that this profit is greater than the sum of Π_1 and Π_2 if each firm were to act independently to maximize its own profit. Find the maximum profit of each firm under Cournot's behavioral assumptions; add them; and compare the sum to the maximum profit of the cartel we just found. Do it now!

A centralized cartel is very attractive theoretically. In practice, however, this monopoly equilibrium (maximization of industry profit) may not be achieved. In other words, there are factors that militate against the achievement of the joint-profit maximization of the cartel. These reasons include:

1. *Mistakes in the estimation of market demand and costs*. The central agent might commit mistakes while estimating and/or aggregating the demands and costs of individual firms. Or, the firms may purposely report incorrect figures to influence their shares in total profit.
2. *Slow process of cartel negotiations*. By the time agreement is reached among firms forming a cartel, market conditions may have changed, and the chosen quantity and price may no more be those that result in monopoly profit.
3. *Fear of government interference*. The cartel may purposely under operate in order not to attract the eyes of the government, particularly if the monopoly price yields too high profits.
4. *Fear of entry*. The fear of attracting new firms to the industry by too high profits is another reason why monopoly profit may not be realized.
5. *Stickiness of the negotiated price*. Even if the cartel is aware of changes in market conditions, it needs new negotiation to change the already agreed on price (and quantity) so that profit deviates from the monopoly profit.

2.3.1.2 Market-Sharing Cartels

In this type of cartel, firms agree to share the market, but keep a considerable degree of freedom concerning the style of their product, their selling activities and other decisions. Thus, this type of collusion is more common than centralized cartel. Unlike the case of a centralized cartel where almost all decisions are delegated to a central agency, in the case of market-sharing cartels each firm has many decisions within its control. There are two basic methods for sharing the market: *non-price competition* and *determination of quotas*.

i. Non-Price Competition Agreements

The member firms agree on a common price, at which each firm can sell any quantity demanded. The price is set by bargaining, with the low-cost firms pressing for a lower price and the high-cost firms for a higher price. The agreed on price must be such as to allow some profits to all members.

If all firms have the same costs, the cartel could be stable. With cost differences, however, the cartel will be inherently unstable. This instability results from the fact that the low-cost firms will have a strong incentive to cheat the other members by cutting their prices secretly. Thus, this form of cartel is very loose.

ii. Sharing of the Market by Agreement on Quotas

Firms may make agreement on quotas, i.e., on the quantity that each firm may sell at the agreed on price. If all firms have identical costs, the monopoly solution will emerge with the market being shared equally among the members. If costs are different, the final quota of each firm depends on the level of its costs as well as on its bargaining skill. During the bargaining process two main statistical criteria are most often adopted: past levels of sales and productive capacity. Firms with better levels of sales in the past and those with a potential of expanding production (high productive capacity) are likely to secure more share of the market.

Another popular method of sharing the market is the definition of the region in which each firm is allowed to sell. In this case of geographical sharing of the market, the price as well as the style of the product may differ.

The following sketch summarizes the types of cartels discussed above.

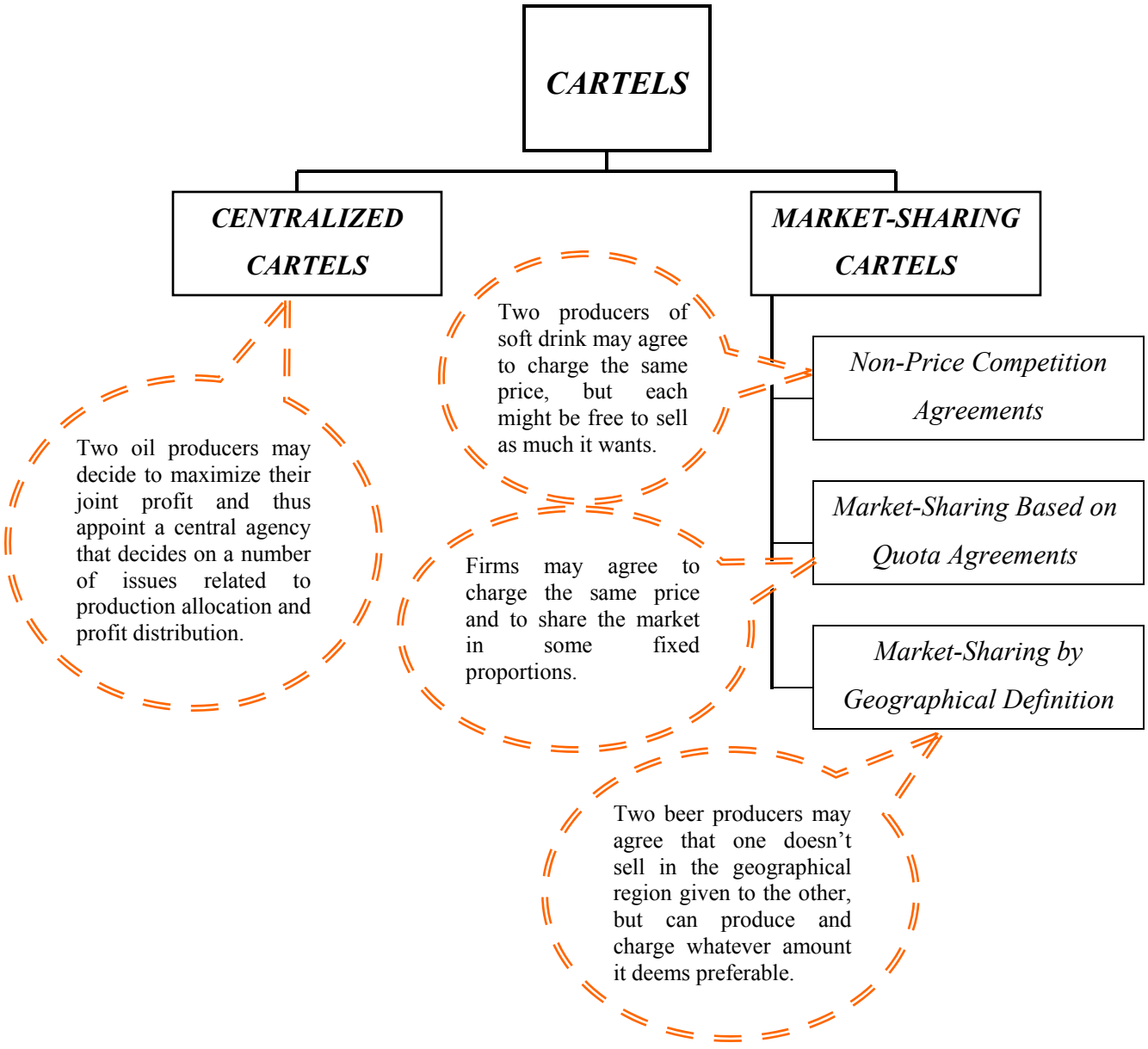


Figure 2.6: Summary Sketch of Cartel Classification

2.3.2 Price Leadership

So far we saw one form of collusion – cartels. The second form of collusion is price leadership. Price leadership is a coordinated behavior of oligopolists where one firm sets the price and the others follow it. The other firms follow the price set by the price leader either because that is advantageous for them or because they prefer to avoid uncertainty about their competitors' reactions (and warfare with their competitor) even if doing so may imply departure of the followers from their profit-maximizing position.

Price leadership allows the members complete freedom regarding their product and selling activities and thus is more acceptable to the followers than a centralized cartel, which requires the surrendering of all freedom of action to the central agency. The prices charged may differ for different firms (particularly if the product is differentiated), but the direction of their changes will generally be the same. That is, even though prices might differ across producers/firms, they do rise and fall together.

There are various forms of price leadership. The most common types are:

A. The low-cost price leadership,
B. The dominant-firm price leadership, and
C. The barometric price leadership.

2.3.2.1 The Low-Cost Price Leadership

The important condition for this model is that firms have unequal costs. It is assumed that there are two firms that produce a homogeneous product at different costs. Clearly, the product (of the two firms) must be sold at the same price. Then, the firm with low cost charges a lower price and its price will be followed by high-cost firm although the price doesn't maximize the profit of the follower. The follower would obtain a higher profit by producing a lower output and selling it at a higher price. But the follower prefers to follow the leader sacrificing some of its profit. The follower prefers to charge whatever price set by the leader (high or low) and does not charge the price which would have

maximized its profit. Why does the follower sacrifice part of its profit and follow the price of low-cost firm? The follower does so in order to avoid a price war, which would eliminate it if price fell sufficiently low as not to cover its long run average cost (LAC).

Although it is stressed that the leader sets the price and the follower adopts it, it is clear that the firms must also enter into a share-of-the-market agreement, formally or informally. Otherwise, the follower could adopt the price of the leader but produce a lower quantity than the level required for maintaining the price in the market, and thus indirectly push the leader to a non profit-maximizing position.

To illustrate the working of this model, suppose the market demand function for a product is defined by: $P = a - bX$ (where $X = X_1 + X_2$), and the cost functions of two firms are: $C_1 = f_1(X_1)$ and $C_2 = f_2(X_2)$ where $C_1 < C_2$.

Given that $C_1 < C_2$, the low-cost price leader in this case will be firm 1. In the simple low-cost firm leadership model, it is also assumed that the two firms share the market equally. That is, firm 1 assumes that the rival firm will produce an equal amount of output to its own, i.e., $X_1 = X_2$.

The leader maximizes its profit on the assumption that $X_1 = X_2$ or equivalently based on the demand function relevant to it. The demand relevant to the leader is found as follows:

$$P = a - b(X_1 + X_2); \text{ but } X_1 = X_2$$

$$P = a - b(X_1 + X_1)$$

$$\boldsymbol{P = a - 2bX_1}$$

The low-cost leader (firm) will then set the price that maximizes its own profit[5].

$$\Pi_1 = R_1 - C_1$$

$$= PX_1 - C_1$$

$$= (a - 2bX_1)\, X_1 - C_1$$

[5] Note that you can substitute $X_1 = X_2$ into the profit function of the leader without defining the demand function relevant to the leader, but that should be done before the derivative is taken.

The first order condition for profit maximization is given by: $\frac{\partial \pi_1}{\partial X_1} = \frac{\partial R_1}{\partial X_1} - \frac{\partial C_1}{\partial X_1} = 0,$ or $MR_1 = MC_1$; and the second order condition is: $\frac{\partial^2 \pi_1}{\partial X_1^{\,2}} < 0$ or $\frac{\partial MR_1}{\partial X_1} < \frac{\partial MC_1}{\partial X_1}$.

The solution to this problem yields the price ***P**** and output $\boldsymbol{X_1^*}$ that the leader produces to maximize its profit. The follower will adopt the same price set by the leader (***P****) and will produce $\boldsymbol{X_2^*}$ (=$\boldsymbol{X_1^*}$ in this case), but does not necessarily maximize its profit.

Numerical example:

Consider two oligopolists (firm 1 and firm 2 as usual) facing the market demand given by: P = 105 – 2.5X (where X = X_1 + X_2), and cost functions: $C_1 = 5X_1$ and $C_2 = 15X_2$. Assuming that the two firms share the market equally,

a. Which of the firms is the low-cost price leader? Why?
b. Determine the profit maximizing level of output and price for the leader.
c. What amount should the follower produce and what price would it charge?
d. What would have been the optimal level of output and price if the follower were to maximize its profit?

Solution:

a. The low-cost leader will be **firm 1**. This is because the cost of production of firm 1 is below that of firm 2 for any level of output. Equivalently, the marginal cost of production is 5 for firm 1 while it is 15 for firm 2.
b. The profit maximizing price and output for the leader are determined as follows:

$$\underset{X_1 | X_1 = X_2}{\text{maximize}} \pi_1 = R_1 - C_1.$$

$$\begin{aligned} \Pi_1 = R_1 - C_1 &= PX_1 - C_1 \\ &= (105 - 2.5X)X_1 - 5X_1 \\ &= [105 - 2.5(X_1 + X_2)]X_1 - 5X_1 \end{aligned}$$

$= [105 - 2.5(X_1 + X_1)]X_1 - 5X_1$ since $X_1 = X_2$ by assumption.
$= (105 - 5X_1)X_1 - 5X_1$
$= 105X_1 - 5X_1^2 - 5X_1$
$= 100X_1 - 5X_1^2$

Then, $\frac{\partial \Pi_1}{\partial X_1} = 100 - 10X_1 = 0. \Rightarrow 100 = 10X_1 \Rightarrow \mathbf{X_1^* = \underline{10}}$.

[Note that the second order condition is satisfied since $\frac{\partial^2 \pi_1}{\partial X_1^2} = -10 < 0.$]

The price it charges is found by substituting $X_1 = 10$ into the demand function:
$P = 105 - 2.5X = 105 - 2.5(X_1 + X_2) = 105 - 5X_1$
$\mathbf{P^*} = 105 - 5(10) = \underline{\mathbf{55}}$.

c. The follower will adopt the same price ($\mathbf{P^* = 55}$) and sell the same amount of output ($\mathbf{X_2^* = 10}$) as the leader**.**

d. If the follower could act to maximize its profit (in the same way as firm 1 did above), it would produce a different level of output and charge a different price. The profit-maximizing output and price of firm 2 is determined as follows:

$$\underset{X_2 | X_2 = X_1}{\text{maximize}} \pi_2 = R_2 - C_2$$

$\Pi_2 = R_2 - C_2 = PX_2 - C_2$
$= (105 - 2.5X)X_2 - 15X_2$
$= [105 - 2.5(X_1 + X_2)]X_2 - 15X_1$
$= [105 - 2.5(X_2 + X_2)]X_2 - 15X_2$ since $X_2 = X_1$ by assumption.
$= (105 - 5X_2)X_2 - 15X_2$
$= 105X_2 - 5X_2^2 - 15X_2$
$= 90X_1 - 5X_2^2$

Then, $\frac{\partial \Pi_2}{\partial X_2} = 90 - 10X_2 = 0. \Rightarrow 90 = 10X_2 \Rightarrow \mathbf{X_2^* = \underline{9}}$.

[Note that the second order condition is satisfied as $\frac{\partial^2 \Pi_2}{\partial X_2^2} = -10 < 0.$]

The price it charges is found by substituting $X_2 = 9$ into the demand function:

$P = 105 - 2.5X = 105 - 2.5(X_1 + X_2) = 105 - 5X_2$

$\mathbf{P^*} = 105 - 5(9) = \mathbf{60}$.

The follower (firm 2) produces 10 units and charges P = 55 while it could have done better by producing 9 units and charging P = 60, just to avoid price war with the leader. If the two firms enter into a price war, firm 1 has the potential to reduce price to a level where firm 2 cannot cover its long run cost.

So far we discussed the low-cost firm price leadership model under the assumption of equal market share. In practice, however, firms may agree to share the market in some constant proportions (different from 50-50). In such cases, we drop our earlier assumption of equal share of the market. Let's shed light on this point with the help of a numerical example.

Numerical example:

Suppose that two firms agreed that firm 1 would be the low-cost leader. If the firms also agreed that their market shares are k_1 and k_2 for the leader and the follower, respectively.

a. How would the quantity of output each firm produces and the market price be determined?

b. With the assumptions that the market shares are $k_1 = \frac{2}{3}$ and $k_2 = \frac{1}{3}$; and the demand and cost functions of firm 1 are represented by $P_1 = 100 - 2X_1 - X_2$, and $C_1 = 2.5X_1^2$, find the equilibrium of the two firms.

Solution:

a. In general, the leader maximizes its profit by substituting the reaction function of the follower into its profit function. Given the information that firm 1 is the low

cost leader, the equilibrium is found by substituting the reaction function of firm 2 into the profit function of firm 1 and then maximizing this profit function.

- ☞ The reaction function of firm 2 is determined as follows:
  - ☛ First, define the market shares: $k_1 = \frac{X_1}{X}$ and $k_2 = \frac{X_2}{X}$.
  - ☛ Divide the market share of the follower (k_1) by that of the leader (k_2) and solve for X_2:

$$\frac{k_2}{k_1} = \frac{X_2/X}{X_1/X} = \frac{X_2}{X}.\frac{X}{X_1}.$$

$$\Rightarrow \frac{k_2}{k_1} = \frac{X_2}{X_1}.$$

$$\Rightarrow X_2 = \frac{k_2}{k_1} X_1 \text{.....Reaction function of firm 2.}$$

- ☞ Now, substitute the reaction function of the follower $(X_2 = \frac{k_2}{k_1} X_1)$ into the profit of the leader and maximize.

$$\underset{X_1 | X_2 = \frac{k_2}{k_1} X_1}{\text{maximize}} \pi_1 = R_1 - C_1$$

The *F.O.C.* $\frac{\partial \Pi_1}{\partial X_1} = 0$ yields the optimal level of $X_1 = \boldsymbol{X_1^*}$. Subsequently, substituting $X_1 = \boldsymbol{X_1^*}$ into the reaction function of firm 2 $(X_2 = \frac{k_2}{k_1} X_1)$ gives us the level of output the follower produces: $\boldsymbol{X_2^*} = \frac{k_2}{k_1} \boldsymbol{X_1^*}$.

- ☞ Finally, using the values found so far, together with the market demand and the profit functions of the two firms, provides the market price and the profit level of each firm.

b. This part will clarify the discussion so far as well as the possibly unattractive calculation in question (a) above.

Step 1: find the reaction function of the follower (firm 2):

$$\frac{k_2}{k_1} = \frac{X_2}{X_1}$$

$$\Rightarrow X_2 = \frac{k_2}{k_1} X_1$$

$$\Rightarrow X_2 = \frac{1/3}{2/3} X_1 = \frac{1}{2} X_1$$

$$\Rightarrow X_2 = 0.5X_1$$.....Reaction function of firm 2.

Step 2: Substitute the reaction function of firm 2, $X_2 = 0.5X_1$, into the profit function of the leader ($\pi_1 = R_1 - C_1$).

$$\pi_1 = R_1 - C_1 = PX_1 - C_1 = (100 - 2X_1 - X_2)X_1 - 2.5X_1^2$$

$$\Rightarrow \pi_1 = (100 - 2X_1 - 0.5X_1)X_1 - 2.5X_1^2$$

$$\Rightarrow \pi_1 = (100 - 2.5X_1)X_1 - 2.5X_1^2$$

$$\Rightarrow \pi_1 = 100X_1 - 2.5X_1^2 - 2.5X_1^2$$

$$\Rightarrow \pi_1 = 100X_1 - 5X_1^2$$

Step 3: Maximize the leader's profit:

$$\frac{\partial \Pi_1}{\partial X_1} = 100 - 10X_1 = 0. \Rightarrow 100 = 10X_1 \Rightarrow \boldsymbol{X_1^* = \underline{10}}.$$

The leader will set the price: $P_1 = 100 - 2X_1 - X_2$

$$\Rightarrow P_1 = 100 - 2X_1 - 0.5X_1$$

$$\Rightarrow P_1 = 100 - 2.5X_1$$

$$\Rightarrow P_1 = 100 - 2.5(10)$$

$$\Rightarrow \boldsymbol{P_1^* = \underline{75}}.$$

Step 4: The quantity which the follower produces is: $\boldsymbol{X_2^*} = 0.5X_1 = \boldsymbol{\underline{5}}$.

The follower sells its output at the price set by the leader, $\boldsymbol{P_1^* = \underline{75}}$.

2.3.2.2 The Dominant-Firm Price Leadership

In this model, it is assumed that there is a large dominant firm that has a considerable share of the market, and some smaller firms, each of them having a small market share. Unlike the previous case where the leader and the follower were defined in terms of cost –comparison, here we define the leader and the follower in terms of their market shares. The model also assumes that the dominant leader knows:

- ☞ the marginal costs of the smaller firms, which it can add horizontally and find the total quantity supplied by the smaller firms at each price, and
- ☞ the market demand curve.

The supply by the smaller firms and the market demand curve for the product are shown as S_1 and D, respectively, in *panel (a)* of the figure below.

With this knowledge (of S_1 and D), the leader can obtain its own demand curve (D_L) by calculating the difference between the total demand (D) and the total supply by the smaller firms (S_1) at each price level.

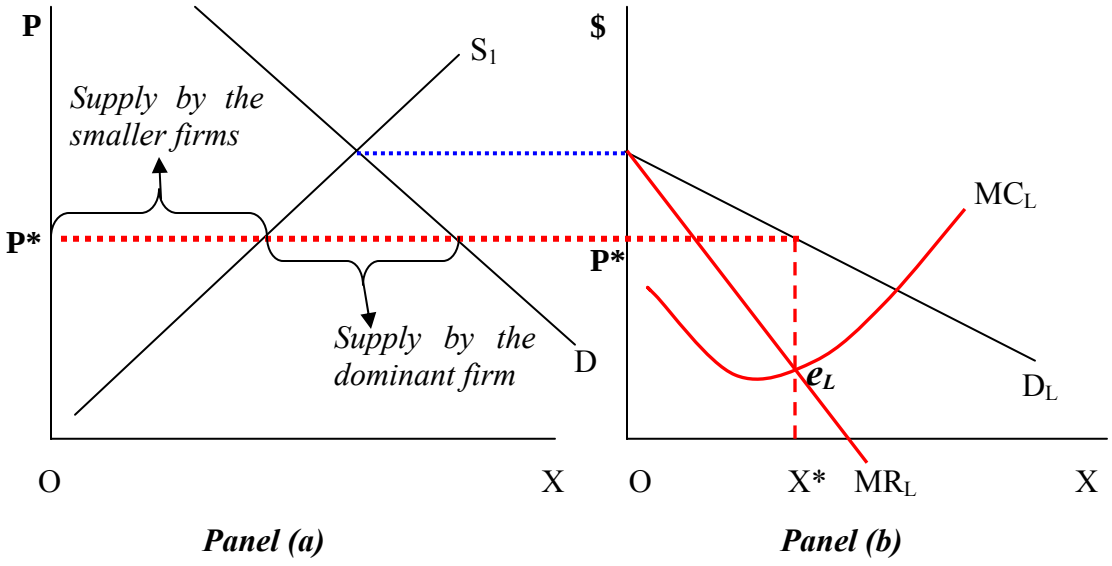


Figure 2.7: The Dominant-Firm Price Leadership

The dominant firm leader maximizes its profit by equating its marginal cost (MC_L) to its marginal (MR_L). Consequently, it sets price $\boldsymbol{P^*}$ and sells a quantity equal to $\boldsymbol{X^*}$ shown in

panel (b) of Figure 2.7 above. The smaller firms are price-takers, and may or may not maximize their profits, depending on their cost structure. If the smaller firms maximize their profits (or if a small firm maximizes its profit), that is by chance, and not by choice.

This model is also called '*the partial monopoly*' model since the large firm (the leader) acts as a monopolist while the small firms are price-takers and act like firms in pure competition.

Numerical example:

Suppose the market demand for a product is given by $Q_D = 50 - 0.3P$; the aggregate supply of the smaller firms by $Q_S = 0.2P$; and the total cost function of the leader firm by $C_L = 2Q_L$. Find:

a. the equilibrium output and price of the leader; and
b. the total output supplied by all the smaller firms at equilibrium.

Solution:

a. equilibrium of the larger (leader) firm:

Step 1: find the demand of the dominant firm. At each level of price (P), the quantity demanded of the dominant firm's product (Q_L) is given by the difference between the total quantity demanded (Q_D) and the quantity supplied by the smaller firms as a whole (Q_S).

That is, $Q_L = Q_D - Q_S$.

$\Rightarrow Q_L = 50 - 0.3P - 0.2P$

$\Rightarrow \boldsymbol{Q_L = 50 - 0.5P}$ or rearranging gives $\boldsymbol{P = 100 - 2Q_L}$.

Step 2: define the profit function of the leader:

$\Pi_L = R_L - C_L = (100 - 2Q_L)Q_L - 2Q_L = 100Q_L - 2Q_L^2 - 2Q_L$

$\Pi_L = 98Q_L - 2\,Q_L^2$.

Step 3: Profit maximization:

$$\frac{\partial \prod_L}{\partial Q_L} = 98 - 4Q_L = 0. \Rightarrow 98 = 4Q_L \Rightarrow \boldsymbol{Q_L^*} = \boldsymbol{24.5}.$$

$\Rightarrow$ The leader will set the price $P = 100 - 2Q_L$

$$\Rightarrow \boldsymbol{P^*} = 100 - 2(24.5) = \boldsymbol{51}.$$

b. The amount supplied by the smaller firms:

At the price set by the dominant firm ($\boldsymbol{P^* = 51}$), the total (market) quantity demanded is: $\boldsymbol{Q_D^*} = 50 - 0.3P = 50 - 0.3(51) = \boldsymbol{34.7}$.

Out of the market quantity demanded, $Q_D = 34.7$ units, the leader supplies $Q_L = 24.5$ units, so the smaller firms produce the remainder ($\boldsymbol{Q_S^*} = 34.7 - 24.5 = \boldsymbol{10.2}$). (Alternatively: $\boldsymbol{Q_S^*} = 0.2(P^*) = 0.2(51) = \boldsymbol{10.2}$).

Similar to the low-cost price leader, the dominant firm leader must make sure that the smaller firms will not only follow its price but also produce the right quantity – the quantity that will not push it to a non-profit-maximizing position. Thus, there should be tight (formal or informal) sharing-the-market agreement between (among) firms where price leadership is the form of collusion.

Although two models were developed one for a low-cost leader and another for a dominant firm leader, in practice, the power of the leader depends both on its costs and its size. If a firm has low costs but is very small (compared to some other firm), it may not find it possible to survive a price, or advertising or product-design war that the dominant firm may start. On the other hand, if the dominant firm loses its cost advantage, it also loses its power to impose an increase in price, since the smaller firms, having lower costs, will normally not follow it in price increases. Hence, in order to have the power to impose its price on others, the leader must be both a low-cost and a large firm.

The leader will lose its position of leadership if it loses its cost advantage or loses a considerable part of its share of the market for whatever reason. In fact, leadership

changes frequently in the real world among competing oligopolists as a consequence of innovation in their products or processes.

In many cases where there are several firms of similar costs and similar size, it may be difficult to agree who will be the leader. Under such circumstances, the leader may be decided by common consent, and may be neither a low-cost nor among the largest firms in the industry. We may then have a case of *barometric price leadership*, the leader being a firm whose price changes are followed by the others for convenience.

2.3.2.3 Barometric Price Leadership

In this model, it is formally or informally agreed that all firms will follow (exactly or approximately) the changes of the price of a firm which is considered to have *a good knowledge of the prevailing conditions in the market* and *can forecast the future developments in the market better than others*. In short, the firm chosen as the leader is considered as a *barometer*, reflecting the changes in economic environment. Usually, it is a firm which has established the reputation of a good forecaster of economic changes from past behavior. A firm belonging to another industry may also be chosen as the barometric leader.

Barometric price leadership may be established for various reasons:

1. Rivalry between several large firms in an industry may make it impossible to accept one among them as the leader. They might instead prefer to follow a firm which had a good forecasting ability.
2. Followers avoid the continuous recalculation of costs, as economic conditions change.
3. The barometric firm usually has proved itself as a 'reasonably' good forecaster of changes in cost and demand conditions in the particular industry (or the economy as a whole), and by following it, the other firms can be 'reasonably' sure that they choose the correct price policy.

According to empirical studies, barometric price leadership occurs frequently as a response to a period of violent price fluctuation and cutthroat (aggressive) competition in an industry, during which many firms suffer and greater stability is widely sought.

2.4 Contestable Markets

In the models discussed so far, we assumed the existence of a few firms (two in most cases). This could be the case if entry is assumed to be impossible or at least difficult. But in some cases the barriers to entry could be so weak that there exists a threat on the incumbent (currently operating) oligopolists. In the extreme case of no barrier to (or free) entry, the existing oligopolists may behave as if they were in perfect competition. Such a behavior is described by the model of contestable markets.

According to the theory of contestable markets developed during the 1980s, even if an industry has only a few sellers (or perhaps only one), it would still operate as if it were perfectly competitive if entry is "absolutely free" and if exit is "entirely costless". The condition of absolutely free entry refers to the situation where other firms can enter the industry and face exactly the same costs as the existing ones. An entirely costless exit refers to the condition in which there are no sunk costs so that the firm can exit the industry without facing any loss of capital.

Firms will sell their products at a price that only covers their average total costs (so that they earn zero economic profit) even if there are only a few firms in the market. In this view, competition within the market is less important than the potential competition for the market. In other words, very crucial to the realization of this zero long run profit is the existence of the potential competition (the threat from potential entrants).

If positive economic profit prevails, a new firm could enter the market, produce the same output (at the same cost) as the firm already in the market, undercut the existing firm's price slightly, and make a profit. Entrants will do this if economic profit is positive, and the price will be pushed down to the point where economic profit is zero. Put differently,

even transitory (temporary) profits will result in entry-exit or hit-and-run behavior on the part of potential entrants into the industry until all profit opportunities are entirely exhausted.

The firms in a contestable market, like those under perfect competition, will produce at minimum cost (at the lowest point on the LAC curve). If they produce at more than minimum cost, firms will enter the industry, produce at lower costs than the existing firms, undercut the existing firms' price, and make a profit. Thus, costs will be pushed down to the minimum level. Also, price cannot exceed marginal cost. If existing firms are charging a price in excess of marginal cost, it is profitable for an entrant to undercut the price of the existing firms. Thus, for an equilibrium to occur price cannot exceed marginal cost.

Is there a way for the existing firms to try to collude with the new entrants to keep price high or to push price up? As both the existing and the currently entering firms know that new firms would enter the market very quickly and undercut their price, they do not collude to push price up.

The extreme assumptions on which the theory of contestable markets is based have been sharply criticized as unrealistic. In reality, of course, entry is not free and exit is not costless. However, even models based on simplifying (and unrealistic) assumptions can be very useful and applicable. The theory is very important in cautioning us against uncritically accepting the view that a market with a few firms (oligopoly) must necessarily be noncompetitive.

2.5 Long Run Efficiency Implications and Welfare Effects of Oligopoly

An oligopolist would leave the industry in the long run if it would incur a loss. On the other hand, if existing firms earn profits, more firms will seek to enter the industry in the long run, and unless entry is blocked or somehow restricted, industry output will expand

until industry profits fall to zero. In other words, an oligopolistic industry may not be oligopolistic for long if entry is easy.

Whether or not the industry remains oligopolistic in the face of relatively easy entry depends on the size of the market for the product relative to the optimum size of firm. Above-average profits will attract new firms. If the market is small relative to the optimum size of firm in this industry, the number of firms will remain sufficiently small so that the industry will still be oligopoly. If the market is large relative to the optimum size of the firm, the number of firms will grow sufficiently large so that the industry will no longer be an oligopoly.[6]

Ease of entry also tends to erode collusive agreements. As long as profits exist in the industry, firms will be tempted to enter and take business away from the collusive group by lowering the price a bit. Once entry of this sort occurs, it becomes more and more difficult to keep a cartel together.

For an industry to remain oligopolistic in the long run, entry must be somewhat restricted. This may result from many reasons, generally the same reasons that led to the existence of oligopoly in the first place (discussed under Section 2.1).

Most oligopoly models presented in this chapter predict infrequent price changes in oligopolistic markets. In many oligopolistic industries, firms tend to use advertising and product variation, more than price, as competitive weapons. They seem to view price-cutting as a dangerous tactic, since it can start a price war that may have grave (serious) consequences. On the other hand, advertising and product variation are viewed as less risky ways of wooing customers away from competitors, though they often result in retaliatory moves by competitors and very great costs.

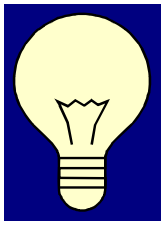
[6] The optimum size of the firm refers to the plant size that enables a firm to produce/operate at the minimum point on the long run average cost curve (the point where the economies of scale is fully used).

Since there is no agreement that any of the models in this chapter is an adequate general representation of oligopolistic behavior, it is difficult to estimate the effects of an oligopolistic market structure on price, output, and profits and hence on social welfare. However, a few words can be said:

1. As in the case of monopoly and monopolistic competition, oligopolists usually do not produce at the lowest point on their LAC curve. Oligopoly, however, often results because of the smallness of the market in relation to the optimum size of the firm, and so it does not make much sense to compare oligopoly to perfect competition. Some products could only be produced at prohibitive (unaffordable) costs under perfectly competitive conditions.
2. As in the case of monopoly, oligopolists can earn long run profits, and so price can exceed LAC. However, some economists believe that oligopolists utilize a great deal of their profits for research and development (R&D) to produce new and better products and to find cheaper production methods. These are the primary sources of growth in modern economies. These same economists point out that, monopolists do not have as much incentive to engage in R&D, and perfect competitors and monopolistic competitors are too small and do not have the resources to do so on a large scale.
3. As in imperfect competition in general, P > LMC at the best level of output. This means that society values an additional unit of the commodity more than the marginal cost of producing it. But again, P = LMC can only be attained under perfect competition, and economies of scale may make perfect competition infeasible/impracticable.
4. Oligopolistic industries tend to spend large amounts on advertising and product variation. The use of some resources for these purposes is certainly worthwhile, since advertising provides buyers with information, and since product variation allows greater freedom of choice. It is difficult, however, to determine exactly how much advertising and product differentiation is socially desirable in the real world.

Thus, the welfare aspects of oligopoly, like that of monopolistic competition, are inconclusive (or indeterminate).

Turning to oligopoly theory itself, all the oligopoly models we have examined are somewhat incomplete and unsatisfactory. This is unfortunate because oligopoly is the most prevalent form of market organization in production in all modern economies. Some progress in oligopoly theory is provided by game theory (to be examined in the next chapter).



2.6 Summary

1. *Oligopoly is the form of market structure in which there are a few sellers of a homogeneous or a differentiated product. Because of the possibility of homogeneous or heterogeneous products, the various ways of reaction among firms, and the uncertainty characterizing the market structure, there is no single widely accepted theory on oligopoly. Different natural and artificial factors are there to act as sources of oligopoly (as well as barriers to entry).*
2. *Cournot's duopoly model is the earliest model of non-collusive oligopoly with the behavioral assumption of each firm deciding on its output level to maximize its profit on the supposition that the competitor will not change its output level. Bertrand's duopoly model modifies the Cournot's duopoly model by changing the variable of competition to price. Stackelberg's duopoly model is an extension of the Cournot's duopoly model where one of the firms is sophisticated enough to recognize the rivalry and to know the reaction function of the naïve follower. Sweezy's model is the final model of non-collusive oligopoly, which explains the price rigidity observed in some oligopoly markets.*
3. *Conditions in oligopolistic industries usually tend to promote collusion, which may take a form of cartel or price leadership. Cartel is an organization of suppliers*

of a commodity aimed at restricting competitions and/or increasing profits. Cartels could be centralized or market sharing. Price leadership is a coordinated behavior of oligopolists where one firm sets the price (or price changes) and others follow. The leader could be a low-cost or a dominant (large) firm or a firm recognized as the best predictor of changes in market conditions.

4. *If entry is easy and exit is less costly, oligopoly markets could be characterized by competition (from potential entrants) pushing costs of production and price down to levels observed under perfect competition. In the extreme case of free entry and costless exit, firms minimize costs of production and charge P = MC (as the theory of contestable markets predicts).*
5. *The social effects of oligopoly are generally difficult to predict. Some positive and some negative impacts are associated with the market structure.*
6. *Each of the oligopoly models examined in this chapter is somewhat incomplete and unsatisfactory. Some progress in oligopoly theory is provided by game theory (a topic to be examined in Chapter Three).*



2.7 Review Questions

I. Multiple-Choice Questions

1. Which of the following concepts apply to oligopoly more than to any other market structure?
   A. Easy entry and more than one firm in the market
   B. Homogeneous product and perfect information
   C. Economies of scale and significant barriers to entry
   D. Concentration and interdependence
   E. Advertising and product differentiation

2. Which of the following statements describes the Cournot's model of oligopoly behavior? The model consists of firms which:
   A. get together to make joint price and output decisions to maximize profit.
   B. make explicit agreements to fix price and quantity.
   C. make implicit agreements to fix price and quantity.
   D. follow the price cuts of a single firm but not its price hikes.
   E. take the output of the other as given, and adjust output until the market finally produces an output level somewhere between competition and monopoly.
3. Which of the following assertions is more closely associated with the kinked demand curve model of oligopoly behavior?
   A. Firms can enter a market in search of profits, but lose nothing if they fail.
   B. Oligopoly prices are more stable than costs.
   C. A large, powerful firm can drive smaller firms out of the market by temporarily selling at an artificially low price.
   D. The quantity of output produced by an oligopolist is somewhere between the output of a competitive market and the output set by a monopoly.
   E. Oligopoly behavior consists of a complex series of moves and anticipated countermoves among rival firms.
4. Which of the following statements describes a movement along the elastic portion of the kinked demand curve?
   A. When a firm increases its price, other firms will not change theirs, so quantity will decrease by a small amount.
   B. When a firm decreases its price, other firms will decrease their prices, so quantity will increase only by a small amount.
   C. When a firm increases its price, other firms will not change theirs, so quantity will decrease by a large amount.
   D. When a firm increases its price, other firms will also increase theirs, so quantity will not decrease very much.
   E. When a firm increases its price, other firms will decrease theirs, so quantity will decrease by a large amount.

5. In contestable markets, large oligopolistic firms tend to behave like:
   A. cartels.
   B. duopolies.
   C. monopolies.
   D. perfectly competitive firms.
   E. monopolistically competitive firms
6. Which of the following statements is false?
   A. The kink in the kinked demand curve follows from the assumption that firms follow price cuts but do not follow price increases.
   B. In contestable markets, entry is free, and exit is costless.
   C. A cartel is generally a form of tacit collusion.
   D. All of the above.
   E. None of the above.
7. There is no widely accepted theory on oligopoly for all of the following but one.
   A. Uncertainty in the market structure
   B. Various ways of reactions among firms
   C. The possibility to have homogeneous or heterogeneous products
   D. All of the above.
   E. None of the above.
8. In which of the following oligopoly models is rivalry not recognized at all?
   A. Stackelberg's model
   B. Price leadership by the dominant firm
   C. The kinked demand model
   D. Barometric price leadership
   E. None of the above.
9. The reaction function of a firm in Cournot's duopoly model:
   A. Shows how much output the firm must produce in order to maximize its own profit for every specific level of output by its rival.
   B. Shows the combinations of prices charged by two firms to maximize their profits.
   C. Is represented by upward sloping curves in quantity axes.
   D. A and C

10. Identify the incorrect one.
    A. Both Bertrand's and Stackelberg's duopoly models are somewhat extensions or modifications of the Cournot's duopoly model.
    B. Sweezy's model explains why firms in oligopoly face a demand curve which is highly elastic for price rises and less elastic for price falls.
    C. Firms producing homogeneous products and competing over prices will not have the incentive to cut prices once their prices are equal.
    D. Firms competing over prices may charge different prices.
    E. None of the above.
11. Which of the following is true?
    A. Stackelberg's disequilibrium may lead to collusion.
    B. Firms in oligopoly enter into collusive agreements for the sole purpose of increasing profits.
    C. Firms in a centralized cartel share the joint profit according to their costs.
    D. A central agency appointed by the firms in a centralized cartel decides on the price to be charged and firms determine their output levels.
    E. None of the above.
12. One of the following is not a form of cartel.
    A. Agreements of firms to charge the same price but to sell any quantity they can.
    B. Agreements on both the price to be charged and the quantity to be sold by each firm.
    C. Sharing the market by defining the region in which each firm sells its product possibly at a different price and any amount.
    D. Agreements to change prices in directions of changes by a firm with good forecasting ability.
    E. None of the above.

Instruction: Answer Questions 13 to 17 Based on the Following Information:

In a duopoly market, the demand for product X is given by: $\mathbf{P = 120 - 3X}$; where $\mathbf{X = X_1 + X_2}$, and the costs of the firms are: $\mathbf{C_1 = 6X_1 + 12}$ and $\mathbf{C_2 = 1.5X_2^2 + 18}$.

13. What will be the Stackelberg's equilibrium quantities if firm 2 is the leader?
    A. $X_1 = 18.5$ and $X_2 \approx 7.2$
    B. $X_1 \approx 7.2$ and $X_2 = 18.5$
    C. $X_1 = 13.75$ and $X_2 = 10.5$
    D. $X_1 = 10.5$ and $X_2 = 13.75$
    E. None
14. What will be the quantity produced by each firm at Cournot's equilibrium?
    A. $X_1 = 27.75$ and $X_2 = 4.1$
    B. $X_1 = 13.75$ and $X_2 = 8.75$
    C. $X_1 = 14.8$ and $X_2 = 8.4$
    D. $X_1 = 18.8$ and $X_2 = 7.16$
    E. None
15. The price charged per unit of output at Cournot's equilibrium is:
    A. 47.25
    B. 50.4
    C. 63
    D. 72
    E. None
16. Assuming that the firms are to form a centralized cartel, the price charged and the joint profit will be (respectively):
    A. 50.4 and 1059
    B. 63 and 645.12
    C. 63 and 1059
    D. 50.4 and 645.12
    E. None
17. If firm 1 is to be the low-cost price leader and if the market shares of the firms are $k_1 = \frac{3}{4}$ and $k_2 = \frac{1}{4}$, then the quantity produced and the price charged by the follower firm will be:
    A. 8 and 72
    B. 9.5 and 63

C. 13.75 and 47.25

D. 4.75 and 63

E. None

II. True or False Questions

18. In the price leadership models, the follower firm never maximizes its profit by choice.
19. There should be a clear share of the market agreement (formally or informally) between firms in both the low-cost and dominant-firm forms of price leadership.
20. Price leadership doesn't require the surrendering of all freedom of action by firms and thus is more acceptable to firms than a centralized cartel.
21. The low-cost price leadership model assumes two firms producing a standardized product at different costs.

III. Discussion Questions and Workout Problems

22. Suppose that two countries are providing port service to Ethiopia: Djibouti (country 1) and Somalia (country 2). The two countries face the following demand curve (ceteris paribus): $P = a + bX$, where P is price, $X = X_1 + X_2$, X_1 is the quantity of services provided by country 1, and X_2 is the quantity of services provided by country 2. These two countries face the same cost structure given by: $C_1(X_1) = F + cX_1$, $C_2(X_2) = F + cX_2$, where F is fixed cost, and c is a constant. Derive the equilibrium price (P) and quantity of services provided by each country (X_1 and X_2) and the aggregate quantity of the service (X) under the Cournot's model of behavioral assumption.
23. Compare the results of question number 22 with the results of monopoly and perfect competition.
24. For question 22, what would be the Stackelberg's solution if Djibouti is the sophisticated leader and Somali is the follower?
25. Redo questions 22, 23 and 24 if the demand curve is given by $P = 100 – 5X$, where $X = X_1 + X_2$, and if the two countries have identical costs given by $C_i = 10X_i$, i = 1, 2.

26. Assume that two firms are providing a differentiated product and are competing over price. Their demand functions are: $Q_1 = 20 – P_1 + P_2$; $Q_2 = 20 + 1.5P_1 – P_2$ where P_1 and P_2 are prices charged by the two firms, respectively, and Q_1 and Q_2 are quantities of output of each firm, and MC = 0. Find the (Nash) equilibrium of the game.
27. Assume that in a purely oligopolistic industry, there is one dominant firm and ten identical smaller firms; the market demand for the commodity is $Q = 44 – 2P$, where P is given in Birr; SMC_d (short run MC for the large firm) = $1.5 + Q/2$, while SMC_s (short run MC for one of the smaller firms) = $1 + Q/4$, input prices remain constant.
    A. What price will the dominant-firm set?
    B. How much will the smaller firms supply together?
    C. How much will the dominant-firm supply?
28. Suppose that there are two firms in an industry. The inverse demand curve for their product is given by $P = 100 – 5X$, where P is price and $X (= X_1 + X_2)$ is sum total of the quantities of output produced by firm 1 (X_1) and firm 2 (X_2). The two firms have costs given by: $C_1 = 10X_1$ and $C_2 = 8X_2$. Assuming that there is a low-cost price leadership type of collusive oligopoly, find the equilibrium price charged and the level of output produced by each firm.
29. What is (are) the difference(s) between:
    A. Cournot's and Bertrand's duopoly models?
    B. Cournot's and Stackelberg's duopoly models?
    C. Cartel and price leadership?
    D. Low-cost and dominant-firm price leadership models?
    E. Dominant-firm and barometric price leadership models?
    F. Centralized and market-sharing cartels?

CHAPTER THREE
INTRODUCTION TO GAME THEORY

Introduction

As we have seen in the previous chapter, a basic feature of oligopoly is that each firm must take account of the possible reactions of its rivals to every one of its actions. For this reason, an oligopolistic firm cannot tell what effect a change in the quantity (price) of its product will have on the price (or quantity) of its product and its profits. The best thing an oligopolist can do is to guess how its rivals will respond to changes in its level of production (or price), and state the possible feedback (repercussion) effects on its price (or output) and profit levels with *probabilities*. The interdependence of firms in oligopolistic markets and the inherent uncertainty about competitors' reactions to any course of action adopted by a firm cannot be analyzed effectively by the traditional tools of economic theory.

One of the most interesting developments in the theory of oligopoly in recent decades has been the appearance and elaboration of the theory of games, which has enriched oligopoly theory considerably. Though this development is still in its infancy, the importance of game theory as a tool of analyzing oligopoly behavior substantiates it as an area to be studied.

Chapter Objectives

This chapter introduces some concepts (to the students) and familiarizes students with how the theory of games is used to analyze the behaviour of oligopolistic firms. Specifically, it would enable students to:

- Point out whether a firm in oligopolistic market structure has a dominant strategy or not;

- Find the dominant strategy equilibrium or the Nash equilibrium (under pure and mixed strategies) of different games;
- Understand and apply the concept of prisoners' dilemma to firms in oligopoly in general and cartel members in particular;
- Describe how repeated games with tit-for-tat strategies help firms overcome the problem of prisoners' dilemma;
- Extend the definition and application of Nash equilibrium to sequential games; and,
- Understand and appreciate the strategic moves firms undertake to deter entry of potential firms (and defend their market shares).

3.1 Some Basic Concepts

This section of the chapter explains some basic concepts that will be used throughout the subsequent sections. Some other concepts will be defined and explained as need arises.

Game: a game is a competitive situation where two or more agents (persons, firms …) pursue their own interests and no agent can dictate the outcome. A game is described in terms of the players, the rules of the game, the payoffs of the game, and the information conditions that exist during the game. These elements, common to all conflict situations, are the fundamental characteristics of a game.

Cooperative versus Non-cooperative Games: the economic games that firms play can be either cooperative or non-cooperative. In **a cooperative game**, players can negotiate binding contracts that allow them to plan joint strategies. In **a non-cooperative game**, on the other hand, negotiation and enforcement of binding contracts are not possible. A centralized cartel is an example of a cooperative game while the Cournot's and Bertrand's models represent examples of non-cooperative games.

Zero-Sum versus Nonzero-Sum Games: we also distinguish between zero-sum and nonzero-sum games. **A zero-sum game** is one in which the gain of one player comes at

the expense of, and is exactly equal to, the loss of the other player. An example of this occurs if firm A increases its market share at the expense of firm B (say, by increasing its advertising expenditure in the face of unchanged state of advertising by firm **B**), ceteris paribus. In such a zero-sum game, the gains of one player equal the losses of the other in magnitude so that total gains plus total losses sum to zero. **A nonzero-sum game** is, on the other hand, one in which the gains of one firm (player) do not come at the expense of an equal benefit of the other. In a nonzero-sum game, even though the gains of one firm (player) may come at the expense of the opponent's, the gain of one and the loss to the other are not equal in magnitude. An example of this might arise:

a. if increased advertising leads to higher profits for both firms – a positive-sum game;
b. if increased advertising raises costs more than revenues and the profits of both firms decline – a negative-sum game; or
c. if one firm gains and the other losses, but the gain is either less than (negative-sum-game) or greater than (positive-sum-game) the loss in magnitude.

Strategy: a strategy is a specific course of action/s with clearly defined values for the policy (or choice) variables or instruments of a player. In other words, it is a complete specification of what a player will do under each contingency in the playing of the game. Simply, a strategy is a rule or plan of actions for playing a game.

Payoff: a payoff is the outcome of each combination of strategies by the rival firms. Alternatively, it is the net gain a strategy will bring to a player (or a firm) for any given counterstrategy of the competitor(s). A payoff is what a game generates as rewards or benefits for the player and it is measured in terms of the goal(s) of the firm.

Payoff matrix: the payoff matrix of a game is a table that gives or summarizes the payoffs from all the strategies open to a firm and the responses of the rival(s).

Game theory: game theory is the theory that examines the choice of optimal strategies in conflict situations. It attempts to study decision-making in situations where there is a

mixture of conflict and cooperation, as in oligopoly. In short, game theory shows how an oligopolistic firm can make strategic decisions to gain a competitive advantage over its rivals or how it can minimize the potential harm from strategic moves by rivals. While game theory is generally concerned with the analysis of strategic interactions, here, we apply it to the study of the economic behavior of firms in oligopolistic markets.

In principle, strategic interactions can involve many players, many strategies, and many rounds. But, in our forthcoming discussions, we will concentrate much on the case of two players (firms or persons), each with two strategies, and mostly on games played for a finite number of times.

3.2 Dominant Strategy and Nash Equilibrium

Let us introduce the concepts of dominant strategy and Nash equilibrium with the help of a hypothetical example. Suppose we have two firms, A and B, and a choice of two strategies for each: *to advertise* or *not to advertise*. Firm A, naturally, expects to earn higher profits if it advertises than if it does not, but the actual level of profit of firm A depends also on whether firm B advertises or not. Thus, each strategy by firm A can be associated with each of firm B's strategies. The payoff matrix below illustrates the four possible outcomes from this simple game.

		Firm B	
		Advertising	Not Advertising
Firm A	Advertising	8, 6	10, 2
	Not Advertising	4, 9	7, 5

Table 3.1: Dominant Strategy and Dominant Strategy Equilibrium

In the payoff matrix above, the first number in each of the four cells (8, 10, 4, or 7) refers to the payoff (in this case, profit) of firm A. Similarly, each of the second entries (6, 2, 9, or 5) is the payoff (profit) of firm B. What strategy should each firm choose?

- Let us first consider the choices of firm A, under various assumptions about B.
  - If we assume that firm B advertises, firm A would earn a profit of 8 units by advertising, and a profit of 4 units by choosing not to advertise. Thus, it follows that: *since a profit of 8 is superior to (better than) a profit of 4, firm A should advertise if B advertises*.
  - If we assume that firm B does not advertise, firm A would earn a profit of 10 units by advertising, and a profit of 7 by choosing not to advertise. In this case as well, *firm A should advertise as 10 > 7*.

  Thus, firm A's profits will always be greater if it advertises than if it does not, regardless of what firm B does.

- Now consider firm B:
  - If firm A advertises, firm B's profits would be 6 units if it advertises, and 2 units if it does not. *So, if firm A advertises, firm B would do better by advertising since 6 > 2*.
  - If firm A does not advertise, firm B's profits would be 9 if it advertises and 5 if it doesn't. *So, in this case as well, firm B should advertise its product since 9 > 5.*

  Firm B's profits will also be greater if it advertises than if it does not, regardless of what firm A does.

Now we are at a point to define one of our concepts – dominant strategy. With reference to the example above, each firm (A or B) will choose to advertise its product regardless of what the other firm does. Consequently, the outcome is that the firms (A and B) will earn profits of 8 and 6 units, respectively. This is given by the top left cell in the payoff matrix above. In this situation, we say that advertising is the dominant strategy for both firms. **Dominant strategy** is, thus, the optimal choice for a player regardless of (or no matter) what the opponent does. The advertising solution (the final equilibrium) for the two firms above holds whether A or B chooses its strategy first; or if both firms decide on

their best strategy simultaneously. This solution or equilibrium where each firm has a dominant strategy (an optimal choice which does not depend on the choice of the opponent) is called a **dominant strategy equilibrium.**

However, not all games have a dominant strategy for each player. An example of this is shown in the following payoff matrix.

		Firm B	
		Advertising	Not Advertising
Firm A	Advertising	8, 6	10, 2
	Not Advertising	4, 9	11, 5

Table 3.2: Nash Equilibrium

Based on the information from Table 3.2,

- ☞ If firm A advertises, B would choose to advertise as 6 > 2; and, if A does not advertise, B would choose to advertise as well since 9 > 5. Thus, B's optimal choice is to advertise whether A advertises or not. That is, *B has a dominant strategy* (which is to advertise).
- ☞ If B advertises, A would choose to advertise because 8 > 4; and if B does not advertise, A would choose not to advertise because 11 > 10. The interpretation is that, A's choice depends on whether firm B advertises or not. Thus, *firm A does not have a dominant strategy*. In order for firm A to determine whether to advertise or not, it must first try to determine what firm B will do. If firm A knows the above payoff matrix, it can figure out that firm B has the dominant strategy of advertising. Consequently, when firm B's choice is revealed, firm A will determine its optimal strategy and this will result in a Nash Equilibrium.

Nash equilibrium is a situation in which each player chooses an optimal strategy, given the strategy chosen by the other player. In relation to a payoff matrix like the ones we saw above, the Nash equilibrium is a cell whose both entries are chosen. The cell (8, 6) –

(advertise, advertise) is the Nash equilibrium of the above game. A pair of strategies is a Nash equilibrium if A's choice is optimal given B's choice, and vice versa. The Cournot's equilibrium examined in Chapter Two is an example of Nash equilibrium.

As a shortcut to find the Nash equilibrium of a game similar to the one we have just seen,[7]

- Fix a row, compare the second entries of the cells across that row, and underline the preferred value (the bigger one if profit, and the smaller one if loss or something undesirable). For the previous example, we compare 6 and 2 in the first row, and underline 6.
- Take the other (second) row, compare the second entries in that row, and underline the preferred one. Since 9 > 5 in our example, 9 is underlined.
- Fix a column, compare the first entries down that column, and underline the preferred value. Since 8 > 4 in our example, 8 is underlined.
- Repeat the third step for the second column. Since 11 > 10 in our example, 11 is underlined.
- The cell whose both entries are underlined gives us the Nash equilibrium.

Note that all dominant strategy equilibria are Nash equilibria; but not all Nash equilibria are dominant strategy equilibria.

There are two major problems with Nash equilibrium of the kind we saw above. These are:

- a game may have more than one Nash equilibrium, or
- a game may not have a Nash equilibrium.

The following two payoff matrices show two such examples: the first one represents a game with no Nash equilibrium, and the second one a game with two Nash equilibria.

[7] Although this might seem a lengthy procedure at your first glance, you will find it very short as you practice.

		Player B	
		Left	Right
Player A	Top	0, <u>0</u>	<u>0</u>, -1
	Bottom	<u>1</u>, 0	-1, <u>3</u>

(a) A Game with No Nash Equilibrium

		Player B	
		Fight	Don't fight
Player A	Fight	<u>350</u>, <u>350</u>	350,300
	Don't fight	300,300	<u>400</u>, <u>400</u>

(b) A Game with Two Nash equilibria:

[The Nash equilibria are *(fight, fight)* and *(don't fight, don't fight)*]

Table 3.3: Non-Existence and Non-Uniqueness of Dominant Strategy Nash Equilibrium

3.3 Mixed Strategies

In Section 3.2 above, we mentioned two problems related to Nash equilibrium: that some games do not have a Nash equilibrium and that some games may have multiple Nash equilibria. The existence of these problems has something to do with the way we define strategies: the adoption of pure strategies. **Pure strategy** is a strategy in which each agent makes a specific choice and sticks to it. Hence, the problems are experienced if each player (firm) chooses one and only one of the alternative strategies available to it with certainty. If we enlarge our notion of strategies, we can find a new sort of Nash equilibrium for games that do not have Nash equilibrium under the use of pure strategy. If the agents are allowed to randomize their strategies – to assign a probability to each choice and to play their choices according to those probabilities – we call this strategy a **mixed strategy**. Under mixed strategies, unlike under pure strategy, a firm (a player) plays all the strategies – each with some probability – rather than sticking to a single strategy.

Nash equilibrium in mixed strategies refers to an equilibrium in which each agent chooses the optimal (relative) frequency with which to play his/her strategies given the frequency choices of the other agent. Every game with a finite number of players and a finite number of strategies/actions has at least one Nash equilibrium.

Consider the following payoff matrix for which no Nash equilibrium exists under pure strategy.

		Player B	
		Left	Right
Player A	Top	2, 3	4, 5
	Bottom	1, 6	7, 0

Table 3.4: A Game without Dominant Strategy Nash Equilibrium but with Mixed Strategy Nash Equilibrium

Suppose player A plays *top* with probability ***a*** and *bottom* with probability ***1 – a***; and player B plays *left* with probability ***b*** and *right* with probability ***1 – b***.

Each player calculates his/her own ***expected payoff*** given the probability with which the opponent plays the strategies. Let's begin with the decision-making process of player A.

From playing top, player A could get a payoff equal to 2 or 4 depending on the actual choice of player B. But, the actual choice of B is not known. What is known about player B is just the probability with which he/she plays the two strategies: *b* and *1 – b*. Hence, player A could get a payoff of 2 with probability *b* and a payoff of 4 with probability *1 – b*. Accordingly,

- ☞ the expected payoff of player A if he/she plays top is given by:

  (A's payoff from playing top when B plays left)(B's probability of playing left) **+** (A's payoff from playing top when B plays right)(B's probability of playing right)

  = 2 (the probability that B plays left) + 4 (the probability that B plays right)

= 2 (b) + 4 (1 – b) = ***4 – 2b***.

- similarly, the expected payoff of player A if he/she plays bottom is given by:
  (A's payoff from playing bottom when B plays left)(B's probability of playing left) **+** (A's payoff from playing bottom when B plays right)(B's probability of playing right)
  = 1 (the probability that B plays left) + 7 (the probability that B plays right)
  = 1 (b) + 7 (1 – b) = ***7 – 6b***.

The decision of player A to switch between the two strategies depends on whether the expected payoff from playing top is less than, equal to, or greater than the expected payoff from playing bottom. That is,

- Player A will be indifferent between playing top and playing bottom (with a different frequency than what he/she is doing now) if the expected payoffs from the two strategies are equal, i.e., if 4 – 2b = 7 – 6b. ⇒ – 2b + 6b = 7 – 4. ⇒ 4b = 3. ⇒ **b = 3/4**; **1 – b = 1/4**. This is interpreted as: if player B plays left with probability b = 3/4 and right with probability 1 – b = 1/4, then the probabilities with which A plays top and bottom will be stable.
- Player A will increase the relative frequency (probability) with which he/she plays a given strategy if the expected payoff from that strategy is greater than the expected payoff from the other one. Specifically:
  - A will increase the relative frequency (probability) with which he/she plays top if 4 – 2b > 7 – 6b. ⇒ – 2b + 6b > 7 – 4. ⇒ 4b > 3. ⇒ **b > 3/4**; and
  - A will increase the relative frequency (probability) with which he/she plays bottom if 7 – 6b > 4 – 2b. ⇒ – 6b + 2b > 4 – 7. ⇒ – 4b > – 3. ⇒ – b > – 3/4. ⇒ **b < 3/4**.

In a similar fashion,

- The expected payoff of player B if he/she plays left is:
  = 3 (a) + 6 (1 – a) = ***6 – 3a***, and
- The expected payoff of player B if he/she plays right is:
  = 5 (a) + 0 (1 – a) = ***5a***.

The decision of player B depends on whether the expected payoff from playing left is less than, equal to, or greater than that from playing right. That is,

- Player B will be indifferent between playing left and playing right (with a different frequency than what he/she is doing now) if the expected payoffs from the two strategies are equal, i.e., if $6 - 3a = 5a. \Rightarrow 6 = 5a + 3a. \Rightarrow 8a = 6. \Rightarrow$ **$a = 6/8 = 3/4$**; **$1 - a = 1/4$**. This is interpreted as: if player A plays top with probability $a = 3/4$ and bottom with probability $1 - a = 1/4$, then the probabilities with which B plays left and right will be stable.
- Player B will increase the relative frequency (probability) with which he/she plays left if the expected payoff from playing left is greater than that from playing right, i.e., if $6 - 3a > 5a. \Rightarrow 6 > 5a + 3a. \Rightarrow 6 > 8a. \Rightarrow 8a < 6. \Rightarrow$ **$a < 3/4$**; and
- Player B will increase the relative frequency (probability) with which he/she plays right if the expected payoff from playing right is greater than that from playing left, i.e., if $5a > 6 - 3a. \Rightarrow 5a + 3a > 6. \Rightarrow 8a > 6. \Rightarrow$ **$a > 3/4$**.

Thus, the (mixed strategy) Nash equilibrium of the game is realized when A plays top and bottom with probabilities 3/4 and 1/4 respectively, and B plays left and right with probabilities 3/4 and 1/4 respectively. Note that we used the probabilities of player B to calculate the expected payoff of player A and vice versa.

Before proceeding to the next section, make sure that you can solve for the mixed strategy Nash equilibrium of the above game by yourself!

3.4 The Prisoners' Dilemma and Cartel Cheating

Recall the two problems with Nash equilibrium we mentioned earlier – the problem of existence (the possibility of a game with no Nash equilibrium) and the problem of uniqueness (the possibility of a game with multiple Nash equilibria). Another problem with Nash equilibrium is that it doesn't necessarily lead to Pareto efficient outcomes. An outcome (or a situation) is said to be Pareto efficient (or Pareto optimal) if it is impossible to make one better off without, at the same time, making another worse off.

To help you understand what Pareto efficiency is, suppose that 100 Birr is available to be shared between you and your friend. Suppose also that, currently, 60 Birr is given to your friend and 40 Birr to you. Is there any way (or potential) to make you better off without reducing the possession of your friend (or the other way round)? Regardless of whether the 60:40 allocation is equitable (or morally fair) or not, no one of you two can be made better off without making the other of you worse off. Thus, the 60:40 allocation represents a Pareto optimal allocation of the available resource. On the other hand, if the share was in such a way that your friend gets 60 Birr and you get 30 Birr, it is possible to make you better off (increase your possession to 40 Birr) without reducing that of your friend. Hence, the 60:30 allocation (out of a total of 100 Birr available for that purpose) is an inefficient (or sub-optimal) allocation. [More on Pareto optimality will be discussed in Chapter Five; the present illustration is sufficient for our purpose in game theory.]

A sub-optimal Nash equilibrium is associated with the concept known as the Prisoners' dilemma. Oligopolistic firms often face such a problem (of inefficient outcome) – the prisoners' dilemma. In relation to oligopolists, prisoners' dilemma refers to a situation in which each firm adopts its dominant strategy, but each could do better by cooperating.

Consider the following situation, the situation from which the name prisoners' dilemma was originally derived. Two suspects are arrested for armed robbery. Each suspect is interrogated separately and no communication is allowed between the two suspects. If convicted, each could receive a maximum sentence of ten years imprisonment. Unless one or both suspects confess (i.e., if both deny), each could receive a maximum sentence of one year in prison. Each suspect is promised that he/she will go free by confessing while the other (who doesn't confess) will receive the full ten-year sentence. If both suspects confess, each gets a reduced sentence of five years imprisonment. This set of information, available to both suspects, is summarized in the payoff matrix given below.

		Suspect B	
		Confess	Deny
Suspect A	Confess	-5, -5	0, -10
	Deny	-10, 0	-1, -1

Table 3.5: The Prisoners' Dilemma

What is the (pure strategy) Nash equilibrium of this game? Let us see.

- If B confesses, A would choose to confess since -5 > -10.
- If B denies (doesn't confess), A would choose to confess since 0 > -1.
  Thus, to confess is the dominant strategy for A.
- If A confesses, B would choose to confess since -5 > -10.
- If A denies (doesn't confess), B would choose to confess as 0 > -1.
  Thus, to confess is the dominant strategy for B.

Hence, the Nash equilibrium of this game is (confess, confess). With each suspect adopting his/her dominant strategy of confessing, each ends up receiving a five-year sentence. But, it would be preferable if each suspect denies and each were imprisoned for a year. Then, why do they confess? Put yourself in their position. If you and your friend were the two suspects, would you deny? No, you won't do that. Not knowing your friend's choice, you fear that you will be in prison for ten years in case your friend confesses but you don't. Your friend faces the same situation. Only if the suspects (you and your friend) can communicate and reach an agreement not to confess and if the agreement could be enforced, would each get a one-year sentence – the Pareto efficient outcome. In the current situation, you cannot communicate and reach an agreement so that you will end up in a sub-optimal position

Now, let us apply this concept of prisoners' dilemma to oligopoly. The concept of the prisoners' dilemma can be used to analyze the incentive to cheat in a cartel (i.e., the tendency to secretly cut prices or to sell more than the allocated quota). Suppose two firms in a centralized cartel face the following situation. At one end, if both firms cheat

(say, each firm chooses to secretly cut the price it charges), each would earn a profit of 8 thousand Birr. If one cheats and the other sticks to the agreed on price, the cheater would enjoy a profit of 20 thousand Birr and the other a profit of 2 thousand Birr. At the other end, if both firms stick to their agreement, each would enjoy a profit of 12 thousand Birr. The payoff matrix below summarizes this information. (All figures in the table are in thousands of Birr).

		Firm B	
		Cheat	Don't cheat
Firm A	Cheat	8, 8	20, 2
	Don't cheat	2, 20	12, 12

Table 3.6: The Prisoners' Dilemma Applied to Cartel Members

Workout the Nash equilibrium of this game! If you have done correctly, you must find that to cheat is the dominant strategy of each firm. Consequently, each firm would decide to adopt its dominant strategy of cheating and end up in earning a profit of 8 thousand Birr each. But, by choosing not to cheat (by sticking to the agreement), each member of the cartel would earn the higher profit of 12 thousand Birr. Hence, the outcome – *(cheat, cheat)* – is sub-optimal (Pareto inefficient), i.e., the firms could have done better. The cartel members then face the problem of prisoners' dilemma. Only if cartel members do not cheat would each earn the higher cartel profit of 12 thousand Birr.

A cartel can prevent or at least reduce the probability of cheating by monitoring the sales of each member and punishing cheaters. However, to do so is not an easy task. Particularly, the larger the cartel and the more differentiated the product, the more difficult it is for the cartel to monitor and punish cheaters and thereby prevent cheating.

3.5 Repeated Games and Enforcing a Cartel

In the preceding section, the players (agents) were assumed to meet only once and play a single-shot prisoners' dilemma game. However, the situation is different if the game is to be played repeatedly. Two firms facing the problem of prisoners' dilemma can increase their profits by cooperating. Such cooperation, however, is not likely to occur in single-move prisoners' dilemma games.

Cooperation is more likely to occur in repeated or many-move games, which are very common in the real world. The best strategy players/firms adopt in such (repeated) games is that of tit-for-tat. **Tit-for-tat** behavior can be summarized as follows: **do to your opponent what he/she has just done to you.** That is, begin by cooperating and continue to cooperate as long as your opponent cooperates. If he/she betrays you, you betray him/her back the next time (you play the game).

For an optimal tit-for-tat to be realized (for firms to cooperate), there must be a hope that cooperation will induce further cooperation in the future. Among other conditions, it must be assumed that the game is *repeated indefinitely* or at least for *a very large and uncertain number of times*. If the game is played for a finite number of times, each firm has an incentive not to cooperative in the final period because it cannot be harmed by retaliation. Each firm knows this and thus will not cooperate on the next-to-the last move. Indeed, in an effort to gain a competitive advantage by being the first to start cheating, the entire situation will unravel (meaning, everything will be clear and there will be no confusion) and cheating begins from the first move.

For instance, if a game is to be played 10 (ten) times, then the players are sure that no one will cooperate on the 10^{th} round because there is no 11^{th} round to punish cheater(s). Then, each player, with no benefit from cooperating on the 10^{th} round thinks of cheating on that round. But, no gain is expected from cheating on the 10^{th} round as every one will cheat on that round. Consequently, each considers the gain from cheating on the 9^{th} round. Again, as each player expects that cheating will occur on the 9^{th} round, each would

consider gaining from cheating on the 8th round; and so on. The final result is that, cheating begins from the first move.

Hence, the outcome of a prisoners' dilemma game to be played for a certain and finite number of times is the same as that of a single shot game with the same payoffs. In both cases, the outcome will be Pareto inefficient.

Coming back to the practical application of tit-for-tat, when would tit-for-tat be an optimal strategy for cartel members? For an optimal tit-for-fat (cooperation) to develop, the following conditions should be met:

- ☞ There should be a stable set of players for the cooperative behavior to develop. There shouldn't be frequent move-ins and move-outs of firms into and out of the cartel.
- ☞ There should be a small number of players to keep track of what each is doing.
- ☞ The market situation should be such that each firm can quickly detect (and is willing and able to quickly retaliate for) cheating by any other firm.
- ☞ Demand and cost conditions must be relatively stable. Instability in demand and cost conditions, coupled with the slow process of cartel negotiation for a different price and quantity, is very likely to result in cheaters.

3.6 Sequential Games and a Game of Entry Deterrence

The games considered so far are games in which both agents (players) act simultaneously. But in many situations one player gets to move first, and the other player responds. An example of this is the Stackelberg's duopoly model, where one player is a leader and the other player is a follower. Another example is that of price leadership where a firm sets a price (or announces a price change) and the others adopt (or follow) this price (or price change).

Oligopolistic firms often adopt strategies to gain a competitive advantage over their rivals, even if it means constraining their own behaviour or temporarily reducing their

own profits. This is what we call a strategic move. A ***strategic move*** can be defined as an adoption of strategies that influences the rival's choice in a manner favorable to oneself (the adopter) by affecting the rival's expectations of how the adopter would behave. For instance, a firm could send a written commitment to customers to match any lower price by competitors – a strategic move. However, for the threat to be credible there must be a commitment that the firm making the threat is ready to carry it out. If the competitor (rival) firms believe that this firms sending commitment will actually do that, then the threat is credible. Consequently, they do not expect to benefit from any price reduction.

Suppose player A chooses top or bottom in the first round; player B gets to observe the first player's (A's) choice; and then (B) chooses left or right. The outcomes of the game between the two players are summarized in the payoff matrix given below.

		Player B	
Player A		Left	Right
	Top	2, 8	3, 7
	Bottom	1, 3	5, 4

Table 3.7: The Normal Form Representation of a Sequential Game

If it were a simultaneous game, this game would have two Nash equilibria: (Top, Left) and (Bottom, Right). However, this **normal form** representation of the game is not convenient to analyze the equilibrium outcome of a sequential game. Sequential games are best analyzed using an extensive form representation. **Extensive form** is the way to represent the game that shows time pattern of choices. Under the assumption that player A is the first-mover, the extensive form representation of the above game looks like the following.

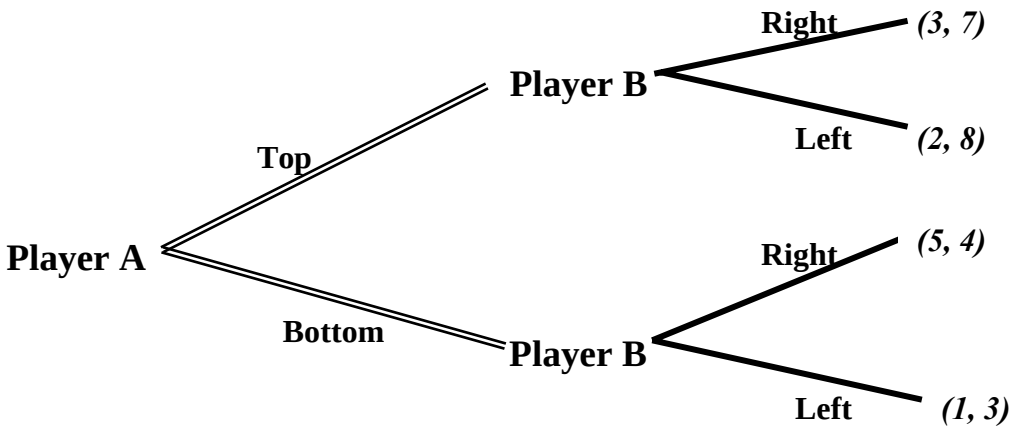


Figure 3.1: The Extensive Form Representation of a Sequential Game

We use the **backward induction** method to analyze this (sequential) game. The name backward induction is an indication that we begin the analysis from the right most tails of the game tree and sequentially go back to the starting point. Put differently, the backward induction method entails starting from the possible responses of the follower and going back to the decision of the leader.

The analysis goes as follows. First, we suppose that player A (the first-mover or the leader) has already made his/her choice and we are sitting in one branch of the game tree. If player A has chosen top, then player B will choose left (as 8 > 7), and the outcome would be (2, 8). If player A has instead chosen bottom, then the sensible thing for player B to do is to choose right (as 4 > 3), and the outcome would be (5, 4).

Now think about player A's initial choice. If A chooses top, he/she knows that the final outcome would be (2, 8) and thus he/she would get a payoff of 2. Besides, he/she also knows that he/she would get a payoff of 5 if he/she chooses bottom. So the sensible thing for player A to do is to choose bottom (since 5 > 2). Thus, the equilibrium choices will be bottom and right with the payoffs of 5 and 4 for A and B, respectively: Nash equilibrium = (Bottom, Right). Figure 3.2 below depicts the backward induction method. Start with the question marks at right most position and follw the dotted block arrows.

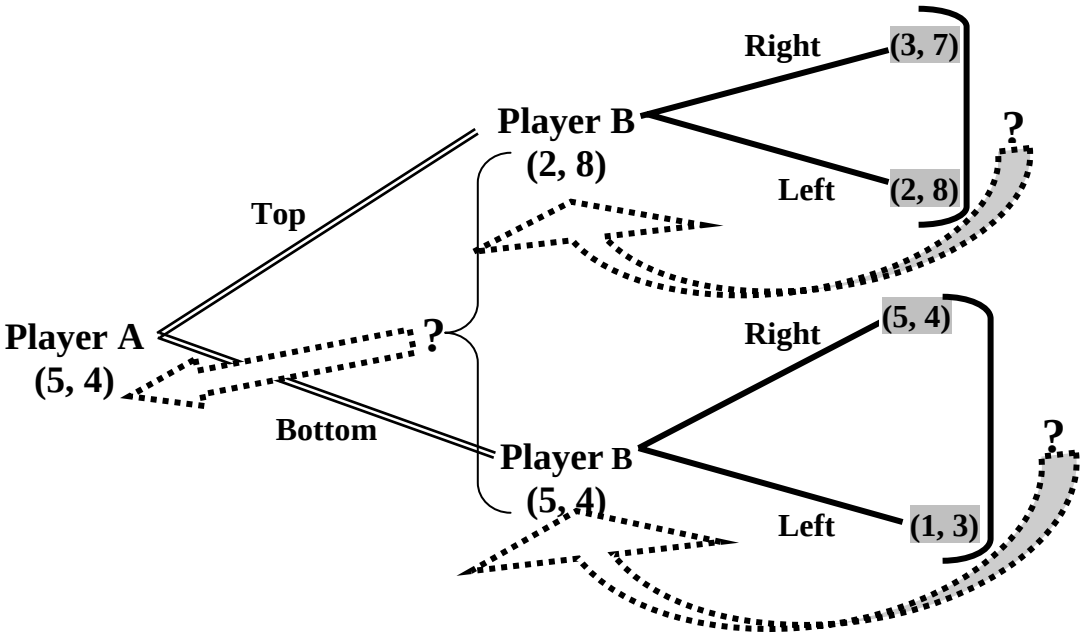


Figure 3.2: The Backward Induction Method

From B's point of view, the above outcome is rather unfortunate since he/she ends up with a payoff of 4 rather than 8. That is, B could have enjoyed a payoff of 8 had A played top. What might he/she do about it? He/she can threaten the leader to play left if the leader (player A) plays bottom. If player A thought that B would actually carryout this threat, he/she would be well advised to play top. Why top? The reason is that top gives player A a payoff of 2 while bottom will only give him/her a payoff of 1 – if B carries out the threat. Under the assumption of rationality, however, if player B chooses to change its response (to A's choice of bottom) from right to left, B would reduce its own payoff from 4 to 1. Hence, player A won't fear that B would play left, which means that player B's threat is not credible: it is an empty threat.

Let us take our analysis of sequential games and the nature of possible threats that could be forwarded from followers to the oligopoly market. Recall the factors that make entry to an oligopolistic market difficult. One important strategy that an oligopolist can use to deter (prevent) market entry is to threaten to lower its price and thereby impose a loss on the potential entrant. Such a threat, however, may or may not be credible and works only if it is credible.

In this situation, the time sequence of the game is that, first, the entrant decides on whether to enter the market or to stay out, and then the incumbent (the existing firm) will decide on its response, i.e., whether to fight (cut its price) or not. In short, first, the potential entrant decides, and then the incumbent makes its decision. The payoffs of this game to the two firms are summarized as follows.

		Potential Entrant	
		Enter	Stay Out
Incumbent	Fight	4, -2	6, 0
	Don't Fight	7, 2	10, 0

Table 3.9: The Normal Form Representation of a Sequential Game between a Potential Entrant and an Incumbent

As you might now be familiar with, the first step to analyze a sequential game is to set up the extensive form representation of the game. The first-mover occupies the starting point of the game tree.

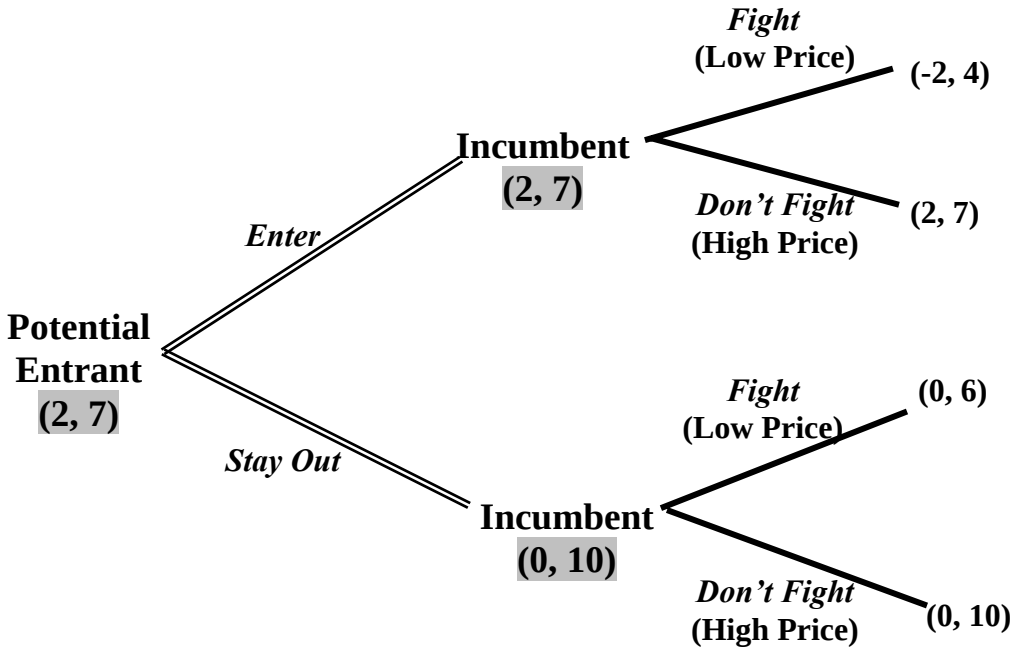


Figure 3.3: A Sequential Game between a Potential Entrant and an Incumbent Firm

Have you noticed the change in the order of the entries in each cell as we moved from the normal to the extensive form representation? For instance, why has the cell (4, -2) in Table 3.9 become (-2, 4) in Figure 3.3? Figure out! (Hint: start by interpreting (-2, 4) in the table!)

From the game tree in Figure 3.3 above, the incumbent firm charges a high price if the potential entrant stays out as 10 > 6. Similarly, if the potential entrant chooses to enter, the incumbent firm will prefer to continue charging the same price (i.e., not to fight) because 7 > 4. This implies that the final outcome would be either (2, 7) or (0, 10). With the knowledge that the final outcome would be either of (2, 7) or (0, 10), the potential entrant chooses to enter. Thus, the Nash equilibrium of this sequential game is: ***(enter, don't fight)***.

The incumbent firm may in fact threaten the entrant that it will fight if the potential entrant enters. However, this threat is not a credible threat. The reason is that the incumbent (assuming it to be rational) is not expected to carryout this threat as its own profit would fall from 7 units to 4 units if it carries out the threat.

Note that, however, the incumbent can make a credible commitment to fight entry at the expense of profits. Although this may seem irrational, a firm could develop a reputation for carrying out its threats even at the expense of its profits (by actually carrying out its threats several times). While we usually judge the credibility or otherwise of a threat on the assumption of rationality, firms in the real world could carryout a threat which is theoretically judged as an empty threat. That is, a rationally empty threat could be made credible with irrationality.

Alternatively, without developing a reputation for irrationality, the incumbent could make a credible threat by expanding its capacity before it is needed, i.e., by building excess capacity. Investment in excess capacity works through changing the payoffs from a game.

Continuing with the above example, the incumbent firm could invest in excess capacity so as to make its threats credible by changing the payoffs from the game as follows.

- ☞ In general, investment in excess capacity increases the cost of the incumbent firm.
- ☞ If the incumbent continues to charge the high price it used to charge, the sales and revenue of the firm do not change and thus its profit falls. In other words, the profits of the incumbent firm are now lower when it doesn't fight because idle or excess capacity increases its costs, without increasing its sales. In our example, this is reflected in the fall of the payoff from its previous level of 7 to (say) 3 if the entrant enters, and from 10 to (say) 8 if the entrant stays out.
- ☞ On the other hand, fighting (charging a lower price) would allow the incumbent firm to increase sales and utilize its newly built capacity. Consequently, both costs and revenues increase. Thus, it can be assumed (for simplicity) that the increase in costs and the increase in revenue exactly offset each other and thus the firm's profit remains unchanged.

The new pairs of possible payoffs are depicted in the extensive form game representation below.

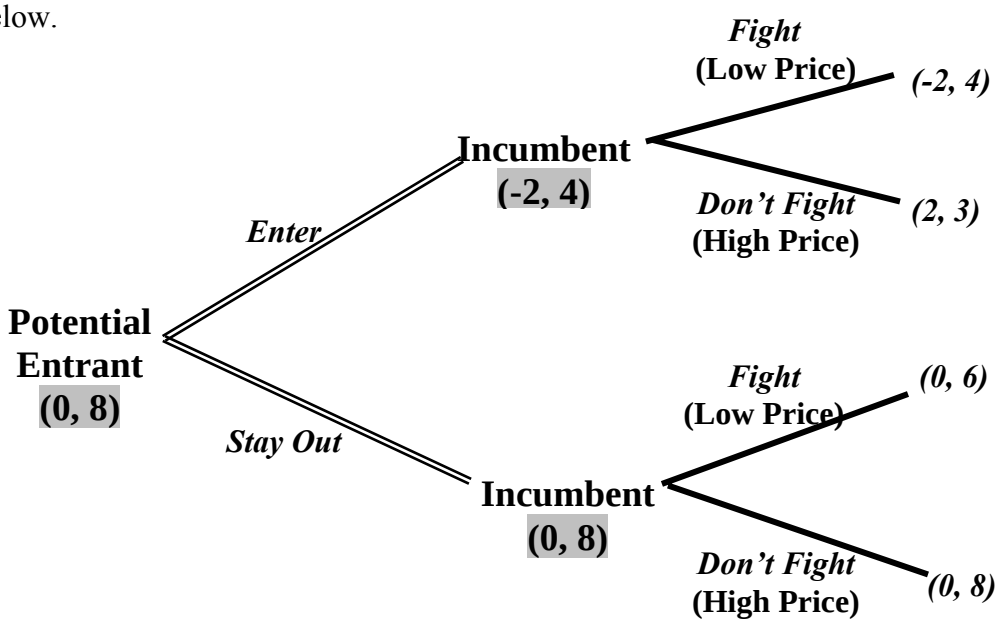
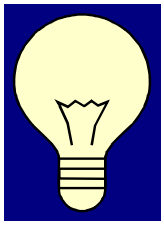


Figure 3.4: A Sequential Game between a Potential Entrant and an Incumbent Firm after Investment in Excess Capacity

With the new set of payoffs, the decision of the potential entrant would be different. The Nash equilibrium of this new sequential game is that the potential entrant stays out and the incumbent does not reduce its price. This means that the threat is credible: the potential entrant knows that if it enters, the incumbent will cut its price and make the entrant lose. But the Nash equilibrium *(stay out, don't fight)* also means that the incumbent will not use the extra capacity. Despite this fact, the investment in excess capacity is meaningful or worthwhile because the incumbent firm has signaled to the potential entrant that it would be able to successfully defend its market.



3.7 Summary

1. *Game theory studies the decision-making behavior of agents in situations where there is a mixture of conflict and cooperation, as in oligopoly. A payoff matrix is a table that gives the payoffs (benefits or gains for players) from all the strategies (specific courses of actions) open to the firm and the responses of the rival(s). Games played by firms could be cooperative or non-cooperative, zero-sum or nonzero-sum games.*
2. *Dominant strategy is the optimal choice for a player no matter what the opponent does. When at least one of the firms doesn't have a dominant strategy, there will be no dominant strategy equilibrium, but a Nash equilibrium. Nash equilibrium is a situation in which each player chooses an optimal strategy, given the strategy chosen by the rival(s). Not all games have a Nash equilibrium; and some games may have more than one. All dominant strategy equilibria are Nash equilibria, but the reverse doesn't hold.*
3. *A game without Nash equilibrium under the use of pure strategies will have a Nash equilibrium under the adoption of mixed strategies.*
4. *Firms are said to face the problem of prisoners' dilemma in cases where it is possible to make at least one party better off without making the other worse off,*

or when their interactions result in Pareto inefficient or Pareto sub-optimal outcomes.

5. *The problem of prisoners' dilemma could be overcome with tit-for-tat strategy (where each firm does to its opponent what the opponent has just done to it) to be played repeatedly for an infinite (or a large and uncertain) number of times.*
6. *Some games in oligopolistic markets involve time pattern of choice – they are sequential games. The outcome of such games is better analyzed using the extensive form representation (than the normal form or payoff matrix representation) of games, and the back ward induction method. Such an analysis (of sequential games) is of great applicability in dealing with the behavior of oligopolistic firms in deterring entry or defending their market shares.*



3.8 Review Questions

I. Multiple-Choice Questions

1. In a prisoners' dilemma game,
   A. the players strive for the attainment of the maximum of their mutual benefits.
   B. each player does have full information about the exact responses of its competitor.
   C. players end up with sub-optimal payoffs.
   D. All are correct answers.
   E. None of the above
2. Which of the following statements is true?
   A. A game that does not have a Nash equilibrium under pure strategies will not have a Nash equilibrium under mixed strategies.
   B. Dominant strategy equilibrium exists for a game in which only one of two firms has a dominant strategy.

C. The difference between pure and mixed strategies lies on the number of strategies available to the players.
D. In a situation where Nash equilibrium does not yield a Pareto efficient outcome, firms face the problem of prisoners' dilemma.
E. B and D are true.

3. In which of the following situations is the maximum possible cartel profit more likely to be attained than the rest?
A. Playing tit-for-tat in a single-move game.
B. Playing tit-for-tat in a game played for an infinite or uncertain number of times, given heterogeneous products.
C. Playing tit-for-tat in a game played for twelve rounds where there is stable set and small number of players.
D. Playing a tit-for-tat in a game played for an infinite number of times where there are stable set and small number of firms produce a homogeneous product.

4. Consider two games (Game1 and Game 2) with exactly the same payoffs which are to be played between two firms. If Game 1 is a simultaneous game and Game 2 is a sequential one, which of the following is true?
A. Games1 and 2 have exactly the same Nash equilibrium.
B. Cournot's equilibrium represents a game like Game 2, while Stackelberg's equilibrium represents a game like Game 1.
C. Game 2 can best be analyzed using the backward induction method, and not easily from the payoff matrix representation directly.
D. The concepts of empty threat and credible threat are associated with Game 1, and not with Game 2.
E. B and D are true. Heuristic

5. The prisoners' dilemma game yields an equilibrium in which:
A. both criminals choose not to confess, but would be better off by confessing.
B. both criminals choose to confess, but would be better off by not confessing.
C. both criminals could get the lightest sentence possible if they both confess.
D. both criminals would receive the maximum sentence if they both confess.
E. both criminals happen to bear the maximum sentence whether they confess or not.

6. In game theory, Nash equilibrium refers to a set of optimal strategies in which:
   A. each player chooses its best strategy, no matter what the opponent does.
   B. each player chooses its best strategy, given the optimal choice of its opponent.
   C. each player chooses a strategy that maximizes the minimum gain that can occur.
   D. each player chooses a strategy that entails doing to its rival whatever the rival has done in the previous round.
   E. each player tries to punish competitors that cheat on a price-fixing agreement.

II. True or False Questions

7. Game theory provides a set of conclusive propositions that can accurately predict the behavior of oligopolistic firms.
8. In a zero-sum game played between two agents, both players are necessarily losers.
9. Entry deterrence is possible only with a credible threat, and not with an empty threat.

III. Discussion Questions and Workout Problems

10. Distinguish between each of the following pairs:
    a) Pure strategies and mixed strategies.
    b) Dominant strategy equilibrium and Nash equilibrium.
    c) Single-shot games and repeated games.
    d) Simultaneous games and sequential games
    e) Normal form representation and extensive form representation of games.
    f) Empty threats and credible threats.
11. Explain the following concepts:
    a) Prisoners' dilemma
    b) Tit-for-tat
12. There are two firms, ***firm I*** and ***firm II***, each examining the feasibility of a particular market. Each has to decide to produce and sell either *chocolate* or *chewing gum*. If both firms choose to produce and sell chocolate, the firms will incur losses amounting to 180 million and 200 million Birr, respectively. If firm I decides to produce and sell

chocolate while the firm II chooses to produce and sell chewing gum, their returns will be 120 million and 220 million Birr profits, respectively. However, if both of them decide to produce and sell chewing gum, they will incur losses of 160 million and 200 million Birr, respectively. And finally, if firm I produces and sells chewing gum while firm II produces and sells chocolate, their profits will be 160 million and 180 million Birr, respectively. Based on this information, answer the following questions.

A. Write the payoff of matrix for this game.
B. Does firm I have a dominant strategy? What about firm II?
C. What is the pure strategy Nash equilibrium of the game if the firms decide their strategy simultaneously?
D. Write the extensive form depiction and find the Nash equilibrium of a sequential game in which firm I is the first-mover.
E. Given your answer to question (D), how could the second threaten the first firm? Will this threat be credible? Why or why not?

13. Given the following payoff matrix of a game, find the mixed-strategy Nash equilibrium if ***Firm 1*** plays *strategy 1* with probability "***t***" and ***Firm 2*** plays *strategy a* with probability "***l***". How do you interpret the equilibrium?

		Firm 2	
		Strategy a	*Strategy b*
Firm 1	*Strategy 1*	(-2, -2)	(-2, -3)
	Strategy 2	(-1, -2)	(-3, 3)

CHAPTER FOUR
FACTOR PRICING AND INCOME DISTRIBUTION

Introduction

In the previous three chapters, we have examined the product market in isolation. In those chapters, the concern has been with the determination of the prices and quantities of final commodities on the assumption that whatever amount of an input needed is available for use at a fixed price. However, just like final consumer goods, inputs (resources) have demanders and suppliers whose interactions determine their prices and quantities. In this chapter, we turn our face to the factor (resource) market and deal with the determination of prices and employment of inputs. Related to this, we will also consider the distribution of income among resource owners.

In many ways, the determination of input prices and employment is similar to the pricing and output determination of commodities. That is, the price and employment of an input are generally determined by the market force of interaction between demand and supply.

There are several important qualifications (differences), however. Two of these are:

1. Whereas consumers demand commodities because of the utility or satisfaction they receive in consuming the commodities, firms demand inputs in order to produce the goods and services demanded by the society. That is, the demand for an input is a 'derived demand'; it is derived from the demand for the final commodities that the input is used in producing. For instance, the demand for labor is explained by the demand for wheat or any other product.
2. While consumers demand commodities, firms demand the services of the inputs. That is, firms demand the flow of input services and not the stock of the inputs themselves.

The pricing of factor of production (inputs) takes place in various markets. We will first examine the determination of factor prices in perfectly competitive product and input markets. Subsequently we will relax the assumption of perfect competition and discuss factor pricing in markets with various degrees of imperfection.

Chapter Objectives

After thoroughly going through this chapter, students would be able to:

- Discuss the equilibrium conditions of profit maximization for firms under different market structures.
- Understand how the demand for and the supply of an input are derived under different assumptions regarding the nature of competition.
- Describe how the demand for and the supply of an input interact to determine the input's price and employment level under perfect and/or imperfect competitions in the output and/or input markets.
- Explain the exploitation of productive factors because of the existence of monopolistic and/or monopsonistic power (of firms).
- Discuss the effects of changes in relative factor prices and technological progress on the relative share of factors (income distribution).

4.1 Factor Pricing under Perfect Competition

This section is devoted to determining the price and employment level of an input when both factor and product markets are perfectly competitive. To this end, we first derive the demand of a firm for the inputs. Two types of demand of a firm for an input are considered: the short run demand of a firm for an input (demand for a single variable input) and the long run demand for an input (i.e., the demand of a firm for a variable factor when there are several variable factors). The analysis of the demand concludes with the derivation of the market demand for a productive factor. The second side underlying the determination of factor price and employment is the supply side. The derivations of the supply of labor and the supply of other factors occupy sub-sections.

Finally, the equilibrium price and level of employment for an input are discussed. This is done by bringing together the market demand for and the market supply of an input. The discussion in this section is very crucial for understanding the subsequent sections.

4.1.1 The Demand for Factors of Production

To reiterate, in this sub-section, we will derive the short run and the long run demands for an input by a firm, and the market demand for an input one after the other.

4.1.1.1 The Demand of a Firm for One Variable Productive Factor

Regardless of the period of analysis, a firm's demand for an input shows the quantities of the input that the firm would hire at different levels of input price.

In deriving the short run demand for an input, we assume that the input under consideration is the only variable factor of production, i.e., the amount used of the other inputs is taken as fixed (cannot be changed). According to the marginal concept, a profit-maximizing firm will continue to hire an input as long as the extra income (receipt) from the sale of the output produced by the input is larger than the extra cost of hiring the input.

Let us first explain some concepts, which are central to our discussion:

- The change in output resulting from the use of an additional unit of a productive factor is known as the *marginal product of the input* (MP_i). That is, $MP_i = \frac{\partial TP}{\partial I}$ where TP stands for total product and I stands for the units of the input. Because this extra contribution is measured in physical units (kilograms of wheat, meters of cloth, etc.), the marginal product is also called the marginal physical product (MPP).
- To convert such an extra contribution to monetary terms, we multiply the marginal physical product of the factor by the market price of the output (P_X). The

monetary value of the contribution of an extra unit of an input is called the *value of marginal product of the input* (VMP_i). That is, $VMP_i = MPP_i \cdot P_X$.

- ☞ The extra income of a firm from the sale of the output contributed by an additional unit of an input is termed the firm's *marginal revenue product of that input*. This extra income is given by the marginal (physical) product of the input (MPP_i) times the marginal revenue of the firm from the sale of the units of the output (MR_X). That is, $MRP_i = MPP_i \cdot MR_X$.
- ☞ The extra expense a firm incurs to purchase (or rent) an additional unit of a factor of production is the firm's *marginal expenditure on the factor*, $ME_i = \frac{\partial TC}{\partial I}$.

When the firm is a perfect competitor in the product market, its marginal revenue is equal to the commodity price ($MR_X = P_X$). Consequently, the firm's marginal revenue product equals the firm's value of marginal product, i.e., $MRP_i = VMP_i$. For concreteness, if the variable input we are dealing with is labor, then MRP_L (= $MPP_L \cdot MR_X$) = VMP_L (= $MPP_L.P_X$).

The marginal expenditure on (or the extra cost of hiring) an additional unit of an input (ME_i) is equal to the input price if the firm is a perfect competitor in the input market. Perfect competition in the input market means that the firm demanding the input is too small to affect the price of the input significantly by itself. In other words, a firm that is a perfect competitor in the factor market can hire any amount of the input (service) at the given market price for the input. Thus, the firm faces a horizontal or infinitely elastic supply (curve) for the input. For example, if the input is labor, this means that the firm can hire any quantity of labor time at the given (market-determined) wage rate.

In general, a profit-maximizing firm in any kind of market structure should hire an additional unit of labor as long as the MRP_i exceeds the marginal expenditure on the input. In the case of a firm that is a perfect competitor in both the product and factor markets, profit is maximized at a point where $MRP_i = VMP_i = P_i$, where P_i is input price. Guaranteeing that the short run demand for any other productive factor is derived in a

similar way, we will carryout our analysis of factor demand with reference to a specific input: labor. With labor as the factor of our concern, the mathematical derivation of the profit-maximizing condition follows. Because of perfect competition, both the price of the firm's product and the price of labor (wage rate) are constants, i.e., $P_X = \overline{P}_X$ and $w = \overline{w}$.

$\max_L \pi = R - C$ where $R = P_X . X$ and $C = \overline{w} L + F$; F = fixed cost.

$$\Rightarrow \frac{d\pi}{dL} = \frac{d(\overline{P}_X . X - \overline{w} . L - F)}{dL} = \overline{P}_X \frac{dX}{dL} - \overline{w} \frac{dL}{dL} = 0.$$

$$\Rightarrow \overline{P}_X . MPP_L = \overline{w}.$$

$$\Rightarrow \boxed{VMP_L = \overline{w}}$$

Recall (from microeconomics I) that there are three stages of production and that a rational firm operates in the second stage of production where the MPP_L is declining but positive. A downward sloping MPP_L multiplied by a fixed output price gives a downward sloping VMP_L curve. Multiplying a downward sloping curve by a constant is just like scaling the curve up (if the constant is greater than 1) or down (if the constant is less than 1). The resulting VMP_L curve is shown in Figure 4.1 below.

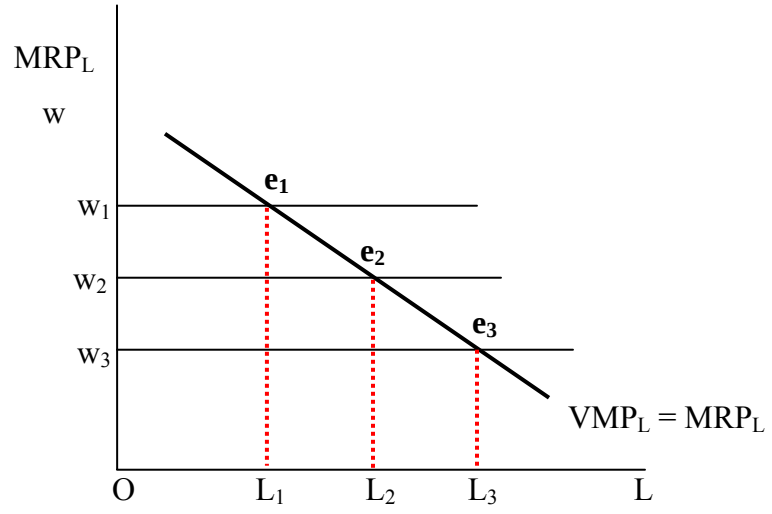


Figure 4.1: Equilibrium of a Firm in Perfectly Competitive Product and Input Markets

With reference to Figure 4.1, for a market wage rate of w_1, the firm's optimal point of operation is point $\mathbf{e_1}$ defined by the condition $VMP_L = w_1$. The firm's profit is at the maximum for wage rate. Given w_1, if the firm operates at a point to the left of $\mathbf{e_1}$, then we witness that $VMP_L > w_1$ (i.e., the extra monetary contribution of hiring one more unit of labor is greater the extra expense on the unit). As a result, the firm would increase its profit by hiring more labor. The opposite holds for operating at any point to the right of $\mathbf{e_1}$. That is, an additional unit of labor adds more to the cost than it adds to the revenue and thus the firm would increase its profit by reducing the amount of labor it uses. This proves that the only equilibrium of the firm is at a point where $VMP_L = w_1$.

Given the VMP_L curve and the equilibrium condition $VMP_L = MRP_L = w$, the firm in Figure 4.1 hires L_1 units of labor if the market wage rate is w_1. Similarly, the firm hires L_2 units of labor when the market wage rate is w_2 and L_3 units for w_3. The graph that shows this relationship between the wage rate and the quantity demanded (hired) of labor is the demand curve (Figure 4.2). Thus, for a firm that is perfectly competitive in both the labor and product markets, the short run demand for labor is given by the VMP_L (which is the same as MRP_L in this case). In general, under perfectly competitive product and labor markets, a firm's demand for a single variable input is the value of marginal product of the input ($VMP_i = MRP_i$)

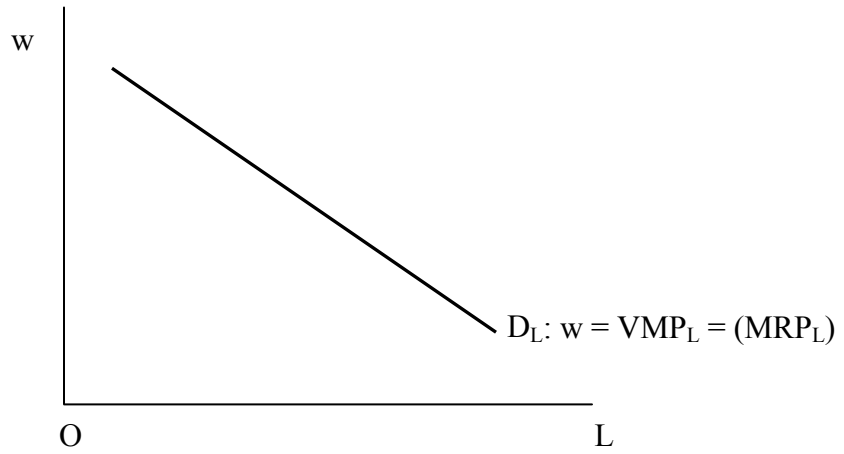


Figure 4.2: The Demand of a Firm for Labor in the Short Run

4.1.1.2 The Demand of a Firm for a Variable Factor when There Are Several Variable Inputs

When there are more than one variable factors of production, the VMP_i curve does not represent the demand for the input. That is, a firm's demand for labor is no more the same as the VMP_i in long run. This is because, in the long run, various resources are used simultaneously in the production process so that a change in the price of one factor leads to changes in the employment (use) of the other factors as well. For instance, a change in wage rate induces not only a change in the quantity of labor a firm uses but also changes in the quantities of other factors (of which capital is one); in the long run, no input is fixed in amount. This change in the amount of other factors in the production process, in turn, shifts the marginal (physical) product curve of the input whose price is initially changed. To continue with the example of labor, a change in wage rate results in changes in the use of labor and capital (all other factors); and the changes in the amounts of these other factors cause the marginal product of labor to shift.

Let us assume that the price of labor (the wage rate) falls. Then this fall in wage rate has three effects: a substitution effect, an output effect, and a profit-maximizing effect.

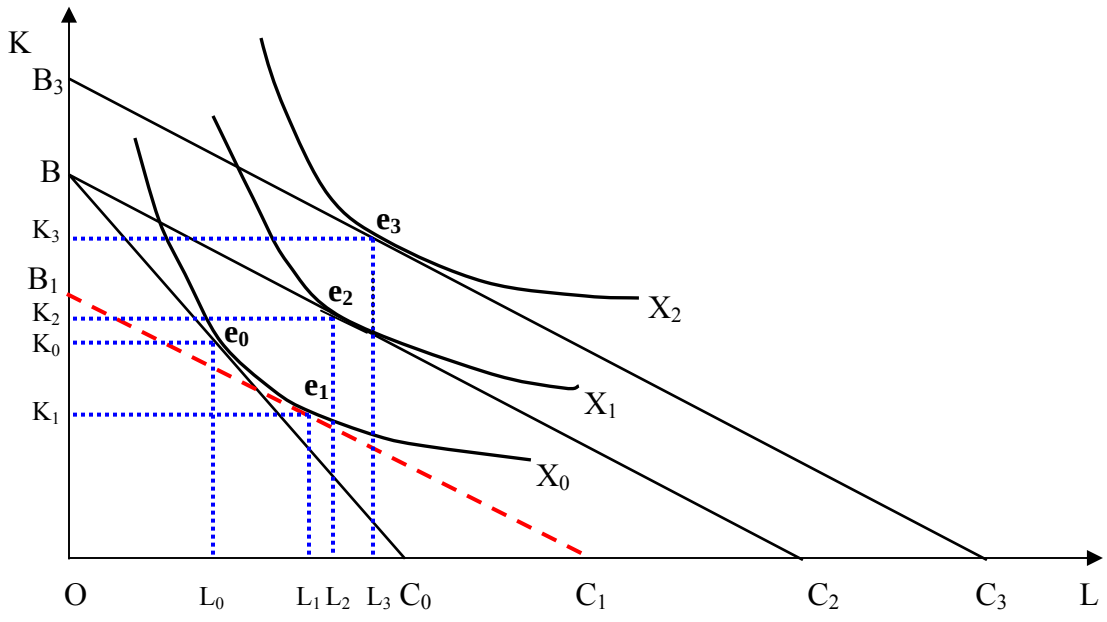


Figure 4.3: Substitution, Output and Profit Effects of a Fall in Wage Rate

Initially, the firm produces a profit-maximizing output X_0 with a combination of L_0 and K_0 given the isocost line BC_0 whose slope is defined by the ratio of factor prices (– w/r) (Figure 4.3). When wage rate falls, the isocost changes from BC_0 to BC_2 and this new isocost (BC_2) is tangent to the isoquant corresponding to output level of X_1 (equilibrium at $\mathbf{e_2}$). Correspondingly, L_2 units of labor and K_2 units of capital are used. This movement from $\mathbf{e_0}$ to $\mathbf{e_2}$ can be split into two: substitution effect and output effect.

To isolate the substitution effect from the output effect, we draw an isocost line (B_1C_1) which is parallel to the new isocost line (BC_2) but tangent to the old isoquant (X_1). The movement from $\mathbf{e_0}$ to $\mathbf{e_1}$ is the substitution effect. This shows that the firm would substitute the cheaper labor for the relatively more expensive capital even if it were to produce the original level of output (X_0). Thus, because of the substitution effect of the wage fall, the employment of labor will rise from L_0 to L_1 while that of capital falls from K_0 to K_1.

However, to produce the same amount of X as before (X_0) – now with more of the cheaper input and less of the relatively expensive one – the firm does not exhaust the total outlay assigned for the two inputs. That is, when wage rate falls, the firm can hire more of the two factors (L and K) with the same expenditure. Hence, the firm produces a higher level of output with more labor and capital (L_2 and K_2) and, therefore, the movement from $\mathbf{e_1}$ to $\mathbf{e_2}$ is the output effect.

Note that the total expenditure of the firm (on L and K) at point $\mathbf{e_2}$ is the same as that at point $\mathbf{e_1}$. Point $\mathbf{e_2}$ is not the final equilibrium of the firm because keeping the total cost/expenditure constant does not maximize its profit. The fall in wage rate results in a shift in the firm's marginal cost curve downward (or to the right). This change in marginal cost is shown by the movement from MC_1 to MC_2 in Figure 4.4 below. With the new marginal cost, the firm's profit maximizing level of output increases from X_1 to X_2.

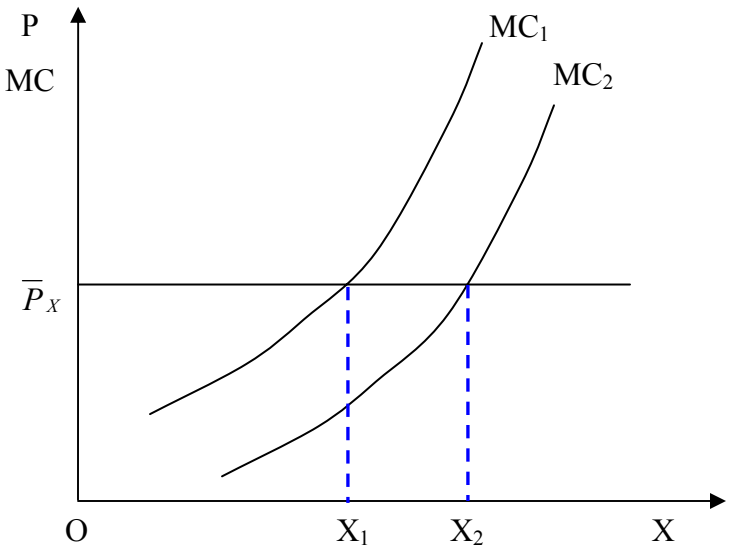


Figure 4.4: Fall in Wage Rate Reduces the Marginal Cost of Production of a Firm

Thus, the isocost line BC_2 in Figure 4.3 must shift upward in a parallel manner. So, the final equilibrium of the firm is attained when isocost B_3C_3 is tangent to the highest possible isoquant (X_2) at point **e_3.** The movement from **e_2** to **e_3** is the profit effect (or the profit-maximizing effect).

Once we have explained the three effects, what is the implication of each of these effects for the demand of labor? The substitution effect of a decline in wage rate increases the units of L and reduces that of K, and thus causes a decline in the marginal physical product of labor. The output and profit effects of the wage fall raise the amounts of both labor and capital. Hence, both the output and profit effects increase the MPP_L at a given level of labor employment and thereby cause the MPP_L to shift upward (to the right).

Assuming that labor and capital are gross complements (more of complementary inputs than they are substitutes), the output and profit effects more than offset the substitution effect. The overall effect of a fall in wage rate is an increase (and rightward shift of) the MPP_L (curve). Given the price of the final commodity, $\overline{P}_X$, the upward shift in the MPP_L curve implies that the VMP_L curve also shifts to the right (as depicted in Figure 4.5 below).

The discussion on the substitution, output and profit effects of a wage change above is of a subsidiary purpose. It just intermediates the change in wage rate and the response in the quantity demanded of the input (labor). Our main interest is on the relationship between the two variables – wage rate and level of employment.

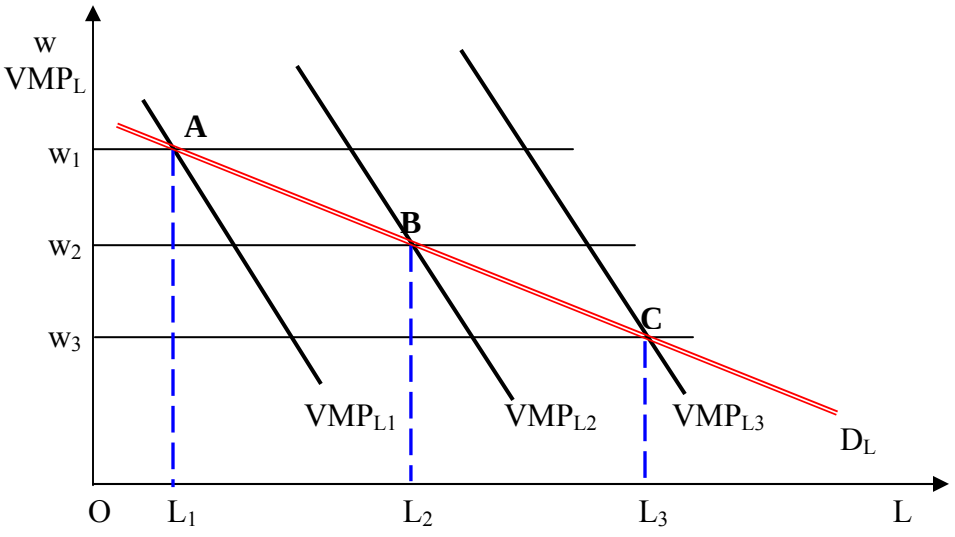


Figure 4.5: Shifts in VMP_L Curve and Equilibrium of a Firm

At the initial wage rate w_1, L_1 units of labor are employed (which is determined by the intersection of VMP_{L1} and the supply of labor the firm faces, w_1). When wage rate falls to w_2, a new optimal quantity demanded of labor will be established at point B on VMP_{L2}. The shift in the VMP_L is explained by the three effects of a wage change in the discussion above. If the wage rate further declines to w_3, the three effects will shift the VMP_L to the right and the new equilibrium will be at point C. The locus of these equilibrium points (points A, B and C) is the demand curve for labor by the firm when several variable factors are used. That is, the long run demand of a firm for labor (for an input in general) is not the same as the VMP curve of the input, but derived from changing (shifting) VMP curves.

4.1.1.3 The Market Demand Curve for an Input

The market demand curve for an input is derived from the demand curves for the input by individual firms. Nevertheless, it is not the simple horizontal summation of the demand curves of the individual firms. This is because when the price of an input (say labor) falls, not only this firm but also other firms will employ more of this factor and other (complementary) inputs to expand production. Thus, the supply of the final commodity increases and consequently its price falls (See Figure 4.6).

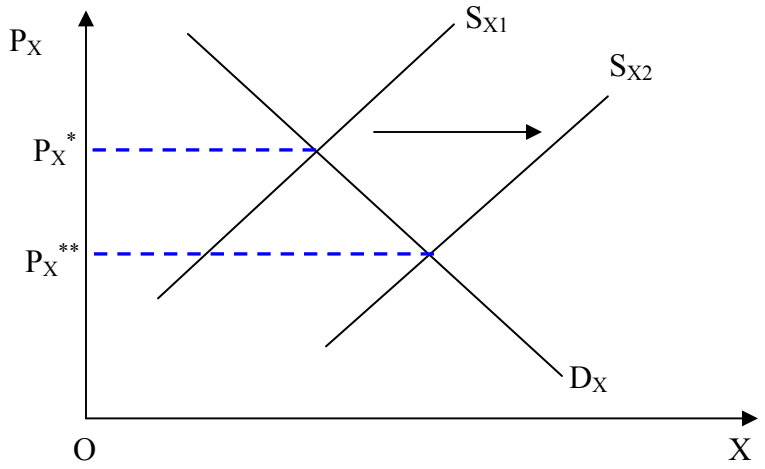


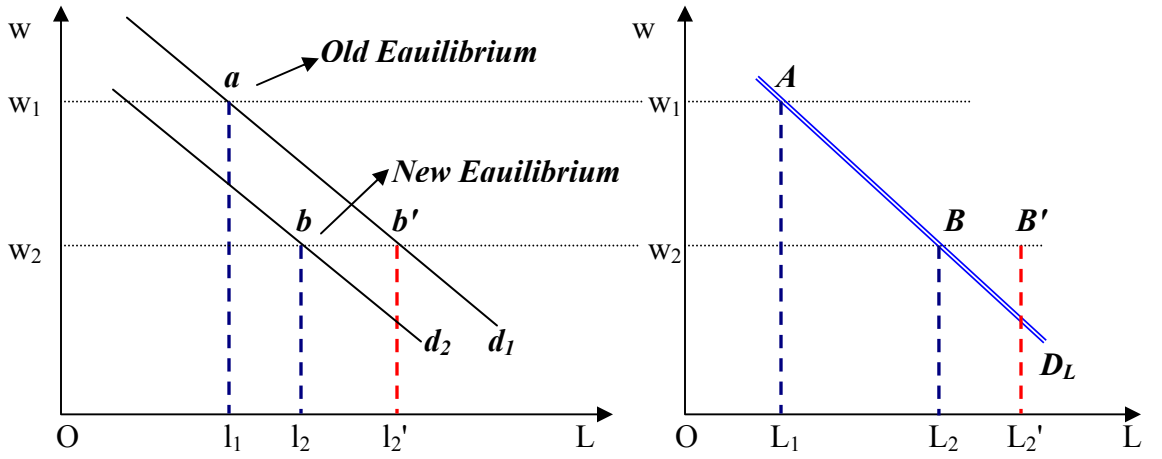
Figure 4.6: The Effect of Fall in Input Price on Commodity Price

For instance, suppose there are 100 identical firms in a market that produce good X. If each of these firms is currently hiring 10 units of labor at a wage rate of $w_1 = 30$ Birr per hour, then the total (market) quantity demanded of labor will be 10x100 = 1,000 units. If the input price (wage rate) falls form $w_1 = 30$ to $w_2 = 20$, then each firm uses more of labor (and other factors) and expand its production of X. As the supply of the commodity increases (the shift from S_{X1} to S_{X2} in Figure 4.6), the equilibrium price of the product will be derived down from P_X^* to P_X^{**}. This fall in price of the final product, in turn, has a negative consequence on the demand for the factor.

Since the $MRP_L = MP_L$ times MR (which is equal to the commodity price in perfectly competitive product market), the reduction in commodity price will cause each firm's

MRP_L (= VMP_L) curve and the demand curve for the input to shift down or to the left. Each firm demands, say, only 12 units of labor rather than what the individual demand curve under the ceteris paribus assumption predicts, say, 15 units. As a result, the market quantity demanded of labor at the wage rate $w_2 = 20$ will be 12x100 = 1200 units instead of 15x100 = 1500 units.

In general, the market demand curve for an input is then derived by the horizontal summation of the individual firms' demand curves for the input after the effect of reduction in the commodity price has been considered (see Figure 4.7 below).



(a) A Firm's Labor Demand Curve **(b) Market Demand Curve for Labor**

Figure 4.7: Deriving the Market Demand for an Input

If the fall in commodity price were not taken into account, and if a simple horizontal summation were taken, it would lead to an overestimation of the market demand for labor (which joins points A and B' in panel (b) of Figure 4.7 above).

We have derived the short run and the long demands for an input by a firm and subsequently the market demand for an input. Now, let us close our demand side analysis of the factor market by describing some factors that affect the demand for a factor of production.

The amount of an input demanded depends on:

- The price of the input under consideration: an increase in input price reduces the quantity demanded of the input, and vice versa. A change in the input's own price results in a change in the *quantity demanded* of the input (movement on a demand curve), and does not change the *demand* for the input.
- The marginal physical product (MPP) of the factor: if the MPP of a factor increases, more of the factor will be demanded, and vice versa. A change in the MPP of a factor implies a shift in the demand for the input either to the left or to the right.
- The price of the output (commodity): when the output the firm produces becomes cheaper, the firm will cut its production and thus demands less of the input used. The opposite happens when the commodity gets more expensive.
- The amount of other factors that are combined with the factor: if, for instance, the amount of a complementary input rises, the productivity (MPP) and thus the demand for the factor under consideration will rise.
- The price of other factors: increase in prices of complementary inputs reduces the amount of the complementary inputs to be used with a given amount of a productive factor, and thus reduces the productivity of (and demand for) the productive factor. An increase in the price of a substitute, on the other hand, increases the demand for the factor of concern.
- Technological progress: technological progress could increase or decrease the demand for an input depending on whether it increases or decreases the MPP of the factor. A technological progress that raises the productivity of labor (more than that of capital), will increase the demand for labor, and vice versa.

4.1.2 Factor Supply

4.1.2.1 The Supply of Labor

In reality, no two individuals possess and supply identical amount of labor. Even though we can compare the quantity of labor supplied by different individuals (say, in terms of

the number of hours each works for), there is difference in the quality of labor supplied. As a reflection of this heterogeneity of labor (and of course some other reasons), we observe different prices (wages) paid to suppliers of labor. However, to simplify things, let us assume that labor is a homogeneous factor: all labor units are identical. This will enable us to derive the supply of labor by an individual and the market supply of labor, and subsequently determine a single equilibrium wage rate.

The ultimate objective of this sub-section is to derive the market supply of labor. In general, the main determinants of the total labor service supplied (in a particular market or economy) include:

- The price of labor (wage rate) that prevails in the market;
- The tastes of consumers, which define their trade-offs between leisure and work;
- Population size in a broader sense (and more specifically, size of the working-age population);
- The labor force activity (participation) rate; and
- The occupational, educational and geographical distribution of the labor force.

The relationship between the quantity supplied of labor and the first factor in the above list – the wage rate – defines the labor supply (function or curve). The other determinants (in the list) are considered as factors that shift the supply curve – shift factors. Since the market supply is derived by summing the quantity supplied by individuals at various wage rates, we begin with the derivation of the supply of labor by a single individual.

A. The Supply of Labor by an Individual

The labor requirements of business firms are generally provided by individuals. These individuals have the choice of using their labor (labor power) either for work or for leisure. Hence, the supply of labor by an individual depends on the individual's preference for leisure and work. The preference of the individual between leisure and work (where work is valued for and in terms of the income it generates) can be represented by a well-behaved indifference map (a set of indifference curves). An

indifference curve is a graph of alternative combinations of goods that provide a given level of satisfaction (utility). The substitutability between income and leisure is assumed imperfect. That is, the marginal rate of substitution of leisure for income ($MRS_{L,Y}$) – the amount of income sacrificed for an extra unit of leisure, holding utility constant – declines as the individual has more and more of leisure.

Let us measure hours available for leisure and work over a given period of time along the horizontal axis and the real income along the vertical axis. In such a set up the $MRS_{LY} = -\frac{dY}{dL}$ = the absolute value of the slope of the indifference curve.

While the individual attempts to achieve the highest possible level of utility, defined by $U = f(Y, L)$, the choice among alternative levels of income (Y) and leisure (L) is, however, restricted due to two constraints: a time constraint and a goods constraint.

- ***A time constraint***: the total amount of time available (T) is divided into hours of work (H) and hours of leisure (L): T = H + L. This time constraint represents all the feasible allocations of time between leisure and work.
- ***A goods constraint***: using the definitions of H, L and T above along with w = wage rate, P = price index, Y = real income (output), and assuming that all the income of the individual comes from labor, the goods constraint is given by: PY = wH. This equation states that total spending (PY) must equal earnings (= wH).
- Rewriting the time constraint as: H = T – L and substituting this into the goods constraint results in: PY = wT – wL. With a little algebraic manipulation, this becomes: wT = PY + wL. This equation is called ***a full-income constraint***. This equation states that full (potential) income (wT) equals the total explicit costs of goods and services (PY) plus the total implicit cost of leisure time (wL). An alternative form of the full-income constraint is given by: $Y = -(\frac{w}{P})L + (\frac{w}{P})T.$

  This equation describes the relationship that exists between hours of leisure and real income. It represents the individual's **budget constraint**. Noting that this budget constraint is expressed in slope-intercept form, the intercept of the budget

constraint on the vertical axis equals $\frac{wT}{P}$ (= the real value of full income), and the slope of the budget constraint equals $-\frac{w}{P}$ (= the negative of real wage rate).

- The equilibrium (optimal) allocation of time between leisure and work (income) is found at the point of tangency of the indifference curve and the budget line. In other words, equilibrium is achieved when the slope of the budget line $(-\frac{w}{P})$ equals the slope of the indifference curve $(-MRS_{LY})$, that is, when $MRS_{LY} = \frac{w}{P}$. This is shown by point ***E*** in Figure 4.8 below. The individual works for 0Q units of time (say, hours) – earning income equal to QE, say Birr – and uses the remainder of his/her time (= QT) for leisure. Note that, in the figure, the time of leisure is measured from left to right and that allotted to work from right to left.

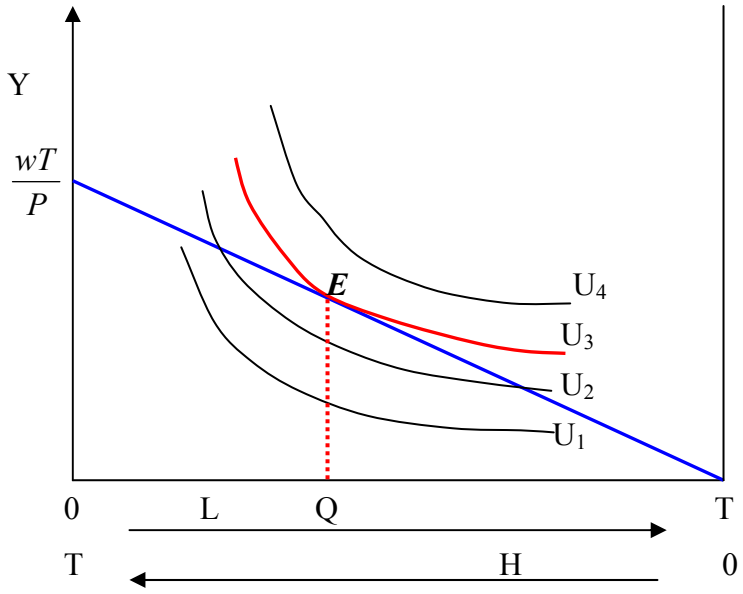


Figure 4.8: An Individual's Allocation of Time between Leisure and Work

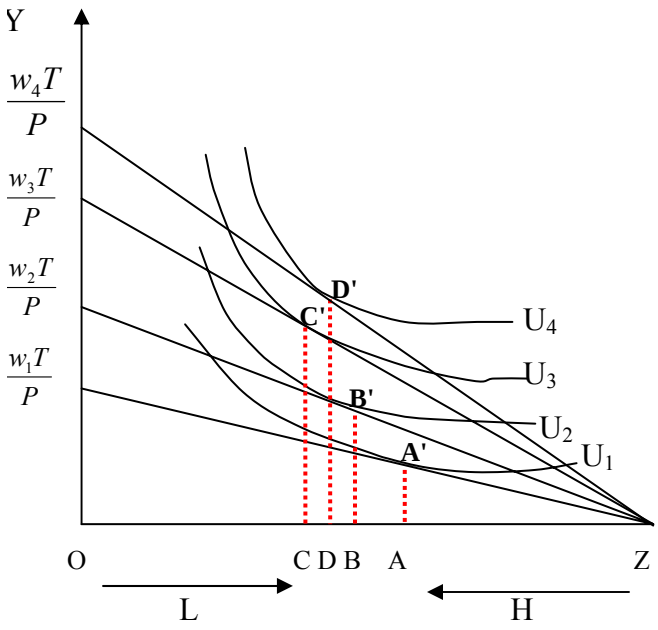
Now, once you understand how the equilibrium of an individual is determined, we will move to the derivation of the individual's labor supply curve. Let us approach the derivation of an individual's labor supply decision graphically. In general, a wage change

has two effects: the substitution effect and the income effect. Consider a rise in the market wage rate.

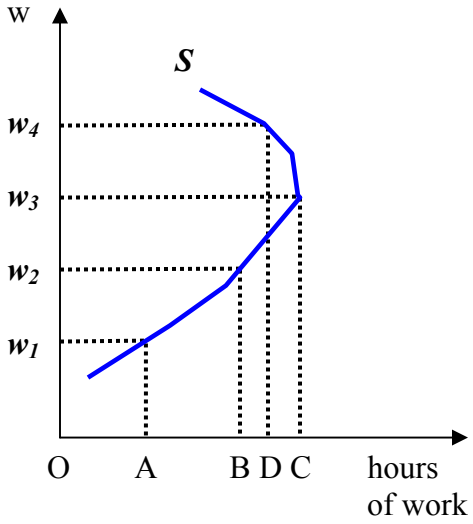
As the wage rate rises, the opportunity cost of leisure time rises. In response to this higher wage, individuals consume less leisure time and spend more time at work (even at the same level of satisfaction). This is the ***substitution effect*** resulting from a higher wage. In short, an increase in wage rate leads an individual to work more, i.e., substitution effect renders wage rate and hours of work to be directly related. Thus, the substitution effect of the wage increase always operates to make the individual's supply of labor curve positively sloped.

An increase in the wage, however, also raises an individual's real income. This leads to an increase in the consumption of all normal goods. Assuming leisure to be a normal good, a higher wage will generally induce individuals to consume more leisure time (and reduce hours of work). Individuals who receive a higher wage can afford to take more time away from work. This is the ***income effect*** resulting from a wage increase. Thus, the income effect of the wage increase always operates to make the individual's supply of labor curve negatively sloped.

With reference to Figure 4.9, at a wage rate equal to w_1, the individual faces a budget line given by $Z\frac{w_1 T}{P}$ and thus at equilibrium by working for AZ hours (uses OA hours for leisure), earning AA' income. If the wage rate increases to w_2, the individual will work for more hours (BZ > AZ), will earn a higher income (BB'), and will have less hours (OB < OA) for leisure. The quantity supplied of labor will also increase when wage rate rises to w_3 (to ZC).



Panel (a)



Panel (b)

Figure 4.9: The Supply of Labor by an Individual

However, at some higher wage rate the hours offered for work may decline. For example, if the wage rate rises to w_4, the individual will work for DZ hours, which is less than the supply at w_3 (CZ) – *Panel (a)*. This pattern of response to higher wage rates produces a backward bending supply curve for labor – *Panel (b)* in the figure.

The reason for which an individual's supply of labor may be backward bending can be explained by separating the substitution effect of wage rate change from the income effect. Both the substitution and income effects operate over the entire possible wage rates. While the substitution effect generally overwhelms the opposite income effect at lower wage rates, the negative income effect overwhelms the positive substitution effect at higher wages implying a backward bending supply of labor curve.

Comparing the supply of labor across people, tastes and preferences of individuals (between leisure and work) differ. Consequently, the wage rate at which an individual's supply curve of labor bends backward is likely to differ from individual to individual.

B. The Market Supply of Labor

The market supply of labor is the summation of individual supplies of labor. Although there is a general agreement that the supply curve of labor by single individuals exhibits the backward bending pattern, it is usually the case that the market supply is upward sloping (not backward bending). This is because, even if some people prefer to work for fewer hours at higher wage rates, some new workers will also be attracted to the market by the higher wages. While the workers attracted could be from a different local labor market (assuming labor mobility) in the short run, population growth also increases the number of individuals entering the labor market in the long run. Figure 4.10 below shows the possibility of simultaneously having backward-bending supply of labor by individuals and an upward sloping market supply of labor.

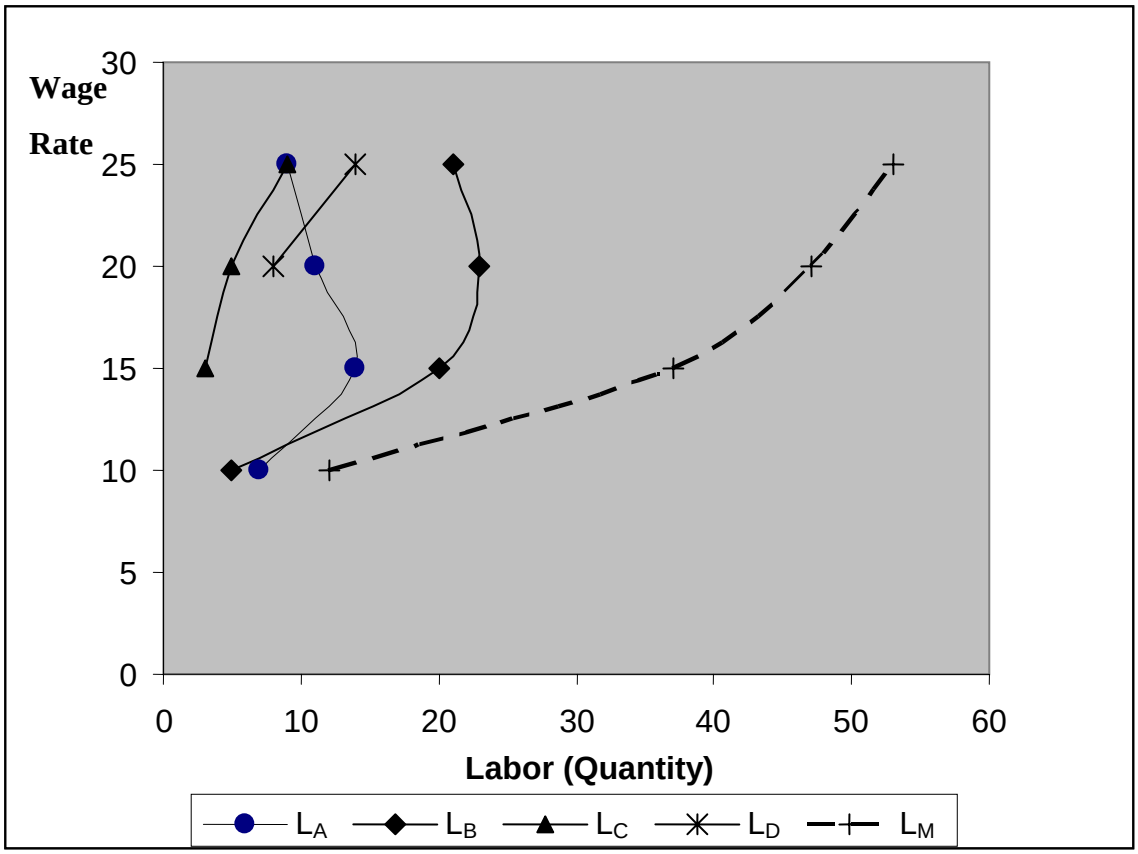


Figure 4.10: Backward Bending Individual Labor Supply Curves and an Upward Sloping Market Supply of Labor

Two of the four individuals (A and B) in the figure above have backward-bending supply curves. However, as the higher (rising) wage rate attracts new workers(C and D) to the market, the total quantity supplied of labor continues to increase with wage rate.

4.1.2.2 The Supply of Other Factors

The supply of an input other than labor, say that of a raw material or capita (an intermediate good), is derived on the same principles as the supply of any commodity. For example, the supply of capital is derived in much the same way as the supply of maize – both are determined by cost of production. That is, the supply of capital is the marginal cost of producing capital above the shutdown point. While the supply of an input to a single perfectly competitive firm is infinitely elastic, the market supply is not

perfectly elastic even. This is because even if a single buyer (a firm) is small to affect the price of the input, all firms taken together do affect the input price.

4.1.3 Factor Pricing

Given the market demand and the market supply of an input, its price is determined by the intersection of the two curves. The following figure depicts the equilibrium wage rate and quantity (employment level) of labor. The equilibrium price and quantity of any other resource is determined in the same way – by the intersection of demand and supply.

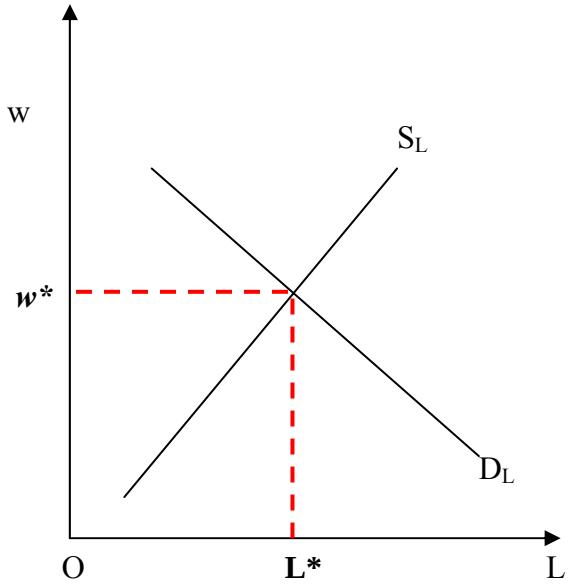


Figure 4.11: Input Price Determination

The equilibrium wage rate is w^* and the employment level is L^* (Figure 4.11). The difference between commodity pricing and factor pricing lies in how we derive the two components – supply and demand. While the demand for a final good is a direct demand (for consumption and utility), the demand for a factor is a derived demand. It is a derived demand in the sense that the demand for the services of the factor is based on the demand of the commodities in whose production the factor is used. On the supply side, the difference between supply of a commodity and factor supply is pronounced for labor. Unlike the supply of commodities, the supply of labor is not determined by the cost of production, but involves the attitudes of individuals toward work and leisure.

4.2 Factor Pricing in Imperfectly Competitive Markets

Throughout our discussion of Section 4.1, we assumed both product and factor markets to be perfectly competitive. In this section, we will relax this assumption of perfect competition. First, we will introduce imperfection to the product market keeping the factor market to be perfectly competitive (Sub-section 4.2.1). Secondly, we will consider the case where some degrees of imperfection characterize both the factor and product markets – where a firm has some power in the market for its product and in the market where it purchases productive factors – Sub-section 4.2.2. Lastly, we consider the case of bilateral monopoly, where a single seller faces a single buyer (Sub-section 4.2.3).

4.2.1 Monopolistic Power in the Product Market

The analysis in this sub-section relies on two basic assumptions:

1. The firm uses a single variable factor – labor – whose market is perfect. That is, the wage rate is given and thus an individual firm faces a perfectly elastic supply of labor.
2. The firm has monopolistic power (is not a price-taker) in the product market. This implies that the demand curve for the product of the firm is downward sloping. Consequently, the marginal revenue curve lies below the demand curve (i.e., MR < P) at all levels of output. This relationship between marginal revenue and price as well as its implication for the relationship between the value of marginal product (VMP) and marginal revenue product (MRP) of a factor are shown in Figure 4.12. We will proceed with the example of labor keeping in mind that a similar analysis holds for any other input. Recalling the definitions of the VMP_L (= $MPP_L.P_X$) and the MRP_L (= $MPP_L.TR_X$) from the previous section, we see that the VMPL > MRPL for the kind of firm we are assuming.

   That is, $MR_X < P_X$

   $$\Rightarrow MR_X \cdot MPP_L < P_X \cdot MPP_L$$

   $$\Rightarrow MRP_L < VMP_L$$

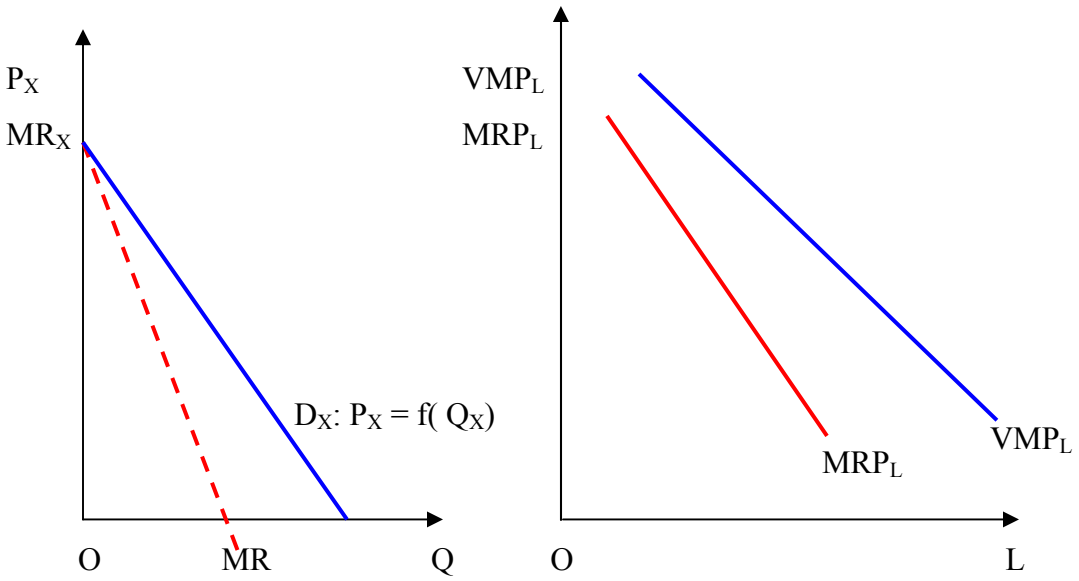


Figure 4.12: $MRP_L \neq VMP_L$ under Imperfectly Competitive Product Market

4.2.1.1 Demand of the Firm for a Single Variable Input

Consider a firm using one variable (labor, L) and one fixed (capital, K) factor of production to produce good X. The firm maximizes its profit with respect to the units of labor it employs.

Given the firm's demand function $P_X = f_1(Q_X)$ and its production function $Q_X = f_2(L, \bar{K})$, the profit function is defined as:

$$\pi = TR - TC$$

$$\Rightarrow \pi = P_X.Q_X - (\bar{w}.L + F)$$

The objective of the firm is: $\max_L \pi = P_X.Q_X - (\bar{w}.L + F)$.

F.O.C.: $\frac{d\pi}{dL} = 0. \Rightarrow \frac{d(P_X Q_X)}{dL} - \frac{d(\bar{w}.L + F)}{dL} = 0$

$$\Rightarrow MRP_L - \bar{w} = 0$$

$$\Rightarrow MRP_L = w$$

S.O.C.: $\frac{d^2\pi}{dL^2} < 0.$

Hence, the firm maximizes its profit by hiring to a point where the marginal revenue product of labor is equal to the wage rate. With reference to Figure 4.13 below, the firm employs L_1^e units of labor if the market wage rate is w_1. Similarly, L_2^e units will be hired at w_2 (and L_3^e units at w_3).

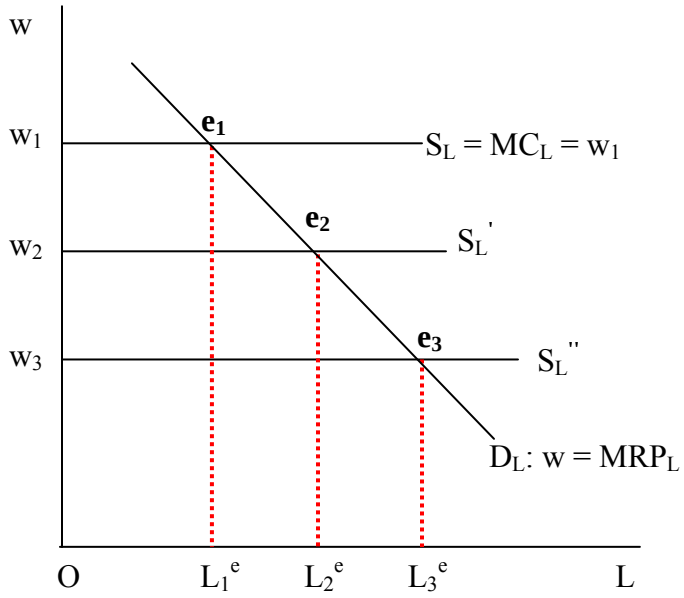


Figure 4.13: Equilibrium of a Firm and the Demand of the Firm for a Single Variable Input (Labor)

Joining the equilibrium points like $\mathbf{e_1}$, $\mathbf{e_2}$ and $\mathbf{e_3}$ (which correspond to different market wage rates) gives MRP_L as a demand curve that relates wage to labor employment.

4.2.1.2 Demand of a Firm for a Variable Factor When There Are Several Variable Factors

When two or more variable factors are used in the production process (i.e., in the long run), the demand for a variable factor is not its MRP curve. Nevertheless, it is formed from equilibrium points on shifting MRP curves. Figure 4.14 facilitates the derivation of the long run demand for an input (specifically for labor).

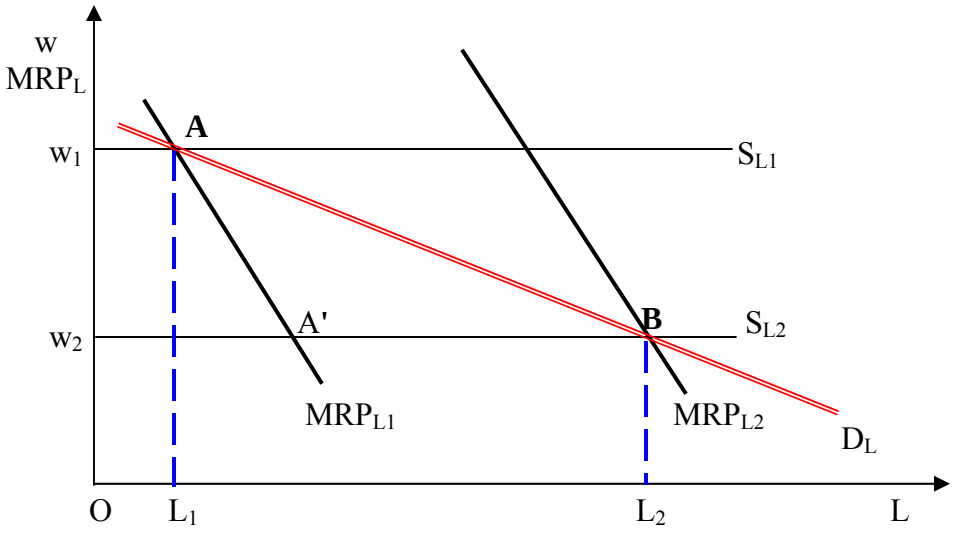


Figure 4.14: Demand of the Firm for a Variable Factor when there are Several Variable Factors

When the wage rate is w_1, the equilibrium of the firm is achieved at point A. If wage rate declines from w_1 to w_2, the firm would move from A to A' along MRP_{L1} if every thing remains constant. However, other things do not remain constant. As seen in the case of the long run demand of a perfectly competitive firm for labor (Section 4.1), the fall in wage rate has three effects: substitution, output and profit effects. The net result of these effects is a shift in MRP_L curve to the right leading to a new equilibrium at B. Then, the curve/line joining A and B in the Figure above is the firm's long run demand for labor.

4.2.1.3 Market Demand and Supply of Factors

The term monopolistic power does not necessarily imply a pure monopolist. A monopolistically competitive firm can influence the price of the product it sells and thus has a monopolistic power in the product market. So is an oligopolist.

The market demand for a factor is the summation of the demand of the individual firms (given that we have more than one firm). In the aggregation, however, we must take into account the shift of the individual demand curves as the price of the factor falls (due to the fall in the price of the final product). The derivation is similar to the one under the

case of perfect competition. The only difference is that, in this case, demand curves of the individual firms are based on MRP_i (and not on VMP_i). If each of the firms using the input is a pure monopolist (the only seller for its product), then the price of the final commodity (unique for each firm) is likely not to be affected, and in such cases the market demand curve is the simple horizontal summation of individual demand curves.

The market supply is not affected by the fact that firms have monopolistic power. This firm and the perfectly competitive firm in Section 4.1 are alike with regard to the resource supply they face. Both are price-takers in the input market; their difference is in the product market. Thus, the market supply of labor is the summation of the supply curves of individuals, as derived earlier. As in the previous case, a single firm faces a horizontal labor supply curve while the market supply of labor is upward sloping. So are the supplies of other factors (for a single firm and for the market).

4.2.1.4 Factor Pricing

The market price of the factor is determined by the intersection of the market demand and the market supply.

When a firm possesses a monopolistic power in the product market, the factor is paid its MRP_i, which is smaller than VMP_i (what the input could have been paid if this firm were a perfect competitor). This effect is called *monopolistic exploitation*. It represents the difference between the amount a factor is paid under perfect competition and the amount the same factor is paid under the imperfection introduced here.

If the two types of firms face the same market price for an input, the firm with monopoly power in the product market would hire less units of the input. Alternatively, if firms under the two scenarios have to use the same amount of labor (L_2 in panel (a) of Figure 4.15), the firm in perfect competition pays a wage rate of w_1 while the other firm pays w_2. The difference, $w_1 - w_2$, measures the level of monopolistic exploitation by the firm. The same concept is depicted in panel (b), but at the market level.

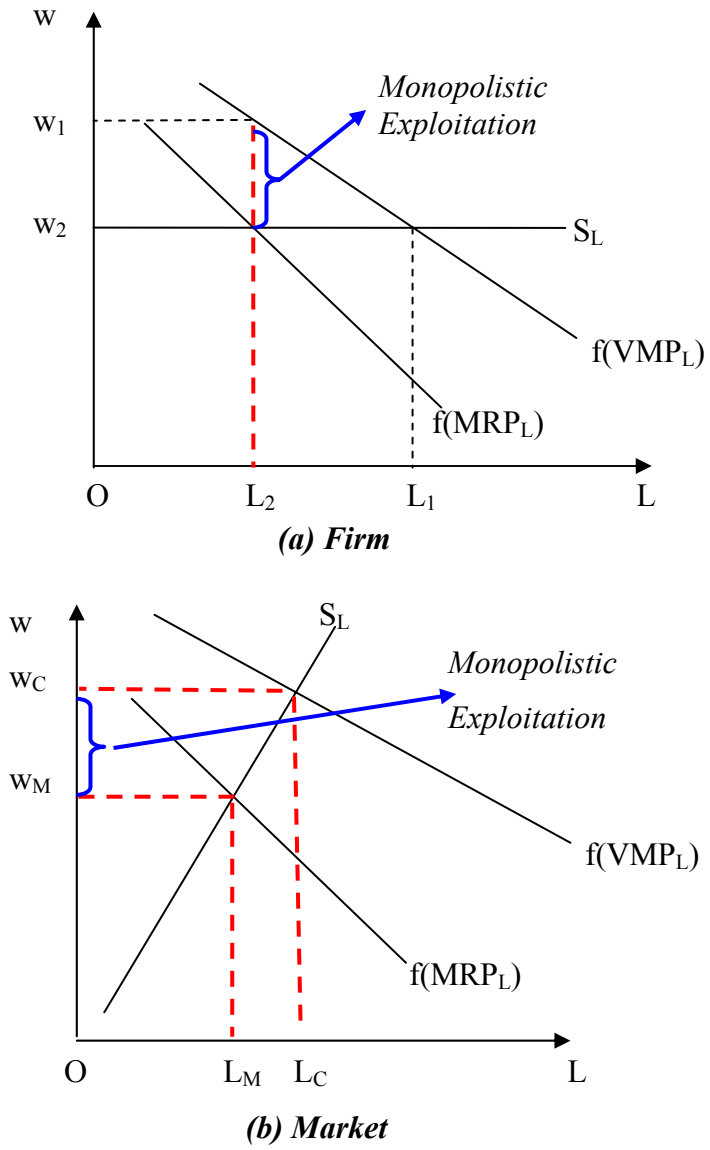


Figure 4.15: Monopolistic Exploitation at the (a) Firm's Level, and (b) Market Level

4.2.2 Monopsonistic Power in the Factor Market

So far, we considered factor pricing under two scenarios. In the first case, we took a firm that is a perfect competitor in both the product and factor markets. Then, we introduced imperfection to the product market (assuming the factor market to be perfectly competitive). Now, we are at a point to introduce imperfect to the factor (resource market). While we could have imperfection in the factor market along with a perfectly

competitive or an imperfect product market, here we will examine the case of a firm that has monopolistic power in the product market and monopsonistic power in the input market.

4.2.2.1 A Monopsonist Using a Single Variable Factor

Assuming the only variable input to be labor, the demand for labor by an individual firm that has a monopolistic power in the product market and a monopsonistic power in the factor market is given by MRP_L. That is, the firm in the second case seen above (Sub-section 4.2.1) and the kind of firm we are analyzing here are the same on the demand side: MRP_L is the short run demand for labor in both cases.

However, unlike the firms in the previous two scenarios, the supply of labor to the individual firm is not perfectly elastic as the firm is large in this case. Suppose that the firm is the only buyer of the input (a monopsonist). The supply of labor this firm faces has a positive slope: as the monopsonist expands the use of labor, it must pay a higher wage rate. The hiring decision of such a firm at a point in time significantly reduces the pool of labor force available in the market thereby increasing the scarcity (and thus the price) of an extra unit of labor.

The supply of labor shows the average expenditure or price that the monopsonist must pay at different levels of employment. Multiplying the price of the input by the level of employment gives the total expenditure of the monopsonist on the input (TE_L = w.L). Then, $AE_L = \frac{TE_L}{L} = \frac{wL}{L} = w.$ Note that *w* is not a constant in this case, but depends on (or a function of) the level of employment (L): w = f(L).

$\Rightarrow$ $w = AE_L = f(L)$ …… The supply of labor the monopsonist faces.

The slope of this labor supply function, dw/dL, is positive (i.e., $dw/dL > 0$).

The relevant magnitude for the equilibrium of the monopsonist is not the wage rate or the average expenditure on labor rather the marginal expenditure of purchasing an additional

unit of the factor. (Recall the general marginal expenditure marginal benefit criterion for profit maximization). The marginal expenditure on labor (ME_L) is derived as follows:

$$ME_L = ME_L = \frac{d(TE_L)}{dL} = \frac{d(wL)}{dL} = w\frac{dL}{dL} + L\frac{dw}{dL}.$$

$$\Rightarrow ME_L = w + L\frac{dw}{dL}.$$

Since $dw/dL > 0$, $L > 0$ and $w > 0$, it follows that the marginal expenditure on labor is greater than the supply of labor the monopsonist faces (i.e., $ME_L > w$) for any level of employment – refer Figure 4.16 below.

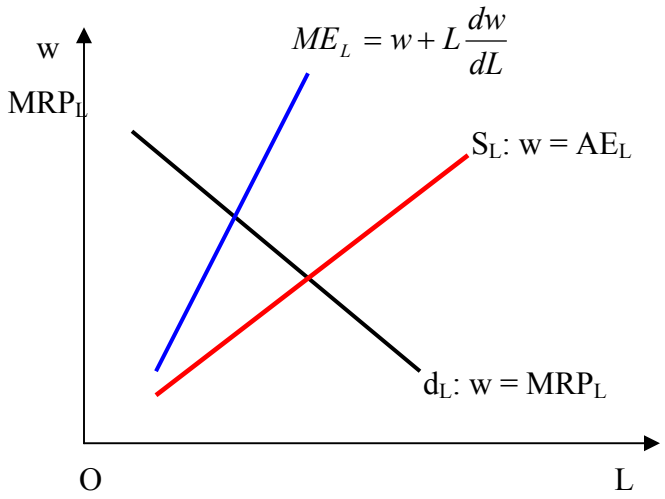


Figure 4.16: A Monopsonist's Demand, Supply, and Marginal Expenditure Curves of Labor When Labor Is the Only Variable Factor

Note also that the ME_L has a steeper slope than the supply curve ($w = S_L$). Assuming a linear supply function, let us prove that the MEL is steeper than the supply of labor curve.[8]

- The slope of the supply function, $w = f(L)$, is: $dw/dL > 0$.
- The slope of the ME_L curve is:

[8] Anyone who prefers to ignore mathematical derivations can safely bypass this proof without any loss of idea.

$$\frac{d(ME_L)}{dL} = \frac{d[w + L\frac{dw}{dL}]}{dL}$$

$$= \frac{dw}{dL} + \frac{d[L\frac{dw}{dL}]}{dL}$$

$$= \frac{dw}{dL} + L\frac{d^2w}{dL^2} + \frac{dw}{dL}.\frac{dL}{dL}$$

$$= 2\frac{dw}{dL} + L\frac{d^2w}{dL^2}$$

For a linear supply function, $\frac{dw}{dL}$ is a constant and consequently $\frac{d^2w}{dL^2} = 0$.

$$\Rightarrow \frac{d(ME_L)}{dL} = 2\frac{dw}{dL}.$$

(Slope of the ME_L curve = $2\frac{dw}{dL}$) > (Slope of the labor supply curve = $\frac{dw}{dL}$)!!!

Coming back to the equilibrium of the firm, the firm realizes the maximum profit when it equates the ME_L to its MRP_L as shown in Figure 4.17 below.

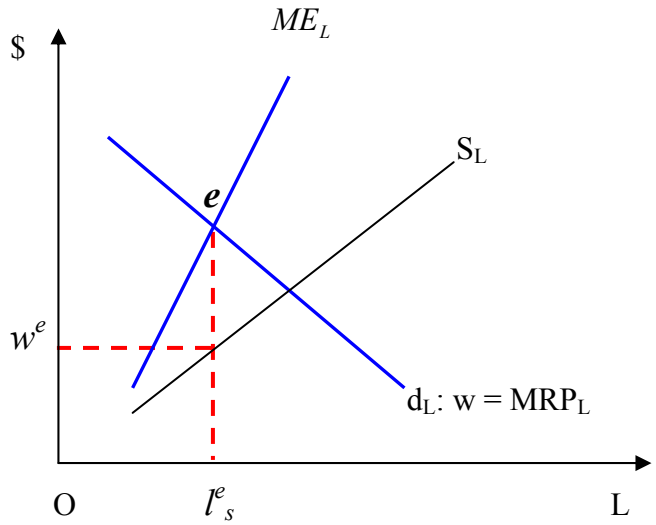


Figure 4.17: Equilibrium of a Monopsonist Using a Single Variable Factor

The firm maximizes its profit by employing l^e_s units of labor a level of employment corresponding to point ***e*** (where $ME_L = MRP_L$). The wage rate that the firm will pay for the l^e_s units of labor is w_e – defined on the labor supply curve.

The wage rate and the employment of labor in the current model are lower than that of perfect competition (discussed in Section 4.1) as well as that of monopoly market (discussed in Sub-section 4.2.1). Figure 4.18 below shows the wage rate and the level of employment in the three models.

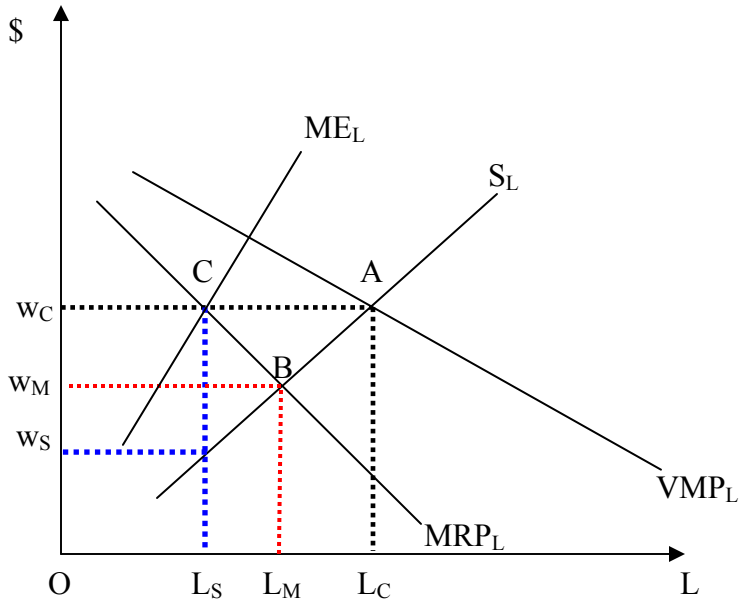


Figure 4.18: Monopsonistic Exploitation

In the figure:

- w_C = wage rate paid by a firm under perfectly competitive product and factor markets.
- w_M = wage rate paid by a firm under perfectly competitive factor market and imperfectly competitive product market.
- w_S = wage rate paid by a firm under imperfectly competitive product and factor markets.
- $w_C > w_M > w_S$ (and $L_C > L_M > L_S$).
- $w_C - w_M$ gives the level of monopolistic exploitation (discussed earlier).

- $w_C - w_S$ is the level of monopsonistic exploitation:
  - $w_C - w_M$ due to the monopolistic power of the firm in the product market, and
  - $w_M - w_S$ solely due to the monopsonistic power of the firm in the factor market.

4.2.2.2 A Monopsonist Using Several Variable Factors

If the input markets are perfectly competitive, a firm minimizes its cost (subject to an output constraint) or maximizes its production (subject to a cost constraint) by using the factor combination at which $\frac{MPP_L}{MPP_K} = \frac{w}{r}$ or $\frac{MPP_L}{w} = \frac{MPP_K}{r}$. If the factor markets are monopsonistic, however, changes in the amounts of factors employed cause changes in factor prices. That is, as the amount of labor (L) and/or capital (K) used in the production process change, both wage rate (w) and rental price of capital (r) do change. Thus, w and r are not constants.

Consequently, a monopsonist who uses several variable factors will use the input combination at which the ratio of the MPP to the ME is equal for all variable inputs. In our simple setting of two variable factors (L and K), the equilibrium of the firm is given by: $\frac{MPP_L}{ME_L} = \frac{MPP_K}{ME_K}$.

Proof: Suppose the demand for the product of the monopsonist is $P_X = f_1(Q_X)$, its production function is $Q_X = f_2(L, K)$ and the supply functions of the factors for the monopsonist are $w = f_3(L)$ and $r = f_4(K)$.

The firm maximizes its profit given by $\pi = P_X.Q_X - wL - rK$ with respect to L and K.

F.O.C:
$$\begin{cases} \frac{\partial \pi}{\partial L} = P_X \frac{\partial Q_X}{\partial L} + Q_X\left(\frac{\partial P_X}{\partial Q_X}.\frac{\partial Q_X}{\partial L}\right) - \left[w\frac{\partial L}{\partial L} + L\frac{\partial w}{\partial L}\right] = 0 \\ \frac{\partial \pi}{\partial K} = P_X \frac{\partial Q_X}{\partial K} + Q_X\left(\frac{\partial P_X}{\partial Q_X}.\frac{\partial Q_X}{\partial K}\right) - \left[r\frac{\partial K}{\partial K} + K\frac{\partial r}{\partial K}\right] = 0 \end{cases}$$

Rearranging, $\begin{cases} [P_X + Q_X \frac{\partial P_X}{\partial Q_X}].[\frac{\partial Q_X}{\partial L}] = w + L\frac{\partial w}{\partial L} \\ [P_X + Q_X \frac{\partial P_X}{\partial Q_X}].[\frac{\partial Q_X}{\partial K}] = r + K\frac{\partial r}{\partial K} \end{cases}$

Note that $\frac{\partial Q_X}{\partial L}$ is the MPP_L, and $w + L\frac{\partial w}{\partial L}$ is the ME_L. Similarly, $\frac{\partial Q_X}{\partial K} = MPP_K$, and $r + K\frac{\partial r}{\partial K} = ME_K$. besides, $P_X + Q_X \frac{\partial P_X}{\partial Q_X}$ is the MR_X. This definition of marginal revenue can be confirmed from the following derivation:

$$MR_X = \frac{d(TR_X)}{d(Q_X)} = \frac{d(P_X Q_X)}{d(Q_X)} = P_X \frac{d(Q_X)}{d(Q_X)} + Q_X \frac{d(P_X)}{d(Q_X)}$$

$$\Rightarrow MR_X = P_X + Q_X \frac{d(P_X)}{d(Q_X)}.$$

Substituting the expressions in the above first order condition equations by the terms MPP_L, MPP_K, ME_L, ME_K and MR_X, yields:

$$\begin{cases} MR_X.MPP_L = ME_L \quad \ldots\ldots(a) \\ MR_X.MPP_L = ME_K \quad \ldots\ldots(b) \end{cases}$$

Dividing (a) by (b) completes our proof:

$$\frac{MR_X.MPP_L}{MR_X.MPP_K} = \frac{ME_L}{ME_K}$$

$$\Rightarrow \frac{MPP_L}{MPP_K} = \frac{ME_L}{ME_K}, \text{or alternatively,} \frac{MPP_L}{ME_L} = \frac{MPP_K}{ME_K}.$$

4.2.3 Bilateral Monopoly

In our discussion so far, we dealt with the implications of the power that firms have in the product and/or input markets for factor prices and employment. We have never said a word on the power that input suppliers (particularly laborers) could have on factor prices. That is, we have not seen the possibility of market imperfection emanating from suppliers of resources. The discussion of bilateral monopoly in this sub-section introduces an

imperfection resulting from the market power of resource suppliers, specifically the role of labor union.

Bilateral monopoly arises when a single seller (a monopolist) faces a single buyer (a monopsonist). In this model, we assume that:

- ☞ all firms are organized in a single body that acts like a monopsonist (the only buyer of labor services); while
- ☞ all laborers are organized in a labor union that acts like a monopolist (a single seller of labor services).

Hence, bilateral monopoly is a model in which the participants are two monopolies, one on the supply side and one on the demand side. Strictly speaking, of course, the one on the demand side is a monopsonist.

The solution to a bilateral monopoly is generally indeterminate. That is, unlike the previous models we examined, market forces do not determine the equilibrium wage rate and the equilibrium level of employment. The model gives only the upper and lower limits within which the wage rate will be determined by bargaining. The outcome of the bargaining cannot be known with certainty. It will depend on the bargaining skills, the political and economic power of the labor union and the firms (or the firms' union), and on many other factors.

Let us clarify the situation of bilateral monopoly with the help of the following diagram (Figure 4.19).

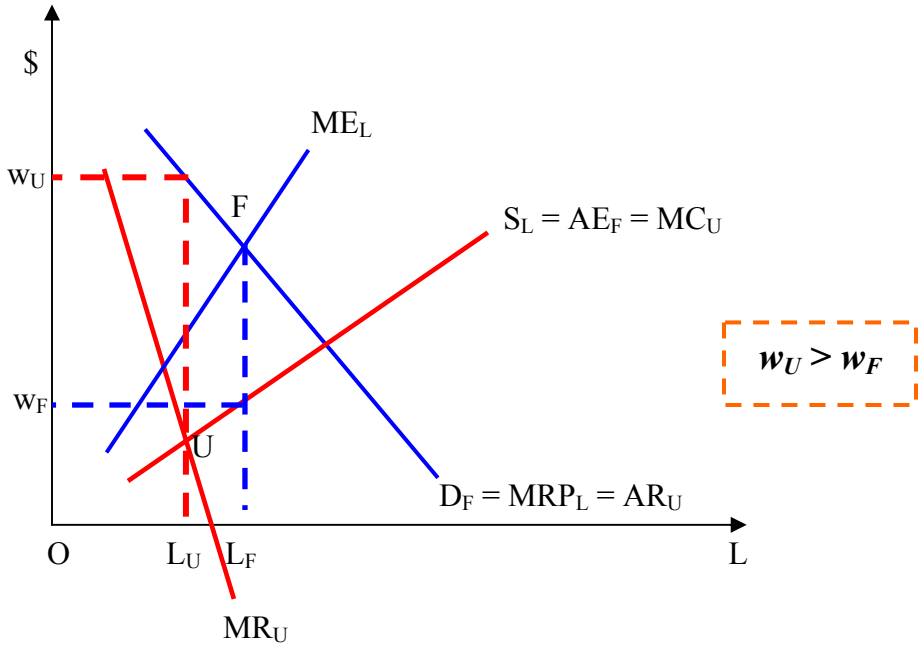


Figure 4.19: Bilateral Monopoly

The monopsonist's (the buyers') labor demand curve is D_F, which is given by the MRP_L. From the point of view of the monopolist (labor union), this curve represents its average revenue (AR_U) curve. Note that what a buyer of an item pays for a unit – the price or the demand function of the buyer – is the per unit receipt – average revenue of the seller. Then, the seller's (or the union's) marginal revenue (MR_U) curve is derived from this AR_U and lies below it.

The supply of labor facing the monopsonist is the upward sloping S_L curve. This shows the average expense (average cost) of the monopsonist on labor. Corresponding to this average cost curve is the marginal expenditure (ME_L) curve. From the point of view of the monopolist (labor union), the S_L may be considered as the marginal cost (MC_U) and the supply curve. This is done for pedagogical purposes and is accomplished by assuming that it behaves as if it were a perfectly competitive seller – it charges a single price for a given level of labor service.

The monopsonist maximizes its profit at point F, where its marginal expense on labor (ME_L) is equal to the marginal revenue product of labor (MRP_L). Thus, the monopsonist desires to hire L_F units of labor at a wage rate equal to w_F. If the buyer had a monopsonistic power but the laborers did not have any market power, the equilibrium would be at point F – with a wage rate of w_F and an employment level of L_F. This is exactly what we saw under Sub-section 4.2.2.1.

The monopolist (labor union), on the other hand, maximizes its gains (profits) at point U, where its marginal cost (MC_U) is equal to its marginal revenue (MR_U). Thus, the monopolist will want to supply L_U units of labor at a wage rate equal to w_U. That is, if the workers were united in a union and had a market power but the firms did not have any power, then the equilibrium would be at point U – with a wage rate equal to w_U and an employment level of L_U.

The price (wage rate) desired by the firms' union (the monopsonist) is the lower limit and that desired by the labor union (the monopolist) is the upper bound of the actual wage rate. The actual wage rate, which is determined by bargaining (negotiation), lies somewhere between w_F and w_U. Its exact location depends on the bargaining skills (power) of the two parties. The stronger the monopolist is, the closer the actual wage rate would be to w_U; and the stronger the monopsonist is, the closer the actual wage rate would be to w_F.

4.3 Elasticity of Factor Substitution, Technological Progress and Income Distribution

The subject matter of income distribution is the study of the determination of the shares of the factors of production in the total output produced in the economy over a given time period. Put differently, income distribution is concerned with how the value of the output produced with the help of different inputs (jointly) is shared among these various inputs.

If we assume that there are two factors of production – L and K – for simplicity, their shares are defined as:

- Share of labor = $\frac{w.L}{V}$, and
- Share of capital = $\frac{r.K}{V}$

Where:

w = wage rate

r = rental price of capital

L = quantity of labor employed

K = quantity of capital used

V = the value of the total output produced in the economy.

The relative factor share is conventionally defined as the ratio of the share of labor to that of capital. That is,

$$Relative\ Factor\ Share\ (RFSh) = \frac{Share\ of\ \ labor}{Share\ of\ capital} = \frac{wL/V}{rK/V}$$

$$\Rightarrow RFSh = \frac{wL}{rK}\ or\ RFSh = \frac{w/r}{K/L}.$$

The factor shares depend on the state of technology that defines the production function, the nature of technical progress and on the relative factor prices, i.e.,

$$\begin{bmatrix} Income \\ distribution \end{bmatrix} = f(\begin{bmatrix} production \\ function \end{bmatrix}, \begin{bmatrix} w \\ \overline{r} \end{bmatrix}, \begin{bmatrix} Technical \\ \Pr ogress \end{bmatrix})$$

There is a strong interrelationship among the three determinants of income distribution. For instance, technology and technical progress do affect the demand for and the supply of factors, which in turn, determine input prices.

In the following two sub-sections, we will look at the effects of changes in relative factor prices (for a given technology) and the effects of technological progress on the distributive share of factors.

4.3.1 Elasticity of Factor Substitution and the Shares of Factors of Production

As factor prices change, firms will substitute the cheaper input for the more expensive one (or the more expensive input for the cheaper one). This response/behavior of profit maximizing firms, induced by changes in factor prices, will result in a change of the K/L ratio. The change in the K/L ratio implies a change in the relative shares of the factors. The size of this effect (of change in factor prices on distributive share of factors) depends on the responsiveness of the change of the K/L ratio to the factor price changes. A measure of this responsiveness (of K/L ratio to changes in w/r ratio) is termed *the elasticity of factor substitution* (σ).

In general, the elasticity of factor substitution is defined as:

$$\sigma = \frac{percentage\ change in\ the\ K/L\ ratio}{percentage\ change\ in\ the\ MRTS_{LK}}$$

$$\sigma = \frac{d(K/L)/(K/L)}{d(MRTS_{LK})/(MRTS_{LK})} = \frac{d(K/L)}{d(MRTS_{LK})} \cdot \frac{MRTS_{LK}}{K/L}$$

In perfectly competitive input markets, the firm is in equilibrium when it chooses the input combination at which $MRTS_{LK} = w/r$. This reduces our definition of elasticity of factor substitution to:

$$\sigma = \frac{d(K/L)}{d(w/r)} \cdot \frac{(w/r)}{(K/L)}$$

Note that the sign of σ is always non-negative. Why? This is because the K/L ratio and w/r ratio move in the same direction. For instance, a rise in (w/r) ratio implies that labor becomes more expensive. As a result, firms will substitute capital for labor (labor by capital), and this raises the (K/L) ratio in production. This shows that (w/r) and (K/L) ratios are positively related, i.e., σ is non-negative.

- If $\sigma = 0$, then it is impossible to substitute labor for capital or vice versa. Such a situation is exemplified by the Leontief type production functions, where inputs are used in fixed proportions and cannot be substituted for one another.
- If $\sigma = \infty$, then there is a perfect substitutability between factors. An example of this is the linear production function, where one input can be traded for another at a constant rate.
- If $\sigma > 0$ (but finite), then factors of production are substitutes but only to a limited degree (extent). An example is the Cobb-Douglas type production functions (where $\sigma = 1$). In this case of finite positive elasticity, there are three categories of substitutability between factors:
  - $\sigma = 1$ implies a unitary substitutability: a percentage change in (w/r) ratio brings about a proportionate change in (K/L) ratio.
  - $\sigma < 1$ implies an inelastic substitutability: a percentage change in (w/r) ratio brings about a less than proportionate change in (K/L) ratio.
  - $\sigma > 1$ implies an elastic substitutability: a percentage change in (w/r) ratio brings about a more than proportionate change in (K/L) ratio.

There is an important relationship between the values of σ and the distributive shares of factors.

- If $\sigma < 1$, firms are not very sensitive to a change in relative factor prices and a given percentage change in (w/r) ratio results in a smaller percentage change in (K/L) ratio. Consequently, a rise in wage rate (in relative terms) will make laborers (when seen collectively) better off as the higher wage rate is beneficial

for them and as the undesirable effect of the rise in (K/L) is weaker. Thus, if there is an inelastic substitutability between labor and capital (i.e., if $\sigma < 1$), an increase in (w/r) ratio raises the relative (distributive) share of labor.

For instance, if $\sigma = 0.5$, a 1% increase in w/r ratio results in a 0.5% increase in K/L ratio, or a 10% increase in w/r ratio results in a 5% increase in K/L ratio. Let us illustrate what happens to the relative factor share using the 10% increase in w/r ratio that induces a 5% increase in K/L ratio. If $\frac{w/r}{K/L}$ is the old (before change) relative factor share (RFSh) and $(\frac{w/r}{K/L})^*$ is the relative factor share after the change in relative factor prices (RFSh*), then:

$$RFSh^* = \left[\frac{w/r}{K/L}\right]^* = \left[\frac{(w/r)(1+0.10)}{(K/L)(1+0.05)}\right]$$

$$\Rightarrow RFSh^* = \left[\frac{1.1(w/r)}{1.05(K/L)}\right] = \frac{22}{21}.[\frac{w/r}{K/L}]$$

$$\Rightarrow RFSh^* \approx 1.048(RFSh)$$

$$\Rightarrow RFSh^* > RFSh.$$

The last line shows that the new relative factor share is greater than the initial relative factor share, or equivalently, the relative factor share rises with a rise in the relative factor price (w/r).

☞ If $\sigma > 1$, a given percentage change in w/r ratio results in a more than proportionate percentage change in K/L ratio, so that the relative share of labor decreases.

For instance, if $\sigma = 2$, then a 10% increase in w/r ratio results in a 20% increase in K/L ratio. Consequently:

$$RFSh^* = \left[\frac{w/r}{K/L}\right]^* = \left[\frac{(w/r)(1+0.10)}{(K/L)(1+0.20)}\right]$$

$$\Rightarrow RFSh^* = \left[\frac{1.1(w/r)}{1.2(K/L)}\right] = \frac{11}{12}.\left[\frac{w/r}{K/L}\right]$$

$$\Rightarrow RFSh^* \approx 0.917(RFSh)$$

$$\Rightarrow RFSh^* < RFSh.$$

The last line shows that the new relative factor share is less than the initial relative factor share, or equivalently, the relative factor share falls as the relative factor price (w/r) rises.

☞ If $\sigma = 1$, a given percentage change in w/r ratio results in an equal percentage change in K/L ratio, so that the relative share of labor remains unchanged.

In general, an increase in the w/r ratio will cause the share of labor (relative to that of capital) to:

1) increase if $\sigma < 1$;
2) decrease if $\sigma > 1$; and
3) remain the same if $\sigma = 1$.

A decrease in w/r ratio will have the opposite effects: reduces the relative share of labor if $\sigma < 1$, decreases it if $\sigma > 1$, and leaves unaffected if $\sigma = 1$.

4.3.2 Technological Progress and Income Distribution

Under Sub-section 4.3.1, we assumed a given production function. In reality, however, technological change takes place continuously and this will cause shifts in the production function, leading to changes in the K/L ratio and the elasticity of substitution. Technological progress makes an isoquant representing a given level of output shift downwards, implying that the same level of output can be produced with smaller quantities of inputs.

Technological progress can be:

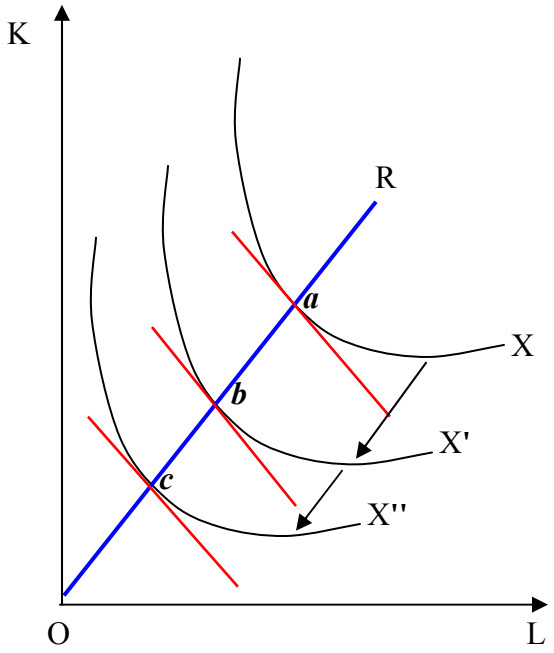
- ☞ neutral,
- ☞ labor deepening (capital-saving), or
- ☞ capital deepening (labor-saving).

1. Technological progress is neutral if, at a constant K/L ratio, the $MRTS_{LK}$ remains unchanged. Since $MRTS_{LK}$ = w/r at equilibrium, it follows that when technological progress is neutral both the K/L ratio and w/r ratio are unchanged. Consequently, the relative factor shares remain unchanged.

2. Technological progress is capital deepening if, at a constant K/L ratio, the $MRTS_{LK}$ declines. This takes place if technological progress raises the productivity of capital (MP_K) more than that of labor (MP_L). That is, $\frac{MP_L}{MP_K}$ decreases causing the $MRTS_{LK}$ to fall. [Slope of the shifting isoquant decreases in absolute value]. This implies that at equilibrium the w/r ratio declines, that is w declines relative to r, while the K/L ratio remains the same. Consequently, the ratio of factor shares declines: RFSh* < RFSh. This is tantamount to saying that a capital deepening technological progress causes the share of labor to decrease and that of capital to increase (in relative terms).

3. Technological progress is labor deepening if, at a constant K/L ratio, the $MRTS_{LK}$ increases. This occurs if technological progress increases the productivity of labor (MP_L) more than that of capital (MP_K). Meaning, the $\frac{MP_L}{MP_K}$ increases, causing the $MRTS_{LK}$ to rise. [Slope of the shifting isoquant increases in absolute value]. This implies that at equilibrium the w/r ratio increases while K/L ratio remains the same. Consequently, the ratio of factor shares increases: RFSh* > RFSh. Equivalently, the relative share of labor increases and that of capital decreases.

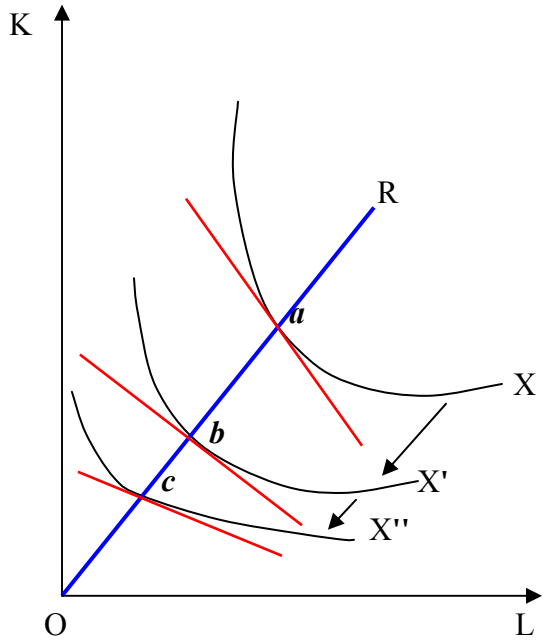
Figure 4.20 sheds light on our discussion of the three types of technological progress.



(a) Neutral Technological Progress:

/Slope of X at a/ = /Slope of X' at b/

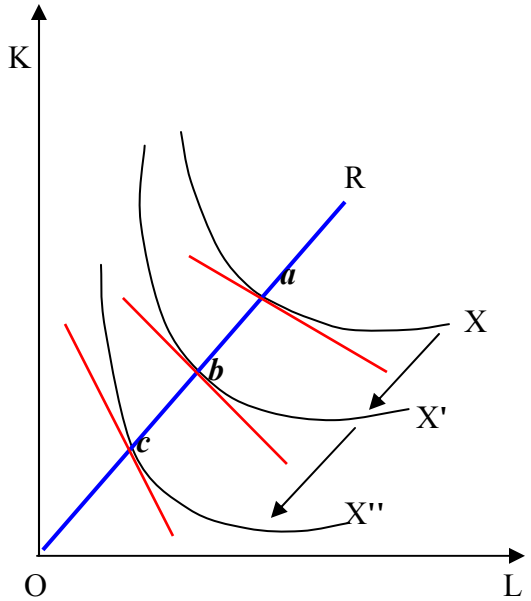
= /Slope of X'' at c/



(b) Capital Deepening Technological Progress:

/Slope of X at a/ > /Slope of X' at b/

> /Slope of X" at c/



(c) Labor Deepening Technological Progress:

/Slope of X at a/ < /Slope of X' at b/ < /Slope of X'' at c/

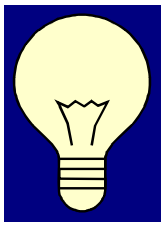
In each of the three cases (panels):

- OR is a ray whose slope shows a constant K/L ratio.
- Points a, b and c show points of production at the given constant K/L ratio as technological progress takes place.
- X, X' and X" represent the same amount of output (under different states of technology).

Figure 4.20: Different Types of Technological Progress

In summary:

- ☞ The relative share of labor increases and that of capital decreases if technological progress is labor deepening;
- ☞ The relative share of labor decreases and that of capital increases if technological progress is capital deepening; and
- ☞ The relative shares of both labor and capital remain unchanged if technological progress is neutral.



4.4 Summary

1. *According to the marginal concept, a profit-maximizing firm will continue to hire an input as long as the extra income (receipt) from the sale of the output produced by the input (MRP of the input) is larger than the extra cost of hiring the input (ME on the input) and to the point* $MRP_i = ME_i$. *This general rule can be modified based on the structure of the factor and product markets. If the firm is a perfect competitor in the product market,* MRP_i *is equal to and can be substituted by the value of marginal product* (VMP_i). *Perfect competition in the input market allows us to substitute ME by input price.*
2. *The demand for an input is a derived demand – derived from the demand for the final commodities that the input is used in producing. For a firm that is a perfect competitor in the product market, its demand curve for an input is either the same as the* VMP_i (= MRP_i) *curve (assuming a single variable factor) or a curve derived from equilibrium points on shifting* MRP_i *curves (in case of several variable factors). In the latter case, substitution, output and profit-maximizing effects of a change in an input's price take a central position in deriving the demand for the input. The market demand for an input is found by summing the individual firms' demand for the input, but the effect of change in input price on commodity price and its repercussion should be considered. Determinants of demand for a productive factor include the input's own price, its marginal*

physical product (MPP_i), the price of final commodities in whose production the input is used, the amount and price of complementary and substitutable inputs, and technological progress.

3. *The supply of labor by an individual depends principally on the individual's preferences for leisure and work (income). As wage rate increases, an individual's opportunity cost of leisure rises and the individual will substitute work for leisure (substitution effect). However, at the same time the individual's income will rise with the increase in wage rate, and the rise in income will increase the consumer's demand for all normal goods (including leisure) – income effect. These two opposing effects operate over the entire wage rate ranges, and the substitution effect overwhelms the income effect at lower wage rates while the negative income effect overwhelms the positive substitution effect at higher wages implying a backward bending supply of labor curve. The market supply of labor curve is not usually backward bending as higher wage rates attract new workers into the market and because of population growth. The supply of an input (other than labor) is derived on the same principles as the supply of a final commodity. The market supply curve of an input is generally upward sloping. On the other hand, the supply curve of an input a firm faces is perfectly elastic if the firm is a perfect competitor in the factor market and upward sloping if the firm has some monopsony power.*
4. *Under perfectly competitive product and factor markets, an input is paid its VMP. When firms have some monopolistic and/or monopsonistic power in the product and/or factor markets, an input is paid less than its VMP and it is said to be exploited. Imperfection in the product market alone results in monopolistic exploitation, and imperfection in both markets results in monopsonistic exploitation. The latter is greater than former. In a bilateral monopoly, where a single seller faces a single buyer, the market solution is indeterminate and input price (and employment) is determined by bargaining of the two parties.*
5. *The shares of factors of production in the total output produced in an economy over a given period of time (income distribution) depend on the production*

function, (relative) factor prices and technological progress. For a given production function, the relative share of a factor (say, labor) falls if its relative price (w/r ratio in case of two factors) rises and if firms are very sensitive in substituting capital for labor, i.e., if there is elastic substitutability between factors ($\sigma > 1$). It rises with rise in w/r ratio if $\sigma < 1$, and remains unchanged if $\sigma = 1$. Technological progress takes place continuously and shifts isoquants downwards. Considering the effects of technological progress on $MRTS_{LK}$ (MP_L/MP_K ratio) at a constant K/L ratio, technological progress can be categorized into neutral, labor deepening and capital deepening. A labor deepening technological progress is the one that increases the marginal productivity of labor more than that of capital (or equivalently, the $MRTS_{LK}$) at constant K/L ratio. This increases the relative share of labor, ceteris paribus. Capital deepening technological progress can be described in the same fashion – it increases $MRTS_{KL}$ (or equivalently reduces $MRTS_{LK}$) at constant K/L ratio and results in increased share of capital, ceteris paribus. Neutral technological progress, which changes the marginal productivities of both labor and capital by the same proportion, leaves the $MRTS_{LK}$ and the relative shares of factors unaffected.



4.5 Review Questions

I. Multiple-Choice Questions

1. The extra benefit or receipt from the sale of the output produced by one additional unit of input is:
   A. Marginal revenue
   B. Marginal physical product of the input
   C. Marginal revenue product of the input
   D. Value of marginal product of the input
   E. Factor price
2. Assuming a firm using several variable factors, what happens to the marginal productivity of labor for a fall in wage rate at a given level of employment if the firm does not want to change the units of output it produces?
   A. Rises
   B. Remains constant
   C. Falls
   D. Any one of the three may happen
3. Which of the following does not necessarily hold true for a profit-maximizing firm in perfectly competitive product and factor markets?
   A. $MRP_L = w$ at equilibrium
   B. $VMP_L = w$ at equilibrium
   C. MRP_L is the demand for labor curve by the firm
   D. $MRP_L = ME_L$ at equilibrium
4. All but one of the following changes cause the demand for capital (curve) of a firm to shift:
   A. Changes in MPP_L
   B. Technological progress
   C. Fall or rise in price of capital
   D. The price of the final commodity in whose production the input is used

E. All

F. None

5. Labor deepening technological progress:
   A. is labor-saving or capital-using.
   B. increases the $MRTS_{KL}$ at a constant K/L ratio.
   C. increases the productivity of labor relative to that of capital.
   D. is desirable for countries with low population growth rate.
   E. All except A.
6. Which of the following is not true?
   A. It is impossible to have a backward bending supply of labor by an individual and a market supply of labor curve which is not backward bending, at the same time
   B. The supply of capital is derived in a way more similar to the derivation of the supply of wheat than that of labor.
   C. Once we derive the market demand for and the market supply of an input, factor pricing is done exactly in the same way as commodity prices are determined.
   D. The relationship between MRP_L and VMP_L are different for perfect competition and pure monopoly, like that between MR and commodity price.
   E. None of the above.
7. In a bilateral monopoly model:
   A. there is no equilibrium wage rate determined by the model.
   B. the monopsonist's desire sets the upper limit of actual wage rate.
   C. the monopolist's desire represents the upper limit of the actual wage rate.
   D. the monopolist determines the lower limit of the wage rate.
   E. A and C.

Answer the next two questions (8 and 9) based on the following information:

The production function of a firm is $Q = 100L^{0.5}K^{0.5}$, K = 100, and the per unit price of the product P = 2 Birr. The firm has monopsonistic power in the labor market and hence it faces the labor supply function w = 4L.

8. The quantity of labor that the firm should hire and the wage rate it should pay in order to maximize its profit are _____ units and _____ Birr, respectively.
   A. 25 and 25
   B. 40 and 158
   C. 25 and 100
   D. 40 and 160
   E. None
9. If the total fixed cost of the firm is 200 Birr, the maximum profit of the firm will be:
   A. 10,000
   B. 2,700
   C. 7,500
   D. 7,300
   E. None
10. A firm with monopolistic power in the product market exploits a laborer more than a firm with monopolistic power in the product market and monopsonistic power in the labor market.
   A. True
   B. False
   C. I can't say
11. Which of the following is correct?
   A. $VMP_L = MRP_L$ if the firm has monopolistic power in the product market and monopsonistic power in the factor market
   B. $VMP_L = MRP_L$ if the firm has no monopolistic power in the product market but has monopsonistic power in the factor market
   C. $VMP_L = MRP_L$ if the firm has no monopolistic power in the product market and no monopsonistic power in the factor market
   D. $VMP_L = MRP_L$ always holds true
   E. $VMP_L = MRP_L$ never holds true

12. A certain firm has monopolistic power in its product market but purchases labor (the only variable input used in the production process) from a perfectly competitive labor market. Which of the following is necessarily true?
    A. The firm's demand for labor is the same as its MRP_L.
    B. The firm is recommended to employ labor till $VMP_L = MRP_L$.
    C. The firm maximizes its profit if VMP_L = wage rate.
    D. The firm maximize its profit if MPP_L = wage rate.
    E. None of the above.
13. If the elasticity of substitution between labor and capital (which are used in production) is greater than unity and wage rate to interest rate ratio increases, then:
    A. the shares of both labor and capital increase.
    B. the shares of both labor and capital decrease.
    C. the share of labor increases but that of capital decreases.
    D. the share of capital increases but that of labor decreases.
    E. None of the above.
14. The XYZ Company is a monopolist for a new, patented food supplement. If the demand for the product is P = 25 – 2Q, and the short-run production function is given by Q = 4L, then the firm's demand for labor can be written as:
    A. w = 100 – 4L.
    B. w = 25 – 4L.
    C. w = 100 – 64L.
    D. w = 25 – 8L.
15. A perfectly competitive firm that hires labor from a perfectly competitive labor market will hire labor until the marginal physical product of labor equals:
    A. w (= nominal wage rate)
    B. P (= output price)
    C. w x P
    D. w/P

16. Suppose $MPP_L = 10/L$. If the firm sells its output in a competitive market at a price of \$8, how much labor will the firm hire if the wage rate is \$5 per unit of labor?
    A. 4 units of labor
    B. 6.25 units of labor
    C. 10 units of labor
    D. 16 units of labor
17. The labor supply curve of an individual slopes upward if:
    A. the income effect is negative.
    B. the substitution effect is positive.
    C. the income effect is larger in magnitude than the substitution effect.
    D. the substitution effect is larger in magnitude than the income effect.
18. If the firm hires to a point where the marginal expense of labor is greater than the marginal revenue product of labor, then:
    A. profits could be increased by increasing employment.
    B. profits could be increased by reducing employment.
    C. profits are maximized.
    D. total cost must be greater than total revenue.
19. If hiring one more worker increases labor costs by \$120 and output by 5 units, and if the firm is a price-taker and output sells for \$25 per unit, then the marginal revenue product of labor is:
    A. \$25
    B. \$100
    C. \$125
    D. \$120
    E. None of the above.
20. If firm X is a monopsonist, then:
    A. it is the only seller of a resource.
    B. it is the only seller of a product.
    C. its marginal expense of hiring labor exceeds the wage rate it pays.
    D. it can cut the wage as it expands its work force.

II. Discussion Questions

21. Why is MRP_L curve not the demand for labor curve when there are several variable factors of production? Explain!
22. Under perfectly competitive product and factor markets, the market demand curve for an input is not the simple horizontal summation of the individual demand curves for the input by individual firms. Do you agree? Argue for or against!
23. Explain each of the following concepts with the help of suitable diagram(s).
    a. Monopolistic exploitation
    b. Monopsonistic exploitation
    c. Bilateral monopoly
24. Discuss the relationship between:
    a. the distributive share of factors and the elasticity of factor substitution, for a given technology of production.
    b. the distributive share of factors and technological progress, for a given K/L ratio.
25. Suppose a firm has monopolistic power in the product market but is a perfect competitor in the labor market. Derive the long run demand for labor curve when the price of labor *increases*. Discuss also the substitution, output and profit effects of the wage increase, graphically.

CHAPTER FIVE
GENERAL EQUILIBRIUM ANALYSIS AND WELFARE ECONOMICS

Introduction

In the previous chapters, we have examined the behavior of individual decision-making units. In the first course of microeconomics, the behavior of individuals (as consumers of commodities and suppliers of inputs) and that of firms (as employers of inputs and producers of commodities) were covered. In the first three chapters of this material, we considered the operation of the market for a single commodity: how a monopolistically competitive firm or an oligopolist makes price and/or quantity decisions. In the fourth chapter, we discussed the workings of individual input markets under various market structures. , missing from all these presentations is the examination of how the various individual pieces fit together to form an integrated economic system.

In this chapter, we take up the topic of interdependence or relationship among the various decision-making units and markets in the economy. This allows us to trace both the effect of a change in any part of the economic system on every other part of the system, and the repercussions from the latter on the former.

Chapter Objectives

After completing this chapter, students would be able to:

- Distinguish between partial and general equilibrium analyses and the conditions under which each type of analysis is appropriate.
- Discuss the conditions required for an economy to be in general equilibrium of: (i) production, (ii) consumption (exchange), and (iii) production and exchange simultaneously.

- Examine the welfare implications of the conditions required for the attainment of general equilibrium.
- Understand the various social decision mechanisms and their implications for social welfare maximization.
- Describe social welfare maximization in relation to the concepts of envy, equity and fair allocation.

5.1 General Equilibrium Analysis

5.1.1 Partial versus General Equilibrium Analysis

Partial equilibrium analysis studies the behavior of an individual decision-making unit and a particular market, viewed in isolation. Examples of such an analysis include the study of:

- how an individual maximizes satisfaction subject to his/her income constraint,
- how a firm minimizes its costs of production and maximizes profits under various market structures,
- how the price and employment of each type of input is determined.

By choosing to undertake a partial equilibrium analysis, one has abstracted from all the interconnections that exist between the market under study and the rest of the economy. In short, the basic characteristic of a partial equilibrium approach is the determination of the price and quantity in a market with the help of demand and supply curves drawn on the ceteris paribus assumption. Only when an industry (or a market) is small and has few direct links with the rest of the economy is partial equilibrium analysis appropriate.

However, a fundamental feature of any economic system is the interdependence among its constituent parts. The markets of all commodities and all productive factors are interrelated, and the prices in all markets are simultaneously determined. For example, consumers' demands for various goods and services depend on their tastes and incomes. In turn, consumers' incomes depend on the amounts of resources they own and factor

prices. Factor prices depend on the demand and supply of the various inputs. The demand of firms for factors depends not only on the state of technology but also on the demands for the final goods they produce. The demands for these goods depend on consumers' income. The message is that there is circular interdependence of activities within an economic system. Figure 5.1 is a sketch of this interdependence among economic activities.

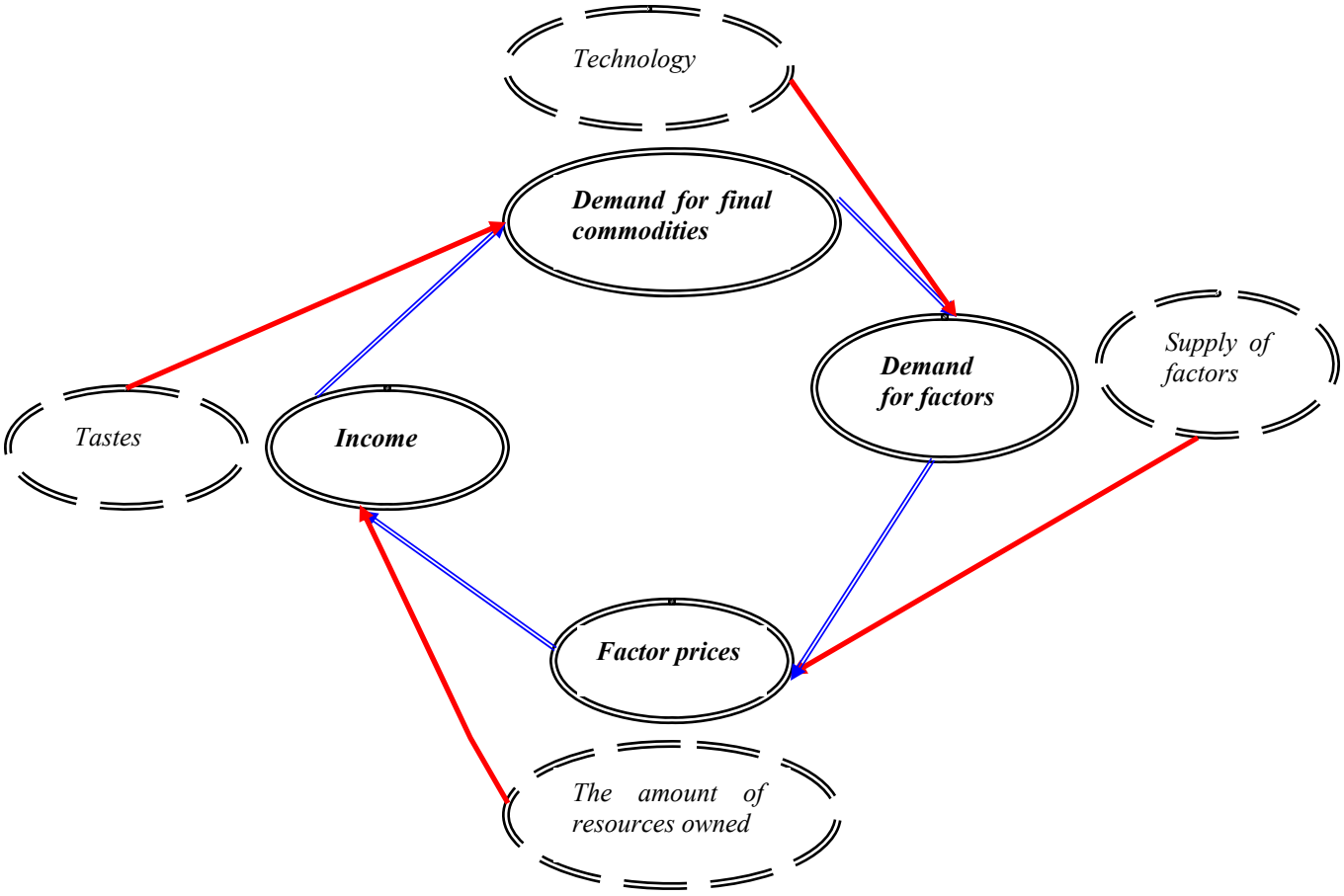


Figure 5.1: A Sketch of the Interdependence of Activities in an Economy

Such interrelated economic variables and multi-directional effects are studied in general equilibrium analysis. General equilibrium studies the interdependence or interconnections that exist among all markets and prices in the economy and attempts to give a complete, explicit, and simultaneous answer to the basic economic questions of *what, how and for whom to produce*. Despite its coverage and comprehensiveness, general equilibrium analysis is not used all the time (in all situations) because dealing with each and all

industries in the economy at the same time – by its very nature – is very difficult, time consuming and expensive.

General equilibrium is a state in which all markets and all decision-making units are in simultaneous equilibrium. A general equilibrium exists if each market is cleared at a positive price, with each consumer maximizing satisfaction and each firm maximizing profit.

The first and simplest general equilibrium model was introduced in 1874 by Leon Walras. In the Walrasian system, all prices and quantities in all markets are determined simultaneously through their interaction with one another. That is, a general equilibrium of an economy is given by the solution to a system of equations. Thus, the first task in establishing the existence of a general equilibrium is to describe the economy by means of a system of equations, in which the number of unknowns should be equal to the number of independent equations.

In the Walrasian system, one of the equations is a redundant (not an independent) equation and deprives the system of a solution since the number of independent equations is less than the number of unknowns. As a result, in this model, the absolute level of prices cannot be determined. In response to this, general equilibrium theorists have adopted the device of arbitrarily choosing the price of one commodity as a numeraire (or unit of account) and expressing all other prices in terms of the price of the numeraire. With this device, each price is given relative to the price of the numeraire.

Even if there is equality of independent equations and unknowns, there is no guarantee that a general equilibrium solution exists. Walras was never able to prove the existence of a general equilibrium. In 1954, Arrow and Debreu provided a proof of the existence of a general equilibrium in perfectly competitive markets, in which there are no indivisibilities and no increasing returns to scale. In 1971, Debreu and Hahn proved the existence of a general equilibrium for an economy with limited increasing returns and monopolistic

competition, without indivisibilities. There is no proof so far that shows the existence of a general equilibrium in oligopoly and monopoly markets.

Apart from the *existence* problem, two other problems are associated with general equilibrium: the problem of its *uniqueness* and the problem of its *stability*. In principle, all markets should clear at positive prices and quantities (existence), and any deviation from a unique equilibrium should bring the economy back to the equilibrium (stability).

5.1.2 General Equilibrium in a Two-Factor, Two-Commodity, Two-Consumer (2x2x2) Economy

To facilitate the discussion and understandability of a general equilibrium analysis, let us construct a simple model/abstraction. Examine each of the following assumptions of (the 2x2x2 model) carefully. You will be expected to appreciate the role of each assumption in the subsequent sub-sections.

1. There are two factors of production (L and K), whose quantities are given exogenously. Both of these factors are homogeneous and perfectly divisible. Homogeneity indicates that all units of labor are identical, and so are the units of capital. Perfect divisibility of factors means that any factor of production could be supplied and used in whatever fraction desired.
2. Only two commodities (X and Y) are produced. The technology of production is given. That is, production at the currently existing technology is considered. The isoquant maps have the usual properties (smooth and convex to the origin implying diminishing marginal rate of technical substitution, MRTS, between factors along any isoquant). Each production function exhibits constant returns to scale. The two production functions are independent (no externalities in production). That is, the production process of good X does not influence that of Y, and vice versa. [More on externalities will be said in the next chapter.]
3. There are two consumers in the economy (A and B) whose preferences are represented by ordinal indifference curves, which are convex to the origin,

exhibiting diminishing marginal rate of substitution, MRS, between commodities. Consumers' choices are independent (no externalities in consumption).

4. The goal of each firm is profit maximization and that of each consumer is utility maximization.
5. There is full employment of factors of production, and all incomes received by their owners (A and B) are spent. Note an implicit assumption in this assumption: the factors of production are owned by the consumers.
6. There is perfect competition in both commodity and factor markets. Thus, both consumers and firms face the same set of prices (P_X, P_Y, w, r).

In this model, a general equilibrium is reached when:

- the four markets (two commodity and two factor markets) are cleared at a set of equilibrium prices (P_X, P_Y, w, r); and
- each economic agent (two firms and two consumers) is simultaneously in equilibrium.

Put differently, three static properties are observed in a general equilibrium solution, reached with a freely competitive market:

- efficient allocation of resources between firms or goods (production equilibrium);
- efficient distribution of commodities between consumers (consumption equilibrium); and
- efficient combination of products (simultaneous equilibrium of production and consumption).

5.1.2.1 Equilibrium of Production (Efficiency in Factor Allocation)

Recall from the previous chapter that a perfectly competitive firm maximizes its profit at a point where the marginal rate of technical substitution between factors equals the input price ratio (i.e., when $MRTS_{LK}^{X} = w/r$). Also, remember that $MRTS_{LK}$ is the slope of an isoquant in absolute terms. Now, we will extend this same condition to the simultaneous analysis of two firms, each trying to maximize its own profit.

The joint equilibrium of production of the two firms in our simple model can be derived by the use of the Edgeworth box of production.

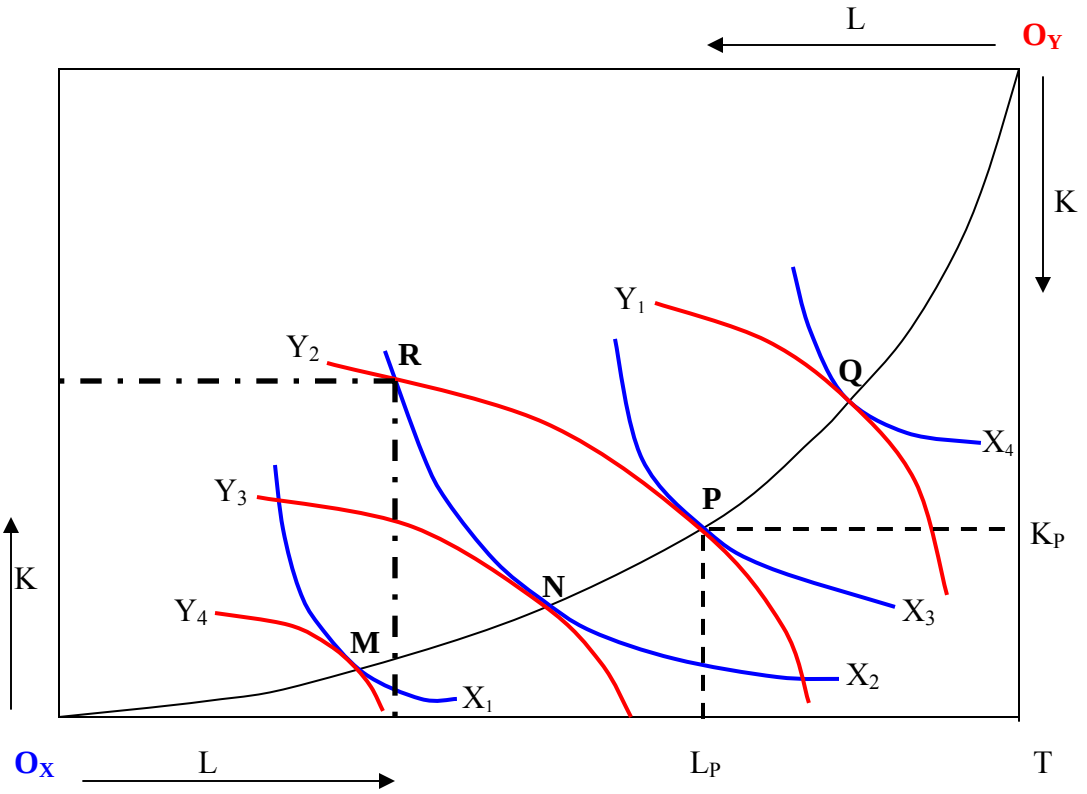


Figure 5.2: Edgeworth Box of Production

With reference to the box in Figure 5.2,

- the size of the box refers to the total amount of L and K available to the economy.
- Any point inside the box indicates how the total amount of each input is utilized in the production of the two commodities.
- O_X is the origin for the production of commodity X. A movement away from O_X (in the northeast direction) implies successively increasing amounts of X ($X_1 < X_2 < X_3 < X_4$, and so on).
- O_Y is the origin for the production of commodity Y. A movement away from O_Y (in the southwest direction) implies successively increasing amounts of Y ($Y_1 < Y_2 < Y_3 < Y_4$, and so on).

If this economy was initially at point R, it would not be maximizing its output of commodities X and Y because, at point R, $MRTS_{LK}^{X} \neq MRTS_{LK}^{Y}$. As the rate at which labor and capital can be substituted for each other differ for the production of the two commodities, this economy is not making the best out of its resources (at point R). The economy can move from point R to point N and increase its output of Y (from Y_2 to Y_3) without reducing its output of X (from X_2). Alternatively, it can move to point P thereby increasing its output of X (from X_2 to X_3) without reducing its output of Y. At points M, N, P and Q, an X-isoquant is tangent to a Y-isoquant (X and Y isoquants have the same slopes) so that $MRTS_{LK}^{X} = MRTS_{LK}^{Y}$.

Curve O_XMNPQO_Y is the *Edgeworth Contract Curve of Production*. It is the locus of tangency points of the isoquants for X and Y (or the locus of points at which $MRTS_{LK}^{X} = MRTS_{LK}^{Y}$). Any point on this contract curve is Pareto efficient (optimal) in the sense that the production of one of these goods cannot be increased without reducing that of another.

The general equilibrium of production occurs at a point that satisfies this Pareto optimality criterion of efficiency in factor substitution. Since this occurs at any point along the Edgeworth contract curve of production, there is an infinite number of potential Pareto-optimal production equilibria. However, with perfect competition, one of these equilibria will be realized, and this equilibrium is the one at which $MRTS_{LK}^{X} = MRTS_{LK}^{Y} = w/r$. If, for instance, this happens at point P in the diagram above, then the economy will produce X_3 units of X and Y_2 units of Y. Correspondingly, O_XL_P units of labor and TK_P units of capital will be used in the production of X while L_PT units of labor and O_YK_P units of capital will be used in the production of Y.

The above discussion is about the production side only. This would enable us to define the equilibrium of production in the input space. However, in a general equilibrium, the amounts of X and Y which maximize the profits of firms must be equal to those which consumers want to buy in order to maximize their utility. Besides, consumers decide their

purchases based on commodity prices, P_X and P_Y. Thus, in order to bring together the production side and the consumption side of the system, we must define the equilibrium of the firms in the product space. This is what we do next.

From each point on the Edgeworth contract curve of production in Figure 5.2, we can read off the maximum obtainable quantity of one commodity, given the quantity of the other. For instance, the economy can produce a maximum of X_2 units of X if it produces Y_3 units of Y (point N in Figure 5.2 above or point N' in Figure 5.3 below). Similarly, it can produce a maximum of X_1 units of X for Y_4 units of Y (point M in Figure 5.2 or point M' in Figure 5.3); zero units of X if all resources are devoted to the production of Y (point O_X in Figure 5.2 above or point O_X' in Figure 5.3).

The locus of all the Pareto-efficient outputs (or the locus of all the maximum attainable combinations of the two commodities, given the resource endowment $\overline{K}$ and $\overline{L}$, and the state of technology) is the production possibilities frontier/curve (PPF/PPC).

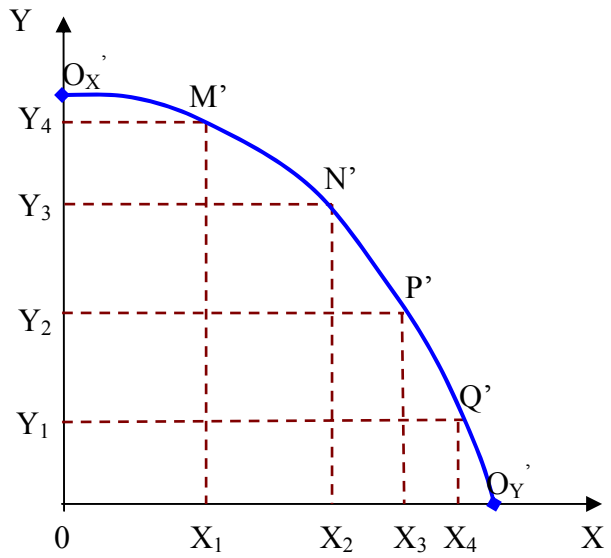


Figure 5.3: Production Possibilities Frontier/Curve

At any point on the curve, all factors are optimally (efficiently) employed. Any point inside the curve is technically inefficient, implying unemployed resources. Points above the curve are unattainable with the currently existing resources and technology. They

could be made attainable only if additional resources or a new technology or both are found.

The PPF is also called *the product transformation curve* because it shows how a commodity is transformed into another, by transferring some factors from the production of one commodity to the other. The negative of the slope of the PPF is called the marginal rate of product transformation ($MRPT_{XY}$) and it shows the amount of Y that must be sacrificed in order to obtain an additional unit of X, keeping efficiency in production. That is, $MRPT_{XY} = -\frac{dY}{dX}$.

When our simple economy is in equilibrium of production, $MRPT_{XY} = \frac{MC_X}{MC_Y} = \frac{P_X}{P_Y}$.

Proof:

- $MC_X = \frac{d(TC_X)}{dX}$ and $MC_Y = \frac{d(TC_Y)}{dY}$ by definition of marginal cost.

$$\Rightarrow \frac{MC_X}{MC_Y} = \frac{d(TC_X)}{dX} \div \frac{d(TC_Y)}{dY}$$

$$\Rightarrow \frac{MC_X}{MC_Y} = \frac{d(TC_X)}{dX} \times \frac{dY}{d(TC_Y)}$$

$$\Rightarrow \frac{MC_X}{MC_Y} = \frac{d(TC_X)}{d(TC_Y)} \times \frac{dY}{dX} \quad \text{.....(1)}$$

- Because $TC_X = w.L_X + r.K_X$ and $TC_Y = w.L_Y + r.K_Y$, and w and r are given,

$$d(TC_X) = w.d(L_X) + r.d(K_X) \quad \text{.....(2)}$$

$$d(TC_Y) = w.d(L_Y) + r.d(K_Y) \quad \text{.....(3)}$$

- Substituting (2) and (3) into (1) gives,

$$\frac{MC_X}{MC_Y} = \frac{d(TC_X)}{d(TC_Y)} \times \frac{dY}{dX} = \frac{w.d(L_X) + r.d(K_X)}{w.d(L_Y) + r.d(K_Y)} \quad \text{.....(4)}$$

- Movement on a given PPF (or full employment of resources) requires that the amount of factors released from the decrease in the production of commodity Y

must be equal to the amount of factors absorbed by the increase in the production of X. That is,

$$d(L_X) = -d(L_Y) \text{ and } d(K_X) = -d(K_Y) \quad \text{.....(5)}$$

☞ Substituting (5) into (4) yields,

$$\frac{MC_X}{MC_Y} = \frac{w.d(L_X) + r.d(K_X)}{w.d(L_Y) + r.d(K_Y)} \cdot \frac{dY}{dX} = \frac{w.[-d(L_Y)] + r.[-d(K_Y)]}{w.d(L_Y) + r.d(K_Y)} \cdot \frac{dY}{dX}$$

$$\Rightarrow \frac{MC_X}{MC_Y} = \frac{-[w.d(L_Y) + r.d(K_Y)]}{w.d(L_Y) + r.d(K_Y)} \cdot \frac{dY}{dX} = -\frac{dY}{dX} = MRPT_{XY}$$

$$\Rightarrow \frac{MC_X}{MC_Y} = MRPT_{XY}.$$

☞ Finally, in perfect competition, the profit-maximizing producer equates the price of the commodity it produces to its marginal cost of production: $MC_X = P_X$ and $MC_Y = P_Y$.

$$\Rightarrow MRPT_{XY} = \frac{MC_X}{MC_Y} = \frac{P_X}{P_Y}.$$

Assuming that the market prices of the commodities define the slope of line AB (in Figure 5.4 below), the general equilibrium of production is given by point T. The two firms are in equilibrium producing the levels of output given by $\boldsymbol{X_e}$ and $\boldsymbol{Y_e}$.

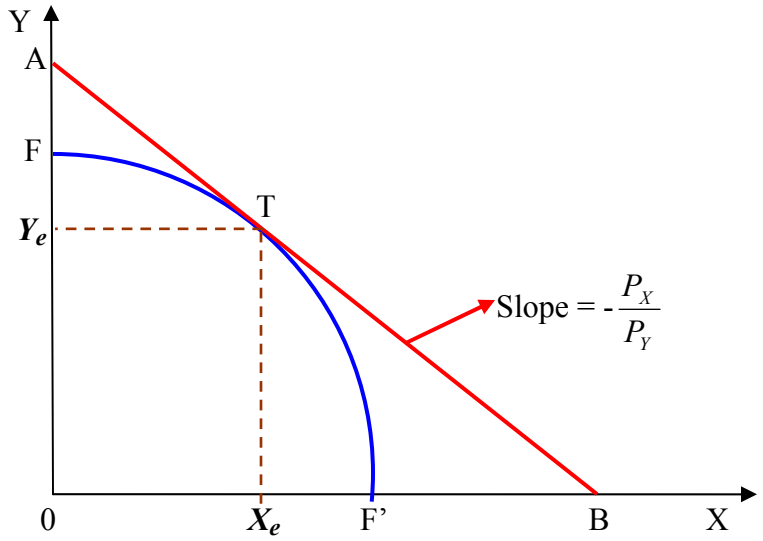


Figure 5.4: Equilibrium of Production

5.1.2.2 Equilibrium of Consumption (Efficiency in Distribution of Commodities)

The joint equilibrium of exchange (consumption) of the two consumers in our simple model can be derived by making use of the Edgeworth box of consumption. Any point in the Edgeworth box of consumption shows six variables: the combination of the two commodities produced (X, Y), and the distribution of each between the two consumers (X_A, X_B, Y_A and Y_B).

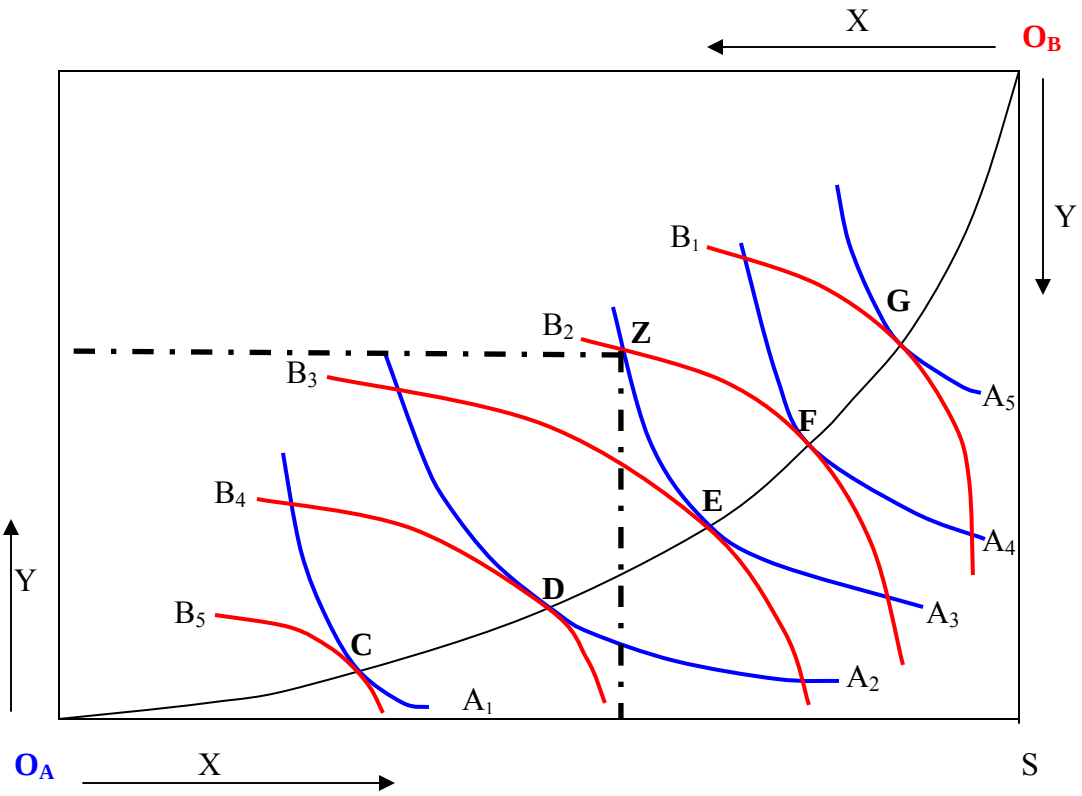


Figure 5.5: Edgeworth Box of Consumption

Given the total quantities of X and Y to be distributed between A and B, there will be an infinite number of Pareto optimal equilibria of distribution. A Pareto-efficient distribution of commodities is one such that it is impossible to increase the utility of one consumer without reducing the utility of the other. However, not all distributions (all points in the Edgeworth box) are Pareto-optimal. Only points of tangency of the indifference curves of the two consumers represent Pareto-efficient distributions. The locus of these points is

called the *Edgeworth Contract Curve of Consumption.* At each point on this curve, $MRS_{XY}^{A} = MRS_{XY}^{B}$.

Any point off the Edgeworth contract curve of consumption (for example point Z) is a point of inefficient consumption because it is possible to increase the utility of at least one consumer without making the other worse off. (Consider movement from Z to E, or to F, or to any point between E and F on the contract curve).

Of the infinite number of points where the condition $MRS_{XY}^{A} = MRS_{XY}^{B}$ holds, only one point is realized in perfect competition. This point is where $MRS_{XY}^{A} = MRS_{XY}^{B} = \frac{P_X}{P_Y}$.

5.1.2.3 Simultaneous Equilibrium of Production and Consumption (Efficiency in Product-Mix)

The general equilibrium of the system as a whole requires the fulfillment of the condition: $MRPT_{XY} = MRS_{XY}^{A} = MRS_{XY}^{B}$. That is, the rate at which the consumers are willing to exchange one good for another (MRS_{XY}) must be equal to the rate at which a good can be transformed into another in production ($MRPT_{XY}$). This is tantamount to saying that the general equilibrium of the system requires that the production sector's plans are consistent with the consumer (household) sector's plans.

In perfect completion, this condition is satisfied since:

- $MRPT_{XY} = \frac{P_X}{P_Y}$ (equilibrium of production);
- $MRS_{XY}^{A} = MRS_{XY}^{B} = \frac{P_X}{P_Y}$ (equilibrium of consumption); and
- it logically follows the two points above that $MRPT_{XY} = MRS_{XY}^{A} = MRS_{XY}^{B} = \frac{P_X}{P_Y}$.

Figure 5.6 below shows the general equilibrium of our simple (*2x2x2*) economy. The economy is in equilibrium (of production) producing X_e units of X and Y_e units of Y. this point (given by point T in the figure) is defined by the equality of the slope of the PPF (= $MRPT_{XY}$) and the commodity price ratio ($\frac{P_X}{P_Y}$).

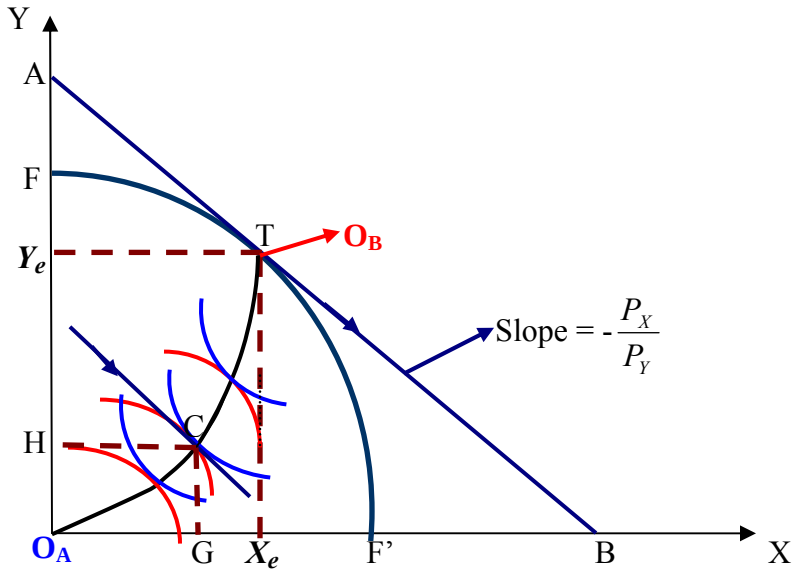


Figure 5.6: Simultaneous Equilibrium of Production and Consumption

Given the X_e and Y_e amounts of the commodities, we can draw the Edgeworth box of consumption with the dimensions O_AY_e and O_AX_e. That is, the amounts of X and Y produced define the size of the Edgeworth box of consumption. Then, the consumers will have an efficient distribution of X and Y between them along the contract curve joining the origins O_A and O_B. The unique equilibrium (of consumption) will be at a point of tangency of the X and Y isoquants (i.e., at a point on the contract curve) with the common slope equal to the slope of the PPF at point T. Suppose this point is point C in the figure. Then, individual A maximizes his/her utility by consuming O_AG units of X and O_AH units of Y; and individual B maximizes his/her utility, consuming GX_e units of X and HY_e units of Y.

Any divergence between MRS_{XY} and $MRPT_{XY}$ implies disequilibrium of the economy. To illustrate, assume that $MRPT_{XY} = 2$ while $MRS_{XY} = 1$ in a hypothetical economy. This

information is interpreted as follows. The economy can produce two units of Y by sacrificing one unit of X (or should sacrifice two units of Y to produce an additional unit of X), while the consumers are willing to exchange one unit of X for one unit of Y. This means firms produce a larger quantity of X and a smaller quantity of Y relative to the preferences of (or the value attached to the two goods by) the consumers – see Figure 5.7 below.

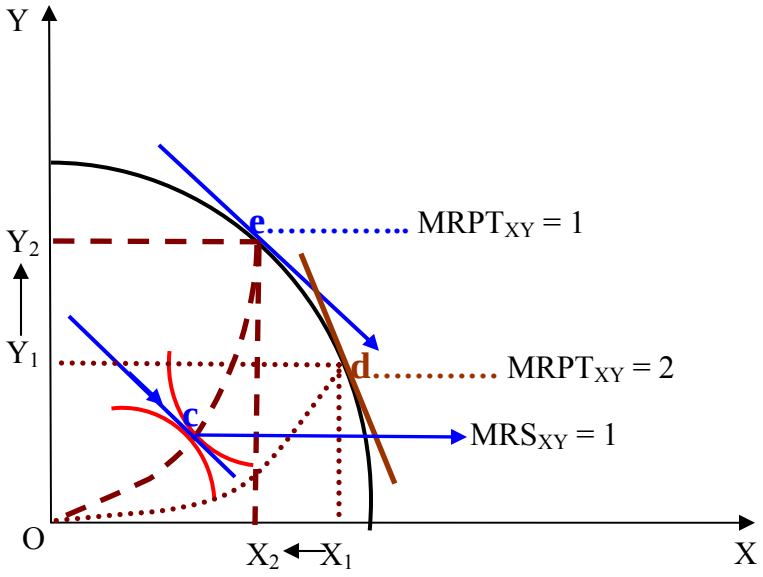


Figure 5.7: Divergence between MRS_{XY} and $MRPT_{XY}$, and Attainment of General Equilibrium

The current situation of the economy is that production is taking place at point ***d*** on the PPF ($MRPT_{XY} = 2$) and equilibrium of consumption is at point ***c*** ($MRS^{A}_{XY} = MRS^{B}_{XY} = 1$). The consumption bundle (being at point ***c***) is off the contract curve of consumption that corresponds to production at point ***d***. producers are placing more value on X than consumers and producing greater units of X than desired. Therefore, firms must reduce the production of X (from X_1 to X_2) and increase that of Y (from Y_1 to Y_2) for the attainment of general equilibrium – movement on the PPF from point ***d*** to point ***e***. With this movement and the production equilibrium established at point ***e***, the producers' and consumers' valuations of the two commodities will be equalized and point ***c*** will become Pareto optimal (a point on a contract curve of consumption).

Now, we have found the total amounts of X and Y the hypothetical economy produces (X_e and Y_e) and how these quantities are divided between the two consumers, A and B. What is next? After determining the equilibrium quantities of the inputs and the commodities, the next step is to determine the prices (P_X, P_Y, w and r).

From the Walrasian general equilibrium framework, we have the following equations:

$$MRTS_{LK}^{X} = \frac{w}{r} = MRTS_{LK}^{Y} \quad \text{.....(1)}$$

$$w = MPP_{LX}.P_X = MPP_{LY}.P_Y \quad \text{.....(2)}$$

$$r = MPP_{KX}.P_X = MPP_{KY}.P_Y \quad \text{.....(3)}$$

$$MRS_{YX}^{A} = MRS_{YX}^{B} = \frac{P_Y}{P_X} \quad \text{.....(4)}$$

Dividing equation (2) by equation (3) gives:

$$\frac{w}{r} = \frac{(MPP_{LX}.P_X)}{(MPP_{KX}.P_X)} = \frac{(MPP_{LY}.P_Y)}{(MPP_{KY}.P_Y)}. \Rightarrow \frac{w}{r} = \frac{MPP_{LX}}{MPP_{KX}} = MRTS_{LK}^{X} = MRTS_{LK}^{Y}$$

which is identical to equation (1). That is, the ratio of equation (2) to equation (3) is identical with equation (1). This implies that there are only 3 (not 4) independent equations with four unknowns. Thus, there will be no unique solution to the above system of equations.

As pointed out earlier in the chapter, we choose one of the prices as a numeraire (unit of account) and express the other prices in terms of (relative to) the numeraire. Let the numeraire be P_X. Then, we will express the other prices (P_Y, w, and r) relative to P_X. as the first step to this end, we solve for the three prices of interest (P_Y, w, and r) from the above equations:

- From equation (1), $w = r.MRTS_{LK}$ .....(5)
- From equation (3), $r = MPP_{KX}.P_X$ .....(6)
- Substituting equation (6) into equation (5) gives:

$$w = MPP_{KX}.P_X..MRTS_{LK} \quad \text{.....(7)}$$

- From equation (4), $P_Y = P_X.MRS_{YX}$ .....(8)

Next, the relative prices are derived from equations (6), (7) and (8) as follows:

- $\frac{r}{P_X} = MPP_{KX}$
- $\frac{w}{P_X} = MPP_{KX}.MRTS_{LK}$ or equivalently, $\frac{w}{P_X} = MPP_{KX}.\frac{MPP_{LX}}{MPP_{KX}} = MPP_{LX}.$
- $\frac{P_Y}{P_X} = MRS_{YX}.$

Note that the change of numeraire leaves the relative prices unaffected.

At its present stage, general equilibrium theory is largely non-operational and unrealistic. Then, why do we need to study it? The following two points present the rationale for studying general equilibrium:

- By viewing the economy as a vast system of mutually interdependent markets, it makes the student aware of the complexity of the real world.
- Moreover, the solution that exists under certain assumptions and its optimality properties can be used as a norm to judge the significance and implications of deviations of the various markets from this ideal state of equilibrium.

5.2 Welfare Economics

What we saw under the previous section (Section 5.1) describes the economic efficiency in production and consumption. An economy in a general equilibrium is characterized by Pareto efficiency in the sense that no one can be made better off without (at the same time) making somebody else worse off. However, an allocation in which an individual (a consumer) gets all the units of the commodities produced and the other gets nothing could be Pareto efficient. That is, the general equilibrium does not care about how fair the outputs produced are shared between/among consumers; it ignores the issue of equity. This section takes up the issue of equity: fairness in the distribution of output produced in an economy.

5.2.1 Definition of Welfare Economics

Welfare economics studies the conditions under which the solution to the general equilibrium model can be said to be socially optimal. It examines the conditions for economic efficiency in the production of output and the exchange of commodities and *equity in the distribution of income*. This definition of welfare economics points out that the maximization of the society's well-being requires not only efficiency in production and exchange but also equity in the distribution of income.

The fact that perfect competition leads to economic efficiency (Pareto optimality in production and exchange) proves Adam Smith's famous *law of the invisible hand*. Smith's law postulates that in a free market economy, each individual by pursuing his/her own selfish interests is led, as if by an invisible hand, to promote the well-being of society more so than he/she intends or even understands. This law leads to the *first theorem of welfare economics*, which postulates that an equilibrium produced by competitive markets exhausts all possible gains from exchange, or that equilibrium in competitive markets is Pareto optimal.

There is also a *second theorem of welfare economics*. This theorem postulates that when indifference curves are convex to their origins, every efficient allocation (every point on the contract curve for exchange) is a competitive equilibrium for some initial allocation of goods or distribution of inputs (income). The significance of the second welfare theorem is that the issue of equity in distribution is logically separable from the issue of efficiency in allocation. This means that whatever redistribution of income that society wants would lead to the exhaustion of all possible gains from exchange under perfect competition. Therefore, Pareto optimality does not imply equity. Society can use taxes and subsidies to achieve what it considers as a more equitable distribution of income.

Under perfect competition, any redistribution of resources or income would lead to a Pareto optimum equilibrium of exchange. The redistribution, however, may discourage work and accumulation of wealth. As a result, there is usually trade-off between

efficiency and equity. As an example, the incentive to accumulate wealth is generally low in a socialist (command) economic system as compared to in a market economy. To the contrary, a command economy is generally better in terms of equitable income distribution. This indicates that any redistribution of income (for the purpose of welfare maximization) should be done in a way that does not bring about scarification of efficiency (or a way implying a minimal loss of efficiency). Put differently, the objective of welfare maximization should be sought (strived for) along with, and should not be at the expense of, efficiency.

5.2.2 The Utility Possibilities Frontier (UPF) and the Grand Utility Possibilities Frontier (GUPF)

The utility possibilities frontier/curve shows the various combinations of utilities received by individuals A and B (i.e., U_A and U_B) when our simple economy is in general equilibrium or Pareto optimum in exchange. Alternatively, it is the locus of the maximum utility for one individual for any given level of utility for the other individual.

The general equilibrium of our simple economy, depicted in Figure 5.6 earlier, is reproduced below (as Figure 5.8) to make our discussion of UPF (UPC) more convenient.

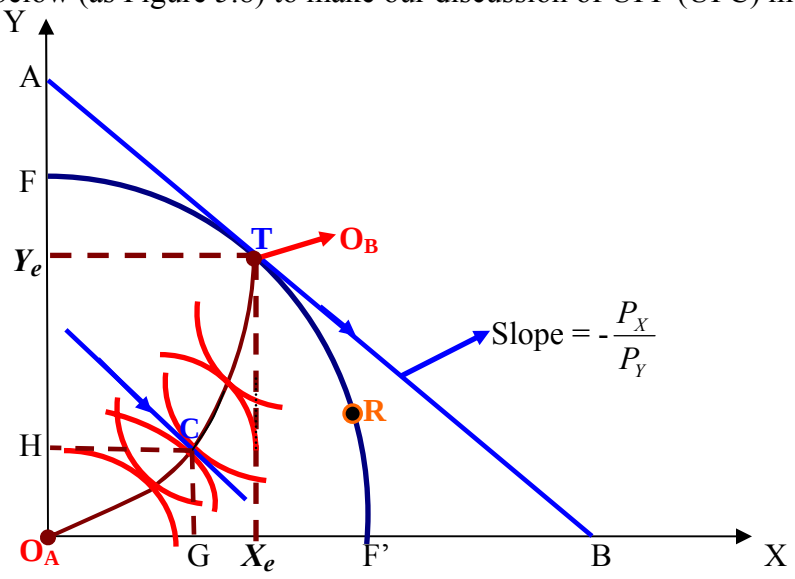


Figure 5.8: General Equilibrium of the Simple 2x2x2 Economy (Reproduced)

The utility possibilities frontier is obtained by mapping a consumption contract curve in Figure 5.8 above from the output or commodity space to utility space. For the equilibrium of production at point T, we have a corresponding consumption contract curve shown as O_ACO_B in the figure above.

At point O_A on the contact curve, consumer A consumes zero (0) units of each good and all that is produced goes to B. The exact opposite occurs at point O_B: consumer B gets nothing and all that is produced is consumed by A. Thus, the utility of A is zero and that of B is at its maximum at point O_A, while the utility of B is zero and that of A is maximized at O_B. In the utility space (portrayed in Figure 5.9 below), these two points (O_A and O_B) are represented by points $\mathbf{O_A'}$ and $\mathbf{O_B'}$, respectively. A movement away from the origin O_A towards O_B (Figure 5.8) implies increase in utility of consumer A and decrease in utility of consumer B. This indicates that the utility levels of the two consumers are inversely related: as the utility of A rises that of B falls and vice versa. This yields a downward sloping UPF in Figure 5.9 below.

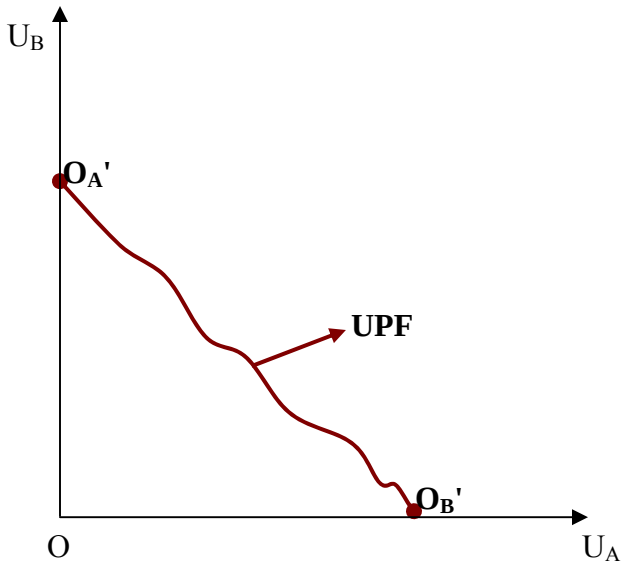


Figure 5.9: The Utility Possibilities Frontier

The utility possibilities frontier $\mathbf{O_A'O_B'}$ in Figure 5.9 shows the various combinations of utilities received by individuals A and B (i.e., U_A and U_B) when the economy composed

of the two individuals is in general equilibrium of exchange. Every point on this UPF has a corresponding point from the contract curve of consumption (O_ACO_B) in Figure 5.8.

Note that this contract curve (and therefore the UPF derived from it) is derived for the equilibrium in production given by T in Figure 5.8. In general, a UPF is associated with Pareto efficient allocations given an equilibrium of production (i.e., given a point on the PPF where $MRPT_{XY} = \frac{P_X}{P_Y}$). This means that we have another UPF for any other point on the PPF. Hence, we can have as many UPFs as there are points on the PPF.

Grand Utility Possibilities Frontier (GUPF) is the envelope of all the UPFs associated with points of Pareto efficiency (optimality) in production and exchange. Figure 5.10 shows two utility possibilities frontiers and the grand utility possibilities frontier.

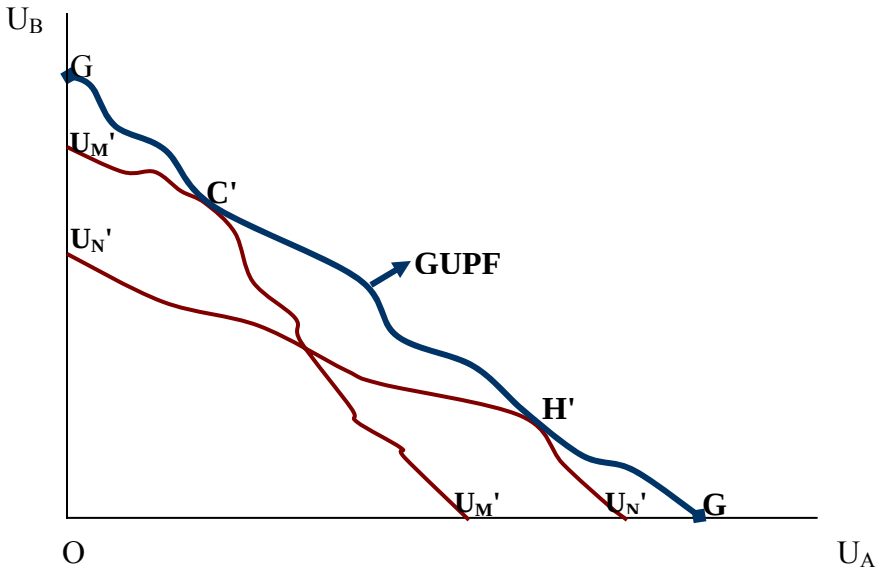


Figure 5.10: Grand Utility Possibilities Frontier

Suppose $U_M'U_M'$ is a utility possibilities frontier (UPF) drawn when the economy is in production equilibrium at point T in Figure 5.8. That is, it is just another name for $O_A'O_B'$ of Figure 5.9. If the economy is in production equilibrium at some other point on its PPF, we will have another contract curve of consumption, and thus a different UPF. Suppose,

as an example, $U_N'U_N'$ is a UPF drawn on the assumption that the economy is in equilibrium of production at point R on its PPF (refer Figure 5.8). $U_M'U_M'$ and $U_N'U_N'$ are just two of the very many possible UPFs. The (upper) envelope of all the possible UPFs is the GUPF. Put differently, by joining points of consumption equilibrium like C', H', and other points from UPFs similarly obtained, we derive the GUPF – curve GC'H'G in the above diagram.

The GUPF indicates that no reorganization of the production-exchange process is possible that makes someone better off without, at the same time, making someone else worse off. Thus, all points on GUPF are points of Pareto efficiency (Pareto optimality) in product-mix. Let us reemphasize the difference between UPF and GUPF. The former one (UPF) refers to the locus of all points of Pareto efficiency in exchange for a particular equilibrium in production. On the other hand, GUPF is the locus of all points of Pareto efficiency in product-mix (or the locus of all points of exchange equilibrium each of which corresponds to a possible equilibrium of production or to a point on PPF). Alternatively, any point on the GUPF is a point of simultaneous equilibrium of production and exchange (consumption) for a particular state of income distribution.

Nevertheless, only at one point on GUPF (corresponding to an income distribution) will social welfare be maximized. Given GUPF of the above sort, at what point would social welfare be maximized? With reference to Figure 5.10 above, is it at point C' or at point H' or at some other point between C' and H'? Determining the Pareto optimum point in production and exchange (or a point on the GUPF) *that maximizes social welfare* needs a social welfare function, to which we now turn our face.

5.2.3 Social Decision Mechanisms

Continuing with our GUPF of Figure 5.10, do you think that point C' represent an unfair distribution because consumer A gets greater utility than B at this point? If you think so, you have made a big mistake! Had our variable been income or quantity of output, we could have easily compared the possessions of A and B. However, what we have in this

case is utility (level of satisfaction), which cannot be compared for different individuals. Remember that utility is subjective: you can only compare the utility you get from a bundle of goods to the utility you get from a different bundle, and not the utility you get from a bundle to the utility someone else (say, your friend) gets from the same bundle. Hence, to tell whether an allocation is fair or not, we have to accept (or use) some kind of value judgment. Put differently, the only way we can decide which of the Pareto efficient points on the GUPF represents the maximum social welfare is to accept the notion of interpersonal comparison of utility.

Before we investigate the implications of this concept, it is worthwhile to consider first how one might go about bringing together the individual consumers' preferences to construct some kind of "social preferences". Two approaches that are of common interest are:

- the Majority Voting Approach, and
- the Rank-Order Voting Approach.

5.2.3.1 Majority Voting Approach (MVA)

Majority voting approach is one way to aggregate individual preferences. In this form of aggregation, allocation X is said to be preferred to allocation Y if majority of the individuals prefer X to Y. For example, in an economy consisting of three individuals, if two of them prefer X to Y and one of them prefers Y to X, using majority-voting approach we conclude that the society prefers X to Y.

However, there is a problem with this method in that it may not generate transitive social preference ordering. Suppose we have the following voting result.

Choice	Person A	Person B	Person C
1st	X	Y	Z
2nd	Y	Z	X
3rd	Z	X	Y

The table shows that:

- ☞ Persons A and C prefer X to Y and person B prefers Y to X. Hence, X > Y as per the majority-voting approach.
- ☞ Persons A and B prefer Y to Z and person C prefers Z to Y. Hence, Y > Z.

If this social preference ordering were transitive we could conclude, from the facts X > Y and Y > Z, that X > Z. However, from the table we see that persons B and C prefer Z to X implying that Z > X. Thus, majority voting can result in social preferences that are inconsistent even if individual preferences are transitive – *the "voting paradox"*.

Since the preferences are not transitive, there will be no "best" alternative from the set of alternatives (X, Y, and Z). Which outcome the society chooses will depend on the order in which the vote is taken. That is, for example, if the three people vote first on:

- ☞ X versus Y and then vote on the winner of this contest versus Z, the best alternative will be Z. *[X versus Y, X wins; then X versus Z, Z wins.]*
- ☞ Z versus X and then vote on the winner of this contest versus Y, the best alternative will be Y. *[Z versus X, Z wins; then Z versus Y, Y wins.]*
- ☞ Y versus Z and then vote on the winner of this contest versus X, the best alternative will be X. *[Y versus Z, Y wins; then Y versus X, X wins.]*

Therefore, majority voting can be manipulated by changing the order in which things are voted to yield the desired outcome. In addition, from the very beginning, we cannot tell whether the gain in utility by the majority is greater (in magnitude) than the loss in utility by the minority.

5.2.3.2 Rank-Order Voting Approach (ROVA)

Here, each person ranks the allocations according to his/her preferences and assigns a number that indicates its rank (and intensity) in his/her ordering. Then, the allocation with the minimum sum of ranks is taken as the socially preferred allocation.

Consider the following table.

Allocation	Person A	Person B	Sum of the Ranks
X	1	2	3
Y	2	1	3

Since the sum of the ranks is equal for both alternatives, the society is indifferent between allocations X and Y.

Rank-order voting has also a limitation. It can be manipulated by introducing new alternatives that change the final ranks of the relevant alternatives. For example, consider the following table. This table shows how what we just saw above changes when a new alternative is introduced.

Allocation	Person A	Person B	Sum of the Ranks
X	1	3	4
Y	2	1	3
Z	3	2	5

Note that the introduction of a new alternative (i.e., alternative Z) resulted in Y being preferred to X despite our earlier conclusion. Earlier, before alternative Z was introduced, we concluded that the society is indifferent between X and Y. Besides, even if each voter indicates his/her intensity of choice by assigning a numerical value to each alternative, it is not possible to compare the meaning of such values across individuals. That is, a value of 4 a person attaches to an allocation may not be the same as a value of 4 another persons attaches to an allocation.

To summarize, both majority voting and rank-order voting have limitations in that astute (shrewd) agents can manipulate their outcomes. The question that normally follows is, "Are there other social decision mechanisms, apart from majority voting and rank-order voting, that are immune to the limitations of these two voting approaches?" In other

words, are there other social decision mechanisms that satisfy a set of desirable properties? The following three are the principal (very desirable) properties on the basis of which social decision mechanisms are judged:

a. Given any set of complete, reflexive, and transitive individual preferences, the social decision mechanism should result in social preferences that satisfy the same properties.
b. If everybody prefers alternative X to alternative Y, then the social preferences should rank X ahead of Y. This is called the *Unanimity* or the *Pareto Efficiency* criterion.
c. The preferences between X and Y should depend only on how people rank X versus Y, and not on how they rank other alternatives. This criterion is termed the *Independence of Irrelevant Alternatives*.

It can be quite difficult to find a mechanism that satisfies all the above properties. In fact, Kenneth Arrow has proved the following remarkable result, which is known as ***Arrow's Impossibility Theorem:***

> *"If a social decision mechanism satisfies properties a, b and c, then it must be a dictatorship: all social rankings are the rankings of one individual."*

Arrow's impossibility theorem shows that three very plausible and desirable features of a social decision mechanism are inconsistent with democracy. That is, under democratic system, there is no perfect way to "aggregate" individual preferences to make one social preference. If we want to find a way to "aggregate" individual preferences to form social preferences, we will have to give up one of the properties of a social decision mechanism described in Arrow's impossibility theorem. Usually, the third property is dropped

5.2.4 Social Welfare Function

Earlier in our discussion, we said that there are as many optimal points as there are points on the PPF. Hence, Pareto efficiency (in product-mix) is a necessary, but not a sufficient condition for social welfare maximization. Each Pareto optimum corresponds to a different income distribution in the economy, and some may involve great inequalities of income than others. Consequently, decision should be made regarding the most desirable point of Pareto optimum.

A number of criteria have been forwarded in order to evaluate the welfare implications of policy measures. Among these are the Pareto criterion (also called Pareto's Unanimity Rule), the Kaldor-Hicks criterion (compensation principle) and the Scitovsky criterion (a double Kaldor-Hicks criterion).

According to the Pareto criterion of social welfare, a situation is optimal only if no individuals can be made better off without making someone else worse off. Stated differently, we can only say that "social welfare" has increased due to a change, if no individual is worse off because of the change and at least one is better off. If one individual is worse off, the fact that interpersonal utilities cannot be added or subtracted prevents economics from saying anything about social utility. Unfortunately, in the real world, any change usually makes some people better off while making others worse off. So, what happens to the welfare of the society in such a case?

To determine whether such an activity is moving the economy towards a higher welfare, two compensation tests have been developed. These tests ask what would happen if the winners were to compensate the losers. Using the Kaldor criterion, an activity will contribute to Pareto optimality if the maximum amount the gainers are prepared to pay is greater than the minimum amount that the losers are prepared to accept. Under the Hicks criterion, an activity will contribute to Pareto optimality if the maximum amount the losers are prepared to offer to the gainers in order to prevent the change is less than the minimum amount the gainers are prepared to accept as a bribe to forgo the change. The

Hicks compensation test is from the losers' point of view, while the Kaldor compensation test is from the gainers' point of view. If both conditions are satisfied, both gainers and losers will agree that the proposed activity will move the economy toward Pareto optimality. This is referred to as Kaldor-Hicks efficiency or the Scitovsky criterion.

The main problem with the compensation-based criteria is that since the compensation principle is supposed to help economists form policy judgments, we are always concerned with actual, and not potential, winners and losers from any change. Whether or not the winners may compensate the losers is therefore irrelevant; the important question is whether the compensation does, in fact take place. The compensation, however, cannot take place. The reason for this is that the (potential) compensators cannot investigate the utility scale of everyone and they cannot rely on the level of change in utility reported by the "losers" – people will report losses in utility in order to receive compensation.

All the above criteria have limitations, and lead to the conclusion that explicit use of personal judgments in interpersonal comparison of utility is inevitable. Determining the most desirable point of the Pareto set in product-mix requires specifying the social welfare function.

The social welfare function embodies value judgments about interpersonal comparison of utility. It is analogous to an indifference map for an individual, except that the social welfare function is a mapping of individual preferences or judgments of everyone in the society as to collective choices. To reiterate, the social welfare function gives a way to rank different allocations that depend on the individual preferences. Thus, a social welfare function is just some function of individual utility functions:

$$W\,[U_1(X), \ldots, U_n(X)]$$

In the above definition, $U_i(X)$ is each individual i's preference over all allocations, summarizing the individual's value judgments about the allocations. For example, person i prefers allocation X to allocation Z if and only if $U_i(X) > U_i(Z)$.

If R and S refer to two points on the consumption contract curve in our simple 2x2x2 economy, then individual A prefers allocation R to allocation S if and only if $U_A(R) > U_A(S)$. Individual A's preference over the two allocations, R and S, is then summarized by $U_A(R, S)$. Generalizing, A's preferences over all the possible allocations could be captured by $U_A(X)$ where X is a vector of all the possible (say, m) allocations [i.e., $X = (X_1, X_2, X_3, \ldots, X_m)$]. By the same token, B's preferences over all the possible allocations could be captured by $U_B(X)$. Hence, the social welfare function is some function whose arguments or components are $U_A(X)$ and $U_B(X)$: $W\{U_A(X), U_B(X)\}$.

How do we construct a social welfare function from the utility functions of individuals (defined over allocations)? There are different ways of formulating social welfare function from preferences of individuals. Some of these are described below.

1. *The (classical) utilitarian or Benthamite welfare function:* this method sums the utility of each individual in order to obtain society's overall welfare. In other words, we add the individual utilities and use the resulting number as a kind of social welfare function. That is,

$$W\{U_1(X),...,U_n(X)\} = \sum_{i=1}^{n} U_i(X).$$

In this form of social welfare function, all people are treated the same, regardless of their initial level of utility. One extra unit of utility for a starving person is not seen to be of a different value than an extra unit of utility for a millionaire. Such a welfare function is represented by linear and downward sloping social indifference (or isowelfare) curves. According to the utilitarian welfare function, we say allocation X is socially preferred to allocation Y if $\sum_{i=1}^{n} U_i(X) > \sum_{i=1}^{n} U_i(Y)$.

2. *The weighted sum-of-utilities welfare function:* this is a slight modification of the utilitarian welfare function. Modification is made in that the utility of all agents are not treated equally in the aggregation of the individual utilities to construct the social welfare function. Accordingly,

$$W\{U_1(X),...,U_n(X)\} = \sum_{i=1}^{n} a_i U_i(X)$$

where a_1, …, a_n are weights, numbers indicating how important each agent's utility is to the overall social welfare.

Somewhat in line with this type of welfare function was the favorite demonstration of the "old welfare economics" of the neoclassical economists brought to realization by Edgeworth and Pigou. Based on the assumption that a man's marginal utility of money diminishes as his money-income increases, they concluded that the marginal utility of a dollar is less to a rich man than to a poor man. Other things being equal, social utility is maximized by a progressive income tax that takes from the rich and gives to the poor.

3. *The Max-Min or Rawlsian utility function*: this type of welfare function considers the welfare of the worst off agent as the one that affects the social welfare of an allocation. That is:

$$W\{U_1(X),...,U_n(X)\} = \min\{U_1(X),...,U_n(X)\}$$

According to the Max-Min criterion, welfare is maximized when the utility of those society members that have the least is the greatest. No economic activity will increase social welfare unless it improves the position of the society member that is the worst off.

The above three ways of aggregating individual utilities into a single social welfare function rely on the assumption that utility is cardinal and interpersonally comparable. With the questionability of the cardinality of utility and the preferred notion of ordinal utility, explicit value judgments started to be incorporated into welfare economics. The criterion that makes use of an explicit set of value judgments in the formation of social welfare function is the Bergson-Samuelson criterion. As welfare improvement could come from the position of some individuals improving at the expense of others, efficiency and distributional issues are different and should be treated separately. While questions of efficiency are assessed with criteria such as Pareto efficiency and the

Kaldor-Hicks compensation tests, questions of income distribution are covered in social welfare function specification. That is, the production and consumption efficiency conditions, which are necessary for welfare maximization, are distinct from the interpersonal ethical values of the social welfare function.

The Bergson-Samuelson's social welfare function is analogous to the individual consumer's utility function. It provides a ranking of alternative states or configurations in which different individuals enjoy different utility levels. For our simple economy of two individuals, the Bergson-Samuelson's social welfare function could be presented by a set of social indifference curves (in utility space) like the ones shown in Figure 5.11 below. Each curve is the locus of combinations of utilities of A and B, which yield the same level of social welfare, and is termed as an isowelfare curve.

Given the welfare function defined as above, we can construct isowelfare curves as shown below.

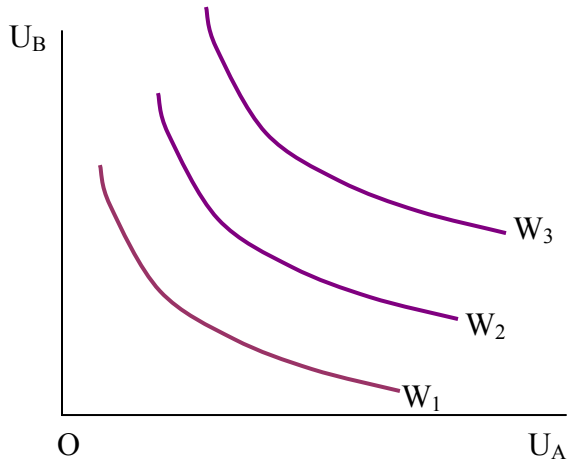


Figure 5.11: Bergson-Samuelson's Social Welfare Curves

Isowelfare curves show those distributions of utility that have constant welfare. The allocation that maximizes social welfare can then be obtained by bringing together our GUPF developed earlier in the chapter and the isowelfare curves.

Social welfare is maximized when GUPF is tangent to the highest possible isowelfare curve.

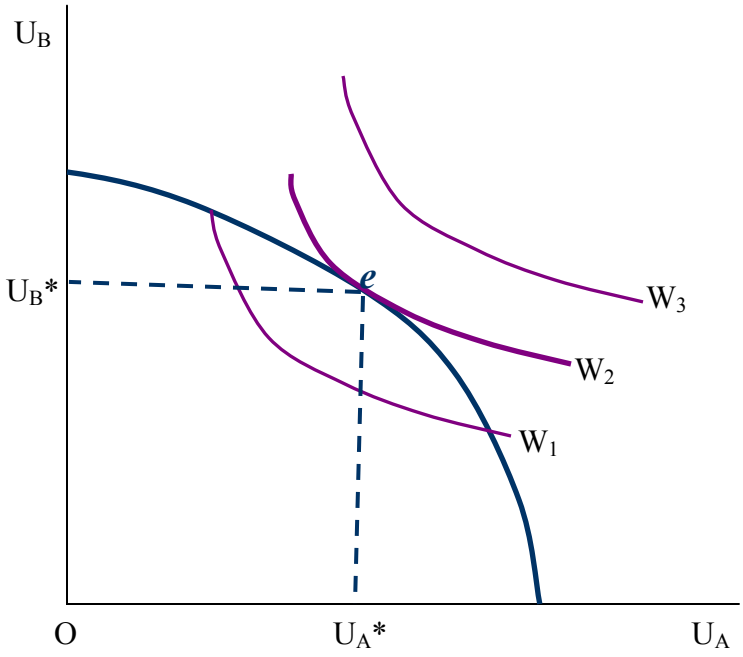


Figure 5.12: Bliss Point

Although all the points on the grand utility possibilities frontier are Pareto efficient, only one point identifies where social welfare is maximized. Point ***e*** in Figure 5.12 represents the point of maximum social welfare – termed the ***Bliss Point***. The economy maximizes social welfare when it achieves general equilibrium with an allocation (state) which generates U_A* and U_B* levels of utility to consumer A and consumer B respectively.

In Sub-section 5.2.5, we will adopt a different approach of welfare analysis. Before that, let us pull over at this point and have a schematic summary of what we did so far (in this chapter).

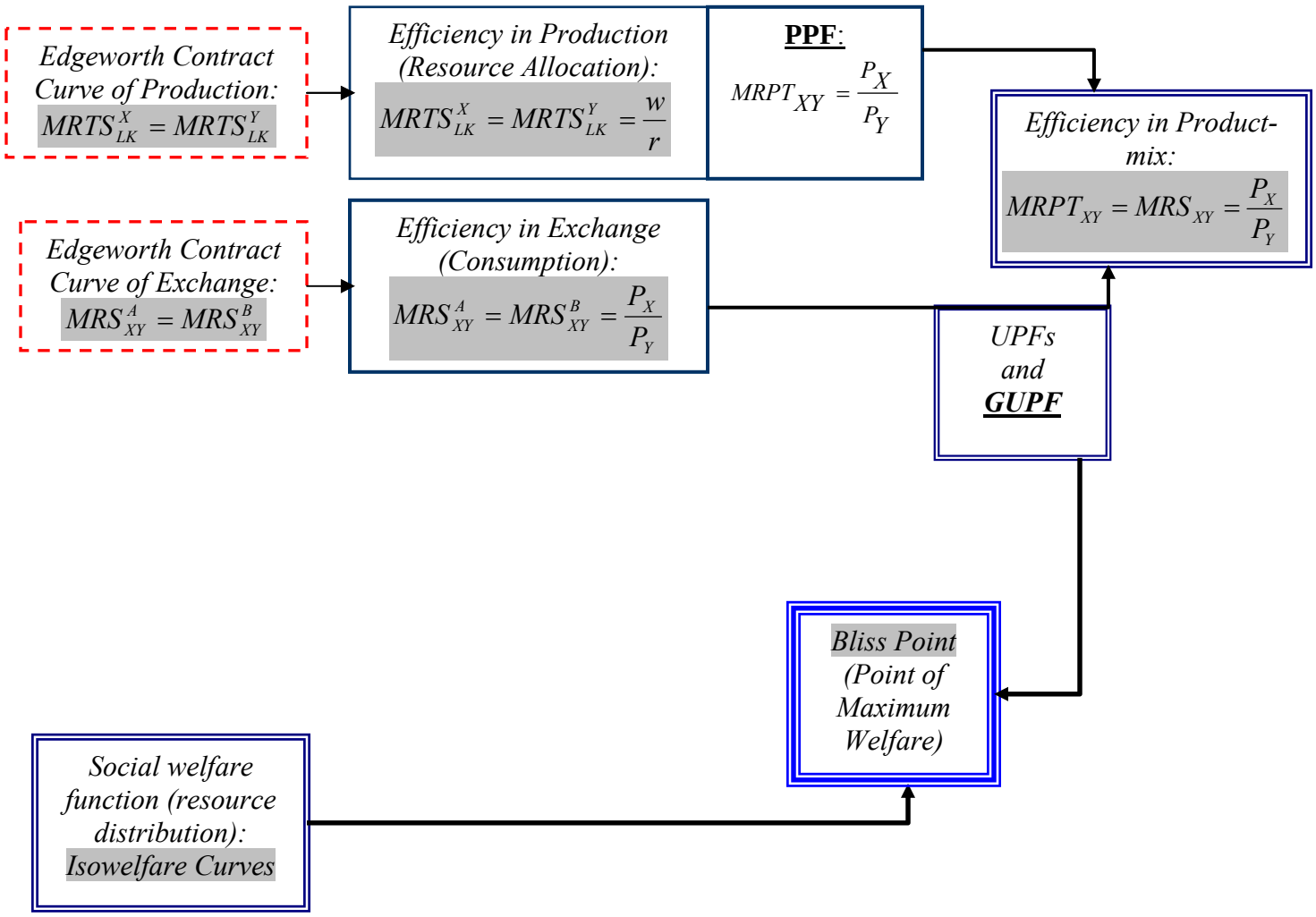


Figure: 5.13: A Schematic Summary of the General Equilibrium Analysis and Welfare Maximization

5.2.5 Fair Allocations, Envy and Equity

The welfare function approach is a very general way to describe social welfare. Because it is so general, it can be used to summarize the properties of many kinds of moral judgments. On the other hand, it is not of much use in deciding what kinds of ethical judgments might be reasonable ones. That is, except assuming that a marginal utility of a

poor man gets higher weight than that of a rich man in a social welfare function,[9] it does not point out specific alternatives on how the total amount of resources should be distributed or redistributed.

An alternative to using the welfare function approach is to start with some specific moral judgments and then examine their implications for welfare maximization. This is the approach taken in the study of *fair allocations*. The forthcoming paragraphs will shed light on how we use a specific kind of ethical judgment regarding resource distribution – equal division – in investigating the implications of redistribution measures for social welfare.

Suppose that we were given some goods to divide fairly among *n* equally deserving people. How should the goods be divided? It is probably safe to say that the goods should be divided equally among the *n* agents. Because the *n* individuals are equally deserving, there is no other appealing alternative.

One pleasing or attractive feature of this equal division is that it is symmetric. Put differently, each agent has exactly the same bundle of goods, and as a result, no agent prefers any other agent's bundle of goods to his/her own. If your friend is given (say) ten (10) units of a good and you are given six (6) units the same good, then you definitely prefer the possession of your friend to yours. However, if each of you (you and your friend) is given eight (8) units of the good, then you do not have any reason to prefer to be at your friend's position. Hence, the later (the eight-eight) allocation/distribution is symmetric.

An allocation is said to be *equitable* if no agent prefers any other agent's bundle of goods to his/her own. If some agent *i* does prefer some other agent *j*'s bundle of goods to his/her own, then we say that *i* envies *j*. thus, an equitable allocation/distribution is characterized by the absence of envy. Recall what a *Pareto efficient* allocation is! By now, you must be

[9] This is implied by the well-behaved isowelfare curves: successively higher levels of MU_A must be accompanied by successively lesser levels of scarification in MU_B for the society consisting of two individual to remain on the same isowelfare curve (or level of social welfare).

in a position to understand and define Pareto efficiency! If an allocation is both equitable and Pareto efficient, then we will call such an allocation a *fair allocation*.

Let us turn back to the case we began previously: the equal division among n individuals. We have seen that this equal division is symmetric and posed a question whether such a symmetric allocation is also Pareto efficient. Unfortunately, an equal division will not necessarily be Pareto efficient. To illustrate, suppose you and your friend are given equal amounts of goods X and Y. You may prefer to have more of good X and less of good Y than you are currently given. To the contrary, your friend might like to have more of Y and less of X than he/she is given. What action would you take? Naturally, you would exchange some units of good X for good Y with your friend whereby you would have more X and your friend more Y.

In general, if agents have different tastes, they will naturally desire to trade away from equal division. Suppose that such a trade takes place and that it moves us to a Pareto efficient allocation. Continuing with our example, suppose you and your friend exchange the two commodities (X and Y) and reach a point at which both of you possess the combination of the two goods you would like to have, given the total units of X and Y available. Such a point is Pareto efficient. The question that arises is, is this Pareto efficient allocation still fair in any sense? Alternatively, does trade from equal division inherit any of the symmetry of the starting point? The answer is, not necessarily. A trade away from equal division resulting in a Pareto efficient outcome may or may not be symmetric (or equitable). Consider the following example.

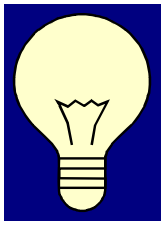
Suppose we have three people: A, B, and C. Also, suppose that A and B have the same tastes, and C has a different taste. Assume further that the starting point is characterized by an equal division of whatever we have among the three individuals, and that A and C has the chance to get together and trade. What do you expect? The logical expectation is that there would be some kind of trade (away from the equal division) between C on one side and B and C on the other side. However, in our hypothetical example, only A and C are given the chance to trade. Then, trade will typically make both A and C better off.

Now B, who has not had the opportunity to trade with C, will envy A, i.e., B would prefer A's bundle to his/her own. To sum up, even though A and B have started with the same allocation, A is luckier in his/her trading and this destroys the symmetry of the original allocation.

Our conclusion from the above paragraph is that arbitrary trading from an equal division will not necessarily preserve the symmetry of the starting point (of equal division). If so, is there any other way that guarantees an outcome (allocation) which is both Pareto efficient and equitable (symmetric) at the same time?

Instead of just any arbitrary way to trade, if we use the special mechanism of the competitive market, a trade away from an equal division will result in a Pareto efficient allocation. Besides, the result of such a trading will be equitable (symmetric). In our example above, since each of the three individuals has complete market information and an equal opportunity to trade, it is impossible for A to envy B. in these circumstances. This implies that a *competitive* equilibrium from equal division must be a fair allocation. Thus, the market mechanism will preserve certain kinds of equity: if the original allocation is equally divided, the final allocation must be fair.

Relating the conclusion of this approach to the approach of adopting a welfare function, we see that whether the price (market) mechanism maximizes social welfare or not depends on the initially prevailing distribution. According to both approaches, a competitive market results in a maximum social welfare (a fair allocation) if the initial distribution of income is equitable (symmetric). The difference between the two approaches is the following. The welfare function approach defines some kind of (aggregate) social utility function. On the other hand, the approach of fair allocations begins with some particular kind of ethical judgment and investigates the implied outcome(s) of such a value judgment.



5.3 Summary

1. *Partial equilibrium analysis studies the behavior of individual decision-making units and individual markets, viewed in isolation. General equilibrium, on the other hand, studies the interdependence or interconnections that exist among all markets and prices in the economy; and attempts to give a complete, explicit, and simultaneous answer to the fundamental questions of what, how and for whom to produce. Only when an industry (or a market) is small and has few direct and significant links with the rest of the economy may partial equilibrium analysis be considered appropriate. Nevertheless, cost and complexity considerations render the general equilibrium analysis impractical or unfeasible.*
2. *A simple economy of two individuals (A and B), two commodities (X and Y), and two inputs (L and K) is in a general equilibrium production when the economy is on its contract curve of production (i.e., when* $MRTS_{LK}^{X} = MRTS_{LK}^{Y}$*). Alternatively, general equilibrium in production is defined by the condition:* $MRPT_{XY} = \frac{P_X}{P_Y}$ *(or when the economy operates on its PPF). This same economy achieves a general equilibrium in exchange when the economy is on its contract curve of consumption (i.e., when* $MRS_{XY}^{A} = MRS_{XY}^{B}$*). For the simple economy to be in general equilibrium of production and exchange simultaneously (general equilibrium in product-mix)),* $MRPT_{XY}$ *in production must be equal to* MRS_{XY} *in consumption for individuals A and B. That is,* $MRPT_{XY} = MRS_{XY}^{A} = MRS_{XY}^{B}$.
3. *A distribution of inputs among commodities and of commodities among consumers is Pareto optimal (or Pareto efficient) if no reorganization of production and consumption is possible, by which some individuals are made better off without (at the same time) making someone else worse off. Thus, the*

conditions for Pareto optimality are the conditions for general equilibrium of production and exchange.

4. *Under perfect competition in input and output markets (with all the assumptions stated in the chapter), all the conditions for Pareto optimality are automatically satisfied. This is captured by the first theorem of welfare economics, which postulates that a competitive equilibrium is Pareto optimal. This theorem is the basic argument in favor of perfect competition and a proof of the Adam Smith's law of the invisible hand. The second theorem of welfare economics conveys that an infinite number of Pareto efficient market equilibria are possible (each corresponding to a different state of resource/income distribution). Accordingly, a point of maximum social welfare is not determined by market forces and thus inevitably involves the use of moral value judgments. In short, the second theorem of welfare economics postulates that equity in distribution is logically separable from efficiency in allocation.*
5. *Welfare economics studies the conditions under which the solution to the general equilibrium model can be said to be (socially) optimal. It examines the conditions for economic efficiency in the production of output and the exchange of commodities plus equity in the distribution of income. A utility possibilities frontier (UPF) is derived by mapping or transferring a contract curve for exchange from output space to utility space. It shows the various combinations of utilities received by two individuals at which the economy is in general equilibrium (or Pareto optimality) of exchange. We can construct an Edgeworth box and a contact curve of exchange from each point on the PPF. From each contact curve of exchange, we can then construct the corresponding utility possibilities frontier and determine on it the point of Pareto optimum in production and exchange. By joining points of Pareto optimality in production and exchange on each utility possibilities frontier, we can derive the grand utility possibilities frontier (GUPF).*
6. *The only way to determine which point on a grand utility possibilities frontier is a point of maximum social welfare is to make use of a normative economics*

(specifically, social welfare function). Kenneth Arrow proved that a social welfare function with some desirable properties cannot be derived by a democratic vote. This is known as the Arrow's impossibility theorem. A criterion that makes use of an explicit set of value judgments in the formation of social welfare function is the Bergson's (Bergson-Samuelson's) criterion. The Bergson-Samuelson's social welfare function is analogous to the individual consumer's utility function. It provides a societal ranking of alternative states or configurations in which different individuals enjoy different utility levels. In the case of two individual consumers, the social welfare function could be presented by a set of social indifference curves – isowelfare curves (in utility space).

7. *The economy maximizes social welfare when the GUPF is tangent to the highest possible isowelfare curve. Alternatively, welfare is maximized when there is a fair allocation. An allocation is fair if it is both Pareto efficient and equitable (if no agent prefers any other agent's bundle of goods to his/her own).*



5.4 Review Questions

I. Multiple-Choice Questions

1. Which of the following would be considered a disadvantage of allocating resources using a market system?
   A. Income will tend to be unevenly distributed.
   B. Significant unemployment may occur.
   C. It cannot prevent the wastage of scarce economic resources.
   D. None of the above.
2. Perfect competition leads to a Pareto optimal outcome because:
   A. resources are allocated among firms efficiently.
   B. final products are distributed among households/individuals efficiently.
   C. the system produces the things that people want.

D. all the points in A, B, and C occur.

3. Which of the following is needed in determining the general equilibrium of the 2x2x2 economy discussed throughout this chapter?
   A. Perfect divisibility (availability in any fractional amount) of factors of production
   B. Homogeneity of inputs
   C. Absence of externalities in both production and consumption activities.
   D. All of the above.
4. One of the following criteria is not applicable to evaluate the change in social welfare in a situation where the welfare position of some individuals improves at the expense of the others'.
   A. The Pareto criterion
   B. The Hicks criterion
   C. The Kaldor criterion
   D. The Scitovsky criterion
   E. None of the above.

II. True/False Questions

5. If one can describe an economy by means of a system of equations, with the number of unknowns equal to the number of independent equations, it is guaranteed that a general equilibrium exists.
6. We map the Edgeworth contract curve of consumption (from output space) to a utility possibilities frontier (in utility space) in the same way as we map the Edgeworth contract curve of production (from input space) to a production possibilities frontier (in output space).
7. Had all units of each input (labor and capital) been equally adaptable to the production of X and Y, the contract curve of production would have been a straight line joining the two vertices of the Edgeworth box of production.
8. All points on a PPF are points of productive efficiency in the face of changing technology given fixed amounts of labor and capital.

9. The marginal rate of product transformation of Y for X, $MRPT_{YX}$, shows the amount of Y that must be given up in order to obtain an additional unit of X, the economy being on the same PPF.
10. An economy is in general disequilibrium if the rate at which the consumers are willing to exchange one good for another equals the rate at which a good can be transformed into another in production.
11. Any point on a GUPF corresponds to point of Pareto optimality in both production and consumption while a point on a UPF corresponds to point of Pareto optimality in consumption.
12. Kenneth Arrow has proved that there is no perfect democratic way to aggregate individual preferences to make a social preference with a set of desirable properties.
13. In drawing isowelfare curves so as to determine a point of social welfare maxima on a GUPF, one has to make use of an explicit set of value judgments, which is not required in deriving the GUPF.
14. Given that the original allocation of resources in an economy is symmetric, perfect competition (with all the assumptions in this chapter) will result in Pareto efficiency in product-mix, preserving the equity in the economy.
15. Society will produce the efficient mix of output if most firms equate price and marginal cost.
16. Real-world markets always produce an efficient outcome.
17. The voting paradox is a demonstration of how majority-rule voting eliminates any excess influence that could be exercised by the person who sets the order of the voting agenda.
18. Whenever there is envy, there is equity; and this equity along with efficiency in product-mix implies fair allocation.

III. Discussion Questions

19. Why do we study general equilibrium theory where it is largely non-operational and unrealistic?

20. Are all points on the contract curve for exchange equally desirable from society's point of view? Why or why not?
21. What is the relationship between Adam Smith's law of invisible hand and Pareto optimality?
22. How will an increase in the quantity of capital available to society affect its Edgeworth box diagram of production? What effect will that have on the PPF?
23. What makes general equilibrium analysis objective while welfare economics is subjective?
24. Suppose MRPTXY = 1.5 and MRSXY = 2 for the two individuals, A and B, in the simple 2x2x2 model.
25. Interpret these two numbers.
26. Is this economy in equilibrium? If not, should the economy produce more of X or more of Y to reach equilibrium in product-mix?

CHAPTER SIX
MARKET FAILURE

Introduction

In the fifth chapter, it was concluded that *all competitive equilibria are Pareto-efficient*, that *all welfare maxima are competitive equilibria*, and that *every competitive equilibrium is a bliss point (a point of welfare maximum) for some welfare function*. One possibility is that the self-interest of the individual economic agents (as Adam Smith argues) results in the maximum social welfare. Alternatively, a redistribution of economic resources via non- or less-distortional mechanisms results in a fair allocation, given perfectly competitive markets.

However, the prerequisites for perfect competition are unlikely to hold in the real world. At least some firms in an economy may (and do usually) have market power and/or there usually exists a problem of asymmetric information. Even a perfectly competitive market may fail because of externalities and/or public goods. Thus, market failure is a commonplace in the real world. Generally, markets fail for four reasons:

1. Externalities,
2. Public goods,
3. Market power, and
4. Asymmetric information.

This chapter is devoted to the discussion of these sources of (reasons for) market failure.

Chapter Objectives

Upon the completion of this chapter, students would be able to:

- Understand the reasons for which the market mechanism fails to maximize social welfare and the rationale for (at least a minimal) government intervention.

- Distinguish between different types of externalities, give some examples for each type, and explain how market failure arises in the presence of externalities.
- Describe what a public good is, give some examples of public goods, and explain how the existence of a public good leads to market failure.
- Discuss how market power in some sector(s) of an economy results in a failure to maximize social welfare through entirely relying on the market mechanism.
- Be familiarized with the problems of adverse selection and moral hazard (which result from asymmetric information) and with the signals and ways economic agents use to overcome (rather, minimize) such problems.
- Apply the concepts of asymmetric information (specifically, adverse selection and moral hazard) and transaction (and monitoring) costs to the problems faced in the financial sector, and understand how group lending helps overcome such problems at least partially.

6.1 Externalities

Externality is said to exist when the production/consumption activities of one party enter the production/consumption activities of other party, without any payment for the effect. The market system (mechanism) is capable of achieving Pareto-efficient allocations when there are no externalities. In the presence of externalities, however, market prices do not reflect the activities of either producers or consumers and thus market outcomes would be Pareto inefficient.

Externalities may be classified as consumption externalities and production externalities. Each of these types may be negative (external costs) or positive (external benefits). Below, we will see examples for the different categories of externalities.

1. *Negative consumption externalities (external diseconomies of consumption)* – refer to uncompensated costs imposed on others (consumers) by the consumption or production activities of some individuals or firms. Examples of negative consumption externalities include:

- ☞ pollution produced by local automobiles,
- ☞ a smoker smoking a cigar next to an individual dining in a restaurant, and
- ☞ playing a loud music in one's neighborhood.

In the examples above, the pollutants (i.e., the users of automobiles), the smoker and the one playing a loud music do not pay compensation for the victims (those who suffer the consequences of the activities).

2. *Positive consumption externalities (external economies of consumption)* – refer to uncompensated benefits conferred on others (consumers) by the increased consumption or production of a commodity by some individuals or firms. Examples for this sort of externalities embrace:
   - ☞ pleasure from observing a flower garden of one's neighbor,
   - ☞ pleasure from listening to music of one's taste from a neighbor (or roommate).

   You do not pay for enjoying your neighbor's flower garden or for listening to some music of your taste from a CD-player of your roommate.

3. *Negative production externalities (external diseconomies of production)* – refer to uncompensated costs imposed on others (producers) by the expansion of consumption or production by some individuals or firms. This type of externality is exemplified by:
   - ☞ water pollution by a steel industry for a fishery down the stream.

   If a steel industry disposes its wastes to a river, this will damage the school of fish in the stream and consequently affect people living on fishing or the fishing industry at large.

4. *Positive production externalities (external economies of production)* – refer to uncompensated benefits conferred on others (producers) by raised consumption or output expansion of some individuals or firms. An example for such an externality is:
   - ☞ an orchard (land area of fruit trees) located next to a beekeeper.

This example reflects a bi-directional externality where both the beekeeper and the owner of the orchard benefit. The beekeeper would benefit since the orchard is a source of ingredients for the production of honey. On the other hand, the bees would help the pollination (and thus the reproduction) of the flower trees.

Note that consumption externalities could be imposed or conferred by consumption or production activities of agents, but the effects are seen on consumption. Similarly, production externalities could result from production or consumption activities of agents, but the effects are seen on production. Hence, in classifying an uncompensated impact of an activity as a production or a consumption externality, we consider the destination or target of the effect rather than the origin (source). Do not bother much with how to classify externalities in line with the above four headings. Rather, focus on understanding how some production or consumption activities impose externalities (costs or benefits) on production or consumption activities.

The crucial feature of externalities is that they are items/goods people care about but that are not sold on markets. It is this lack of markets for externalities that causes problems. For instance, we do not have market for a neighbor with the hobby of gardening flowers.

The discussion so far being sufficient to introduce and clarify the concept and nature of externalities, now, how does the presence of externalities (of whatever kind) result in market failure? We will shed light on this issue of how the existence of externalities could explain market failure soon.

Perfect competition leads to a Pareto optimum outcome only when private costs equal social costs and private benefits equal social benefits. However, if there are externalities, there is divergence between private and social benefits and/or private and social costs. For a private decision-maker, the optimality (equilibrium) condition is given by the equality of marginal benefit and marginal cost of the agent. That is, optimal decision of a private agent requires the condition ***MPB = MPC*** (where MPB = marginal private benefit and MPC = marginal private cost). For the society at large, on the other hand, an efficient

allocation requires that MSB = MSC, (where MSB = marginal social benefit and MSC = marginal social cost).

In general, external economies (positive externalities) result in production or consumption of commodities to a level less than socially optimal. To the contrary, external diseconomies (negative externalities) result in production or consumption of commodities to a more than socially optimal level.

Let us illustrate market failure resulting from externalities with the help of one of the categories of externality discussed earlier. Figure 6.1 below depicts the case of a production activity that generates a positive externality.

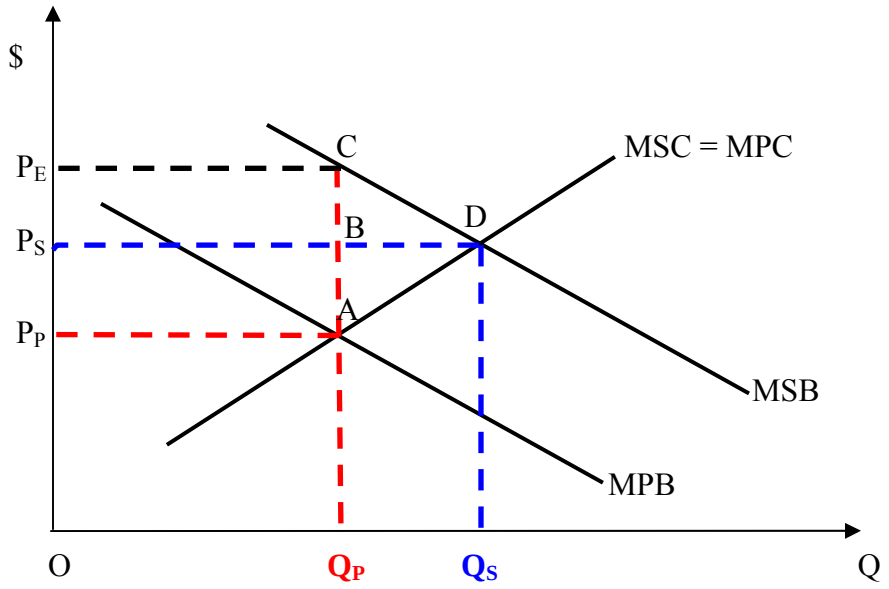


Figure 6.1: A Production Activity with Positive Externality

The firm compares its marginal benefit and marginal cost of production and defines its equilibrium at point where MPB = MPC (point A in Figure 6.1). For this particular type of externality, the society does not incur any additional cost (above that incurred by the firm) and thus MSC = MPC. The firm produces Q_P units of output and sells the product at a unit price of P_P. Given that this activity confers an uncompensated benefit on others, the benefit of the production activity is above what the firm recognizes (i.e., MSB > MPB).

Taking this into account, the extra benefit to the society is larger than the extra social cost (i.e., MSB > MSC) at Q_P units of output. This difference between the MCB and MSC, given by the distance from A to C in the figure, measures the marginal external benefit (MEB) of the production activity. Since MSB (= MPB + MEB) > MPB at this point, it would be better if output were expanded to the level where MSB = MSC (to point B). In other words, from the viewpoint of the society, optimality is defined at point D, where Q_S units would be produced and each unit would be sold at a price equal to P_S. However, the firm is concerned only with its (private) marginal benefits and costs. As a result, it will not expand its production to Q_S since MPC > MPB at Q_S. Rather, it produces Q_P units only as shown above. In a nutshell, a commodity the production of which entails a positive externality will be undersupplied.

By the same token, if the production of a certain good is responsible for external diseconomies, more of this good is likely to be produced than is socially optimal. Similarly, if an individual takes an action that contributes to the welfare of the society but which results in no payment for him/her, he/she is certainly likely to take his/her action less frequently. Likewise, if an economic agent takes an action that results in costs that this agent is not forced to pay, then it is certainly likely this action is undertaken more frequently than is socially desirable.

With external diseconomies of consumption and production, governments might try to intervene by directly limiting the level of consumption and production (*direct control*) or with a tax on consumers and producers (*effluent fee*). As an example for the direct control, pollutants might be forced to produce a lesser amount of pollution than they could do if based on their (private) marginal benefits and costs. Alternatively, economic agents whose actions imply negative externalities may be required to pay taxes for producing the externality. Imposing taxes would increase the private marginal cost of the agents thereby making them internalize the effects of their actions and rendering a lower level of production/consumption to be optimal. On the other hand, with negative externalities (external economies) of production and consumption, Pareto optimality in production and consumption would require a subsidy for the producers and consumers.

However, the corrective tax or subsidy (for correcting positive or negative externalities) may have some unintended side effects that lead to inefficiency. That is, an attempt to correct inefficiency resulting from externalities may itself result in inefficiency.

In the previous paragraph, we considered two methods for correcting externalities: through direct control and through charging effluent fee. Alternatively, *assignment of clearly defined and transferable property rights* could avoid externalities (and promote economic efficiency) – Coase theorem. According to the Coase theorem, well-defined property rights will work in one of the following two ways. If the property right is granted to the firm producing an externality, then the victims will pay the firm in order to avoid the external cost. On the other hand, if the property right is granted to the consumers, then the firm producing the externality will compensate the victims (consumers). The outcome of both approaches will be the same – the externality will be internalized. The problem with this approach is that, this solution to externalities requires zero (or insignificant) transaction costs. Transaction costs (comprising of the legal, administrative, and informational expenses of drawing up, signing, and enforcing contracts) may, however, be so high in practice that negotiations are not feasible. Even if it may not lead to the absolute maximization of social welfare, trying to define unambiguous and transferable property rights does help to promote Pareto optimality.

6.2 Public Goods

In concluding that a perfectly competitive economy results in an optimal allocation of resources, excludability and rivalry in consumption were assumed. Excludability was assumed in the sense that a person consumes a good only if he/she pays the price. Rival consumption was assumed in the sense that if one person consumes a particular good (say, drinks a bottle of beer), someone else cannot consume it too. In the real world situation, however, not all goods have these two characteristics. Indeed, many very important goods like the quality of the environment and national security do not have them – they are non-excludable and non-rival.

If consumption of a commodity by one individual does not reduce the amount available for others (i.e., if the commodity is non-rival in consumption), the commodity is a public good. Note that even though non-rival consumption and non-exclusion often go hand in hand, a public good is defined in terms of non-rivalry in consumption only.

Public goods are goods that are non-rival: the marginal cost of provision to an additional consumer is zero. That is, once provided (made available), the addition of consumers up to capacity constraint does not reduce the availability of the good. Examples of public goods are national defense, law enforcement, national TV, streetlights, sidewalks, etc.

Though many public goods are non-excludable (i.e., even if it is technically difficult or prohibitively costly to exclude others from using these goods), some are excludable. For example, it is possible to allow only paying customers to view cable TV; it is also possible to charge a fee for using an uncrowded bridge and to prevent people who do not pay from crossing it. Still, these items are public goods as additional customers could consume them with no extra (marginal) cost.

The market mechanism will not work properly for a public good even if it conforms to the exclusion principle. Because there is non-rivalry in consumption (the marginal cost of the additional consumer is zero), there is welfare improvement if people are added to the consumption of the good without any payment. That is, excluding those who do not pay reduces the satisfaction of the excluded but does not increase the satisfaction of the others (those who paid).

Moreover, it is frequently impossible to prevent people from consuming a public good whether or not they pay for it. Put more explicitly, many public goods are characterized by non-excludability in consumption. There is no way to prevent someone from benefiting from national defense. In such situations, usually, people are attempting to free ride on each other. The free-rider problem arises when a consumer or producer does not pay for a non-excludable good in the expectation that others will do. As an example, consider cleaning your room (dormitory). Each of you may prefer to see the living room

clean and is willing to do his/her part. However, each may also be tempted to free ride so that no one ends up cleaning the room, with the result being untidy room. This "letting the other person do it" may be optimal from an individual's point of view, but it is Pareto inefficient from the viewpoint of the society.

In general, the larger the group size, the more acute will become the free-rider problem. This problem can be and is generally overcome by the government taxing the general public to pay for the public goods. Sometimes governments go to great lengths to induce people to pay their fair share of taxes to provide public goods. It is, however, practically impossible for the government to know exactly the optimal amount of the public good that it should provide or induce the private sector to provide. This is because each individual has no incentive to reveal accurately his/her preference or demand for the public good. Who will be ready to tell his/her willingness to pay for a streetlight? Besides the difficulty of estimating the demand for a public good, there is also a problem of possible government inefficiency in providing public goods and in intervening in the market in a broader sense. Government intervention should be considered in line with the degree of market inefficiency relative to government inefficiency: government should not intervene if the inefficiency it creates is not less than inefficiency due to market failure.

6.3 Market Power

Inefficiencies arise when a producer (supplier) of a commodity has market power. This might result from the fact that markets for some goods and services are characterized by economies of scale operating over a large range of output implying the inevitability of having a single firm or a few firms supplying the entire market – technical externality. Recall that such a barrier to entry has been discussed in Chapter Two in relation with oligopoly

How could the existence of a monopolistic power explain market failure? In the first two chapters (as well as in Chapter Four) of this teaching material, we saw that the demand

for the product of a firm with some monopolistic power is downward sloping. Below, we will reconsider the efficiency implication of such a downward sloping demand curve.

For example, suppose there is a monopoly power in one of the two industries in a hypothetical economy. At equilibrium (where MR = MC), a firm in the industry with monopoly power will produce and sell less units of output at a higher price as compared to a firm in a perfect competitive industry (i.e., P > MR). If this industry characterized by some monopoly power is labeled as industry X, then we would have the condition $MC_X = MR_X < P_X$ at equilibrium. If the other industry, call it industry Y, is perfectly competitive, then a firm in this industry will produce at a point where $MC_Y = MR_Y = P_Y$. Thus, bringing the two industries together and recalling the definition of the marginal rate of product transformation, we will have: $MRPT_{XY} = \frac{MC_X}{MC_Y} < \frac{P_X}{P_Y}$. Consequently, this economy will not reach a Pareto-optimum point. Pareto efficiency being a necessary condition for social welfare maximization, our hypothetical economy will not maximize the social welfare.

Inefficiency (market failure) may also result from market power or imperfection in input market. For instance, an economy with perfectly competitive labor market but imperfectly competitive capital market cannot maximize social welfare. In general, as we have seen in Chapter Four, the least-cost input combination in production is given by the condition:

$$\frac{MPP_L}{ME_L} = \frac{MPP_K}{ME_K} \text{ or } \frac{MPP_L}{MPP_K} = \frac{ME_L}{ME_K}.$$

With the information given regarding the two factor markets,

$ME_L = w$ and $ME_K > r$.

Then, $\frac{MPP_L}{MPP_K} = \frac{ME_L}{ME_K} < \frac{w}{r}$. Since $\frac{MPP_L}{MPP_K} = MRTS_{LK}$ by definition, this implies that $MRTS_{LK} < \frac{w}{r}$ – a violation of the condition for Pareto optimality (general equilibrium)

in resource allocation (or production). Hence, this economy does not reach a point of Pareto-optimality, and thus social welfare will not be maximized.

In sum, the existence of firms with market power in some sector of an economy is one reason for the failure of market in producing or yielding maximum social welfare .

6.4 Asymmetric Information

Asymmetric information is a situation in which a buyer and a seller of an item possess different amounts of information about a transaction. There are certainly many real world markets in which it may be very costly or even impossible to get accurate information about the quality of the goods being sold.

The following situations (examples) are intended to clarify how economic agents on the two sides of a transaction may possess differing level of information:

1. Labor is assumed to be a homogeneous factor in almost all cases we referred to it. For instance, in Chapter Four, we assumed that everyone had the same "kind" of labor and supplied the same amount of effort per hour worked. In reality, it may be very difficult for a firm to determine how productive its employees are. On the other hand, each employee has a better (or full) knowledge of his/her own labor quality.
2. When a consumer buys a used car, it may be very difficult for him/her to determine whether the car is a good car or a poor quality or defective one (a lemon). By contrast, the seller of the used car possibly has a pretty good idea of the quality of the car.
3. An insurance company will not get an unbiased selection of customers, and/or cannot observe the action of insured customers. That is, the insurance company faces a difficulty in choosing good-risk customers from the very beginning, and also has a difficulty in observing the level of care each customer takes. Yet, the customer knows his/her risk category as well as the level of care he/she is actually taking.

4. A financial institution will not get an unbiased selection of borrowers and/or cannot observe the action of the borrowers while using the fund. This situation is similar to the situation of the insurance company in the third example except that now the agent facing the problem is a financial institution.

Asymmetric information may cause significant problems to the efficient functioning of a market. In the presence of asymmetric information, there may be too few transactions. Besides, some transactions may completely fail. Next, we will discuss two common problems associated with asymmetric (imperfect) information: adverse selection and moral hazard. After we deal with adverse selection, we pause to look at a related concept of signaling before we discuss moral hazard.

6.4.1 Adverse Selection (The Hidden Information Problem)

Adverse selection is the situation where low-quality products drive high-quality products out of the market because of the existence of asymmetric information between buyers and sellers. It arises when one side of the market cannot observe the type or quality of the goods on the other side of the market. The problem of adverse selection could arise in all markets characterized by asymmetric information. The market for “lemons” (defective products such as used cars) and all insurance markets represent good examples. In general, equilibrium in a market involving hidden information will typically involve too little trade taking place because of the externality between the "good" and "bad" types. Such equilibrium will be inefficient relative to the equilibrium that could have prevailed with full information.

Consider the used-car market example mentioned earlier. In this market, sellers of used-cars know exactly the quality of the cars they are selling while prospective buyers do not. As a result, the market price for used-cars will depend on the quality of the *average* used-car available for sale. The owners of “lemons” would then tend to receive a higher price than their cars are worth, while the owners of high-quality used-cars would tend to get a lower price than their cars are worth. The owners of high-quality used cars would

therefore withdraw their cars from the market. This withdrawal of high-quality cars, in turn, lowers the average quality and thus the price of the remaining cars available for sale. Subsequently, sellers of the now above-average quality cars would withdraw their cars from the market, further reducing the quality and price of the remaining used-cars offered for sale. This process continues until only the lowest-quality cars remain in the market, to be sold at the appropriate very low prices. Thus, the ultimate outcome of the process is that low-quality cars drive high-quality cars out of the market.

Consider another example, the insurance industry. Suppose that an insurance company wants to offer insurance for bicycle theft. In some areas, there is a high probability that a bicycle will be stolen, and in other areas, thefts are quite rare. Suppose that the insurance company decides to offer the insurance based on the average theft rate. Who is going to buy the insurance at the average rate? Not the people in the safe communities – they do not need much insurance anyway. Instead, the people in the communities with a high incidence of theft will want the insurance. However, this means that the insurance claims will mostly be made by the consumers who live in the high-risk areas. The insurance company will not get an unbiased selection of customers; rather it will get an adverse selection. Thus, if insurance companies must charge a single premium because they cannot distinguish between high-risk and low-risk groups (individuals), more high-risk groups (individuals) will insure, making it unprofitable to sell insurance.

Let us see another example from the insurance market. When a health insurance company sets the insurance premium for the average individual (i.e., an individual of average health), the unhealthy people are more likely to purchase insurance than the healthy people. Because of this, the insurance company is forced to raise the insurance premium, thus making it even less advantageous for healthy individuals to purchase insurance. This will increase even further the proportion of unhealthy people in the pool of insured people, thus requiring still higher insurance premiums. In the end, insurance premiums would have to be so high that even unhealthy people would stop buying insurance.

Insurance companies try to overcome or reduce the problem of adverse selection by requiring medical checkups, charging different premiums for different age groups and occupations, and offering different rates of coinsurance, amounts of deductibilities, length of contracts, and so on. Because there will always be some variability in risk within each group, however, the problem of adverse selection cannot be fully eliminated. The only way to avoid the problem completely is to provide *compulsory insurance* to all the people in the group. In fact, there are social institutions that help to solve this market failure (due to adverse selection in the market for health insurance). It is commonly the case that employers offer health plans to their employees as part of the package of fringe benefits. Then, the insurance company can base its premium rate on the average over the set of *all employees* as such a mechanism assures that all employees participate in the program. This will eliminate the problem of adverse selection. Alternatively, it is generally possible to require the purchase of insurance that reflects the average risk in the population by everyone. This alternative for overcoming adverse selection in the insurance market is termed *compulsory purchase plan*.

Credit companies significantly reduce the adverse selection problem that they face (because of the externality between good and bad borrowers) by sharing credit histories with other credit companies. This allows the credit market to operate and keep the interest rate to acceptably low levels.

6.4.2 Market Signaling

Market signaling is the process by which sellers send messages (or signals) to buyers conveying information about the quality of their goods and services. What do you expect to happen if sellers of high-quality products, lower-risk individuals, better-quality borrowers, or more-productive workers can somehow inform or send signs of their superior quality their buyers (customers)? The logical outcomes are:

- ☞ individuals would be able to identify high-quality products;
- ☞ insurance and credit companies would be able to distinguish between low- and high-risk individuals and firms; and

- ☞ firms would be able to identify higher productivity workers.

As a result,

- ☞ sellers of high-quality products would be able to sell their products at higher prices;
- ☞ lower-risk individuals could be charged lower insurance premiums;
- ☞ better quality borrowers would have more access to credit; and
- ☞ higher-productivity workers could be paid higher wages.

Market signaling can thus overcome the problem of adverse selection, for the most part.

Recall our example of the used-car market: the owners of the used car knew the quality but the purchaser had to guess at the quality. We saw that this asymmetric information could cause problems in the market. In some cases, the adverse selection problem would result in too little transactions being made. However, the owners of the good used-cars have an incentive to try to convey the fact that they have a good car to the potential purchasers. They would like to choose actions that signal the quality of their cars to those who might buy it.

One sensible signal in this context would be for the owners of good used-cars to offer a warranty. This would be a promise to pay the purchaser some agreed upon amount if the car turned out to be a lemon. Owners of the good used-cars can afford to offer such a warranty while the owners of the lemons cannot afford this. This is a way for the owners of the good-quality used-cars to signal that they have good cars. In general, a firm can signal the higher quality of its products to potential customers by offering assurances (guaranties and warranties), by a policy of exchanging defective items, by adopting brand names (say through franchising) and so on.

Similarly,

- ☞ The acceptance of coinsurance and deductibles by an individual or a firm sends a powerful message to insurance companies indicating that the individual or the firm is a good-risk.
- ☞ The credit history of a potential borrower (indicating that he/she has repaid past debts in full and on time) sends a strong signal to credit companies that he/she is a good credit risk.
- ☞ Higher levels of educational accomplishments (such as years of schooling, degrees awarded, grade point average achieved, etc.), which show the individual's intelligence and determination to complete college, serve as a powerful signal to an employer of the greater productivity of a potential employee. Firms usually rely on the market signal provided by education for lack of a better signaling device when they first hire employees. [After they had adequate opportunity to determine the employees' true productivity on the job, they rely on actual performance.]

In each of the above cases (and in the presence of adverse selection in general), signaling helps to make markets with asymmetric information perform better.

6.4.3 Moral Hazard (The Hidden Action Problem)

Moral hazard refers to a situation where one side of the market cannot observe the action of the other side. It exists, for instance, when an insured party whose actions are unobserved can affect the probability or magnitude of a payment associated with an event. In general, it arises whenever an externality is present in that an economic agent can shift some of its costs to others.

Reconsider the bicycle theft insurance market we described earlier. Suppose for simplicity that all of the consumers live in the areas with identical probabilities of theft, i.e., suppose there is no problem of adverse selection. On the other hand, the probability of theft may be affected by the actions taken by the bicycle owners.

If it is impossible to buy bicycle-theft insurance, consumers have an incentive to take the maximum possible amount of care. In this case, the individual bears the full cost of his/her actions and accordingly he/she wants to "invest" in taking care until the marginal benefit from more care just equals the marginal cost of doing so. However, if the consumer can purchase bicycle-theft insurance, then the individual's incentive to take care will become less. If the amount of care is observable, then there is no problem because the insurance company can base its rates on the amount of care taken. However, insurance companies cannot observe all the relevant actions of those they insure. Therefore, a trade-off will be involved: too little insurance means that people bear a lot of risk, too much insurance means that people will take inadequate care. The fact that the consumer can choose the amount of care he/she takes implies that the insurance company will not allow the consumer to purchase as much insurance as he/she wants if the company cannot observe the level of care.

The amount of a good traded in a competitive market is determined by the condition that demand equals supply: the marginal willingness to pay equals the marginal willingness to accept (sell). In the presence of moral hazard in insurance market, the market equilibrium has the following property. Each consumer would like to buy more insurance and the insurance companies would be willing to provide this amount of insurance only if the consumers continued to take the same amount of care. However, this trade cannot occur because if the consumers were able to purchase more insurance they would rationally choose to take less care.

The method by which insurance companies try to overcome the problem of moral hazard is by specifying the precautions that an individual or firm must take as a condition for buying insurance. For example, the insurance company might require yearly physical checkups as a condition for continuing to provide health insurance to an individual, increase insurance premiums for drivers involved in accidents, and require the installation of a fire detector before providing fire insurance to a firm. By doing this, the insurance company tries to limit the possibility of illness, accident, or fire, and thereby reduce the number and amount of possible claims it will face.

Another method used by insurance companies to overcome or reduce the problem of moral hazard is *coinsurance*. This refers to insuring only part of the possible loss or value of the property being insured. The idea is that if the individual or firm shares a significant portion of a potential loss with the insurance company, then the individual or the firm will take more precautions to avoid losses from illness or accidents.

As a second example of a market where the problem of moral hazard arises, consider the labor market. It is difficult or impossible for firms to monitor workers' productivity accurately. If workers are paid the equilibrium wage, they are likely to shirk or slack off on the job, as their actions are not accurately observed by their employers. In such situations, firms try to overcome the problem by inducing employees to work more productively and not to shirk through paying a higher-than-equilibrium wage (called efficiency-wage). At the efficiency-wage, there is a great deal of unemployment. Thus, employees would not shirk and not risk being fired because it is not easy to find another similarly rewarding job in view of such unemployment. The theory, which postulates that firms willingly pay higher than equilibrium wages to induce workers to avoid shirking or slacking off on the job is termed *efficiency wage theory*.

6.4.4 Information Asymmetry, Monitoring Costs and Microfinance

When a lender wants to provide credits to the borrowers, he/she faces certain problems. These problems include:

1. It is costly (or impossible) for the lender to know whether a borrower is able to repay his/her debt, i.e., whether the borrower is a hard-working person or not. This is the problem of adverse selection.
2. It is costly to monitor the borrower on what is being done with the loan. The borrower might use the money borrowed for consumption (or unproductive activities), which is not easily transferred into monetary repayment. Alternatively, the borrower might use the loan in risky activities such as gambling. This refers to the problem of moral hazard.

3. If enforcement (of the court procedure) is costly, then the borrower may *voluntarily* default even if he/she has the potential to payback.

The first two reasons result in *involuntary default*. That is, because the lending might have been made to a lazy (bad-risk) borrower or because necessary follow-ups have not been adequate, the borrower would finally be unable to repay his/her loan. The third reason is a source of voluntary default: even if the borrower is a hard-worker, invests the loan in a productive area, and become successful, he/she might choose not to payback the loan if the legal framework is weak.

Let us now compare the above sets of problems facing lenders in relation to three categories of moneylenders: big financial institutions, village moneylenders, and micro-financial institutions.

Formal Financial Institutions

For big financial institutions like banks (say, Commercial Bank of Ethiopia – CBE), verifying the borrowers credit-worthiness and monitoring what is being done with the loan are very costly, especially for a large number of small-scale borrowers. Thus, formal banks do not provide credit for small borrowers (say, farmers) because of *adverse selection, moral hazard* and *huge transaction costs*. The last factor – huge transaction costs – is explained by dealing with a large number of small farmers. However, social welfare can be increased if it is possible to provide credits for those who demand it. Then, what should be done, or what alternative is there?

Village Moneylenders

Village moneylenders, as compared to banks (or big financial institutions), have a comparative advantage in these small-scale loans for the following reasons:

1. They can deal with small-scale lending more effectively. As village moneylenders have location advantage, there would be less costs of transaction.

2. They have better access to information about who are good and who are bad credit risks than outsiders (say, banks).
3. They are in a better position to monitor the progress of the loan to insure repayment.

This means that the village moneylenders are in a better position for dealing with the problems of *transaction costs*, *adverse selection* and *moral hazard*. The problem with these village moneylenders is that they have limited funds and thus can provide credit for limited number of people only. This results in a very high interest rate. Consequently, many profitable projects, which could have been undertaken by the potential borrowers (peasants), will not be realized. Improved access to credit could lead to an increase in investment, and a corresponding increase in the standards of living (social welfare).

Microfinance Institutions

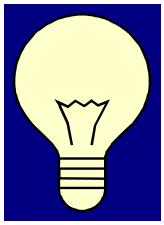
Muhammed Yunas, an American-trained economist from Bangladesh, has developed an institution known as the ***Grameen Bank*** (village bank) to address some of the above problems. This kind of village bank is widely known as microfinance institution.

The Grameen Bank (or microfinance) works as follows. In the Grameen plan, (five) entrepreneurs with separate projects get together and apply for loan as a group. If the loan is approved, two members of the group get their loan and commence (start) their investment activity. If they are successful in meeting the repayment schedule, two more members get loans. If they are also successful, the last member (the group leader) will get a loan. In doing so, the Grameen Bank avoids the problems of both formal financial institutions (banks) and local (village) moneylenders (at least partially):

i. The interest rate is much lower than that of local moneylenders, implying the possibility of more investments realized.
ii. Since the quality of the group influences whether or not individual members will get loans, potential members are highly selective about whom they will join with. As a result, the problem of adverse selection is avoided to some extent.

iii. Since members of the group can only get loans if other members succeed in their investments, there are strong incentives to help each other out and share expertise. This way they avoid or at least minimize the problem of moral hazard.
iv. The activities of choosing candidates for loans and monitoring the progress of the repayments are all done by the peasants themselves, not directly by the loan officers at the bank. This way, microfinance institutions reduce the transaction costs of selecting and monitoring borrowers.

Empirically, the Grameen bank has been very successful in encouraging investment and raising the loan-recovery rate (from 30 – 40% to 98% in Bangladesh as an example). This success has led to the widespread adoption of the system throughout the third world, though with some modifications to suit particular realities.



6.5 Summary

1. *The market mechanism sometimes fails to maximize the welfare of a society. The reasons behind such a failure are the prevalence of market power and asymmetric information in at least some parts of an economy, and the existence of externalities and public goods even in otherwise perfectly competitive markets.*
2. *Externalities are harmful and beneficial side effects borne by those not directly involved in the production or consumption of a commodity. Market prices do not correctly reflect the activities of either producers or consumers (there will be divergence between MSB and MPB or MSC and MPC) in the presence of externalities. With positive externalities (external benefits) of production or consumption, the benefit to the society (as measured by the price the society is willing to pay) exceeds the full social cost of the commodity (on marginal basis) and too little of the commodity is produced or consumed. With negative*

externalities (external costs) of production or consumption, the MSB falls short of the MSC and too much of the commodity is produced or consumed.

3. *Public goods are goods that are non-rival in consumption. Non-rivalry in consumption means consumption of a good by some individual does not reduce the amount available for others (at zero marginal cost). Some public goods, such as national defense, are also non-excludable, i.e., once provided, it is impossible or prohibitively costly to confine its use to only those paying for it. Because of the free-rider problem, public goods are usually under produced or under consumed.*
4. *Whenever there is a firm (or firms) with some monopoly power in either the product or input market, there will be no Pareto-optimality in the economy in general. Monopoly in the product market implies divergence between price and marginal cost at equilibrium [P > MR (= MC)]. Assuming perfect competition in the market for Y and imperfect competition in the market for X,* $MRPT_{XY} = MC_X/MC_Y < P_X/P_Y$ *(because* $P_X > MC_X$ *and* $P_Y = MC_Y$ *at equilibrium). Analogous arguments can be made for imperfect competition in the market for an input.*
5. *Asymmetric information is a situation in which the two sides of a market transaction possess different levels of information. It gives rise to problems like adverse selection and moral hazard, which in turn result in too few (or no) transactions for goods and services. Adverse selection (the hidden information problem) is the situation where low-quality products drive high-quality products out of the market due to the existence of asymmetric information between buyers and sellers. The markets for used-cars, insurance markets, and the labor market are the usual examples where such a problem is faced. Sellers of goods or services usually send messages (or signs) to buyers in order to convey information about the quality of their goods or services. The process of conveying such information is termed market signaling and helps much in making markets perform better. Moral hazard refers to a situation where one side of the market cannot observe the action of the other side as in the insurance market where the insured party (without being observed by the insurance company) can affect the probability or*

magnitude of a payment associated with an event. The problems of adverse selection and moral hazard, together with huge transaction and monitoring costs, are responsible for malfunctioning of financial institutions. These problems in the financial sector have largely been reduced by the introduction of group lending (Grameen Bank).



6.6 Review Questions

I. Multiple-Choice Questions

1. For every unit produced beyond the level at which P = MC:
   A. society uses up resources that have a value (or cost) in excess of the benefits that consumers place on that unit.
   B. society uses up resources that have a value (or cost) that is less than the benefits that consumers place on that unit.
   C. society begins to enjoy the benefits of additional production.
   D. society stops enjoying the benefits of additional production.
2. Which of the following is the direct approach to solving externality problems?
   A. Government-imposed taxes and subsidies.
   B. Private bargaining and negotiation.
   C. The sale or auctioning of rights.
   D. Government regulation.
3. The use of a tax to solve an externality problem is an attempt to accomplish all of the following except one. Which one is it?
   A. Force decision makers to consider the full cost of their decisions.
   B. Fully compensate the damaged parties.
   C. Provide an incentive for firms to use a different technology.
   D. Force firms to stop production of externality-producing activities.
   E. None of the above.

4. Adverse selection in the labor market arises when employers are unable to tell the difference between (and pay the same wages to) more productive and less productive workers. As a result:
   A. the market is unable to find jobs for less productive workers.
   B. less productive workers are replaced by more productive workers.
   C. only the least productive workers will eventually remain in the market.
   D. the distribution of income between productive and unproductive workers tends to become more equal.
   E. the labor market must establish a minimum wage to be paid to less productive workers.
5. Moral hazard is an information problem that arises when:
   A. a buyer or a seller enters into an exchange with another party who has more information.
   B. high quality goods tend to disappear, leaving only low quality goods in the market.
   C. a party to a contract passes the cost of his or her behavior on to the other party to the contract.
   D. only social pressure or moral responsibility forces people to make voluntary payments.
   E. The rules of morality are absent from market transactions.
6. Inefficient allocation of resources that occurs because single firms have some control over price and potential competition is termed market failure due to:
   A. imperfect markets.
   B. imperfect information.
   C. externalities.
   D. public goods.
   E. Pareto optimality.
7. The sources of market failure are all of the following except:
   A. imperfect market structure.
   B. the existence of public goods.
   C. the presence of external costs and benefits.

D. imperfect information.

E. None of the above.

8. Inefficient allocation of resources that occurs because households do not know all there is to know about product availability, quality, and price and firms do not know all there is to know about factor availability, quality, and price is termed market failure due to:

A. imperfect markets.

B. imperfect information.

C. externalities.

D. public goods.

E. Pareto optimality.

9. Externalities are:

A. always negative.

B. costs or benefits bestowed on the individual or group that is part of a transaction.

C. a problem only if decision makers do not take them into account.

D. automatically considered by the market mechanism.

E. not a problem in planned economies.

10. The total cost to society of producing an additional unit of a good or service is its:

A. marginal private cost.

B. marginal social cost.

C. price.

D. marginal-cost price.

E. marginal damage cost.

11. Externalities may be internalized by:

A. bargaining and negotiation.

B. government-imposed taxes and subsidies.

C. legal rules and procedures.

D. Regulation of production level.

E. All of the above.

12. A good is non-rival in consumption when it:
    A. has spillover effects.
    B. is priced at its marginal social cost.
    C. is sold at auction.
    D. can be enjoyed by one person without that interfering with another's consumption.
    E. is non-excludable.
13. The reasons for government failure include all of the following except the:
    A. problems with measuring social costs and damages.
    B. difficulty in determining citizens' preferences.
    C. inefficiency of the government itself due to the lack of market discipline.
    D. likelihood that elected and appointed officials will act in their own self-interest and not necessarily for the good of society.
    E. None of the above.

II. True/False Questions

14. Taxes on externality-producing activities are designed to eliminate the externality.
15. Sometimes, the most efficient solution to an externality problem is for the damaged party to avoid the damage.
16. According to Coase, bargaining will bring the contending parties to the right solution regardless of where the rights are initially assigned.
17. A modern solution to the problem of pollution may involve the sale of the right to pollute.
18. The government can hardly obtain enough information to measure society's demands for public goods accurately.
19. When high-quality goods are mixed with low-quality goods in the same market and with imperfect information, the high quality goods tend to disappear.

III. Discussion Questions

20. State the factors that are responsible for market failure, and briefly explain how each of these factors contributes to the failure.
21. Explain the concept of adverse selection with the help of an example and discuss how the relevant economic agents (in your example) try to overcome the problem.
22. The well-being of the small-scale farmers in rural Ethiopia would be improved if they could get credit at a reasonable rate. However, most of the time big financial institutions like Commercial Bank of Ethiopia are not expected to supply loans to small- scale farmers (borrowers).
    a) What economic reasons do these institutions have in deciding not to provide loans to the farmers? (Or, what problems do they face?) Explain each!
    b) How could each of these problems be solved by the introduction of group lending?

ANSWERS TO MULTIPLE-CHOICE AND TRUE/FALSE REVIEW QUESTIONS

CHAPTER ONE	***CHAPTER TWO***	***CHAPTER THREE***	***CHAPTER FOUR***	***CHAPTER FIVE***	***CHAPTER SIX***
1. C	*1. D*	*1. C*	*1. C*	*1. A*	*1. A*
2. D	*2. E*	*2. D*	*2. C*	*2. D*	*2. D*
3. A	*3. B*	*3. D*	*3. C*	*3. D*	*3. D*
4. C	*4. C*	*4. C*	*4. C*	*4. A*	*4. C*
5. B	*5. D*	*5. B*	*5. C*	*5. FALSE*	*5. C*
6. D	*6. C*	*6. B*	*6. A*	*6. TRUE*	*6. A*
7. C	*7. E*	*7. FALSE*	*7. E*	*7. TRUE*	*7. E*
8. B	*8. E*	*8. FALSE*	*8. C*	*8. FALSE*	*8. B*
9. B	*9. A*	*9. TRUE*	*9. D*	*9. FALSE*	*9. C*
10. A	*10. C*		*10. B*	*10. TRUE*	*10. B*
11. A	*11. A*		*11. C*	*11. TRUE*	*11. E*
12. TRUE	*12. D*		*12. A*	*12. TRUE*	*12. D*
13. TRUE	*13. C*		*13. D*	*13. TRUE*	*13. E*
14. TRUE	*14. C*		*14. C*	*14. TRUE*	*14. FALSE*
15. FALSE	*15. B*		*15. D*	*15. FALSE*	*15. TRUE*
16. TRUE	*16. C*		*16. D*	*16. FALSE*	*16. TRUE*
17. TRUE	*17. D*		*17. D*	*17. FALSE*	*17. TRUE*
	18. TRUE		*18. B*	*18. FALSE*	*18. TRUE*
	19. TRUE		*19. C*		*19. TRUE*
	20. TRUE		*20. C*		
	21. TRUE				

REFERENCES

Dwivedi, D.N. (1997). *Microeconomic Theory*. 3rd ed. Vikas Publishing House Pvt Ltd, New Delhi.

Ferguson, C.E. and Gould J.P. (1989). *Microeconomic Theory*. 6th ed. Irwin Publications.

Hyman, D. (2005). *Public Finance: A Contemporary Application of Theory to Policy*. 8th ed. South-Western, Thompson: USA.

Henderson, M. and Quandt E. (1980). *Microeconomic Theory: A Mathematical Approach*. 3rd edition. McGraw Hill.

Koutsoyiannis, A. (1981). *Modern Microeconomics*. 2nd edition. St Martins Pr.

Mansfield, E. (1988). *Microeconomics: Theory and Applications*. Shorter sixth edition. W.W. Norton & Company: New York, London.

McConnell C. R., S. L. Brue and D.A. Macpherson. (2003). *Contemporary Labor Economics*. 7th edition. McGraw-Hill, Irwin.

Pindyck, R. S. and D.L. Rubinfeld (1995). *Microeconomics*. 3rd ed. Prentice-Hall International, Inc.

Prentice-Hall (2000-2001). Distance Learning at Prentice Hall. Prentice-Hall, Inc. Retrieved from http://www.prenhall.com/mischtm/distlearn.html.

Salvatore, D. (2003). *Microeconomics: Theory and Applications.* 4th ed. Oxford University Press, New York.

Varian, Hal R. (1992). *Microeconomic Analysis*. 3rd ed. W.W. Norton & Company: USA.

Varian, Hal R. (2002). *Intermediate Microeconomics: A Modern Approach*. 6th ed. W.W. Norton & Company.