E5 Managerial Economics

Commonwealth Executive Master of Business Administration
Commonwealth Executive Master of Public Administration

Code 5048

Department of Business Administration
ALLAMA IQBAL OPEN UNIVERSITY

Commonwealth of Learning Executive MBA/MPA

E5 Managerial Economics

Code 5048

Tutor Guide to the Case Studies

Acknowledgment

In preparing this course, I was assisted by Dan Kennedy, a recent International MBA graduate, with distinction, of Schulich School of Business at York University, who wrote all the eight cases for this course and who helped me with valuable suggestions and comments.

Farrokh Zandi, Ph.D.

Course author: Farrokh Zandi, Ph.D.
Professor, Schulich School of Business
York University
Toronto, Canada

Course designer: Douglas R. Cronk, M.A., ABD
Vancouver, BC, Canada

Copyright © Commonwealth of Learning, 2003
All rights reserved. No part of this course may be reproduced in any form by any means without prior permission in writing from:

The Commonwealth of Learning
1285 West Broadway
Suite 600
Vancouver, BC V6H 3X8
CANADA
e-mail: info@col.org

Contents

1 Introduction 1

2 Case Studies 2

2.1 Monas Jute: Rising to the Challenge in Bangladesh 3

2.2 The Canadian Halal Poultry Business – Fatima Foods 4

2.3 Flores de Cali: The Cut Flower Industry of Colombia 5

2.4 The Manjula Plantation's Monsoon Malabar Dilemma 6

2.5 Siddiqi Fine Leather Unlimited – Competing for Greater Exports in Pakistan 7

2.6 Tan Bien Cashews 8

2.7 Warung Pojok: Going from Corner Stall to Mainstream in Indonesia 9

2.8 Building a Business in Bhutan – Where to Begin? 10

Tutor Guide to the Case Studies

1 Introduction

The ongoing challenge of business schools has been to effectively translate the theory and knowledge of a textbook and classroom setting, as well as by distance education, into skills useful for decision making in an office environment. The answer to this challenge has relied partially on the use of businesses cases – examples from real business situations that have been developed to aid students in applying their learning in a controlled environment. Cases vary from discipline to discipline, but very often then put the student in the role of consultant or advisor to answer a specific problem.

This, of course, is the trick of cases. There is rarely one specific problem to answer, and there is never a correct answer to be developed.

A question on price elasticity of demand has one correct answer – if you have the variables, or the information by which to determine the variables, you should, with training, answer the question correctly. Cases are not nearly as straightforward. A good answer to a case focuses on working through the information that is available, identifying what is not available, making reasonable and stated assumptions where possible, and developing alternatives as a result. One student, or set of students, may develop a series of alternatives and recommendations that are completely different than another set. As long as both sets are substantiated, developed, and can be implemented, they are both equally correct.

Cases ask a variety of questions that are meant to both lead, and mislead, the students. As cases have no cut-and-dried answers, they are written with a purpose to challenge students to work with the pertinent information, identifying and setting aside information that is secondary to the main task at hand. Very often, at the root of a problem, the first question that students may decide to ask is, 'Is this business venture or business situation a viable one?' Once they have answered this satisfactorily, they will be in a good position by which to develop their alternatives and recommendations.

A note on the development of alternatives: it is the opinion of the course authors that each alternative developed should be a separate and individual alternative from the others. That is to say, a third alternative developed should not be a combination of the first and second alternatives. It is important to identify as many possible paths for a company to take, and to test these paths, before deciding on a course of action. A recommendation can certainly be built on two or more alternatives, as long as, strategically and economically, this is the best business decision possible.

The following cases have been developed to help put the student in the traditional role of advisor to a given specific business situation, to draw on the tools they have learned through the ten blocks of this course. Not all ten blocks are equally applicable to each case – the students will have to approach each case separately, and identify what tools

are, and are not, needed. The one unifying factor of each case is the necessity to use business strategy (Block 10) to help approach each case, and even here, not all strategic models are needed for each case.

2 Case Studies

The eight cases listed below draw on real data from situations, industries, and companies around the world. Each case has been developed so that students understand both the internal challenges facing the company and the socio-economic environment within which the company operates. The industries examined, including flower exports, leather work, coffee growing, and textile exports, have all been framed by such countries as Colombia, Pakistan, India, and Bhutan, respectively. Each company has been given a fictitious name and fictitious data, with the national data available being as current and correct as resources allow. The data is purposefully incomplete, as students will never have all the data they need. Therefore, they are encouraged to round out the information with additional research, if they and the instructor so choose, or make reasonable assumptions upon which to craft responses. Note that it is the authors' opinion that the best learning from cases comes in groups – group discussion helps foster teamwork and builds a common learning platform. Groups of no more than four should be used to analyse each case, unless the instructor deems otherwise. Note that two of the eight cases are required assignments; it is recommended that one of these assignment cases is group based. However, if both assignment cases are individual based, this still leaves you the possibility of using one or more of the remaining six cases as practice cases for group based activities.

The 'Key Lessons To Learn/Models To Use' for each case share common elements, with attention given to specific learning opportunities. These, and the 'Discussion Points' have been developed to guide the tutor down *possible* paths of discussion and learning. However, these cases are intended to draw from all ten of the chapters in this course, and therefore a complete answer should involve considering economic theory as well as strategic possibilities.

It is important to restate that the best answer is a well developed, well-supported, realistic answer considering the viability of the company and its position within an industry.

2.1 Monas Jute: Rising to the Challenge in Bangladesh

Key Lessons to Learn/Models to Use

1. Porter's Five Forces Model analysis
    - Use to display the competitiveness of the Jute industry, concentrating on Bangladesh and India.
    - What are the economic drivers that help define the Jute industry?
2. SWOT analysis for Monas Jute
    - Are the company's strengths and opportunities in line with the industry's drivers?
3. Ansoff Product/Market Mix analysis
    - What options for Monas Jute can be identified with this model?
    - Would they help make the company more competitive? How?

Discussion Points

1. Little financial information is available about the company, or the costs of its products. However, given the current atmosphere of the jute industry and the information provided, what is the long-term sustainability of the company's current business model?
2. Does Bangladesh have a competitive advantage, as a nation, in jute? If so, what are its factors? If not, can it develop one?
3. Short of government support, is there anything that the industry itself could do to make itself more competitive?
    - Develop common jute standards, or focus the industry as a whole in a specific direction, thus creating an industry-wide competitive advantage?

2.2 The Canadian Halal Poultry Business – Fatima Foods

Key Lessons to Learn/Models to Use

1. Porter's Five Forces Model analysis
    - Pay particular attention to substitutes – non-halal versus halal chicken, beef, and lamb.
    - What are the drivers for this industry?
2. SWOT analysis for Fatima Foods
    - Are the company's strengths and opportunities in line with the industry's drivers?
    - Analysed objectively, is the market large enough to sustain two definitions of halal chicken?

Discussion Points

1. Suppose the mechanized slaughter of chickens will eventually cut the cost of producing this version of halal meat to half of the cost of producing halal meat through non-mechanized means. The economics of the situation would dictate that only those consumers able to pay twice the price for the second type of halal meat would do so. Is this enough demand to sustain a business on?
    - What other markets are potential markets for this more strict, and more costly, definition of halal meat? What additional information would you need to analyse these potential markets?
2. While a factor of differentiation, the strict definition of halal slaughter may prove cost ineffective in the long run. What factors would help the company analyse what their future profitability opportunities are?

2.3 Flores de Cali: The Cut Flower Industry of Colombia

Key Lessons to Learn/Models to Use

1. Porter's Five Forces Model analysis
    - Use the model to understand the competition within the industry worldwide, not just Colombia.
    - Is the Colombian 'brand' of flowers in a defensible position? What is its greatest threat?
    - What are the drivers for this industry?
2. PEST Analysis for Colombia
    - What factors within Colombia may weaken its industry's competitiveness?
    - What could the industry do to strengthen its position, either by dealing with this weakness, or strengthening the industry in another way?
3. Ansoff Product/Market Mix analysis
    - Is there a solution to the industry's challenges when its current products are viewed through this model?

Discussion Points

1. The greatest demand for cut flowers occurs at very few, very specific times of year. What defines this demand? Could it be increased? Are the drivers of this demand 'rational' (predictable and repeatable) drivers?
2. In advising Manuel Torres, what would you describe as the single greatest factor inhibiting the growth of the cut flower industry? What could an industry association like ASOCOFLORES propose that might draw support from ProExport?

Given the growing trends in Fair Trade practices around the world, how might this affect the industry in Colombia? (Fair Trade can be briefly defined as paying a farmer a 'living wage' for his/her crop. The consumer agrees to pay a slight increase in price as they are interested in supporting this 'living wage.')

2.4 The Manjula Plantation's Monsoon Malabar Dilemma

Key Lessons to Learn/Models to Use

1. Porter's Five Forces Model analysis
    - Use the model to understand the competition within the industry worldwide, not just India.
    - Examine both the tangible (taste and country of origin) aspects of substitutes (coffee, tea, other beverages) and the intangible (brand, association with lifestyle, etc.).
    - What are the drivers for this industry?
2. SWOT analysis for Fatima Foods
    - Are the company's strengths and opportunities in line with the industry's drivers?
    - What are Manjula plantation's competitive advantages? Are they only coffee specific, or are they transferable to other crops?
3. Use Game Theory to analyse what other rival producers of coffee may do, given the current price of coffee?
    - Is there an issue of Prisoner's Dilemma here?
    - What is the best course of action for Manjula to consider, if prices stay depressed for the foreseeable future?

Discussion Points

1. Consider the other crops Manjula produces. What impact do they have on Manjula's current operations? Should Manjula continue to produce multiple crops?
    - If more information is needed, students are encouraged to find factual market information not already provided.
2. What could the plantation do to reduce the high costs of production? How would this effect the quality of the coffee?
3. Should the company consider refocusing on exporting pre-roasted, pre-packaged coffee under its own label? What costs would the company have to consider for such a venture? Do they have any competitive strengths that would allow them to do this effectively?

2.5 Siddiqi Fine Leather Unlimited – Competing for Greater Exports in Pakistan

Key Lessons to Learn/Models to Use

1. Porter's Five Forces Model analysis
    - Use the model to understand the competition both within Pakistan and the industry worldwide.
        - Examine, in particular, the components of the industry that could affect costs of production.
    - What are the drivers for this industry?
2. PEST Analysis for Colombia
    - What factors affect Pakistan's competitiveness on the global stage?
    - What could the industry do to strengthen its position, either by dealing with this weakness, or strengthening the industry in another way?
3. SWOT Analysis for Siddiqi Fine Leather
    - Are the company's strengths and opportunities in line with the industry's drivers?
    - What are Siddiqi's competitive advantage? Would you suggest that they expand and acquire additional competitive advantages, or develop them internally?

Discussion Points

1. What is the most potentially successful component of the Pakistani leather industry in the near future?
2. Given the PEST analysis, is Pakistan in a position to comply with an increase in environmental regulations internally? Increased international environmental regulations?
3. As a manager, what are the two biggest components to future success for Siddiqi Fine Leather?

2.6 Tan Bien Cashews

Key Lessons to Learn/Models to Use

1. SWOT Analysis for Tan Bien Cashews
    - Are the company's strengths and opportunities in line with the industry's drivers?
    - Would becoming part of the Vietnamese Stock exchange strengthen, or weaken, its short-to-midterm position?
    - What are the drivers for joining the stock market?
2. PEST analysis of Vietnam
    - Consider the current business environment in Vietnam – does it suggest that additional investment in private industry will take place?
    - Is the interest in the stock market long-term?
    - Aside from the Vietnamese stock exchange, what other opportunities exist for a large, successful firm like Tan Bien Cashews?

Discussion Points

1. What is the most promising market for Vietnamese cashews?
    - What forms should these cashews be in?
2. Where are Tan Bien's competitive advantages? What can they do to protect their position in the industry while building a global clientele?
3. As a consultant, would you consider listing on the Vietnamese Bourse an advisable move? Should the company stay private in the short-term? Why or why not?

2.7 Warung Pojok: Going from Corner Stall to Mainstream in Indonesia

Key Lessons to Learn/Models to Use

1. PEST analysis of Indonesia
    - Consider the current business environment in Indonesia – does it suggest that additional investment in private industry will take place? Will entrepreneurs continue to be successful as they have in the recent past?
    - What are the single greatest economic challenges Warung Pojok may face in the short-term?
2. SWOT Analysis for Warung Pojok
    - Are the company's strengths and opportunities in line with the industry's drivers? Are they as clear here as in other cases?
    - What is the competitive advantage Warung Pojok has?
3. Use Game Theory to assess what Mr. Indra should do in the current environment.
    - Use a strong economy and a weak economy as the two players in a matrix, which will affect Warung Pojok. Is there a no-lose situation Mr. Indra can take, given your analysis?

Discussion Points

1. Although he appears to have little formal training in managerial economics or strategy, Mr. Indra is certainly very successful. Is Mr. Indra a creation of his circumstances? Could his success have been duplicated in a more stable economy?
    - If yes, what are the key components of his success, and how should they be replicated?
    - If no, consider whether he can take his current situation and maintain it in better economic circumstances.
2. As a consultant, what would be the greatest threat to the expansion of Warung Pojok?
    - Consider what impacts the price elastic ties of demand.
    - Consider your PEST analysis (hint – look at a map of Indonesia).

2.8 Building a Business in Bhutan – Where to Begin?

Key Lessons to Learn/Models to Use

1. Porter's Five Forces Model analysis
    - Use the model to understand the environment in Bhutan for start-up businesses. What are the barriers to entry?
    - Structure your impression of the business environment as an ongoing analysis of whether Dasho Lody should ultimately start a business, or buy an existing one (that would not be weaving?)
2. Game Theory concepts
    - What are the benefits and draw-backs of first mover advantage?
    - How does the involvement of a UN agency change the Prisoner's Dilemma for Dasho?
3. PEST analysis for Bhutan
    - Consider whether Bhutan is ready for an internet-based business.

Discussion Points

1. Draw up a list of economic variables Dasho would need to consider as he reviews his option to start a company. Are these variables inhibiting, or can Dasho's education deal with, and overcome, them?
2. How strong is the demand for Bhutanese textiles at the moment? What could be done to promote an increase in demand? (In other words, what is the potential market size for Bhutanese textiles?)
    - This will require additional research on the part of students.
    - What would Dasho have to consider if he wanted to open, or purchase, an existing Internet café? Is this a better alternative for him than the textiles opportunity?

Commonwealth of Learning Executive MBA/MPA

E5 Managerial Economics

Block 1

Introduction to Managerial Economics

Course author: Farrokh Zandi, PhD
Professor, Schulich School of Business
York University
Toronto, Canada

Course designer: Douglas R. Cronk, M.A., ABD
Vancouver, BC, Canada

Copyright © Commonwealth of Learning, 2003
All rights reserved. No part of this course may be reproduced in any form by any means without prior permission in writing from:

The Commonwealth of Learning
1285 West Broadway
Suite 600
Vancouver, BC V6H 3X8
CANADA
e-mail: info@col.org

Contents

1 Block One: Objectives and Introduction 1
1.1 Introduction 1
2 The Nature of Managerial Economics 1
2.1 Economic Theory and Managerial Economics 2
2.2 Positive versus Normative Economics 2
2.3 The Use of Models in Managerial Economics 3
3 The Theory of the Firm 4
3.1 The Objective and Value of the Firm 5
3.2 Certainty, Risk, and Uncertainty 7
3.3 Limitations of the Profit-Maximization Model 8
3.4 Sales (Revenue) Maximization Model 9
3.5 The Agency Model 9
3.6 Marginal Analysis 10
3.6.1 Demonstration Problem 11
4 Summary and Review 12
5 Review Problems 13
6 Answer Key to Review Problems 14
7 References and Suggested Further Reading 15

1 Block One: Objectives and Introduction

After working through Block One of this course, you should be able to:

1. Understand the scope of Managerial Economics.
2. Distingush between two different approches to Economics: positive and normative.
3. Calculate the value of the firm.
4. Understand the nature of decision making under risk and uncertainty.
5. Understand the implicatons of the firm's different objectives.

1.1 Introduction

The application of economic theory and its stock of analytical tools to everyday activities is no easy matter, as many frustrated students of economics have found out. 'The economic theory is very elegant, but how do I apply it to real-life problems?' Because this is a legitimate question for which there are no easy answers, the branch of economic study called managerial economics emerged in the 1940s to bridge the gap between theory and practice. Managerial economics today is an integral part of both undergraduate and graduate programs in business administration. It is also a popular offering in the curricula of other professional schools, such as those in engineering, public administration, public health, law, medicine, and urban and regional planning. This is because the principles and techniques of managerial economics apply not only to profit-seeking organizations, but to any organization that wants to optimize the use of its resources, whether that organization is domestic or international.

2 The Nature of Managerial Economics

This section introduces you to the subject and then proceeds to build the foundation necessary for decision making under uncertainty. Decision making under uncertainty is a central theme of managerial economics discussed in section 4. In this section, we first examine the definition and scope of managerial economics. This leads to a discussion of the use of 'models' in managerial economics. We will see that verbal, graphical, and algebraic models depicting the behaviour of consumers and business firms can be of substantial assistance in decision analysis.

We will then discuss the issue of decision making under uncertainty. In business the outcome of a decision is usually far from certain (at the time the decision is taken) because the decision maker has incomplete information and the outcome depends on the simultaneous behaviour of rival firms and other factors influencing the managerial cost

and demand conditions. When the outcome of a decision is not predictable with certainty, we say that the decision is made under conditions of risk or uncertainty.

To some, managerial economics is the application of economic theory (particularly microeconomic theory) to problems concerning the optimal allocation of economic resources. Other scholars view managerial economics as applied microeconomics, an approach requiring the integration of principles and practices from other functional areas such as accounting, finance, marketing, and management. Still other scholars see economics as a link between economic theory and decision science. In this capacity it is to contribute to sound decision making not only in business-oriented profit-making organizations, but also in government agencies and non-profit organizations.

2.1 Economic Theory and Managerial Economics

Economic theory traditionally is divided into two broad subfields: microeconomics and macroeconomics. *Macroeconomics* focuses on aggregate economic conditions. Those are conditions that set the environment within which a business operates. On the other hand, *microeconomics* focuses on the economic forces that influence the decisions made by individual consumers, firms, and industries.

Macroeconomics is concerned with the economy as a whole. It is thus concerned with aggregate demand and aggregate supply. Aggregate demand is the total amount of spending in the economy, whether by consumers, by overseas customers for our exports, by the government, or by firms when they buy capital equipment or stock up on raw materials. Aggregate supply is the total national output of goods and services.

Microeconomics is concerned with the individual parts of the economy. It is concerned with the demand and supply of *particular* goods and services and resources: cars, butter; clothes, and haircuts; electricians, secretaries, blast furnaces, computers, and coal. The crux of microeconomic influences on business decision-making is the answer to the two-part question: How much should the firm produce, and how much should it charge for this output?

Since microeconomic theory deals with topics that are more closely related to the firm's behaviour in resource allocation (namely, consumer behaviour, production and cost analysis, market structure and pricing, profit planning, and finance), it makes a major contribution to managerial economics. However, it does not operate in a vacuum. Its environment is in the global setting, which both affects and is affected by the activities of individuals who are producers and consumers all over the world.

2.2 Positive versus Normative Economics

In economics, the distinction is made between those areas that are positive and those that are normative. Positive economics is descriptive: it describes how economic agents or economics systems *do operate* within the economy or society. Normative economics, on the other hand, is prescriptive: it prescribes how economic agents or systems *should operate* to attain desired objectives.

Managerial economics is primarily normative, since it seeks to establish rules and principles to be applied in decision making to attain the desired objectives. But managerial economics must start from a positive perspective: we must always be mindful of the actual practices in the business or institutional environment. For example, if firms choose their price level by applying a mark-up to their direct costs rather than by equating marginal revenue and marginal costs as implied by microeconomic principles, managerial economists should seek the optimal level of the mark-up rather than making speeches to deaf ears about marginalist principles. The approach taken in this course is to integrate business practice with economics. By reference to the microeconomic principles, the business practice can be evaluated in terms of its efficiency in attaining the desired objectives.

Central to the study of managerial economics is the concept of profit. Profit is defined as the excess of revenues over costs. For a non-profit institution, an excess of revenues over costs is called a surplus. If costs exceed revenues there is a loss, known as a deficit in the non-profit sector. No firm or organization, not even the national government, can incur losses or deficits forever. The objective of a business firm usually requires profit maximization over its time horizon, which may be short term or long term depending on the firm and its circumstances. Non-profit institutions typically do not seek to make a surplus but wish to spend their available funds to maximum effect. The decision problems facing business and non-business institutions are therefore essentially similar, involving revenue enhancement if possible and cost control wherever possible. The attainment of a profit (or the avoidance of a deficit) is seen as a measure of managerial effectiveness.

2.3 The Use of Models in Managerial Economics

Economic theories seek to predict and explain economic behaviour. Economic theories usually begin with a *model*. This abstracts from the details surrounding an event and seeks to identify a few of the most important determinants of the event. For example, the theory of the firm assumes that the firm seeks to maximize profits, and on that basis it predicts how much of a particular commodity the firm should produce under different forms of market structure or organization. While the firm may have other (multiple) aims, the profit-maximization model accurately predicts the behaviour of firms, and, therefore, we accept it. Thus, methodology of economics (and science in general) is to accept a theory or model if it predicts accurately and if the predictions follow logically from the assumptions.

Managerial economics is also closely related to the decision sciences. These utilize the tools of mathematical economics and econometrics to construct and estimate decision models aimed at determining the optimal behaviour of the firm (i.e., how the firm can achieve its goals most efficiently). Specifically, mathematical economics is used to formalize (i.e., to express in equational form) the economic models postulated by economic theory. Econometrics then applies statistical tools (particularly regression analysis) to real-world data to estimate the models postulated by economic theory and for forecasting.

Throughout this course you will see hundreds of special words that comprise the jargon of economics, dozens of graphs, and more than a few algebraic equations. The jargon, graphs, and equations are used to model relationships that exist between and among variables and to facilitate analysis and decision making.

In mathematical economics, economic relationships are expressed in mathematical form thereby rendering them amenable to empirical testing or other modelling techniques. In addition to making the research for solutions easier, mathematical representations often permit insights into problems that would easily be missed in a purely descriptive approach.

3 The Theory of the Firm

In this section, we examine first the reason for the existence of firms and their principal functions. Then we define the value of the firm and the constraints under which it operates. Finally, we discuss the limitations of the theory of the firm. This is a very important section since the theory of the firm is the centrepiece and central theme of managerial economics.

A *firm* is an organization that combines and organizes resources for the purpose of producing goods and/or services for sale. There are hundreds of thousands of firms in a country such as Canada. These include proprietorships (firms owned by one individual), partnerships (firms owned by two or more individuals), and corporations (owned by stockholders). Firms produce almost 80 per cent of all goods and services consumed in Canada. The remainder is produced by the government and not-for-profit organizations, such as museums and foundations.

Firms exist because it would be very inefficient and costly for entrepreneurs to enter into and enforce contracts with workers and owners of capital, land, and other resources for each separate step of the production and distribution process. Instead, entrepreneurs usually enter into longer-term, broader contracts with labour to perform a number of tasks for a specific wage and fringe benefits. Such a general contract is much less costly than numerous specific contracts and is highly advantageous both to the entrepreneurs and to the workers and other resource owners. By internalizing many transactions (i.e., by performing within the firm many functions), the firm also saves on sales taxes and avoids price controls and other government regulations, which apply only to transactions among firms.

On the other hand, firms do not continue to grow larger and larger indefinitely because of limitations on management ability to effectively control and direct the operation of the firm as it becomes larger and larger. It is true that up to a point, a firm can overcome these internal disadvantages of large size or diseconomies of scale by establishing a number of semi-autonomous divisions (i.e., by decentralizing). Eventually, however, the

increased communication traffic that is generated coupled with the further and further distancing of top management from the operation of each division imposes sufficient diseconomies of scale to limit the growth of the firm. Furthermore, the firm will reach a point where the cost of supplying additional services within the firm exceeds the cost of purchasing these services from other firms. An example is provided by some highly technical (legal, medical, or engineering) service that the firm may need only occasionally.

The function of firms, therefore, is to purchase resources or inputs of labour, capital, and raw materials in order to transform them into goods and services for sale. Resource owners (workers and owners of capital, land, and raw materials) then use the income generated from the sale of their services or other resources to firms to purchase the goods and services produced by firms. The circular flow of economic activity is thus complete. In the process of supplying the goods and services that society demands, firms provide employment to workers and pay taxes that government utilizes to provide services (such as national defence, education, and fire protection) that firms could not provide at all or as efficiently.

3.1 The Objective and Value of the Firm

A number of models have been advanced to explain the behaviour of business organizations and their managers in terms of their goals and objectives. These models include the profit maximization model, the sales-maximization model, and the agency model.

It has long been customary for economists to assume that the primary goal of any type of organization is to maximize the benefits provided by the organization's operations in relation to its costs. In the case of a business organization, the benefits it seeks are profits. Since our economic system permits firms in unrelated industries to earn all they can, it was originally proposed that the firm is best described by a profit-maximization model derived from the theory of the firm set forth in microeconomic studies.

Earlier versions of the profit-maximization model concentrated on decision making to maximize short-run profits, that is, to maximize short-run total revenue minus total cost. In later versions of the model, the goal of the firm is assumed to be the maximization of the *value of the firm* or *wealth* in the long run. This is so because many decisions involve cash flows extending beyond the present period. When choosing among alternative decisions, it is important to distinguish between revenues that are received immediately and those that are received at some later date. Since the firm's value in the long run depends upon a stream of future profits that may or may not actually materialize as expected, the model has been expanded to include the time value of money (the present value of future profits) and the concepts of risk. To understand the model, the student must have a clear understanding of the concept of present value.

Note that a dollar received today is worth more than a dollar received next year, because a dollar held today may be deposited in a bank and at the end of one year it will be worth the original dollar plus the interest earned on that dollar. Hence, if the interest rate is, say,

10%, a dollar held today will be worth $1.10 one year from today. Looking at this from the reverse aspect, a dollar earned one year from today has *present value* of only about 91 cents because 91 cents would grow to a dollar (at 10% interest) within the year.

The concept of present value is based upon the principle of compound interest. Suppose that $1 is invested today at i per cent interest per annum: then one year from today the accumulated amount B_1, which is principal plus interest, will be

$$B_1 = \$1 + (\$1 x i)$$
$$B_1 = \$1(1+i)$$

If the investment accumulates interest over two years, two years from now, it will be B_2, where

$$B_2 = \$1(1+i)(1+i) = \$1(1+i)^2$$

The reverse of compounding is called discounting. If we expect B_2 to equal $\$1(1+i)^2$ two years from now, it is worth $1 now when discounted at i per cent; that is

$$\frac{\$1(1+i)^2}{(1+i)^2} = \$1$$

or

$$\frac{B_2}{(1+i)^2} = \$1$$

Therefore, the firm's value in the long run is the present value of a stream of future (expected) profits as follows

$$PV = \frac{\pi_1}{(1+i)} + \frac{\pi_2}{(1+i)^2} + \ldots\ldots\ldots\ldots\ldots + \frac{\pi_n}{(1+i)^n} = \sum_{t=1}^{n} \frac{\pi_t}{(1+i)^t} \qquad (1)$$

where PV is the present value of all expected future profits of the firm, π_1, $\pi_2 \ldots \pi_n$ represent the expected profits in each of the n years considered, and i is the appropriate discount rate used to find the present value of future profits, $\sum$ refers to 'the sum of,' and t refers to time taking on values from 1 to n.

The introduction of the time dimension in equation (1) allows for the consideration of uncertainty. For example, the more uncertain is the stream of expected future profits, the higher is the discount rate that the firm will use, and, therefore, the smaller is the present value of the firm.

Since profits are equal to total revenue (TR) minus total costs (TC), this equation can be rewritten as

$$\text{Value of firm} = PV = \sum_{t=1}^{n} \frac{TR_t - TC_t}{(1+i)^t} \qquad (2)$$

This equation provides a unifying theme for the analysis of managerial decision making and, indeed, for this entire text. Specifically, TR depends on sales or the demand for the firm's output and the firm's pricing decisions. These are the major responsibilities of the marketing department. The TC depends on the technology of production and resource prices. These are the major responsibilities of the production and personnel departments. These are discussed in detail in the following section. The discount rate (i) depends on the perceived risks of the firm and on the cost of borrowing funds. These are the major responsibilities of the finance department.

Note that an important question is: what is the appropriate discount rate? The decision maker must choose the discount rate quite carefully, since use of an inappropriate discount factor could cause a poor decision to be made when the future profit streams of alternative decisions differ markedly.

If the same amount of profit is expected every year, the problem of maximizing the stream of profits reduces simply to the problem of maximizing π. The present value of a stream of equal payments may be calculated as follows

$$PV = \pi \left[\frac{1 - (1+i)^{-n}}{i} \right] \tag{3}$$

where n is the number of years. The expression in brackets is called the discount factor.

3.2 Certainty, Risk, and Uncertainty

The state of information under which a decision is made has important implications for the predictability of the outcome of that decision. If there is full information, the outcome of a decision will be foreseen clearly and unambiguously. In this situation (of certainty) the firm can accurately predict the outcome of each of its decisions. When there is less than full information, however, the decision maker may foresee several potential outcomes to a decision and, therefore, will be unable to predict consistently which outcome will actually occur. In this case we say that the individual or firm is operating under conditions of risk and uncertainty.

Certainty exists if the outcome of a decision is known in advance without a shadow of a doubt. Under conditions of certainty, a decision leads to a single possible outcome, which is perfectly foreseen. *Uncertainty* is involved when a decision might lead to one of several possible outcomes and the exact outcome is not known in advance. Instead, there will be a probability distribution of possible outcomes, which the decision maker must identify.

When the future return is certain, the risk-free discount rate *i* is appropriate. If the future return is not certain, we must use the risky discount rate, say *r*, where *r* equals *i* plus a risk premium to compensate for the risk involved. The magnitude of *r* reflects the degree of risk. It is the firm's capitalization rate or cost of capital, that is, the rate of return

required by an investor after giving due consideration to the firm's business and financial risk.

Risk can be regarded as a subcategory of uncertainty in which the probabilities of each outcome can be assigned on an objective basis. Risk is defined as a state of knowledge in which each alternative leads to one of the set of specific outcomes, with each outcome occurring with a probability that is known objectively to the decision maker. For example, risk is involved when one flips a coin, throws dice, or plays a hand of poker. The probability of flipping a coin and having it land 'heads' is 1/2, since there are only two possible outcomes (ruling out the coin landing on its edges), and each is equally likely to occur, given an unbiased coin. Similarly, when one throws two dice, the probability that they will turn up 'snake eyes' or any other pair of numbers, is 1/6 x 1/6 = 1/36.

In each of the preceding illustrations, the probability of each outcome is known a priori. That is, on the basis of known mathematical and physical principles, we can calculate the probable incidence (or proportion) of each possible outcome in the total number of outcomes associated with a number of trials. That is, if we toss a coin a number of times, we expect to observe 'heads' about half the time. Although 'heads' might appear three or even four times out of the first four tosses of the coin, given a sufficiently large number of trials, the proportions will converge on 1/2 for each of the two possible outcomes.

3.3 Limitations of the Profit-Maximization Model

The principles governing the profit maximization model may tell us much about how firms would like to behave, how they would like to make investment decisions, and how they would prefer to time their activities. In a world in which the timing of investments is crucial to success, the profit maximization model can give important insights into the probable behaviour of firms. But the profit-maximization model, like all models, is a simplified, abstract version of reality. In the real world, there are a number of complexities that limit its descriptive adequacy. This is because in addition to a lack of complete information, (1) it requires the firm to predict accurately the magnitude and timing of a stream of profits to be realized in the future – something that is difficult to do under the best of circumstances and impossible under the worst – and (2) there are many legal, ethical, and social constraints that limit a firm's all-out pursuit of profit. Within this network of constraints, the firm seeks optimal profit.

This means that firms do not necessarily seek maximum profits but instead seek to balance their desire for profit with other goals and objectives – short-run and long-run, economic and non-economic. Achievement of these other goals as well as profits provides maximization of benefits, and this is not necessarily the same as maximization of profit.

Recognition of this fact has led to a number of alternative models that state that the motivation of the firm is something other than maximization of profit. These models,

which are of considerable importance to the understanding of the behaviour of firms, can be grouped into several categories including: S*ales-maximization* models and the *Agency* model.

3.4 Sales (Revenue) Maximization Model

The sales-maximization model is probably the best known alternative to the profit maximization model. This is because it is easily understood and because it can be supported by intuitively appealing anecdotal examples. Rigorous empirical tests, however, have failed to support the strict sales-maximization hypothesis, particularly in terms of the long-run objectives of the firm. However, a number of reasons as to why firms might give greater consideration to sales revenues are identified including that the firm's management may feel that lack of growth in sales will impair the company's reputation and dealings with its customers, distributors, financial institutions, and employees, and that it will reduce the company's influence in the marketplace and make it more vulnerable to competitors.

The proponents of sales maximization recognize that some minimum level of profit is necessary, but they hold that the sales-maximizing firm is willing to sacrifice some or all profit above this minimum in order to increase sales. Casual observations often have been made of this type of behaviour. Japanese firms, in particular, have drawn considerable attention to their efforts to increase their share of the world market by allegedly 'dumping,' i.e., selling their product abroad at prices lower than those charged in their own country, a tactic that is objectionable to the World Trade Organization.

However, the nature of the evidence suggests that such firms may be sacrificing some short-run profits in order to maximize profits in the long run. Sales maximization is part of their strategy to secure an advantageous competitive position that will produce more profit in the long run.

3.5 The Agency Model

The agency model is concerned with contractual relationship between the *principals* (stockholders) and their *agents* (the managers). We have stated that the objective of the firm is to maximize its net worth in (expected) present-value terms. By *the firm* we really mean the owners or shareholders of the firm, whose interests presumably lie in the maximization of their personal net worth. However, in modern corporations that are owned jointly by thousands of stockholders, the owners of the firm are typically excluded from the day-to-day process of making decisions, having delegated the authority to trained managers whose job is to make these decisions. In order for managers to function, the stockholders, acting through their board of directors, must delegate decision-making authority to a chief executive officer (CEO).

There are a number of things that executives might do in pursuit of their own interest rather than maximization of the owners' wealth. For example, managers might enact highly visible policies that make themselves look good and thus insure promotions.

Managers might be more interested in expensive perquisites (perks), such as beautifully decorated offices, luxury automobiles, and memberships in exclusive country clubs, than in reducing costs.

The problem is that the stockholders have no way of knowing whether or not they are doing their best to maximize the owners' wealth. Only in the case of owner-managed firms can we expect the objectives of the owners and the managers to coincide perfectly.

This difficulty has been called the *principal-agent problem*. The manager is an agent of the shareholders (the principals), making decisions on their behalf. It is theoretically possible for stockholders to evaluate the managers' efforts by establishing an elaborate and costly intelligence system to monitor the managers' actions. In reality, of course, the stockholders' interests lie in avoiding such costs. Some agency theorists hold that monitoring management action is unnecessary. They point to empirical data that show a strong relationship between a company's profitability and executive compensation and offer these data as proof that market dynamics are sufficient to lead managers to give their best effort.

Other theorists argue that some costs to the stockholders are inevitable and are best incurred in the form of incentives for managers to give their best efforts for maximization of the owners' wealth. This usually takes the form of an employment contract that provides for an annual bonus directly related to the firm's profits. The bonus may be related to the current year's profits or to the firm's performance over several years in the immediate past.

Such incentive contracts, however, do not guarantee the manager's unflagging pursuit of shareholder objectives, because the problem of asymmetric information remains.

3.6 Marginal Analysis

Under conditions of certainty, the decision maker has complete knowledge of the various states of nature that are relevant to the decision and knows which one will occur. The decision maker simply chooses the strategy, course of action, or project that yields the greatest payoff.

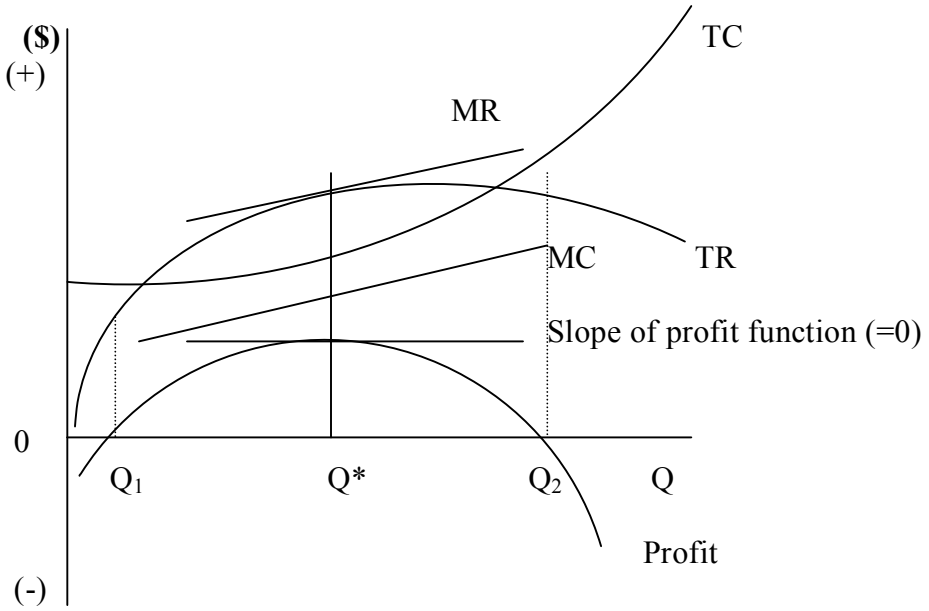
In general, decision making under certainty requires that we find the largest payoff either by maximizing benefits (such as revenue, profit, or utility) or by minimizing costs. We can refer to these activities as optimization analysis. For this purpose, we will focus on marginal analysis, one of a few optimization techniques available to the decision maker.

Here, the objective is to find the level of output and sales that will maximize profit. Marginal analysis is a tool that can be used to find this optimal relationship. It utilizes the concepts of marginal revenue and marginal cost as illustrated by Figure 1-1. The figure shows typical revenue, cost, and profit curves of microeconomic theory.

Marginal revenue (MR) is defined as the additional revenue (change in total revenue) gained from the sale of one more unit of the product. It is graphed as the slope of the total revenue *(TR)* curve. Marginal cost (MC) is defined as the additional cost (change in total

cost) of producing one more unit of output. It is graphed as the slope of the total cost (TC) curve.

Figure 1-1



At output levels Q_1 and Q_2 *TR* is equal to TC, and therefore profit is zero. For $Q > Q_1$ or $Q < Q_2$ the firm faces a loss (negative profit). When output is between Q_1 and Q_2, profit is positive. Marginal analysis shows that profit is maximized at Q*, where MR = MC. This is shown as the largest gap between the TR and TC curves, which coincides with the peak of the profit function. Note that at Q*, the line tangent to the TR curve (MR) is parallel to the line that is tangent to the TC curve (MC). Therefore at the point of profit maximization, MR = MC.

3.6.1 Demonstration Problem

Assume that a family-operated frozen yogurt outlet that sells its product in a competitive market is facing the following price and cost information

Market Price = \$2 (per frozen yogurt)

$TC = 100 - 2Q + 0.02Q^2$, Q = daily units

What is this firm's maximizing quantity and what is its maximum profit?

Answer:

Total Revenue function = TR = P x Q = 2Q MR = 2

Total cost = $TC = 100 - 2Q + 0.02Q^2$ MC = –2 + 0.04Q

At the point of optimality:

MR = MC

2 = –2 + 0.04Q

Q = 100 frozen yogurt per day

π = TR –TC = \$2 (100) – [100 – 2(100) + 0.02(100)^2]

= \$100 per day.

4 Summary and Review

This course takes a prescriptive approach to managerial decisions; that is, it focuses on how managers can use economic analysis to arrive at optimal decisions. The aim of the prescriptive approach is to aid in solving important and difficult real-world decisions. One often hears the complaint, 'That's fine in theory, but it wouldn't work in practice.' There is some validity to this objection; yet, in our view, the criticism misses the main point. To be useful, decision-making principles must be applicable to actual practice. In fact, most of the analytical methods we discuss were developed in response to important, recurring managerial decision problems. Of course, one can expect that different decisions will call for varying degrees of analysis in each step. The degree to which a decision is analyzed (or the amount of information to be gathered) is itself a choice the manager must make. The key is to use the analytical approach flexibly.

In the course of our discussion, we will make frequent reference to the actual practice of managerial decision making – the customary methods by which business and government decisions are made. We need hardly point out that managerial practices frequently differ from our prescriptions. After all, if managers (and future managers like yourself) were always able to analyze perfectly the complex choices they face, there would be little need for courses like this one. Actual managerial practice changes slowly. Many methods and practices accepted as essential by today's managers were unknown or untried by managers of earlier generations. These include many of the core decision methods of this course: information and uncertainty, the Attribute approach to consumer behaviour, cost minimization, and optimal pricing and market segmentation. The challenge of the prescriptive approach is to improve current and future practices.

5 Review Problems

1. What is the relationship between the field of managerial economics and (n) microeconomics and macroeconomics?

2. How does the theory of the firm differ from short-term profit maximization? Why is the former superior to the latter?

3. Find the present value of $1,000 due in one year if the discount rate is 5 per cent, 10 per cent, 15 per cent, and 20 per cent. Find the present value of $1,000 due in two years if the discount rate is 5 per cent, 10 per cent, 15 per cent, and 20 per cent.

4. The owner of a firm expects to make a profit of $10,000 for each of the next five years and to he able to sell the firm at the end of the fifth year for $50,000. The owner of the firm believes that the appropriate discount rate for the firm is 15 per cent. Calculate the value of the firm.

5. Suppose that the total revenue and total cost from an activity are, respectively, given by the following equations: $TR = 200 + 30Q - 5Q^2$ and $TC = 100 + 10Q$.

   a. Write out the equation for the marginal revenue and marginal cost.

   b. What is the equation for the marginal profit?

   c. What level of Q maximizes profits?

   d. At that level of Q, what is the level of profit?

6. A manager makes the statement that output should be expanded so long as marginal revenue exceeds marginal cost. Does this strategy make sense? Explain.

7. The revenue function for the microchip producer is $TR(Q) = 170Q - 20Q^2$. Derive the expression for marginal revenue, and use it to find the output level at which *revenue* is maximized. Confirm that this is greater than the firm's profit-maximizing output, and explain why.

8. Some economists have advanced a theory of the firm that postulates that firms seek to maximize sales rather than profits or the value of the firm.

   a. Explain what motivations managers might be seeking to maximize sales rather than profits.

   b. Explain the reasons why we retain the theory of the firm in terms of value maximization rather than sales maximization.

6 Answer Key to Review Problems

1. Macroeconomics focuses on aggregate economic conditions. Those are conditions that set the environment within which a business operates. On the other hand, microeconomics focuses on the economic forces that influence the decisions made by individual consumers, firms, and industries. Managerial Economics is the application of Microeconomics, but draws on other fields such as Finance, Accounting, Marketing, etc.

2. The theory of the firm is different from the short-term profit maximization in that it accounts for the future as well as the current revenues and costs. A business decision maker typically does not view his or her business as a series of one-year operations, but rather from a long-term perspective based on a continuous stream of revenues and costs that need to be viewed collectively not separately.

3. For a payment one year due:

$$PV = \frac{\$1000}{(1+0.05)} = \$952.38, PV = \frac{\$1000}{(1+0.1)} = \$909.09, PV = \frac{\$1000}{(1+.15)} = \$869.56$$

For a payment due in two years:

$$PV = \frac{\$1000}{(1+0.05)^2} = \$907.03, PV = \frac{\$1000}{(1+0.1)^2} = \$826.44, PV = \frac{\$1000}{(1+.15)^2} = \$756.14$$

4.

$$PV = \frac{\$10000}{(1+0.15)} + \frac{\$10000}{(1+0.15)^2} + \frac{\$10000}{(1+0.15)^3} + \frac{\$10000}{(1+0.15)^4} + \frac{\$10000}{(1+0.15)^5} + \frac{\$50000}{(1+0.15)^5}$$

$$PV = \$8695.65 + \$7561.43 + \$6575.16 + \$5717.53 + \$4971.76 + 24858.83 = \$58380.37$$

5. a. MR = 30 – 10Q, MC = 10.

   b. Marginal profit = (30 – 10Q) –10 = 20 – 10Q

   c. Marginal profit = 0, 20 – 10Q = 0, Q = 2

   d. Profit = TR (= 200 + 30Q - $5Q^2$) – TC (= 100 + 10Q) = 240 – 120 = 120

6. As long as MR exceeds MC, the marginal profit is positive, and hence, it pays to produce more output to capitalize on the opportunity.

7. MR(Q) = 170 – 40Q. In order to maximize total revenue, MR must be set equal to zero: MR = 0, and 170 – 40Q = 0, Q = 4.25. Profit maximization requires MR = MC. Since MC is normally positive, and not zero, the level of output that maximizes profit is, therefore, smaller than that maximizes revenue.

8. Some economists have advanced a theory of the firm that postulates that firms seek to maximize sales rather than profits or the value of the firm.

    a. There are a number of reasons as to why firms might give greater consideration to sales revenues are identified including that the firm's management may feel that lack of growth in sales will impair the company's reputation and dealings with its customers, distributors, financial institutions, and employees, and that it will reduce the company's influence in the marketplace and make it more vulnerable to competitors. The proponents of sales maximization recognize that some minimum level of profit is necessary, but they hold that the sales-maximizing firm is willing to sacrifice some or all profit above this minimum in order to increase sales.

    b. Despite its imperfection, the theory of the firm, which is based on value (wealth) maximization, is essential in explaining the behaviour of firms under conditions in which severe competition, perfect competition, and monopolistic competition prevail. Furthermore, the nature of the evidence suggests that even though firms may be following a different path when sacrificing some short-run profits, they would be doing so in order to maximize profits in the long run.

7 References and Suggested Further Reading

Baumol, W. J. 'What Can Economic Theory Contribute to Managerial Economics?' *American Economic Review*, May 1961.

Baumol, W. J., *Business Behavior, Value and Growth***,** Rev. Ed. New York: Harcourt Brace Johanovich, 1967.

Baye, M., *Managerial Economics and Business Strategy*, McGraw Hill, Irwin, 2002.

Wong, R. E. 'Profit Maximization and Alternative Theories: A Dynamic Reconciliation,' *American Economic Review*, September 1975.

Commonwealth of Learning Executive MBA/MPA

E5 Managerial Economics

Block 2

Decision Making Under Risk and Uncertainty

Course author: Farrokh Zandi, PhD
Professor, Schulich School of Business
York University
Toronto, Canada

Course designer: Douglas R. Cronk, M.A., ABD
Vancouver, BC, Canada

Copyright © Commonwealth of Learning, 2003
All rights reserved. No part of this course may be reproduced in any form by any means without prior permission in writing from:

The Commonwealth of Learning
1285 West Broadway
Suite 600
Vancouver, BC V6H 3X8
CANADA
e-mail: info@col.org

Contents

1 Block Two: Objectives and Introduction 1
1.1 Introduction 1
2 Risk and Uncertainty 2
3 Measuring Risk with Probability Distributions 2
3.1 Probability Distributions 3
3.2 Expected Value 4
3.3 Measures of Risk (Standard Deviation) 5
3.3.1 Demonstration Problem 7
4 Utility, Risk Aversion, and Risk Premium 8
4.1 Risk and Diminishing Marginal Utility 9
4.2 Risk Premium 12
4.2.1 Demonstration Problem 14
4.2.2 Demonstration Problem 14
5 Risk Adjustment in Decision Making 15
5.1 The Risk-Adjusted Discount Rate 15
5.1.1 Demonstration Problem 15
5.2 The Certainty-Equivalent Approach 17
5.2.1 Demonstration Problem 17
6 Asymmetric Information 18
6.1 Adverse Selection 18
6.2 Moral Hazard 20
7 Summary and Review 20
8 Review Problems 21
9 Answer Key to Review Problems 23
10 References and Suggested Further Reading 25

1 Block Two: Objectives and Introduction

After working through Block Two of this course, you should be able to:

1. Explain the difference between risk and uncertainty.
2. Measure the expected return and the measure of risk.
3. Explain the concept of cardinal utility as it pertains to risk.
4. Identfy attitudes towards risk identified by the behaviour of the marginal utility of income.
5. Determine the risk premium.
6. Explain how decision makers adjust for risk in their estimation of projects rates of return.
7. Explain the concepts of asymmetric Information as reflected by adverse selection and moral hazard.

1.1 Introduction

Until now, we have examined managerial decision making primarily under conditions of certainty. In such cases, the manager knows exactly the outcome of each possible course of action. However, in business the outcome of a decision is usually far from certain (at the time the decision is taken) because the decision maker has incomplete information and the outcome depends on the simultaneous behaviour of rival firms and other factors influencing the managerial cost and demand conditions. When the outcome of a decision is not predictable with certainty, we say that the decision is made under conditions of risk or uncertainty. Most strategic decisions of the firm are of this type. Therefore, it is essential to extend the basic model of the firm presented in Block 1 to include risk and uncertainty.

In the first Block, we distinguished between risk and uncertainty and introduced some of the concepts essential for risk analysis. Building on that discussion, in this Block, we will examine methods for measuring risk and for analyzing the manager's attitude toward risk.

2 Risk and Uncertainty

In simple microeconomics, economists assume full information, or certainty. That is, they assume they know the exact shape and location of demand and cost curves, such that they know exactly how much will be demanded at each price and exactly what the cost of production will be at the chosen output level. In the business world, however, firms typically operate under conditions of incomplete information, or uncertainty, and must estimate the quantity demanded and the costs of production based on the limited information they have at hand or can obtain by conducting information-search activity.

The state of information under which a decision is made has important implications for the predictability of the outcome of that decision. If there is full information, the outcome of a decision will be foreseen clearly and unambiguously. In this situation (of certainty), the firm can accurately predict the outcome of each of its decisions. When there is less than full information, however, the decision maker may foresee several potential outcomes to a decision and, therefore, will be unable to predict consistently which outcome will actually occur. In this case, we say that the individual or firm is operating under conditions of risk and uncertainty.

Certainty exists if the outcome of a decision is known in advance without a shadow of a doubt. Therefore, in this case, there is only *one possible outcome* to a decision and this outcome is known precisely. On other hand, *uncertainty* is the case when there is *more than one possible outcome* to a decision and where the probability of each specific outcome occurring is *not* known or even meaningful.

Risk can be regarded as a subcategory of uncertainty in which the probabilities of each outcome can be assigned on an objective basis. Risk is involved when one flips a coin, or throws dice. The probability of flipping a coin and having it land ‘heads’ is 1/2, since there are only two possible outcomes, and each is equally likely to occur, given an unbiased coin. Similarly, when one throws two dice, the probability that they will turn up ‘double six’ or any other pair of numbers, is 1/6 x 1/6 = 1/36.

3 Measuring Risk with Probability Distributions

As the above examples suggest, the greater the variability, i.e., the greater the number and range of possible outcomes, the greater is the risk associated with the decision or action. In the example, the probability of heads or tails is 1/2, whereas the probability of dice landing a pair is 1/36. In more complex business decisions, such as drilling for oil, it is possible that due to insufficient information or instability in the structure of the relevant variables, the investor will not know either the possible oil outputs or their probability of occurrence.

In the previous section, we defined risk as the situation where there is more than one possible outcome to a decision and the probability of each possible outcome is known or can be estimated. In this section we examine the meaning and characteristics of probability distributions, and then we use these concepts to develop a precise measure of risk.

3.1 Probability Distributions

When faced with outcomes that involve risks, a primary task of managers is to develop techniques that will enable them to calculate and subsequently minimize the risks inherent in particular problem. One method used to accomplish this is to calculate the probability distribution of possible outcomes from a set of sample observations, and then compute an *expected value*. That is, if several different levels of profit (or loss) are perceived as possible and each of these is assigned a probability of occurring. How does the decision maker summarize all these data so they can be compared with other potential solutions to the same problem?

The probability of an event is the chance or odds that the event will occur. For example, suppose you have just made an investment of $100 value in shares of an auto making (Low Technology) company believing that one of the three following possibilities might occur to your investment:

1. its value increases by 10%, to $110
2. its value stays the same, ($100)
3. its value falls by 10%, to $90 (see Table 2-1)

Furthermore, assume that, based on your prior investigation of this company's past performance as well as the current market conditions, you have come to conclude that the probability (likelihood) of outcome (a) is 25%, i.e., 1 out of 4 chances that this outcome occurs. For outcome (b) and (c) these probabilities are 50% and 25%, respectively.

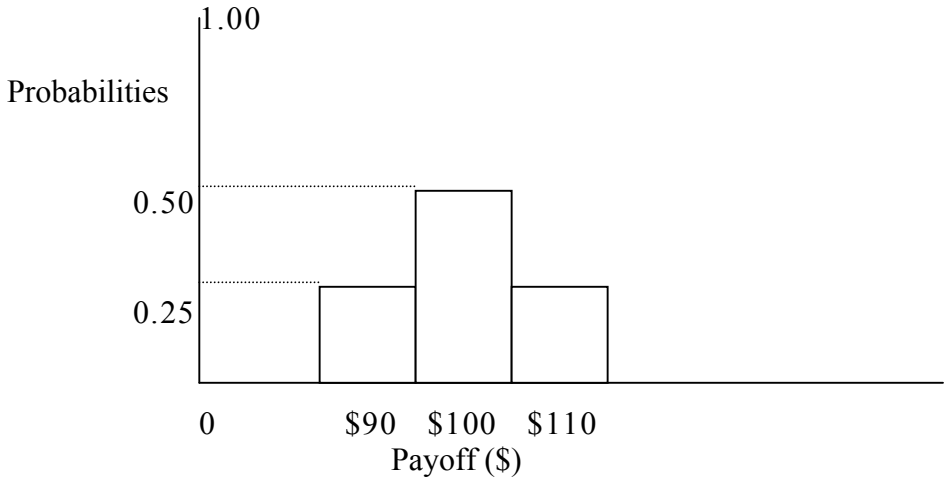
Table 2-1 Probability Distribution of States of Alternative Scenarios of Share Prices (Low-Tech Company)

Sate of the Affairs	**Probability of Occurrence**
Share price increases	0.25
Share price stays unchanged	0.50
Share price decreases	0.25
Total	**1.00**

Note that the sum of the probabilities is l, or 100 per cent, since one of the three possible scenarios of the share prices must occur with certainty.

The probability distribution depicts all possible payoffs and their associated probabilities. Figure 2-1, below, shows the probability distribution of your auto-making company's stock price. Each bar represents a different possible payoff from investing in the company's shares.

Figure 2-1



Learning Tip

Not all risky events are conducive to easy deduction of probabilities of particular outcomes, such as flipping a coin, which offers a chance of 1/2 for heads or tails. In our example of the auto making company, this deduction is difficult and challenging. The question is: how would you know that the chance of share price increase is 25% and not 20%? Your evaluation of the various scenarios is therefore *subjective*, and the probabilities that reflect subjective beliefs about risky events are called *subjective probabilities.*

3.2 Expected Value

Given the probabilities associated with the possible outcomes of your risky investment, how much can you expect to make? The answer to this question is the *expected value.* The expected value of a lottery is a measure of the average payoff that the lottery will generate. The expected value of an outcome is the value of that outcome multiplied by the probability of that outcome occurring. Since several outcomes are possible under risk and uncertainty, the expected value of a decision is the sum of the expected values of all the possible outcomes that may follow the decision. We can illustrate this with your automaker stock example:

$$E(X) = P_1X_1 + P_2X_2 + ...P_nX_n$$
$$E(X) = \sum_{i=1}^{n} P_iX_i \quad (1)$$

where X_i is the value of the ith payoff, and P_i is the probability of the ith state of nature. Or

Expected value = Probability of (a) x Payoff if (a) occurs +

= Probability of (b) x Payoff if (b) occurs +

= Probability of (c) x Payoff if (c) occurs.

Applying this formula we get:

Expected value = (0.25 x 110) + (0.50 x 100) + (0.25 x 90) =100

Learning Tip

Note that the expected value of a probability distribution need not equal any of the possible outcomes (although it does in this case). The expected value of your automaker's shares is a weighted average of the possible payoffs that you believe might occur, where the weight associated with each payoff equals the probability that the payoff occurs if the lottery is repeated many times. If you made the same investment over and over again and took an average of the payoffs you made from this investment, that average would be nearly indistinguishable from the lottery's expected value of $100.

3.3 Measures of Risk (Standard Deviation)

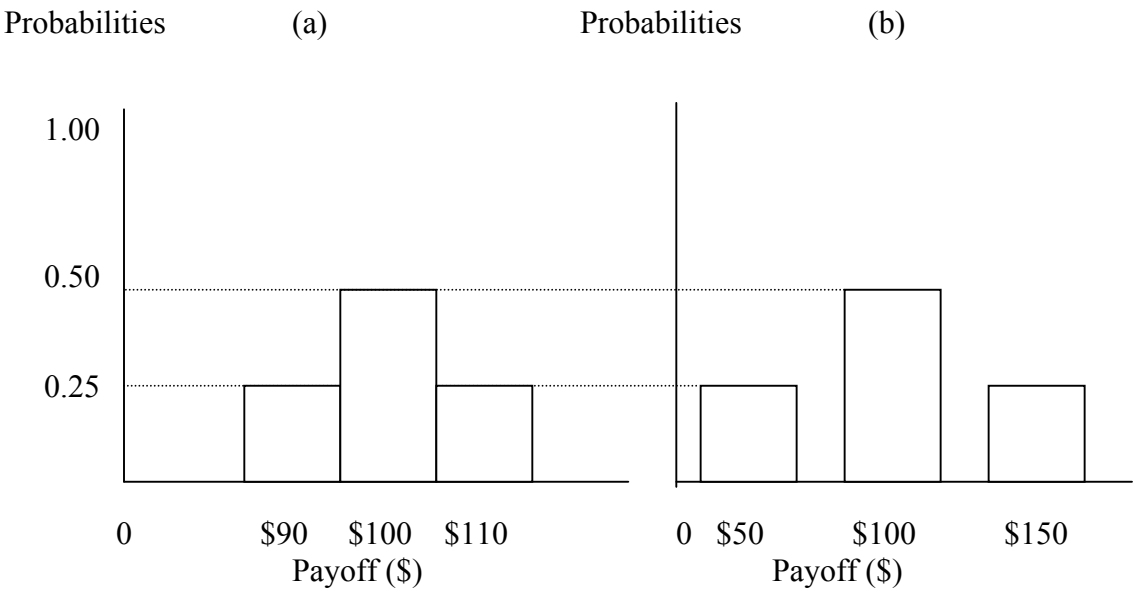
Suppose you had a choice of two investments, Table 2-2 presents the payoffs associated with two investments: L and H. Investment L is the same as the investment in the shares of the (Low Tech) auto-making company in Table 2-1, whereas H represents the outcome of investment in another company's shares, e.g., a (High Tech) telecommunication company. The expected value of each project is assumed to be $100, but the range of outcomes for project L (from $90 to $110) is smaller than for project H. The latter is assumed to range from $50 to $150. These are obtained from a greater share price increase or decrease, which is assumed to be 50%, instead of 10% (as in project L). The 50% increase or decrease in share prices translates into the value of your investment increasing or decreasing from $100 to $150 or $50, respectively. The probabilities by which the three outcomes are expected to occur are kept the same in both projects.

Table 2-2 Probability Distribution of States of Share Prices (High and Low Tech companies)

Project	Sate of Share Prices	Probability of Occurrence	Outcome	Expected Value
	Increase	0.25	$110	$27.5
L	Unchanged	0.50	$100	$50
	Decrease	0.25	$90	$22.5
	Expected Value from Project H			**$100.0**
	Increase	0.25	$150	$37.5
H	Unchanged	0.50	$100	$50
	Decrease	0.25	$50	$12.5
	Expected Value from Project L			**$100.0**

Figure 2-2 depicts the expected value and the variability in the outcomes of investment L, in panel (a) and H, in panel (b). Note that the expected values of the two investments are the same: $100. However, the telecommunication stock is riskier than the automaker's stock because while the stock of automaker will probably remain at its current value of $100, the telecommunication stock has a greater likelihood of going up or down, panel (b).

Figure 2-2



Again, the height of the each bar measures the probability that a particular outcome (measured along the horizontal axis) will occur. Since both investments have the same expected outcome ($100) but the relationship between the payoffs (outcomes) is less dispersed in investment L, panel (a), than in investment H, panel (b), investment L is *less risky* than H. In other words, with the telecommunication stock, the investor stands to gain more or lose more than with stock in the auto making company.

Intuitively, we sense that farther away from the mean the actual payoff can be, the riskier the investment. Hence, one way of measuring risk is to calculate the *range,* which is difference between the most extreme payoff values. In our example, as noted above, the range of investment L is 20 (from a low 90 to a high 110) while the range of investment H is 100 (from a low 50 to a high 150).

The range, however, is useful in preliminary evaluation, but it considers only the extreme values and gives no weight to values in between. A more common and more accurate measurement of risk is the statistic called *standard deviation,* which is a measurement of variation of payoffs from the expected value. The higher the standard deviation, the greater the deviation of possible payoffs, and therefore, the greater the risk.

The standard deviation is calculated as follows

$$\sigma = \sqrt{\sum [X_i - E(X_i)].P_i} \qquad (3)$$

where σ is the standard deviation. This expression suggests a three-step procedure in calculating the standard deviation: (a) first calculate the expected value, also called the mean, (b) then take the difference between each outcome (payoff) and the mean, and then square the result. And (c) finally multiply each squared deviation in step (b) by the associated probability and add them up.

3.3.1 Demonstration Problem

Find the standard deviation of the two investment alternatives L and H, in the example above.

Answer:

Investment L: Expected value = $100

Payoff (X_i)	Prob.(%)	$[X_i - E(X)]^2$	$[X_i - E(X)]^2.P_i$
90	0.25	$[90 - 100]^2 = 100$	100 x 0.25 = 25
100	0.50	$[100 - 100]^2 = 0$	0 x 0.50 = 0
110	0.25	$[110 - 100]^2 = 100$	100 x 0.25 = 25

Standard Deviation $= \sqrt{50} \approx 7$

Investment L: Expected value = $100

50	0.25	$[50 - 100]^2 = 2500$	2500 x 0.25 = 625
100	0.50	$[100 - 100]^2 = 0$	0 x 0.50 = 0
150	0.25	$[150 - 100]^2 = 2500$	2500 x 0.25 = 625

Standard Deviation $= \sqrt{1250} \approx 35$

Based on these calculations, investment H, with a greater standard deviation, is far more risky than investment L.

An alternative measure of the riskiness of a risky investment is the *variance.* Variance is the square of standard deviation.

4 Utility, Risk Aversion, and Risk Premium

Suppose in the process of job-hunting in the field of management consulting a job hunter is faced with the following two options. The first option is a job offer from a large multinational company that promises to pay $50,000 a year. The second one is an offer from a small but growing local company that promises to pay $20,000 a year plus a hefty $60,000 in commission assuming that the job hunter meets a million dollar sales quota per year. Her assessment, however, shows that there is a 50% chance that she meets this quota and 50% that she does not. The expected value of this lottery is as follows:

$$E(X) = 0.5 (\$60,000) + 0.5 (0) = \$30,000 \quad (4)$$

Based on this calculation, which job would she choose? If she chooses the job with the large multinational company, she is guaranteed $50,000, and if she accepts the job with the small local company, her expected income would be 50,000 ($20,000 salary plus $30,000 in expected commission). On the surface, it appears that both jobs offer the same income. However, it is very likely that she would choose the 100 per cent salary job (certain outcome) with the multinational company over the local company's (risky) commission job. The reason is that the seemingly more exciting job with the local company may in fact end up paying her just $20,000, if our job hunter is unable to meet her sales quota. This suggests that most people, as that goes for firms' managers, faced with two alternative projects of equal expected value of profit but different coefficients of variation or risk, will generally prefer the less risky project (i.e., the one with the smaller coefficient of variation). While it is true that some managers may very well choose the more risky project (i.e., are *risk seekers*) and some may be indifferent to either choice (i.e., are *risk neutral*), most managers are *risk averters*. The reason for this is to be found in the principle of diminishing marginal utility of money. The meaning of diminishing, constant, and increasing marginal utility of money will be explained with the aid of a

reward structure that helps explain transformation of dollar payoffs into a more meaningful measurement. Utility is such a measurement, and it can be expressed in conceptual units called *utils.* Although difficult to establish a standard util by which one can perform a cardinal measurement of utility, it is nonetheless a useful concept.

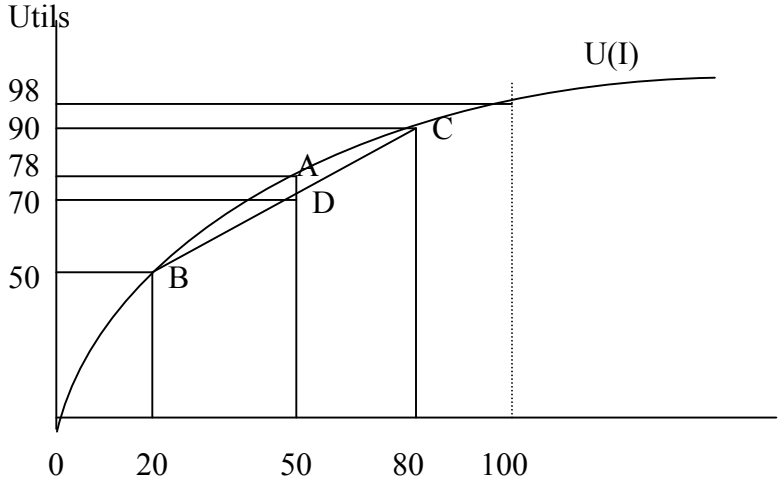
4.1 Risk and Diminishing Marginal Utility

At this point, it is necessary to explain the relationship between risk and utility in a formal manner. To do so, profit and loss must be measured in terms of marginal utility rather than absolute dollar values. Marginal utility is defined as the change in total utility that takes place when one more unit of money is gained or lost.

The three ways in which utility may theoretically relate to income are depicted in Figures 2-3, 2-4, and 2-5. These Figures depict behaviour of different types of investors when investment yield or income is increased by equal increments. Money income or wealth is measured along the horizontal axis while the utility or satisfaction of money (measured in utils) is plotted along the vertical axis. Each curve represents utility as a function of income, U = U(I), where U stands for utils and I for income. The slope of each curve represents marginal utility, which is where our interest lies.

The most common behaviour, depicted in Figure 2-3, is a risk avoider (averter). The reason for risk aversion is *diminishing marginal utility*. It shows that with no investment there is no return. A given increment ($20,000) to income when income is low, zero, increases utility by 50 (vertical axis), U($20,000) = 50, whereas the same increment to income when income is, say $80,000, increases utility by a smaller amount, U($100,000) –U($80,000) = (98–90) = 8.

Figure 2-3



If, therefore, the total utility of the money curve is concave (or facing down), doubling money income less than doubles utility. This is the basic explanation of risk and can be used to illustrate the behaviour of our job hunter.

Remember that our job hunter's salary at the established multinational company is $50,000, and as depicted in Figure 2-3, the level utility associated with this income, U($50,000), equals 78, point A. The job-hunter's income at the less established local company, however, is one of the two cases. She either makes $20,000, in case she fails to make any commission income, in which case the corresponding utility would be 50, point B, or she makes $80,000, if she succeeds in meeting her quotas, where the corresponding utility would be 90, point C. Therefore, the job hunter's *expected utility* at the local company is the expected value of the utility levels she could receive if she worked for the local company:

$$0.50 \times U(\$20,000) + 0.50 \times U(\$20,000 + \$60,000)$$

$$= 0.5 \times 50 + 0.5 \times 90$$

$$= 70. \qquad (5)$$

This is depicted by point D in Figure 2-3.

The above analysis shows that although the local company offers the same expected salary as the established multinational company, the job hunter's expected utility at the local company, 70, is lower than the utility she would receive from the job with the multinational company, 78.

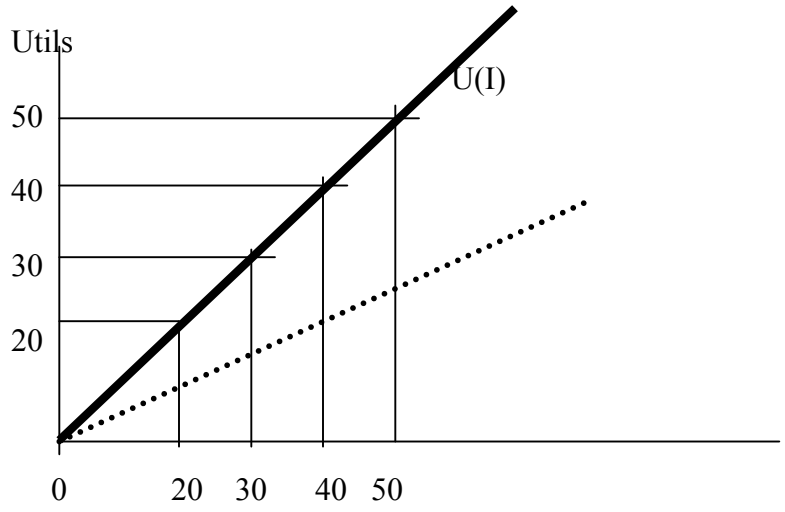
Thus, we see that the utility from the 100 per cent salary job (risk-free) is greater than the expected utility from a commission based job (risky) with equal expected value of income. Therefore, if our job hunter personality fits that represented by Figure 2-3, she will prefer the risk-free to the risky job. This is the preference of a decision maker who is *risk averse.*

In Figure 2-4, the utility function is a straight line, implying that doubling income doubles utility so that the marginal utility of money is constant. The straight-line utility function characterizes a person who is indifferent to risk, for whom marginal utility of a dollar lost is equal to that of a dollar gained.

Learning Tip

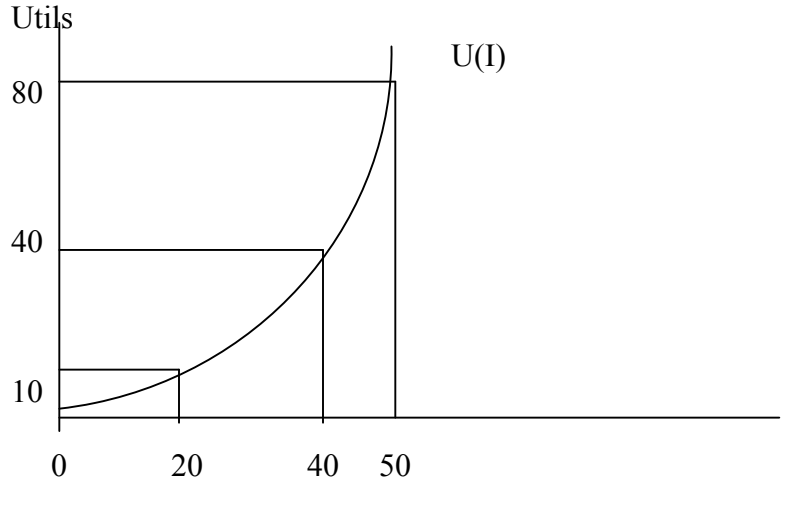
Note that indifference to risk does not imply indifference to profit. The individual represented by the heavy line utility function places a greater utility value on a dollar gained or lost than does the individual who is represented by the dotted-line utility function, in Figure 2-4.

Figure 2-4



Finally, in Figure 2-5, if the total utility of money curve is convex or faced down, doubling income more than doubles utility, so that the marginal utility of money income increases. This represents the case of compulsive gamblers, who place higher utility on dollars won than dollars lost. The more they win, the more important winning becomes.

Figure 2-5



Most individuals are risk averters because their marginal utility of money diminishes, i.e., they face a total utility curve that is concave or faces down. To see why this is so, consider the offer to engage in a bet to win $10,000 if a head turns up in the tossing of a coin or to lose $10,000 if a tail comes up. The expected value of the money won or lost is

Expected value of money income = E(I)

= 0.5($10,000) + 0.5($-10,000) = 0

Learning Tip

Even though the expected value of such a *fair game* is zero, a risk averter (risk averse) would gain less utility by winning $10,000 than he or she would lose by losing $10,000. In such a case, the individual will refuse a 'fair bet.' From this, we can conclude that risk-averting managers will not necessarily accept an investment project with positive expected value or a positive net profit. To determine whether or not the manager will accept the project, we need to know his or her utility function of income.

4.2 Risk Premium

Aversion to risk by managers and investors is manifested in many ways. The following are a few examples of how risk may be averted. Grade AA bonds sell for a higher price than grade B bonds. Investors diversify either by creating individual portfolios or by investing in mutual funds. People deposit their money in government treasury bills at low rates of interest rather than in bonds that may earn substantially more interest. And people buy all kinds of casualty and life insurance.

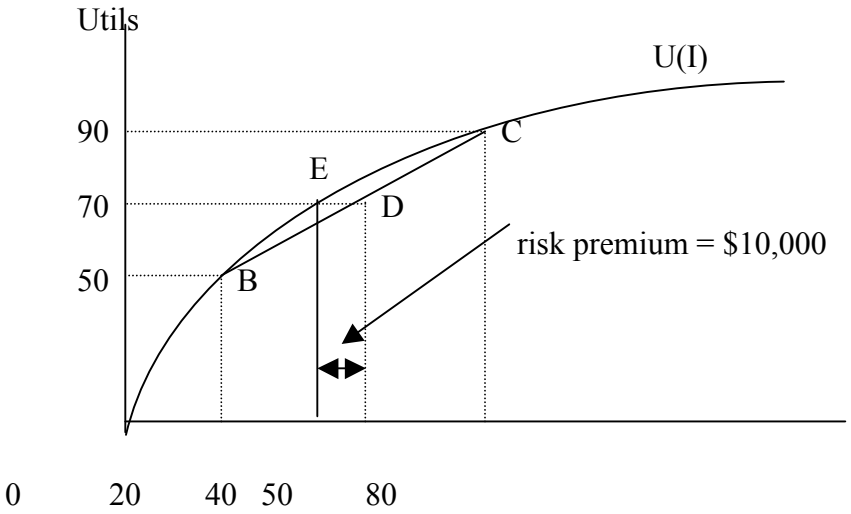
Why, then, if investors are averse to risk, do they put their money into common stocks, commodities, precious metals, collectibles, and other risky investments? The answer is that they do not do so unless they receive a risk premium. The investor wants to be compensated not only for the use of his or her money, but also for the risk that it may be lost. In other words, the investor demands a higher rate of return when risk is involved.

Learning Tip

A risk-averse investor (decision maker) might prefer a risky investment (gamble) to a sure (risk-free) investment if the expected return from the latter is sufficiently larger than the return from the sure investment. This sufficiently larger return is what it takes for a risk averse individual to bear risk – risk premium.

To illustrate this, we recall the example of our job hunter in Figure 2-6 below. In that example we showed that the job hunter preferred the risk-free (with the multinational company) to the risky job (with the local company).

Figure 2-6



Accordingly, the *risk premium* is the minimum payment (compensation) to the risk averse decision maker (our job hunter) to make her indifferent between the risky and risk-free events. In order to find the risk premium for our job seeker, let us ask the following question: at what level of sure (risk-free) income (with the multinational company) would the resulting level of utility be equal to the expected utility of the risky (commission-based) income? In Figure 2-6, the expected utility of the risky job that is expected to pay $50,000 is 70, point D. Therefore, the risk-free income whose corresponding level of utility is also equal to 70 has to be about $40,000, at point E. Note that E and D correspond to the same level of utility. Therefore, our job hunter would be indifferent between a $40,000 (all salary) job with the multinational firm and a $50,000 risky (partly commission based income) job with the local company. Hence, the risk premium of the risky local company's offer is $10,000.

Learning Tip

Note that $10,000 is the minimum reward that you require in order to give up the certainty of a risk-free event and bear risk.

Learning Tip

Investors' attitudes toward risk depend first of all upon the shape of their utility curves, which in turn depend upon the investors' personalities, and second, upon each one's current position on the curve, which is determined by current income. If current income is low, marginal utility of income (measured by the slope of the curve) is high, and one will accept risk for a much lower premium than when income is higher.

4.2.1 Demonstration Problem

Suppose you have a utility function that is concave (faced down). Also suppose that you bet $100 on the flip of a coin at even odds. The probability of winning is 0.5 and the probability of losing is also 0.5. If you win, you will get $100 and if you lose you will lose $100. Should you take the bet?

Answer:

Note that if you win you get $100 and if you lose you pay $100. We also know that if you win you gain fewer utils of utility than you sacrifice utils if you lose $100. This is because of the shape of the utility function, diminishing marginal utility. Since the probability of winning or losing is the same, the expected value in utils is inevitably negative: 0.5(utils gained) + 0.5(utils sacrificed) <0. This is so because utils gained are fewer than utils sacrificed. Clearly, the investor should not take this bet.

4.2.2 Demonstration Problem

Suppose you have a utility function that is concave (faced down). Also suppose that you bet $100 on the flip of a coin at even odds. The probability of winning is 0.5 and the probability of losing is also 0.5. This time if you win, you will get $120 and if you lose you will lose $100. Should you take the bet?

Answer:

It depends on the shape of your utility function. The $20 premium may or may not be sufficient to make you indifferent between the two possibilities. It may take more or less than $20 to take this bet. The curvature of the utility function speaks to this issue. The steeper the utility curve, the smaller the required risk premium, and vice versa.

5 Risk Adjustment in Decision Making

In estimating the payoffs for a particular strategy, the decision maker must consider both the present value of future returns and the degree of risk. In this section we examine two of the most commonly used methods: the risk-adjusted discount rate and the certainty-equivalent approach that a risk-averse decision maker employs to compare decision alternatives on a risk-adjusted basis.

5.1 The Risk-Adjusted Discount Rate

Under conditions of risk and uncertainty the present value of future returns are not known with certainty. Therefore, in estimating the payoffs for a particular strategy, the decision maker requires to maximize the (*expected) Net* Present Value (NPV), which combines the present value calculation with expected-value analysis.

$$NPV = \sum_{t=1}^{n} \frac{R_t}{(1+i)^t} - I_0 \tag{6}$$

where R_t represents the expected net return (cash flow) in each of the n years considered, and i, as discussed in Block One, is the appropriate discount rate, and I_0 is the amount of the initial investment.

One popular method of adjusting the NPV criterion of equation (6) to deal with an investment project subject to risk is using higher discount rates for more risky decision alternatives. We may define the risk-adjusted discount rate as the required rate of return from a proposed investment after due consideration of the risk involved:

$$NPV = \sum_{t=1}^{n} \frac{R_t}{(1+r)^t} - I_0 \tag{7}$$

where r is the risk-adjusted discount rate, r = i + risk premium. As discussed earlier, every firm has a required rate of return reflecting its perception of its normal risk (normal business risk plus financial risk).

5.1.1 Demonstration Problem

For example, suppose a firm's normal business and financial risk requires a 20 per cent rate of return. The firm is considering an investment strategy that initially costs $100,000 and is expected to yield $50,000 cash inflow per year for the next three years.

a. Calculate the net present value of the investment at a discount rate of 20 per cent. Should the firm accept this project?

b. Suppose that the risk were such that management feels it should get a 25 per cent return. Calculate the NVP for the adjusted discount rate. Should the firm accept the investment project?

Answer:

a.

$$NVP = \frac{\$50{,}000}{(1.20)} + \frac{\$50{,}000}{(1.2)^2} + \frac{\$50{,}000}{(1.2)^3} - \$100{,}000 = \$5{,}324.$$

Since the net present value is positive, a risk-neutral firm should accept this project. A risk-averse firm may not necessarily accept this project. It depends on the firm's degree of risk aversion.

b.

$$NVP = \frac{\$50{,}000}{(1.25)} + \frac{\$50{,}000}{(1.25)^2} + \frac{\$50{,}000}{(1.25)^3} - \$100{,}000 = \$ -2{,}400.$$

Here the NVP is negative. The project fails to provide a 25 per cent discounted return and should be rejected by both a risk averse and a risk neutral firm.

Thus we see that in the risk-adjusted discount-rate approach to evaluation of proposed investments, risk is wholly reflected by the discount rate and discounting process. There are, however, at least three limitations to this approach to incorporation of risk:

1. How do we determine the appropriate discount rate? Clearly, the introduction of a new product is riskier than buying government bonds, but how much riskier? It is very difficult to resolve this question consistently and objectively, particularly when there is no historical evidence on which to base an estimate.

2. This method does not consider the probability distribution of future cash flows information that could be of great value.

3. The risk-adjusted discount rate does not offer any consistent method for evaluation of risk, an evaluation that may be quite subjective. This limitation may be overcome by the certainty-equivalent approach.

5.2 The Certainty-Equivalent Approach

The risk-adjusted discount-rate approach discussed in the preceding section accounts for risk by simply modifying the discount rate, appearing in the denominator of the valuation model. In contrast, the certainty-equivalent approach accounts for risk in the numerator of the valuation model and uses a risk-free discount rate, i (such as the rate of return on government bonds) in the denominator to account for the time value of money. The degree of risk is reflected in the numerator by multiplying the expected risky return by a certainty-equivalent coefficient.

$$NPV = \sum_{t=1}^{n} \frac{\alpha_t R_t}{(1+r)^t} - I_0 \tag{8}$$

where α is the certainty equivalent coefficient, a number between 0 and 1. 1 means the project is risk free, and 0 means too risky to be considered.

Certainty equivalent of a decision alternative is the sum of money available with certainty that would make the manager indifferent between taking that decision and accepting the certain sum of money:

$$\alpha R = R^* \tag{9}$$

where R^* is the risk-free equivalent cash flow.

5.2.1 Demonstration Problem

The manager of a company regards the sum of $80,000 with certainty as equivalent to the expected (risky) net cash flow or return of $100,000 per year for the next three years, what is the value of α?

Answer:

$$\alpha = \frac{80{,}000}{100{,}000} = 0.8$$

Learning Tip

The risk-adjusted discount approach accounts for risk by adding a risk premium to the risk-free rate to obtain the risky rate of return, whereas the certainty equivalent approach accounts for risk in the valuation of the certainty equivalent coefficient, R*. They will yield same NPV if:

$$\frac{\square R}{(1+i)^t} = \frac{R^*}{(1+r)^t}$$

or (10)

$$\square = \frac{(1+i)^t}{(1+r)^t}$$

6 Asymmetric Information

Asymmetric information refers to situations in which one party knows more about its own actions or personal characteristics than another party. When some people in the market have better information than others, the people with the least information may choose not to participate in a market. For example, the market for used cars is a market characterized by information asymmetry. The seller always knows better than the buyer about the quality of the car including whether the car has been regularly serviced and inspected or whether it has been involved in an accident. That is why some people tend to shy away form used cars, unless some form of warranty is tacked on.

6.1 Adverse Selection

In the above example, while warranties may reduce the financial cost of having a lemon, they do not eliminate the bother, such as the time it takes to bring the car into the shop. Of course, the owners know they have a lemon and would like to pass it along to someone else. Those with the worst lemons are going to be the most willing to sell their car, whatever the price. But at a high used-car price, they will be joined by owners of better-quality cars. As the price drops, more of the good cars will be withdrawn from the market as the owners decide to keep them. And the average quality of the used cars for sale will drop. We say there is an *adverse selection* effect. The mix of those who elect to sell changes adversely as price falls.

Learning Tip

Adverse selection generally arises when an individual has *hidden characteristics* that he or she knows, but that are unknown by the other party.

Asymmetric information affects many other managerial decisions, including automobile insurance, employing workers, and issuing credit to customers. Private automobile insurers do not have complete information about the risk level of their clients, but they know that, as a group, young single males demonstrate the highest claim frequency. The problem of adverse selection arises because the insurers cannot identify the individuals who are high risk. One way to rectify this matter, at least partially, is to segment the market by age or gender charging higher premiums to high-risk individuals.

This type of discriminating policy, however, is bound to be an imperfect way to overcome the adverse selection problem. There will undoubtedly be some drivers from certain categories, say, young males, who are cautious drivers but who have to pay high premiums because of the accident record of others of their gender and age group. On the other hand, a non-discriminatory policy, i.e., charging the same premium to all drivers will also discriminate between good and bad drivers by not rewarding the good drivers. The end result is that some drivers are less insured than they would like to be, because of high prices.

Another area where asymmetric information affects managerial decisions is the job market. Job applicants have much better information about their own capabilities than does the person in charge of hiring new workers. A job applicant who claims to have excellent skills may be lying or not telling the whole truth; the personnel manager has less information than the applicant. This is why firms spend considerable amounts of time and money setting up several interviews, designing tests to evaluate job applicants, doing background checks, and the like. The basic reason for these types of expenditures is to provide the firm with better information about the capabilities and tendencies of job applicants.

Learning Tip

One of the consequences of asymmetric information is that there may be relatively few buyers and sellers, far fewer than there would be with perfect information.

6.2 Moral Hazard

A second problem faced by insurance companies is an incentive problem. Insurance reduces people's incentives to attempt to avoid a loss and encourages them to take excessive risk. If there existed some form of business bankruptcy insurance, entrepreneurs would take riskier steps than necessary or warranted. However, people buy insurance for their house, cars, and valuable belongings. For example, a person who has no fire insurance on a house may choose to limit the risk by buying smoke alarms and home fire extinguishers and by being especially cautious. However, if he has fire insurance, he might not be so as careful. Therefore, *moral hazard* generally occurs when one takes *hidden* actions that it knows another party cannot observe.

When moral hazard problems are strong, insurance firms will offer limited or even no insurance. The limitations often take one of two forms: deductibility provisions and co-insurance. Insurance policies may pay damages only above some initial amount, referred to as a deductible. For example, your car insurance policy may require you to pay the first $500 of damages before insurance benefits kick in. This reduces the moral hazard problem associated with small claims; drivers might be much more vigilant about avoiding minor accidents than major ones. Alternatively, insurance policies may pay only some specified proportion of damages. This is referred to as co-insurance. It forces those who are insured to bear some cost of any accident and so to behave with greater care.

7 Summary and Review

This chapter has dealt with methods and approaches to decision making under conditions of risk and uncertainty. Under the condition of risk, the primary decision criterion for selecting the optimum strategy is expected value. The degree of risk is indicated by the *standard deviation*.

How decision makers choose to deal with risk depends upon their attitudes. Some may try to seek risk, some may be indifferent toward it, but most businesspeople try to avoid risk. Their attitudes are based upon utility functions in which increasing increments of income (profits) bring decreasing increments of satisfaction (utility).

Risk aversion is based on the principle of diminishing marginal utility of money, which is reflected in a total utility of money curve that is concave or face down. A risk-averse decision maker will accept risk only if there is a commensurate risk premium. Every business firm and individual investor has in mind some required rate of return that reflects the perceived risk. As the degree of risk increases, the required rate of return also increases along a market-indifference curve that depicts the investor's risk-return trade-off function.

The profit-maximization model can be simultaneously adjusted for both risk and the true value of money by several techniques. Two of the most common are the risk-adjusted discount rate and the certainty-equivalent approach. The former involves adding a risk premium to the risk-free rate of interest or discount used to find the present value if the

net cash flow or the return of the investment. A better method is the latter, which uses a risk-free discount rate in the denominator and incorporates risk by multiplying the net cash flow or return in the numerator of the valuation by the certainty-equivalent coefficient.

8 Review Problems

1. Suppose the required rate of return by a firm is 20 per cent while the risk on government Treasury bills is 8 per cent. The firm is considering an investment of $500,000 in a venture that promises to yield $150,000 per year for the next five years.

   a. Calculate the NPV of the proposed venture by the risk adjusted discount method.

   b. Calculate coefficient of θ that will cause the certainty-equivalent approach to yield the same result.

2. Suppose a firm is considering an investment of $100,000 that is expected to yield a cash flow of $50,000 per year for three years. Suppose that management's perception of risk is such that it considers risk-free returns of $45,000 in the fist year, $40,000 in the second year, and $35,000 in the third year to be equivalent to the risky return of $50,000 for each year. Calculate θ for each of the three years.

3. Why is the certainty-equivalent approach to risk adjustment considered to be superior to the risk-adjusted approach?

4. Explain the concept of risk premium. What causes the magnitude of risk premium change?

5. 'For a risk averse consumer the expected utility of a gamble is greater than the utility of the expected value of that gamble.' True or False? Explain.

6. 'For a risk lover consumer the expected utility of a gamble is greater than the utility of the expected value of that gamble.' True or False? Explain.

7. Explain why a used car that is only six months old and has been driven only 5,000 km typically sells for 20 per cent less than a new car with the same options.

8. Sketch each of the following utility functions. Based on the obtained information determine the decision maker's attitude toward risk in each case.

   a. $U = I^{1/2}$

   b. $U = I$

   c. $U = I^2$

   where U is the utility and I represents income.

9. A trucking firm pays its drivers based on how quickly they deliver goods from one point to another. Unfortunately, to obtain the higher payoff, the truck drivers must break the speed limit. Suppose that a driver who makes a fast delivery obtains a payoff of $100 and a driver who makes a slow delivery obtains a payoff of $81 per shipment. A slow driver has no chance of getting a ticket for speeding, but a faster driver receives a ticket with probability 50%. A driver who receives a ticket is responsible for paying the fine himself. Two types of drivers are employed at the firm: type A drivers with utility $U(I) = I^2$ and type B drivers with utility $U(I) = I^{1/2}$, where I is income.

    a. If the fine for speeding is $36, which type of driver speeds?

    b. If the fine for speeding rises to $51, which type of driver speeds?

    c. If the fine remains at $36, but the probability of receiving a ticket rises to 75%, which type of driver speeds?

    d. Show graphically that either a rise in the amount of the fine or the probability of receiving a ticket can be used to deter speeding.

10. Consider a person with an initial wealth level of $100 who faces a chance to win $20 with probability 1/2 and to lose $20 with probability 1/2. If this person's utility function is given by $U(I) = I^2$, where U is measured in utils, will she accept this gamble?

11. Consider a person with an initial wealth level of $100 who faces a chance to win $20 with probability 1/2 and to lose $20 with probability 1/2. If this person's utility function is given by $U(I) = I$, where U is measured in utils, will she accept this gamble?

12. Distinguish between adverse selection and moral hazard.

9 Answer Key to Review Problems

1. Suppose the required rate of return by a firm is 20 per cent while the risk on government Treasury bills is 8 per cent. The firm is considering an investment of $500,000 in a venture that promises to yield $150,000 per year for the next five years.

   a.

$$NPV = 150{,}000\left[\frac{1-(1.20)^{-5}}{0.20}\right] - 500{,}000 = -\$51{,}408.18$$

   b. Uncertainty-equivalent approach, α would be:

   Year 1 : $\alpha = 1.08/1.20 = 0.9$

   Year 2 : $\alpha = (1.08)^2/(1.20)^2 = 0.81$

   Year 3 : $\alpha = (1.08)^3/(1.20)^3 = 0.729$

   Year 4 : $\alpha = (1.08)^4/(1.20)^4 = 0.656$

   Year 5 : $\alpha = (1.08)^5/(1.20)^5 = 0.590$

2. Year 1 : $\alpha = \$45{,}000/\$50{,}000 = 0.9$

   Year 2 : $\alpha = \$40{,}000/\$50{,}000 = 0.8$

   Year 3 : $\alpha = \$35{,}000/\$50{,}000 = 0.7$

3. If there are any time periods in which perceived risk is more or less than what is represented by the risk-adjusted discount rate, then the certainty-equivalent approach can provide a better estimate of the net present value of a proposed investment. For example, the return from an investment for the introduction of a new product might be more uncertain in its earlier years, while the firm is struggling for product recognition and market share, than in later years when the product's market has become established. The certainty equivalent approach can easily handle this situation during the process of establishing α for each separate time period. The certainty-equivalent approach enables managers to specify directly the degree of risk for and then discount the cash flow.

4. The risk premium is the minimum payment (compensation) to the risk averse decision maker (our job hunter) to make her indifferent between the risky and risk-free events. The magnitude of risk premium changes with changes in attitude towards risk. The higher the degree of risk aversion the greater the risk premium required by the decision maker.

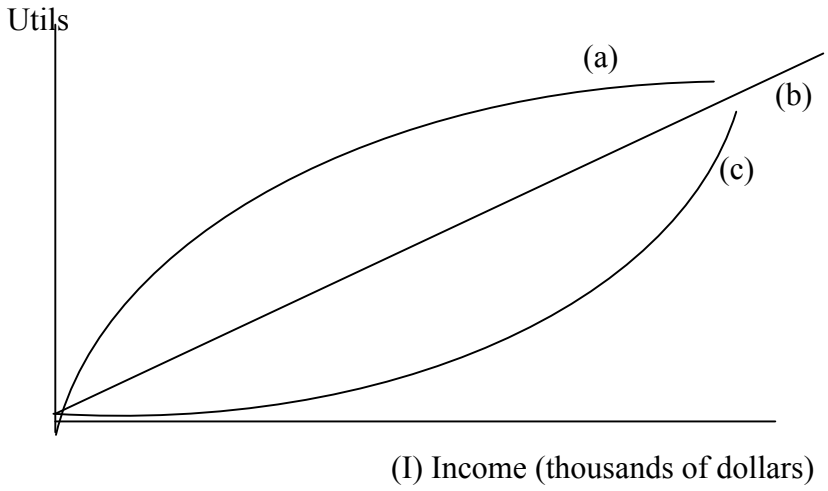
5. FALSE: If two events (one risky and one safe) have the same outcome, due to diminishing marginal utility of income, the expected utility of the risky event will always be smaller than the utility of the safe asset (concave utility function). You can show this by drawing the marginal utility of income curve under alternative risk attitudes.

6. TRUE. See answer to question 5.

7. This is due to asymmetric information. The buyer of the used car is not aware of the behaviour and driving habits of the first owner – adverse selection (hidden characteristics). The first owner knows how the car has been kept up, while the buyer does not.

8. Sketch each of the following utility functions. Based on the obtained information determine the decision maker's attitude toward risk in each case.

   a. $U = I^{1/2}$

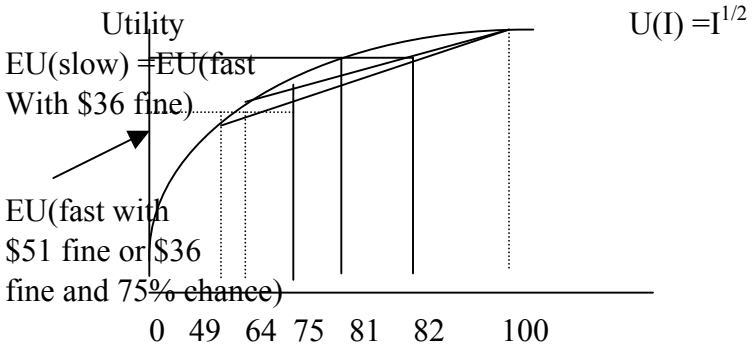


   a. If the fine is \$36, then the expected utility of speeding is EU(fast) = .5U (\$100 – \$36) + .5U(\$100). The expected utility of driving slowly is EU(slow) = U(\$81). Hence, for type A drivers, we have EU(fast) = $.5(\$64)^2 + .5(\$100)^2$ = 2048 + 5,000 = 7,048 and EU(slow) = $(\$81)^2$ = 6,561. Hence, type A drivers prefer to speed. For type B drivers, we have EU(fast) = $.5(\$64)^{1/2} + .5(\$100)^{1/2}$ = 9 and EU(slow) = $(\$81)^{1/2}$ = 9. Hence, the type B drivers are indifferent between speeding and not speeding. Any slight increase in the fine would make them prefer not to speed, in fact.

   b. If the fine rises to \$51, we can recompute the above expected utilities to obtain for type A: EU(fast) = 6,200.5 and EU(slow) = 6,551 so that the type A drivers do not wish to speed. For type B we have: EU(fast) = 8.5 and EU(slow) = 9 so that the type B drivers do not wish to speed, either.

c. If the probability of receiving the fine increases to 75%, we can recompute the above expected utilities to obtain for type A: EU(fast) = 5,572 and EU(slow) = 6,561 so that type A does not wish to speed. For type B, we obtain EU(fast) = 8.5 and EU(slow) = 9, so that type B does not wish to speed either.

d. Driver B



10. Expected value of the gamble is: 1/2 (+20) + 1/2 (-20) = 0. Based on the utility function specified here, we can see that the person's utility curve is concave upward (a gambler). For a gambler, the marginal utility a $20 gain is greater than the marginal utility of a $20 loss. Therefore, the person will accept the bet since he enjoys more from gaining $20 than losing $20 of the same probability, and that the expected utility of the gamble is greater than the utility of the expected return.

11. Expected value of the gamble is: 1/2 (+20) + 1/2 (–20) = 0. Based on the utility function specified here, we can see that the person's utility curve is a straight line. For a risk neutral, the marginal utility a $20 gain is the same as the marginal utility of a $20 loss. Therefore, the person will be indifferent between gaining $20 than losing $20 of the same probability, and that the expected utility of the gamble is equal to the utility of the expected return.

12. Adverse selection generally arises when an individual has hidden characteristics that she knows, but that are unknown by other party. Moral hazard generally occurs when one takes hidden actions that it knows another party cannot observe.

10 References and Suggested Further Reading

Arrow, K. J. 'Risk Perception in Psychology,' *Economic Inquiry*, 1982.

Baye, M., *Managerial Economics and Business Strategy*, McGraw Hill, Irwin, 2002.

Graham, D., 'Cost- Benefit Analysis under Uncertainty,' *American Economic Review*, September 1981.

Commonwealth of Learning Executive MBA/MPA

E5 Managerial Economics

Block 3

Quantitative Demand Analysis

Course author: Farrokh Zandi, PhD
Professor, Schulich School of Business
York University
Toronto, Canada

Course designer: Douglas R. Cronk, M.A., ABD
Vancouver, BC, Canada

Copyright © Commonwealth of Learning, 2003

All rights reserved. No part of this course may be reproduced in any form by any means without prior permission in writing from:

The Commonwealth of Learning
1285 West Broadway
Suite 600
Vancouver, BC V6H 3X8
CANADA
e-mail: info@col.org

Contents

1 Block Three: Objectives and Introduction 1
1.1 Introduction 1
2 Elasticities of Demand 1
2.1 The (Own) Price Elasticity of Demand 2
2.2 Point versus Arc Elasticity 4
2.2.1 Demonstration Problem 5
2.3 Arc Price Elasticity 5
2.3.1 Demonstration Problem 6
2.4 Elasticity, Marginal Revenue, and Total Revenue 7
2.5 Factors That Determine Price Elasticity 10
2.5.1 Demonstration Problem 11
2.5.2 Demonstration Problem 12
2.6 Elasticity of a Product versus Elasticity of a Brand 12
2.6.1 Demonstration Problem 13
3 Other Types of Elasticity 13
3.1 Income Elasticity 13
3.2 Business Implications of Income Elasticity 14
3.3 Cross-Price Elasticity 15
3.4 Advertising Elasticities 16
3.5 Obtaining Elasticities from the Demand Function 17
3.5.1 Demonstration Problem 18
3.6 Elasticities for Nonlinear Demand Functions 19
3.6.1 Demonstration Problem 20
4 Summary and Review 20
5 Review Problems 20
6 Answer Key for Review Problems 23
7 References and Suggested Further Reading 26

1 Block Three: Objectives and Introduction

After working through Block Three of this course, you should be able to:

1. Explain the concept of elasticity.
2. Distinguish between the point and the arc elasticity of demand and apply this concept to various decsion-making situations.
3. Demonstrate the business application of income and other elasticities.
4. Explain the relationship between the price elasticity and pricing decisions.
5. Demonstrate the relationship between elasticity, marginal and total revenue.

1.1 Introduction

The term *elasticity* refers to the number of units of a particular product that consumers are willing to buy under specifically formulated conditions. In this Block, we examine demand theory, or the forces that determine the demand for a firm's product, as well as the concept of elasticity. Demand is a function of a number of independent variables or determinants. In this Block, we express demand as a mathematical expression. This will facilitate determining the demand elasticity with respect to a whole set of variables such as price, income, adverting, etc. Measurements of elasticity can help business managers to understand the market characteristics of the goods or services they are selling, and thus can help them in planning their marketing strategies, especially the pricing of their products.

2 Elasticities of Demand

If we lower a product's price, we know that sales will increase, but by how much? And what will happen to total revenue? What will happen to sales if consumers' disposable income increases? What will happen to sales if the advertising budget is increased? Will a change in the price of butter affect the sales of margarine? If so, by how much? These are important questions in the business world, and they all can be answered by understanding the concept and measurement of *elasticity*.

In general, the elasticity of any function is defined as the percentage change in the dependent variable Y that is caused by a one per cent (or relatively small) change in the independent variable X while all other independent variables are held constant. This general concept of elasticity is applicable to any function. In theory, the demand function has elasticity for each of its many independent variables. However, we shall confine our discussion to the four demand elasticities that are most widely discussed in the literature of demand theory. These are:

1. price elasticity of demand, which measures the responsiveness of sales to changes in price;

2. income elasticity of demand, which measures the responsiveness of sales to changes in consumer income;

3. cross elasticity of demand, which measures the responsiveness of sales of a product to changes in price of another price; and

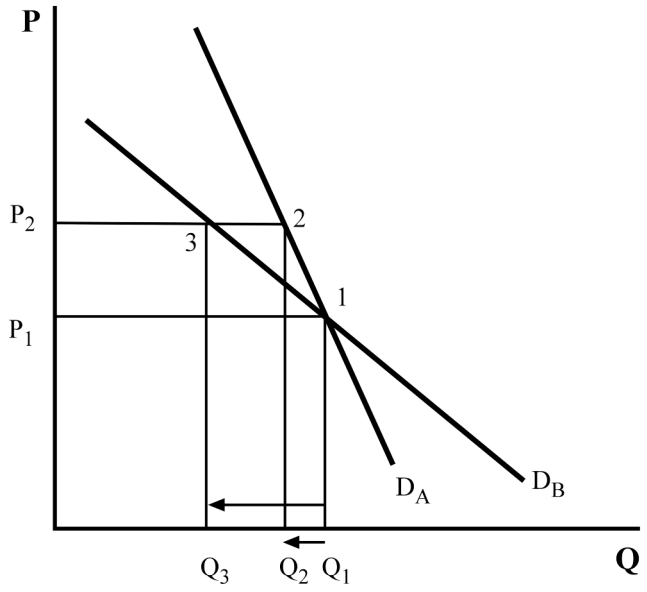
4. advertising elasticity, which measures the responsiveness of sales to changes in the amount spent on advertising and promotion.

2.1 The (Own) Price Elasticity of Demand

Look at a firm that is considering a price increase. The firm understands that according to the law of demand, the increase in price will result in the loss of some sales. This may be acceptable if the loss in sales is not too large. If sales do not suffer much, the firm may actually increase its sales revenue when it raises its price. If sales drop substantially, however, sales revenues may decline, and the firm could be worse off.

Figure 3-1 illustrates the implications of the firm's pricing decision when its demand curve has one of two alternative shapes, D_A and D_B. Suppose the firm is currently charging P_1 and selling Q_1, and is considering an increase in price to P_2.

Figure 3-1 Effect of Pricing with Two Demand Curves



When the demand curve is D_A, a change in price from P_1 to P_2 has only a small effect on the quantity demanded. However, when the demand curve is D_B, the same change in price results in a large drop in quantity demanded. When D_A is the demand curve, we would hypothesize that the increase in price would increase sales revenues since the price increase swamps the quantity decrease. Here, the quantity demanded is not very sensitive

to price. But when D_B is the demand curve, the price increase would reduce sales revenues since the price increase is swamped by the quantity decrease. Here, the quantity demanded is very sensitive to price. We would expect that the price increase would decrease sales revenues.

As this analysis shows, the shape of the demand curve can strongly affect the success of the firm's pricing strategy. The concept of the own price elasticity of demand summarizes this effect by measuring the sensitivity of quantity demanded to price. The own price elasticity of demand, commonly denoted by E_p, is the percentage change in quantity brought about by a 1 per cent change in price. Letting subscript '1' represent the initial situation and '2' represent the situation after the price changes, the formula for elasticity is:

$$E_P = \frac{\%\Delta inQ}{\%\Delta inP} = \frac{Q_2 - Q_1}{Q_1} / \frac{P_2 - P_1}{P_1} \tag{1}$$

where Δ (the Greek letter delta) is the conventional symbol for 'change.' Suppose the price increases, in Figure 3-1, from P_1 (=\$5) to P_2 (=\$6). In case of D_B, the quantity demanded drops to 1,500 from 2,000, whereas in case of D_A, it drops to 1,800. Then for D_B,

$$E_P = \frac{(1500 - 2000)}{2000} / \frac{6 - 5}{5} = -\frac{500}{2000} / \frac{1}{5} = -\frac{5}{4} = -1.25 \tag{2}$$

In this case, percentage decrease in quantity is 33%, whereas the percentage increase in price is 20%. Using the same formula for the second demand curve, D_A, we see that the decrease in price from \$6 to \$5 indicates an elasticity coefficient of –0.50:

$$E_P = \frac{(1800 - 2000)}{2000} / \frac{6 - 5}{5} = -\frac{200}{2000} / \frac{1}{5} = -\frac{5}{10} = -0.50 \tag{3}$$

The percentage increase in quantity in this case would be 10%, (1800 – 1000)/2000. Thus, over the range of prices between \$5.00 and \$6.00, quantity demanded falls at a rate of 1.25 per cent in case of D_B, and at the rate of 0.50 in case of D_A.

The demand is said to be *elastic*, if the percentage change in the quantity demanded exceeds the percentage change in price; the elasticity coefficient E_p will have a value greater than one, which is the situation along demand curve, D_B. If the percentage change in the quantity demanded is less than the percentage change in price, then E_p will be less than one. This, by definition, is an *inelastic* demand, which is the situation along demand curve D_A. If E_p is equal to one, demand is unitary elastic.

Note that the minus sign in front of the coefficient is due to the fact that ΔQ and ΔP have opposite signs, i.e., price, and quantity move in opposite directions along a demand curve. For purposes of our analysis, however, we drop the negative sign and consider only the absolute value of the coefficient. These points are summarized in Table 3-1 below.

Table 3-1: Different Types of Elasticity

Demand responsiveness may be classified in *absolute* terms as:		
a.	perfectly elastic	Ep = ∞(infinity)
b.	elastic	Ep >1
c.	unitarily elastic	Ep = 1
d.	inelastic	Ep <1
e.	perfectly inelastic	Ep = 0

Price elasticities can be estimated using statistical techniques, and economists and marketers have estimated price elasticities for many products. But in most practical situations, managers will not have the benefit of a precise numerical estimate of elasticity based on statistical techniques.

2.2 Point versus Arc Elasticity

We distinguish between point elasticity of demand and arc elasticity of demand on the basis of the size of the change in price and quantity represented by Δ in the preceding elaboration of the elasticity concept. We use point elasticity when the changes in price and quantity are infinitesimally small, since such a small price change represents a virtual point on the demand curve. For more substantial price changes, we speak of arc elasticity, (example above) since we are considering a discrete movement along, or an arc of, the demand curve. For point price elasticity, the formula is modified to read

$$E_P = \frac{dQ_X}{dP_X} \cdot \frac{P_X}{Q_X} \qquad (4)$$

where the letter d is substituted for to reflect the infinitesimally small changes in the variables P_x and Q_x.

Two components of the price elasticity are (a) the inverse of the slope, $\frac{dQ_X}{dP_X}$, which for a linear demand function is constant and (b) $\frac{P_X}{Q_X}$, the coordinates of a point on the demand curve, which is a variable. Therefore, the absolute value of the price elasticity gets larger as we move upward along a given demand curve with the price increasing and the quantity decreasing. Thus, the price elasticity of demand varies along a linear demand curve.

Point price elasticity is the appropriate concept for finding the elasticity value at a particular price level if the demand curve is known. That is, if you already know the

slope of the demand curve, you simply weight the reciprocal of that slope by the appropriate price-quantity ratio to find the elasticity.

2.2.1 Demonstration Problem

Assume that $Q_X = 10 - 2P_X$, where Q_X is the quantity demanded and P_X is the price of good X. What is the price elasticity at the point on the demand curve where $P_X = \$1$?

Answer:

Plugging in the point elasticity equation we obtain,

$$E_P = -2x\frac{1}{8} = -0.25$$

where $\frac{dQ_X}{dP_X} = -2$. This may be interpreted to mean that when the price is \$1, one per cent change in price will cause a 0.25 per cent change in quantity demanded.

2.3 Arc Price Elasticity

This concept is defined as the relative responsiveness of quantity demanded to a discrete change in price, as opposed to an infinitesimal change. As previously discussed, the point formula of Equation (4) reflects a marginal concept, and is valid only for very small movements from point to point along a demand curve. Furthermore, Equation (4) requires that the precise change in Q_X generated by a very small change in P_X be known, which requires that the demand function be known. There are many instances, however, when we are interested in measuring elasticity when the demand function is not known or when our interest lies in a larger segment of the demand curve. For this, we need the formula for arc elasticity, which calculates the average elasticity between two points on the demand curve.

Suppose the price of pineapple juice in the local supermarkets is reduced from \$3.50 per litre to \$3.00 per litre, and this causes the average sales of pineapple juice to increase from 200 litres to 300 litres per day. Assuming that all other factors entering the demand function have remained constant, we expect that two price-quantity combinations above are points on the demand curve for the juice.

P_X ($ Per Litre)	**Q_X** (Litre)
\$3.50	200
\$3.00	300

If we use the upper point as our base point, the average (arc) price elasticity between these two points will be

$$elasticity = \frac{\%\Delta Q_X}{\%\Delta P_X} = \frac{\frac{Q_2 - Q_1}{Q_1}}{\frac{P_2 - P_1}{P_1}} = \frac{\frac{300 - 200}{200}}{\frac{3 - 3.5}{3.5}} = \frac{\frac{100}{200}}{-\frac{0.5}{3.5}} = -3.5 \quad (5)$$

If on the other hand the lower point is used as the base point,

$$elasticity = \frac{\frac{Q_1 - Q_2}{Q_2}}{\frac{P_1 - P_2}{P_2}} = \frac{\frac{200 - 300}{300}}{\frac{3.5 - 3.0}{3.0}} = \frac{\frac{100}{300}}{-\frac{0.5}{3}} = -2 \quad (6)$$

As expected, the point price elasticity is lower for the lower point than the higher point on the demand curve. This is because the elasticity at the top of a demand curve is lower than at the bottom of the curve.

Therefore, the question is what to do. Since the arc elasticity is meant to provide an average measure of elasticity over the range rather than at a specific point, e.g., either the upper or the lower point, the solution is to change our base for calculating elasticity to an average base (between the two points). These average coordinates designate a point half way between them along a straight-line demand curve.

$$elasticity = \frac{\frac{Q_1 - Q_2}{(Q_1 + Q_2)/2}}{\frac{P_1 - P_2}{(P_1 + P_2)/2}} = \frac{\frac{200 - 300}{(200 + 300)/2}}{\frac{3.5 - 3.0}{(3.0 + 3.5)/2}} = \frac{\frac{100}{250}}{-\frac{0.5}{3.25}} = -2.6 \quad (7)$$

Three points need to be emphasized:

1. The arc elasticity is in fact the point elasticity at the midpoint of the arc, and hence it is a better summary measure of elasticity over the arc than either extreme point.

2. The arc elasticity value becomes increasingly less accurate when we move toward the end of the arc.

3. Finally, the wider the range (arc), the less useful this concept will be.

2.3.1 Demonstration Problem

Suppose a demand function is represented by $Q_X = 100 - 10P_X$. What is the price elasticity of demand at the point where the price is \$4? What is the price elasticity at the point where the price is \$5? What is the arc elasticity between these two points?

Answer:

At P_X = \$4, $Q_X = 100 - 10\,(4) = 60$, and at P_X = \$5, $Q_X = 100 - 10\,(5) = 50$. Hence for the first point (\$4, 60):

$$E_P = \frac{dQ_X}{dP_X}.\frac{P_X}{Q_X} = -10\left(\frac{4}{60}\right) = -2/3$$

and for the second point (\$5, 50):

$$E_P = \frac{dQ_X}{dP_X}.\frac{P_X}{Q_X} = -10\left(\frac{5}{50}\right) = -1$$

The arc elasticity between these two points is:

$$elasticity = \frac{\frac{Q_1 - Q_2}{(Q_1 + Q_2)/2}}{\frac{P_1 - P_2}{(P_1 + P_2)/2}} = \frac{\frac{50 - 60}{(60 + 50)/2}}{\frac{5 - 4}{(5 + 4)/2}} = \frac{\frac{10}{55}}{-\frac{1}{4.5}} = -\frac{45}{55} = -.82$$

Learning Tip

Note that the average elasticity at the midpoint of the arc is NOT an average between the elasticities of the end point, i.e.,

$\{(-2/3) + (-1)\}/2 = -.83 \neq -.82$

2.4 Elasticity, Marginal Revenue, and Total Revenue

Table 3-2 shows the hypothetical prices and quantities demanded of software, the own price elasticity, and the total revenue ($TR = P_xQ_x$) for the linear demand function, $Q_X^d = 10 - P_X$. Column 4 and 5, in Table 3-2, show total revenue and marginal revenue. Column 6 shows the price elasticity, $E_P = \frac{dQ_X}{dP_X}.\frac{P_X}{Q_X}$, for a linear demand curve. Given the constant slope, and hence the constant inverse of the slope, for this linear relationship (-1), the absolute value of the price elasticity gets larger as we move upward along a given demand curve with the price increasing and the quantity decreasing. Thus, the price elasticity of demand varies along a linear demand curve.

For points A through E in Table 3-2, an increase in price increases total revenue. Note that we have inverted the demand curve and therefore the inverse demand curve is $P_X^d = 10 - Q_X^d$. For example, an increase in price from \$1 to \$2 per unit increases total

revenue by $7, (MR = $7). Notice that for these two prices, the corresponding elasticity of demand is less than 1 in absolute value.

For points G through K, an increase in price leads to a reduction in total revenue. For example, when the price increases from $7 (where the price elasticity is –2.33) to $8 (where the price elasticity is –4), we see that total revenue decreases by $5 (MR = –$5). The price-quantity combination that maximizes total revenue is at point F, where the price elasticity equals one.

Table 3-2 Total Revenue, Marginal Revenue, and Elasticity $P_X^d = 10 - Q_X$

	Price	Quantity	Total Revenue	Marginal Revenue	Price Elasticity
	($/$Q_X$)	(Q_X)	(TR)	(MR)	(E_P)
A	0	10	0	0	0.00
B	1	9	9	9	-0.11
C	2	8	16	7	-0.25
D	3	7	21	5	-0.43
E	4	6	24	3	-0.66
F	5	5	25	1	-1.00
G	6	4	24	-1	-1.50
H	7	3	21	-3	-2.33
I	8	2	16	-5	-4.00
J	1	9	9	-7	-9.00
K	0	10	0	0	-∞

Note that the mathematical expression representing MR function is found as follows:

$$TR = P_X x Q_X = (10 - Q_X)Q_X = 10Q_X - Q_X^2, \text{ and} \tag{8}$$

marginal revenue that is the derivative of total revenue, with respect to Q, is as follows:

$$MR = \frac{dTR}{dQ_X} = 10_X - 2Q_X \tag{9}$$

Note that the marginal revenue equation has the same vertical intercept as the demand equation, but a slope (–2) that is double the slope of the demand curve, (–1).

Learning Tip

Notice that mathematically speaking, MR is the slope of the TR curve. Hence when TR is increasing (between A and E), MR is positive, Figure 3-2 below, whereas when TR is decreasing (between G and K), MR is negative. Finally, when TR is maximized (at point F), MR is zero.

Price elasticity, total revenue, and marginal revenue are all functionally related. Figure 3-2 shows the demand curve corresponding to the table in the upper panel, while the relationship between price and total revenue is graphed in the lower panel. The upper portion of the demand curve, where the price is higher, the magnitude of price elasticity (in absolute terms) is greater than one, hence the demand is said to be elastic. In the lower portion of the curve, the magnitude of the elasticity is less than one, hence demand is said to be inelastic. At the midpoint, where the price is \$5, the magnitude of elasticity is equal to one, and demand is said to be unit elastic.

Figure 3-2 Demand Curve Corresponding to Table 3-2

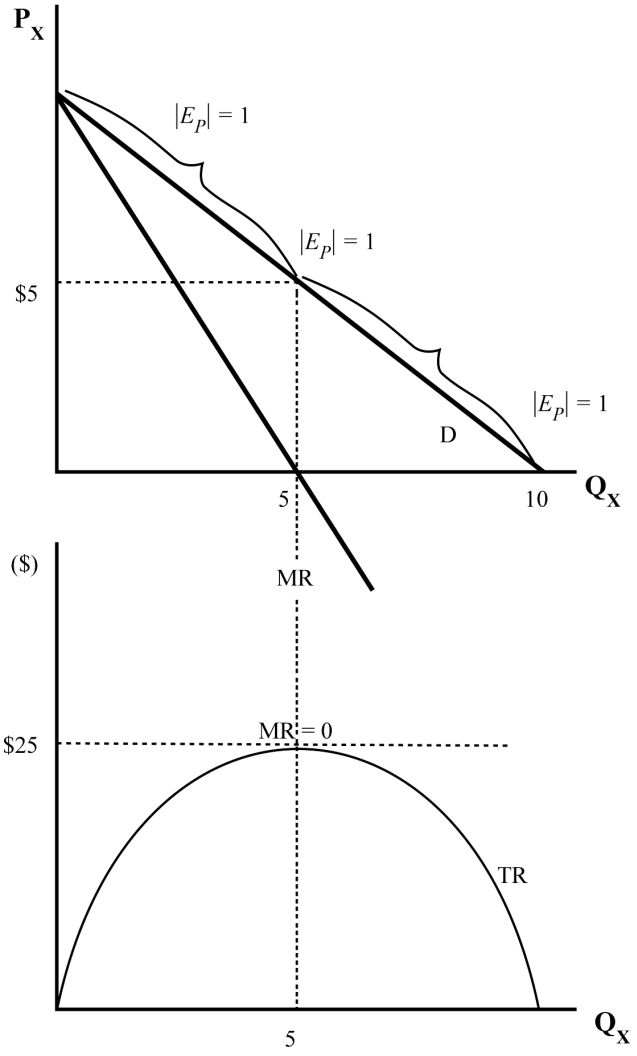


Figure 3-2 shows that when demand is elastic, reducing the price causes total revenue to rise. Why? Because a magnitude of elasticity greater than 1.0 means that if the price is cut by some percentage, the quantity demanded will rise by a greater percentage. The increase in the number of units sold will more than make up for the lower price, and total revenue will increase. By the same reasoning, when demand is inelastic, a price increase will cause total revenue to rise, even though fewer units will be sold. Maximum revenue occurs when the magnitude of elasticity equals 1.0. At this point, any change in price in either direction will cause a reduction in total revenue.

Managers pursuing the 'cut price and make it up in volume' strategy are, in effect, using the total revenue test.

The relationship depicted in the demand curve discussed above can be generalized for all linear demand curves. Furthermore, the relationship between the price, marginal revenue, and the price elasticity can be captured by the following simple formula:

$$MR_X = P_X\left(1 - \frac{1}{|E_P|}\right) \tag{10}$$

2.5 Factors That Determine Price Elasticity

Four key characteristics of a product influence its elasticity:

1. The degree to which it is viewed as luxury or necessity.
2. The number of substitutes that are available to buyers.
3. The price of the product in relation to buyers' incomes.
4. The amount of time allowed for buyers to react to price changes.

Consider the following examples.

Electricity has few substitutes; therefore, the demand for it is inelastic. As more substitutes become available, demand becomes more elastic. AZT, the drug that combats AIDS, at one time had no substitutes; therefore, demand was inelastic. The emergence of substitutes will make AZT's demand more elastic.

Suppose the price of salt or pepper doubled. It is such a small portion of expenditures for most people in North America that the price increase there would pass almost unnoticed, and quantity demanded would respond only slightly: inelastic demand. In contrast, a doubling in the price of a good that is important in one's budget (gasoline, perhaps) will provoke a great response. This is a good example of the role the price plays in the household budget.

The demand for oil offers a good example of short- and long-run elasticity. When OPEC conspired to raise the price of oil in 1973 and again in 1979, consumers, especially in industrialized oil importing nations, responded by reducing their purchases by a relatively small amount. One economic study at that time pointed out that the short-run elasticity of demand for gasoline was about –0.10, a coefficient indicating a very inelastic demand. In the 1980s, however, these consumers had changed their pattern of consumption by car pooling, driving their cars at lower speeds, using more fuel-efficient automobiles, and turning their thermostats down.

Producers in these countries complemented this response by using more fuel-efficient machinery. Thus, the long-run response to increased oil prices was much more elastic than the short-run response.

Learning Tip

In many cases, managers cannot acquire a precise value of elasticity due to unavailability of information or complexity of the product. In those instances, they rely on their own knowledge of the product and the nature of the markets. The following are examples of such cases.

a. The product has *unique features* that differentiate it from competitive products, and buyers are aware of the prices and features of rival products. Airline service is a good example of a product that is hard to differentiate and where consumers can easily inform themselves of the range of prices that exist in a particular market. This makes airline services more elastic.

b. The product is an *input* that buyers use to produce a final good whose demand is itself sensitive to price. In this case, if buyers tried to pass through to their customers even small changes in the price of the input, demand for the finished good could decrease dramatically. Demand for the input would be very elastic.

c. Buyers would incur significant costs if they switched to a substitute product. *Switching costs* could arise if the use of a product requires specialized training or expertise that is not fully transferable across different varieties of the product.

2.5.1 Demonstration Problem

How would you categorize the demand in the market for components and materials by a personal computer manufacturer?

Answer:

This is the market for inputs. Therefore, demand for the input of component and materials is likely to be highly price elastic because consumer demand for personal computers is highly price elastic.

2.5.2 Demonstration Problem

How would you categorize the demand for upgrades in the following case? A consumer develops expertise in using a particular word processing package that is incompatible with available alternatives.

Answer:

Since switching costs will be high, the price sensitivity for upgrades will be low.

Learning Tip

Comparisons among substitute products are difficult. This could be because the product is complex and has many performance dimensions, because consumers have little or no experience with substitute products and thus would face a risk if they purchased them, or because comparison shopping is costly. Items sold door-to-door, such as Avon cosmetics, have traditionally been price inelastic because, at the time of sale, most consumers lack good information about the prices of alternatives.

2.6 Elasticity of a Product versus Elasticity of a Brand

Students often mistakenly suppose that just because the demand for a product is inelastic, the demand facing each seller of that product is also inelastic. Consider, for example, wine. Many studies have documented that the demand for wine is price inelastic, with elasticities well below l. This suggests that a general increase in the price of all types of wine from all sources would only modestly affect overall consumption of wine. However, if the price of only one specific winery or vintage of wine increases, the demand for that type would probably drop substantially because consumers would switch to the now lower-priced substitutes. As you might expect, responsiveness is generally greater for a brand than for the product category, for the simple reason that competing brands within the category offer consumers more substitutes.

Brand-level elasticities are higher than industry-level elasticities because consumers can purchase other brands when only one brand raises its price. Brand-level elasticities should also increase as more firms enter the market and more brands are offered.

2.6.1 Demonstration Problem

Should a firm use an industry-level elasticity or a brand-1evel elasticity in assessing the impact of a price change?

Answer:

The answer depends on what the firm expects its rivals to do. If a firm expects that rivals will quickly match its price change, then the industry-level elasticity is appropriate. If, by contrast, a firm expects that rivals will not match its price change, then the brand-level elasticity is appropriate.

3 Other Types of Elasticity

In addition to price elasticity, there are three other important types of elasticity that economists track: *income elasticity*, *cross-price elasticity*, and *advertising elasticity*. These elasticities measure, respectively, the responsiveness of demand to changes in consumers' income, the price of substitute goods, and advertising.

3.1 Income Elasticity

You can reasonably expect that when income rises, consumers will buy more of a particular product, less when their income falls. In fact, goods and services that exhibit such a relationship are called *normal.* However, where there is an inverse relationship between changes in income and consumer demand, the products are called *inferior.* Examples of inferior products or services are less expensive means of transportation (bus versus plane), low quality rice, and no-name products. As people's incomes rise, they start to replace these products with their higher-priced substitutes: brand-name products, for example.

Income elasticity is measured in the same way as price elasticity. The percentage change in the quantity demanded is compared with the percentage change in income. The arc income elasticity formula is:

$$E_M = \frac{\%change(quantity)}{\%change(income)} = \frac{\%\Delta Q_X}{\%\Delta M} \tag{11}$$

where E_M denotes income elasticity. We can categorize the results of this computation as follows.

If the income elasticity coefficient is positive ($E_M > 0$), it indicates that a move in the same direction is occurring with both income and changes in quantity demanded. Products with coefficients greater than zero are called *normal.* As your income increases, you will probably increase your spending on soft drinks, books, clothes, CDs, etc.

If the income elasticity is negative, it indicates that quantity demanded and level of income move in opposite directions. Therefore, the product is inferior. Potatoes, beans, and generic aspirin are good examples of inferior goods. As your income increases, you will probably decrease your spending on such goods (negative elasticity, $E_M < 0$). Conversely, if real income levels decline, the quantity demanded of an inferior good will increase.

If the income elasticity coefficient is greater than one, it indicates that demand is very sensitive to changes in income. In this case, we can refer to the product as a luxury or superior product.

Home ownership might be a luxury. If your income is low, you can only rent. If your income rises, you may qualify for mortgage loans and enter the house-purchasing market. Expenditure on house purchases rises more than the increase in income in such a case.

Alternatively, the point-elasticity formula for income is

$$E_M = \frac{\%\Delta Q_X}{\%\Delta M} = \frac{\Delta Q_X}{\Delta M} \cdot \frac{M}{Q_X} \tag{12}$$

3.2 Business Implications of Income Elasticity

The implications of income elasticity of demand to the business decision-maker are considerable. If the income elasticity for your product exceeds one, the demand for your product will grow more rapidly than does total consumer income. As well, it will fall more rapidly than does total consumer income when income levels are generally falling. Hence, while income elasticity greater than one in a growing economy indicates a growth industry, it also indicates a greater vulnerability to downturns in the level of aggregate economic activity.

Contrastingly, if the income elasticity of demand for your product is positive but less than one, the demand for your product will grow more slowly than the gross national product or consumer income. (However, it will be relatively recession-proof, in the sense that the demand will not react in the volatile fashion of luxury goods.) Third, if your product is regarded as an inferior good by the market as a whole, you must expect the quantity demanded of your product to decline as the gross national product rises.

Therefore, knowledge of a product's income elasticity can help managers in several different ways. First, it can alert them to the impact on demand caused by movements in the macroeconomy. A recession can be expected to reduce the demand for normal or superior products. In an economic recovery or expansion, these same products should experience rising demand. For example, during the sustained economic expansion of the

1980s, companies that sold luxury consumer products with high-status designer names did very well. In the 1990s, however, many of the same companies experienced sluggish sales because of the slowdown in the economy.

To offset the impact of the business cycle on product demand, a manager might do well to select a portfolio of goods and services with a variety of income elasticities. Thus, in a recession, the demand for a company's inferior or low-income-elasticity products will be sustained and may even increase. In expansionary economic times, the company's high-income-elasticity products would take the lead in sales.

Learning Tip

No good is automatically inferior, normal, or a luxury. For a poor person gaining income in a distant neighbourhood, a bus ride may be a luxury, while that same bus ride might be inferior for a millionaire. You should also realize that a luxury — e.g., bubblegum, need not be equated with *high-priced*, nor necessity (dental care) with *low-priced.* Clearly, there may be a link between income elasticity and price elasticity. Whether a good has many substitutes may be linked with whether it is considered a luxury or a necessity.

3.3 Cross-Price Elasticity

Cross-price elasticity is a measure of the responsiveness of consumers to changes in the price of a particular good, Good X, relative to changes in the price of substitute or complementary products, Good Y. The cross-elasticity of demand provides a measure of the degree of substitutability, or complementarity, between product X and some other product.

Cross-elasticity of demand is defined as the percentage change in quantity demanded of product X, divided by the percentage change in the price of some product Y.

The arc formula is

$$E_C = \frac{\%\Delta Q_X}{\%\Delta P_Y} \tag{13}$$

And the point-elasticity formula is

$$E_C = \frac{\Delta Q_X}{\Delta P_Y} \cdot \frac{P_Y}{Q_X} \tag{14}$$

The main point here is the sign (positive or negative) of the relationship rather than the magnitude. If it is a positive relationship, the goods are substitutes; if it is negative, they are complements. As a secondary issue, the larger (in absolute value) the coefficient, the more related are the two goods. For instance, a small decrease in the price of Pepsi may cause a sizeable decrease in the demand for Coke (close substitutes) but a smaller decrease in the demand for, say, tea.

Learning Tip

For *substitutes*, an increase in the price of product A (Pepsi) will lead to a decrease in the quantity demanded of that product or brand and an increase in the demand for B (Coke), a positive coefficient. For *complements*, an increase in the price of CDs will lead to a decrease in the demand for CD players, negative coefficient.

Knowledge about cross-price elasticity with respect to substitute products is particularly useful to assess the impact on sales of changes in a competitor's prices. For example, what impact will a reduction in the price of Microsoft's Word have on the sales of Word Perfect? To minimize the cross-price elasticity of a product with respect to changes in the price of a substitute, companies spend considerable sums on advertising designed to establish or strengthen brand loyalty. There appears to be increasing cross-price elasticity in consumer goods markets, as evidenced by the growing market share of lower-priced, private-label consumer products. This is becoming quite worrisome for the makers of leading premium brands of consumer products such as Procter & Gamble, Colgate-Palmolive, and Philip Morris.

The cross-price elasticity of a product with respect to complementary products is also important for managers to understand. For example, a seller of computer products can reduce the price of its PCs to stimulate demand for its software. If the profit margin is high for the product whose demand is affected by the cut in the price of the complementary products, this pricing tactic is particularly appealing. For example, a clothing store might reduce the price of its men's suits to stimulate the demand for high-profit margin items such as ties, shirts, and socks. Furthermore, the degree of complementarity between suits and the fashion accessories can be stimulated by the friendly persuasion of the salespersons.

3.4 Advertising Elasticities

We know that advertising has an impact on the quantity of output sold. Specifically, the quantity demanded of product X will typically show a positive response to the advertising in support of product X, a negative response to the advertising of substitutes, and a positive response to the advertising of complements.

The advertising elasticity of demand for product X measures the responsiveness of the change in quantity demanded to a change in the advertising budget for product X. We expect a positive relationship between advertising and quantity demanded, but we also expect that the responsiveness of sales to advertising will decline as advertising expenditure continues to increase.

Similarly, cross-advertising elasticity of demand measures the responsiveness of quantity demanded of product X to a change in the advertising efforts directed at another product, Y. As stated earlier, one expects cross-advertising elasticity to be negative between substitute products and positive between complementary products. For example, increased advertising efforts for a particular movie would be expected to reduce the quantity demanded of admission tickets to other movies and attractions but to increase the sales at the refreshment kiosk in the lobby of that particular movie theatre. In effect, the increased advertising would have shifted the demand curves to the left for all substitute attractions, while shifting the demand curve to the right for the refreshment kiosk.

$$E_{adv} = \frac{\%\Delta Q_X}{\%\Delta adv} \tag{15}$$

It is clear that we might calculate the elasticity of demand with respect to any variable that influences the demand for a product.

3.5 Obtaining Elasticities from the Demand Function

We will now turn our attention to how to calculate various elasticities from a demand function. In order to do this, we need to recall the demand function. As discussed earlier, a demand function refers to the relationship that exists between the quantity demanded of a particular product and the determinants of demand. In notational shorthand, such is represented as follows:

$$Q_X^d = f(P_X, P_Y, M, adv.,...) \tag{16}$$

If the demand function is linear, then

$$Q_X^d = a_0 + a_X P_X + a_Y P_Y + a_M M + a_{adv} adv \tag{17}$$

where a_0, a_X, a_Y, a_M, and a_{adv} are the coefficients of the demand function indicating the marginal effect of each right hand side (independent) variable on the quantity demanded (dependent variable). Note that a_0 shows the marginal effect of other variables not accounted for here.

Learning Tip

Note that the positive and negative sign preceding each coefficient indicates the direction of the influence that variable has on Q_X. Given this specification of the demand function, we can calculate the elasticity of demand with respect to all the independent variables.

The relationship between elasticities and these coefficients, as discussed earlier, are summarized as follows:

Own price elasticity: $E_P = a_X . \frac{P_Y}{Q_X}$

Cross price elasticity: $E_C = a_Y . \frac{P_Y}{Q_X}$

Income elasticity: $E_M = a_M . \frac{M}{Q_X}$

Advertising elasticity: $E_{adv} = a_{adv} . \frac{adv}{Q_X}$

3.5.1 Demonstration Problem

Suppose the demand function for product X has been estimated to be

$$Q_X^d = 20 + \square 5P_X + 4P_Y + 0.2M + 3adv,$$

where the symbols for the independent variables are the same as those introduced above. Suppose X sells for \$20 per unit, the price of alternative product is \$30 per unit, average consumer income is \$25,000, and the advertising budget is 40 units. Calculate the demand elasticity with respect to all the independent variables. Interpret your results.

Answer:

Plugging in this equation:

$$Q_X^d = 20 \square 5(\$20) + 4(\$30) + 0.2(25{,}000) + 3(40) = 5{,}160$$

hence $E_P = \square 5 . \frac{20}{5160} = \square 0.19$, $E_C = 4 . \frac{30}{5160} = 0.23$

$$E_M = 0.2.\frac{25{,}000}{5160} = 0.97 \text{, and } E_{adv} = 3\frac{40}{5160} = 0.23$$

The following conclusions can be drawn. (a) Since the coefficient of P_y is positive, X and Y are substitutes. (b) The own-price elasticity of demand is inelastic (–0.19). (c) The product under study is almost unit income elastic. And (d) advertising does some good, as indicated by the positive coefficient to adv.

Learning Tip

For a linear demand curve, the value of elasticity depends on the particular price and quantity at which it is calculated. This means that the own price elasticity is not the same as the slope of the demand curve. In fact, for a linear demand function, demand is elastic at high prices and inelastic at lower prices.

3.6 Elasticities for Nonlinear Demand Functions

Managers frequently encounter situations where a product's demand is not a linear function of prices, income, advertising, and other demand shifters. In this section we demonstrate that the tools we developed can easily be adapted to these more complex environments.

Suppose the demand function is not a linear function but instead is given by

$$Q_X = a_0.P_X^{a_X}.P_Y^{a_Y}.M^{a_M}.adv^{a_{adv}} \tag{18}$$

In this case, the variables have a multiplicative impact on Q_X, and the coefficients are transformation is used. Expressed in the natural logarithm (log):

$$\ln Q_X = \ln a_0 + a_X.\ln P_X + a_Y.\ln P_Y + a_M.\ln M + a_{adv}.\ln adv \tag{19}$$

When the demand function is log-linear, as in equation (19), the coefficients are the elasticities.

3.6.1 Demonstration Problem

Suppose the demand for product X is given in log-linear as follows:

$\ln Q_X = 50 - 0.75 \ln P_X + 1.2 \ln P_Y + 2 \ln M$. What are the demand elasticities of product X with respect to income and its own price?

Answer:

$E_P = -0.75$

$E_M = 1.2.$

4 Summary and Review

The demand for a product that a firm faces depends on the price of the product, consumers' income, the price of related products, advertising and promotions. Such a relationship is referred to as a demand function. Elasticity is the most commonly used measurement of the sensitivity of demand to any of its determinants. In general, elasticity is defined as a percentage change in quantity (the dependent variable) of output demanded caused by a one percentage change in an independent variable. Elasticity is a general concept that applies to any function, price of the product, income, price of related products, advertising, etc. The demand function is unit elastic at the point where marginal revenue is zero and total revenue is at a maximum. Below this point, demand is inelastic, marginal revenue is negative, and a price increase total revenue. Above this point, demand is elastic and the firm can increase total revenue by reducing price.

5 Review Problems

1. For the demand function Q = 10 – 0.5P
   a. Find the point elasticity of demand. Briefly explain why the value is negative.
   b. Find the point elasticity of demand at P = 5.

2. For the inverse demand function P = 5 – 2Q
   a. Find the revenue function (i.e., revenue as a function of quantity).
   b. Find the elasticity of demand.
   c. At which point the elasticity of demand is one? What will be the effect on revenue if you change the price by 1% at this point? Explain.

3. The cross-price elasticity of tea and coffee is estimated to be 0.15. Explain why the value is positive. If the coffee price decreases by 25%, what will be the effect on sales of tea?

4. For the demand function $Q = 10P^{-0.5}$

   a. Find the elasticity of demand.

   b. Find the marginal revenue function.

   c. Find the relationship between marginal revenue and elasticity of demand. Briefly explain.

5. Find the income elasticity for a good with the demand function $Q = 10MP^{-1}$, at M = 50,000 and P = 100. (Here, Q = Quantity, M = income, and P = price). Is this good normal or inferior? Explain.

6. The demand for yogurts of brand 'A' is estimated to be,

   $Q = 50 - 5P_a + 4P_b + 0.05M$

   Q: Quantity, P_a: Price of brand 'A', P_b: Price of brand 'B', M = Income

   a. Find the own price elasticity of demand.

   b. Find the cross price elasticity of demand.

   c. Find the income elasticity of demand.

   d. Are 'A' and 'B' substitutes or complements? Explain.

   e. Is 'A' an inferior good? Explain why or why not.

7. The demand for good X is given to be $Q = 100M^{-0.5}P_x^{-0.5}P_y^{-0.25}$. Find own price, cross price, and income elasticities of demand. Interpret the relationship between X and Y. (Q: Quantity; P_x: Price of X; P_y: Price of Y; M: Income). What sort of good is product X? Explain.

8. A store selling clothing and shoes has monthly revenue of $100,000 from clothing and $50,000 from shoes. Own price elasticity of demand for clothes is –2 and cross price elasticity of demand between clothes and shoes is –0.5. If the store increases the price of clothes by 5%, what will be the change in revenue?

9. The demand function for good Y is as follows:

   $Q = 50P_Y^{-2}P_X^{0.5}MA^{0.75}$

   Q: Quantity; P_x: Price of X; P_y: Price of Y; M: Income, A: Advertising Expenditure

At $P_y = 25$, $P_X = 16$, M = 20,000 and A = 75,000, find the own price, cross price, and income elasticities. If the advertising expenditure is increased to 100,000, how will it affect the revenue?

10. If the log-linear demand function of a good is

    $\text{Log } Q = 200 + 0.5\log P_y + 0.25\log P_x + 0.05\log A + 0.01\log M$

    Q: Quantity; P_x: Price of X; P_y: Price of Y; M: Income, A: Advertising Expenditure

    Find the own price, cross price, and income elasticities of demand. Can we get the answers by looking at the log-linear function? Explain.

11. For the demand function $Q = 200 - 2P + 0.5M$, where Q: Quantity; P: Price; M: Income:

    a. Find the point elasticity of demand.

    b. Find the elasticity of demand.

    c. Find the ranges of price at which the elasticity of demand is inelastic, elastic, and unit elastic.

12. The demand function for good Y is as follows:

    $Q = KP_y^a \times P_x^b \times M^c$

    Q: Quantity; P_x: Price of X; P_y: Price of Y; M: Income, A: Advertising Expenditure; K: constant (>0). Find the own price, cross price, income, and advertising elasticities of demand. If X is a normal good, and X and Y are complements, then find the restrictions for the parameters a, b, & c.

13. Define mathematically the relationship between marginal revenue and elasticity of demand.

6 Answer Key to Review Problems

1. $Q = 10 – 0.5P$

   Elasticity = $P/Q \times dQ/dP = P/Q(–0.5)$

   Since $P > 0$ and $Q > 0$, therefore, elasticity <0. Elasticity is negative because with increase in price, quantity demanded goes down.

   At $P = 5$, $Q = 7.5$. Elasticity = $5/7.5\ (–0.5) = –0.33$

2. $P = 5 – 2Q$, $Q = 2.5 – 0.5P$

   Revenue function $R(Q) = P \times Q = 5Q – 2Q^2$

   Elasticity = $P/Q \times dQ/dP = P/Q \times (–0.5)$

   $|\text{Elasticity}| = 1$ …… $|-0.5P/Q| = 1$, $Q = 0.5P$, $2.5 – 0.5P = 0.5P$, $P = 2.5$, and $Q = 1.25$

   At this point if the price is changed by 1% then the quantity demanded will also change by 1%, leaving revenue unchanged.

3. The cross price elasticity of tea and coffee is 0.15. The positive value says that if the price of coffee goes up, then the quantity demanded of tea will also go up. Thus tea and coffee are substitutes.

   $\%\Delta Q_t / \%\Delta P_t = 0.15 \Rightarrow \%\Delta Q_t / 25 = 0.15 \Rightarrow \%\Delta Q_t = 3.75$

   Quantity demanded of tea will increase by 3.75%

4. $Q = 10P^{-0.5}$…… $P = 100Q^{-2}$

   Elasticity = $P/Q \times dQ/dP = P/Q \times 10(-0.5)P^{-1.5} = (–0.5)(10P^{-0.5})/Q = –0.5$

   Revenue function $R(Q) = PQ = 100Q^{-2} \times Q = 100/Q$

   Marginal Revenue = $dR/dQ = –100Q^{-2}$

   Marginal Revenue = $dR/dQ = d(PQ)/dQ = P + Q_x\ dip/dQ = P[1+Q/P \times dP/dQ] = P[1+ 1/(P/Q \times dQ/dP)] = P[1 + 1/E_d]$, where E_d = elasticity of demand.

5. $Q = 10MP^{-1}$

   $\ln Q = \ln 10 + \ln M – \ln P$

   $d \ln Q / d \ln M = 1$ = income elasticity of demand

   With a positive change in income, the quantity demanded for the good will also change positively. Therefore, the good is normal. Therefore, at $M = 50{,}000$ and $P = 100$, the elasticity will be 1.

6. $E(Q,P_a) = -5 \times P_a/Q$

$P_a = 50$, $P_b = 50$, $M = 10{,}000$. Plugging in, $Q = 50 - 5(50) + 4(50) + 0.05(10{,}000) = 500$

$E(Q,P_a) = -5.\ 50/500 = -0.5$

$E(Q,P_b) = +4 \times P_b/Q = 0.4$

$E(Q,M) = 0.05 \times M/Q = 1$

$E(Q,P_b) > 0$, therefore, A and B are substitutes.

$E(Q,M) > 1$, therefore A is a normal good.

7. $Q = 100.\ M^{-0.5}.\ P_x^{-0.5}.\ P_y^{-0.25}$

$\ln Q = \ln 100 - 0.5 \ln M - 0.5 \ln P_x = -0.25 \ln P_y$

$d \ln(Q) / d \ln(M) = -0.5 = E(Q,M)$

$d \ln(Q) / d \ln(P_x) = -0.5 = E(Q,Px)$

$d \ln(Q)/d \ln(P_y) = -0.5 = E(Q,Py)$

Cross price elasticity < 0, hence x and y are complements.

Income elasticity > 0, hence x is normal.

8. A store selling clothing and shoes.

$$\Delta R = [R_c(1 + E_{QcPc}) + R_s(E_{QcPc})]x(\%\Delta Pc)$$
$$\Delta R = [100(1 \Delta 2) + 50(\Delta 0.5)]x0.05$$
$$= \Delta 6.25$$

Revenue will decrease by $6,250

9. $Q = 50P_y^{-2}.P_x^{0.5}.M.A^{0.75}$

$\ln Q = \ln 50 - 2\ln P_y + 0.5\ln P_x + \ln M + 0.75\ln A$

Therefore,

$$E_{Q,A} = 0.75$$
$$\frac{\%\Delta Q}{\%\Delta A} = 0.75$$
$$\%\Delta Q = 33(0.75) = 25$$

With Q increasing by 25%, revenue will also increase by 25%, if price remains the same.

10. Since this is a log linear function, therefore, the coefficients of log(P_y)and log(P_x) and log(M) will give us the elasticities (cross price, own price, and income, respectively).

    E (Q,Px) = 0.5

    E (Q,Py) = 0.25

    E (Q,M) = 0.01

11. $E_{QP} = -2 \times P/Q$

    $E_{QM} = 0.5 \times M/Q$

    Own price elasticity is inelastic

    $| E_{QP} | < 1$

    $2 \times P/Q < 1$

    $P < Q/2$

    Own price elasticity is unit elastic

    $|E_{QP}| = 1$

    $2 \times P/Q = 1$

    $P = Q/2$

    Own price elasticity is elastic

    $E_{QP} | > 1$

    $x\ P/Q > 1$

    $> Q/2$

12. Own price elasticity = $d\ lnQ/d\ lnP_y$

    $= K \times P_y^a \times P_x^b \times M^c$

    $n\ Q = ln\ K + a\ ln(P_y) + b\ ln(P_x) + c\ lnM$

    Therefore, own price elasticity = a, cross price elasticity = b, and income elasticity = c.

13. Without attempting to derive it here, the mathematical expression relating MR to P is as follows:

    $MR = P(1-E_p)$

7 References and Suggested Further Reading

Baye, M., *Managerial Economics and Business Strategy*, McGraw Hill, Irwin, 2002.

Besanko, D. and R. Braeutigam, *Microeconomics: An Integrated Approach*, Wiley, 2002.

Douglas, J. E., *Managerial Economics: Analysis and Strategy*, Prentice Hall, 1992.

Pindyck, R. and Rubinfeld D., *Microeconomics*, Fifth Edition, Prentice Hall, 2001.

Commonwealth of Learning Executive MBA/MPA

E5 Managerial Economics

Block 4

The Theory of Consumer Behaviour

Course author: Farrokh Zandi, PhD
Professor, Schulich School of Business
York University
Toronto, Canada

Course designer: Douglas R. Cronk, M.A., ABD
Vancouver, BC, Canada

Copyright © Commonwealth of Learning, 2003
All rights reserved. No part of this course may be reproduced in any form by any means without prior permission in writing from:

The Commonwealth of Learning
1285 West Broadway
Suite 600
Vancouver, BC V6H 3X8
CANADA
e-mail: info@col.org

Contents

1 Block Four: Objectives and Introduction 1
1.1 Introduction 1
2 Indifference Curves 2
2.1 Budget Constraint 5
2.2 Changes in Income 6
2.3 Changes in Price 6
2.4 Consumer Equilibrium 7
2.4.1 Demonstration Problem 9
2.5 Price Changes and Consumer Behaviour 9
2.6 Income Changes 10
2.7 The Income and Substitution Effects of a Price Change 11
2.7.1 Demonstration Problem 13
2.7.2 Demonstration Problem 13
2.8 Corner Solutions 14
3 Application of Indifference Curves 15
3.1 Quantity Discounts 15
3.2 Voucher versus Cash Subsidies 15
4 Relationship Between Indifference Curve Analysis and Demand Curves 17
4.1 Individual Demand 18
4.2 Market Demand 19
5 The Attribute Approach to Consumer Choice 20
5.1 Depicting Products in Attribute Space 20
5.2 Utility Maximization from Attributes 22
6 Summary and Review 23
7 Review Problems 24
8 Answer Key to Review Problems 26
8 References and Suggested Further Reading 28

1 Block Four: Objectives and Introduction

After working through Block Four of this course, you should be able to:

1. Explain the essence of ordinal approach to utility analysis.
2. Explain the concept of utility maximization and consumer's equilibrium.
3. Describe the applications of indifference curves analysis and apply the analysis to a variety of consumer situations.
4. Explain how individual demand curve is derived from the indifference curve analysis.
5. Explain the concept of attribute approach to consumer choice.
6. Explain the difference between the traditional theory and the Attribute approach to consumer choice.
7. Describe the application of the Attribute approach in the marketing field and apply this approach to marketing situations.

1.1 Introduction

This Block develops tools that help a manager understand the behaviour of individual consumers and the impact of alternative incentives on their decisions. Despite the complexities of human thought processes, managers need a model that explains how individuals behave in the marketplace and in the work environment.

As we have illustrated, consumers' efforts to get the most for their money are reflected in their individual demand functions. Market-demand curves are simply the aggregates of all individual consumer-demand curves. Since utility is based on individual taste and preferences, each market-demand curve reflects the aggregate market preferences of all consumers in that market. Therefore, a market-demand curve is a powerful signal to producers about what and how much to produce.

The Utility approach offers useful insights into consumer behaviour. In this context, consumer preferences, which are *subjective,* are connected to changes in prices, incomes, and other variables in the marketplace, which are *objective.* This approach, which is referred to as the ordinal approach, requires that combinations of products be ranked in order of preference. The ordinal approach to consumer behaviour (demand analysis) assumes that consumers are able to rank all conceivable bundles of commodities; that is, when confronted with two or more bundles of goods and services, consumers can determine an order of preference among them.

Order of preference does not require consumers to estimate how much utility will be attained from a bundle of goods. Only the ability to rank is fundamental. Furthermore, the degree of preference is irrelevant. It is quite enough for the consumer to think, subjectively and idiosyncratically, that one product or bundle of goods is better than another. In more precise terms, we assume that the consumer's preference pattern possess the following properties.

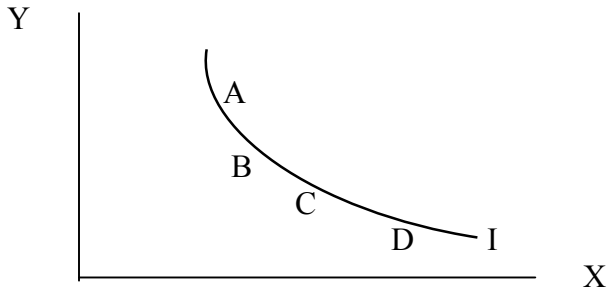
Property 1: *Completeness*. If an individual can rank any pair of bundles chosen at random from all conceivably bundles, he or she can rank all conceivable bundles. Put differently, for any two bundles, say A and B, either A > B or B > A, or A ~ B.

Property 2: *More is better than less*. If bundle A contains at least as many units of each good as bundle B, and more units of at least one commodity, A must be preferred to B. If more is better than less, the consumer must view the products under consideration as 'goods' instead of 'bads.'

2 Indifference Curves

Let us consider some implications of these assumptions about preference orderings. Most important, they enable us to generate a graphical description of consumers' preferences. An *indifference curve* defines the combinations of two goods (or two bundles of goods) that give the consumer the same level of satisfaction, i.e., they are equally preferred. Along an indifference curve, the consumer is indifferent between alternative combinations of goods X and Y. This is shown as the curve labeled I, in Figure 4-1. Bundles A, B, C, and D are all equally preferred combinations of X and Y located on the indifference curve providing the consumer with the same level of satisfaction.

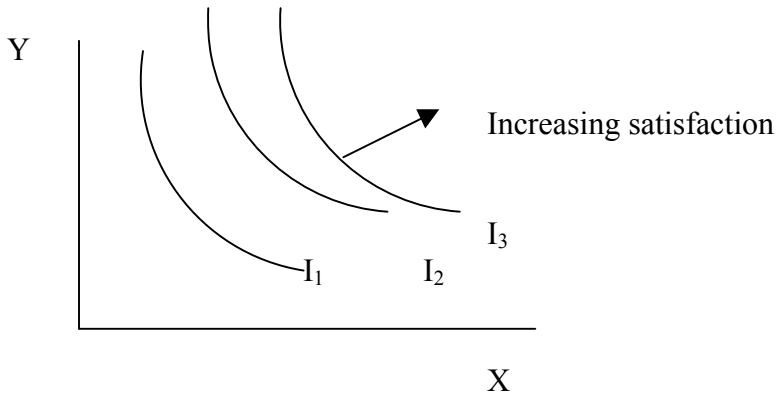
Figure 4-1



An indifference curve also allows us to compare the satisfaction implicit in bundles that lie along it with those that lie either above or below it. The completeness property of preferences implies that there is an indifference curve that passes through every possible bundle. Therefore, we can represent a consumer's preference with an *indifference map.*

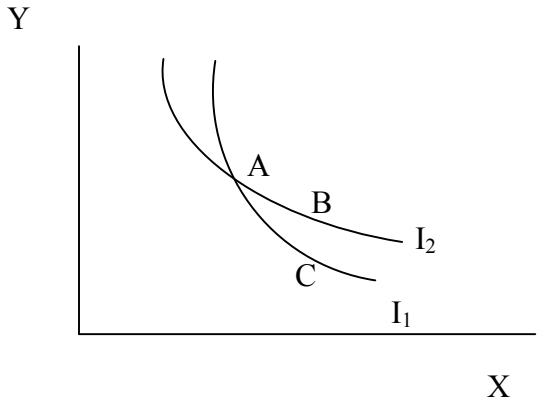
I_1, I_2, and I_3, in Figure 4-2, are index values used to denote the order of preference that corresponds to the respective indifference curves.

Figure 4-2



Property 3: *Transitivity.* Given three bundles of goods (A, B, and C), if an individual prefers A to B and B to C, he must prefer A to C. Similarly, if an individual is indifferent between A and B and between B and C, he must be indifferent between A and C. Finally, if he is indifferent between A and B and prefers B to C, he must prefer A to C. This assumption obviously can be carried over to four or more different bundles.

Figure 4-3



The assumption of transitive preferences, together with the 'more is better than less' assumption, implies that indifference curves (from the same indifference map) cannot cross. To see why, since A and B are on the same indifference curve (I_2), A ~ B, in Figure 4-3, and since A and C are on the same indifference curve (I_1), A ~ C, the consumer must be indifferent between B and C also. This, however, can happen if these

points are on the same indifference curve. This situation implies that the consumer cannot make a choice.

Property 4. Indifference curves are concave from above (convex to the origin). In the preceding discussion, it was seen that indifference curves are concave from above and downward sloping. These characteristics arise from the assumption of *diminishing marginal utility* that was built into the concept from which the indifference curves were derived; the more of a product you consume, the less will be the additional satisfaction (utility) obtained from additional consumption of that product. Since diminishing marginal utility plays a crucial role in the consumer behaviour model, it must be thoroughly understood.

As previously noted, different combinations of goods can provide equal levels of total utility. When a consumer remains on a particular indifference curve, one commodity can be substituted for the other so that the consumer remains as well off as before. The rate at which a consumer is willing to make such a substitution is a matter of great interest and importance. We call it the *marginal rate of substitution (MRSxy)* of X for Y, defined as the number of units of Y that must be given up to acquire one additional unit of X while satisfying the condition of constant total utility.

MRS_{xy}, or simply MRS, is measured as the slope of the indifference curve, which is different at each point along the curve. Since each point represents a different combination of goods X and Y, it follows that each combination has a different MRS. Is there a relationship between marginal utility and the marginal rate of substitution? Indeed there is, since the slope of the indifference curve is the direct result of the law of diminishing marginal utility.

Consider what happens when we move down the indifference curve between any two points. Consumption of Y is reduced by ΔY unit causing a loss of utility, which is equal to the change in Y, (ΔY), times the loss of utility associated with this change in Y (marginal utility of Y = MU_y). That is, $-\Delta Y \times MU_y$. But, since total utility is unchanged as we move down the curve, the loss of utility from consuming less Y is precisely offset by a gain from consuming more X, $\Delta X \times MU_x$:

$$-\Delta Y \times MU_y = \Delta X \times MU_x \qquad (1)$$

Rearranging this equation, we get

$$-\frac{\Delta Y}{\Delta X} = \frac{MU_X}{MU_Y} \qquad (2)$$

The left hand side is the slope of an indifference curve, therefore,

$$MRS = \frac{MU_X}{MU_Y} \qquad (3)$$

The continuously declining *MRS* is the logical result of the assumption that the marginal utility of a product decreases as we obtain more of it. It follows, then, that the more of a product one has, the more willing one is to trade it for another product.

This willingness to exchange what we value less for what we value more is true whether the owners of the commodities are individuals, firms, or nations. Thus the marginal rate of substitution governs both domestic and foreign trade.

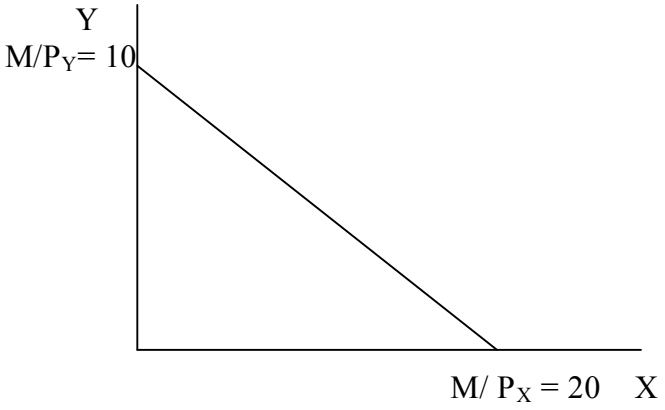
2.1 Budget Constraint

Indifference curves reflect the consumer's personal feelings and the relative values regarding the consumption of various combinations of any two products. They show what combinations of X and Y the consumer is willing to accept as equally satisfactory and they are totally independent of the consumer's income and market prices. Whereas indifference curves show what the consumer is *willing* to do, income and market prices determine what a consumer is *able* to do. In other words, the budget constraint determines the combinations of X and Y that are affordable. The consumer's expenditure on X plus his or her expenditure on Y cannot and does not exceed the consumer's income. Assuming that the consumer's income is entirely spent on these two goods, we can write a budget line relationship as follows:

$$P_X \times X + P_Y \times Y = M \qquad (4)$$

where P_X and P_Y are the price of good X and Y, respectively, and M is the income or the budget constraint. Suppose that the available money income is \$100, P_x = \$5, and P_Y = \$10. If the entire \$100 is spent on good X, a maximum of 20 units (100/\$5) of X can be bought, whereas if the entire \$100 is spent on Y, as much as 10 units (100/\$10) of Y can be purchased. These combinations are illustrated in Figure 4-4. A straight line joining these two points on the graph shows *all* the other combinations of X and Y that the consumer's income will allow him or her to purchase at these prices.

Figure 4-4



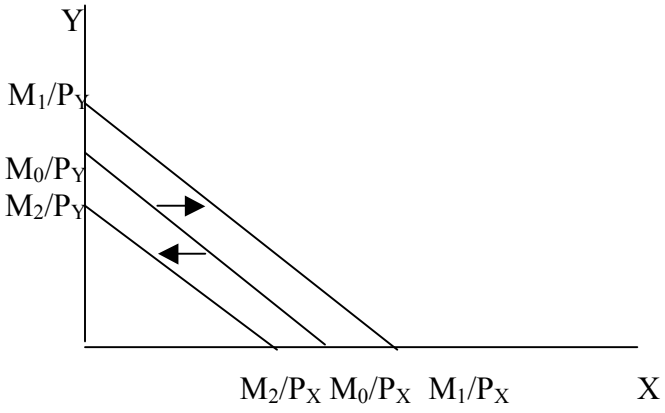
The points inside the budget line are attainable while those outside are not. The slope of the budget line is given by $-P_X/P_Y$ (= the rise (M/P_Y divides by the run M/P_X) that represents the rate at which the X can be exchanged for Y in the market place. In the example above, the slope equals –1/2.

2.2 Changes in Income

The consumer's opportunity set depends on market prices and the consumer's income. As these parameters change, so will the consumer's opportunities. Let us now examine the effects on the opportunity set of changes in income by assuming prices remain constant.

Suppose the consumer's initial income in Figure 4-5 is M_0. What happens if M_0 increases to M_1 while prices remain unchanged? Recall that the slope of the budget line is given by $-P_X/P_Y$. Under the assumption that prices remain unchanged, the increase in income will not affect the slope of the budget line. However, the vertical and horizontal intercepts of the budget line both increase as the consumer's income increases, because more of each good can be purchased at the higher income. Thus, when income increases from M_0 to M_1, the budget line shifts to the right in a parallel fashion. This reflects an increase in the consumer's opportunity set, because more goods are affordable after the increase in income than before. Similarly, if income decreases to M_2 from M_0, the budget line shifts toward the origin and the slope of the budget line remains unchanged.

Figure 4-5

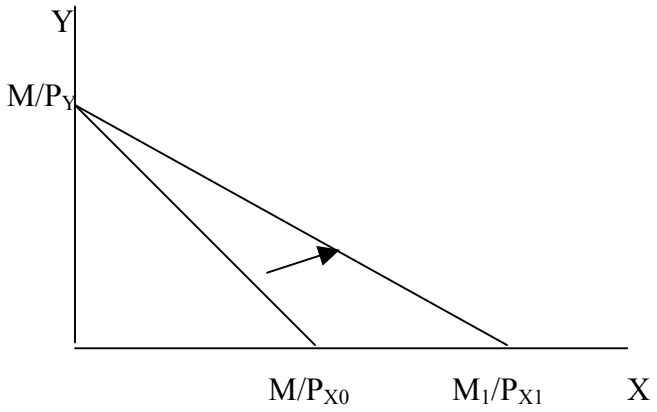


2.3 Changes in Price

Now suppose the consumer's income remains fixed at M, but the price of good X decreases to P_1 from P_0. Furthermore, suppose the price of good Y remains unchanged. Since the slope of the budget line is given by $-P_X/P_Y$, the reduction in the price of good X changes the slope, making it flatter than before. Since the maximum amount of good Y that can be purchased is M/P_Y, a reduction in the price of good X does not change the Y intercept of the budget line. But the maximum amount of good X that can be purchased at

the lower price (the X intercept of the budget line) is M/P_{X1}, which is greater than M_0/P_{X0}. Thus, the effect of a reduction in the price of good X is rotate the budget line counterclockwise, as in Figure 4-6.

Figure 4-6



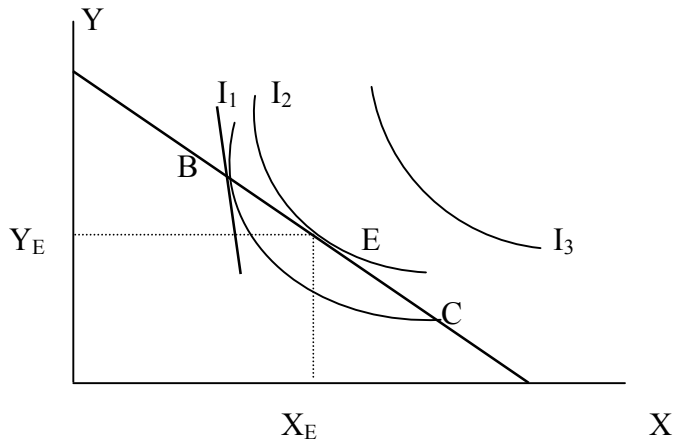
Similarly, an increase in the price of good X leads to a clockwise rotation of the budget line.

2.4 Consumer Equilibrium

The principal assumption upon which the theory of consumer behaviour rests is that consumers attempt to allocate their limited money incomes to purchase available goods and services so as to maximize their satisfaction (utility). In other words, consumers' indifference map provides a diagrammatic representation of a consumer's taste and intensity of desire for different product combinations while the consumer's purchasing power (and thus ability to satisfy material wants) is reflected by the line of attainable combinations. Putting the two together shows all the desired product combinations on the indifference curves that are also attainable.

As illustrated in Figure 4-7, the point of equilibrium can be located by drawing the budget line on the indifference map for goods X and Y. Since the number of indifference curves upon the indifference map is infinite, one curve will be tangent to the budget line regardless of where the budget line lies. The point of tangency is the point of equilibrium, representing the attainable combination of X and Y that gives the highest level of utility. As indicated in this Figure, the budget line is tangent to the indifference curve, I_2, at point E, where the consumer acquires Y_E units of Y and X_E of X. Since indifference curves may not intersect, it is clear that I_2 represents the highest level of utility that can be obtained with the budget available. An indifference curve representing a higher level, I_3, does not touch the budget line, and hence is unattainable, whereas an indifference curve representing a lower level, I_1, intersecting the budget line at two places is inferior to I_2. It should not be difficult to see why point E is preferable to any other point on the budget line. For example, point B also exhausts the income, but it clearly offers less utility than point E because it lies on a lower indifference curve.

Figure 4-7



Note, at the point of tangency between the indifference curve and the budget line, the slope of the indifference curve equals the slope of the budget line, that is,

$$MRS = \frac{P_X}{P_Y} \tag{5}$$

The condition for maximizing satisfaction requires that the consumer allocate his or her purchasing power so that the marginal rate of substitution of X for Y is equal to the ratio of the price of X to the price of Y. The interpretation of this condition is straightforward. The MRS defines the rate at which the consumer is willing to exchange X for Y. The price ratio (P_X/P_Y) shows the rate at which the consumer can exchange X for Y. Unless the two rates are equivalent, it is possible for the consumer to alter purchases of X and Y and achieve a greater degree of satisfaction.

Suppose that at the current purchase combination, MRS = 4, meaning that the consumer is willing to exchange 4 units of Y for 1 more unit of X. If P_X = \$6 and P_Y = \$2, then the consumer need only give up 3 units of Y at \$2 each to obtain the dollars needed to buy another unit of X. This situation is reflected by point B in Figure 4-7, where the slope of the indifference curve is steeper than the slope of the budget line. This means the consumer is willing to give up more of good Y to get an additional unit of good X than he or she actually has to give up, based on market. Confronted with these circumstances, it is in the consumer's interest to consume less of good Y and more of good X. This substitution continues until ultimately the consumer is at point E, in Figure 4-7, where the MRS is equal to the ratio of prices.

In general, then, consumer maximization of satisfaction requires equality between the marginal rate of substitution for any pair of products and the ratio of their prices, otherwise some exchange can be made which will increase consumers' overall satisfaction.

2.4.1 Demonstration Problem

Suppose an individual uses exactly two pats of butter on each piece of toast. If toast costs \$0.20/slice and butter costs \$0.10/pat, find this individual's best affordable bundle if she has \$12 per month to spend on toast and butter.

Answer:

The individual is willing to exchange one piece of toast for two pats of butter, i.e., MRS = 2. At the same time, the price ratio (P_{toast}/P_{butter}) = 2. Therefore, the individual is already maximizing her satisfaction since MRS = (P_{toast}/P_{butter}). Therefore, her \$12 budget should be allocated in such a way that for every piece of toast that she consumes she will consume two pats of butter, i.e., 60 pats of butter and 30 toasts.

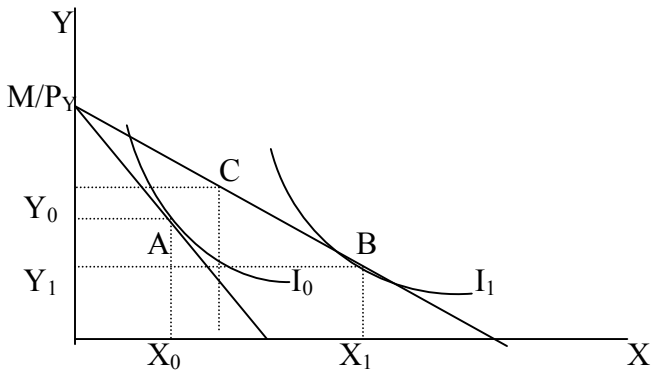
2.5 Price Changes and Consumer Behaviour

A change in the price of a good will lead to a change in the equilibrium consumption bundle. To see this, assume that the price of X is reduced. This, as explored earler, results in a counterclockwise rotation of the budget line. The consumer responds by moving from point A to point B on a higher level of utility.

Precisely where the new equilibrium point lies along the new budget line after a price change depends on whether the two goods are *substitutes* or *complements*. Recall from Block 3 that goods X and Y would be called substitutes if an increase (decrease) in the price of X leads to an increase (decrease) in the consumption of Y. On the other hand, they would be called complements if an increase (decrease) in the price of good X leads to a decrease (increase) in the consumption of good Y.

Figure 4-8 shows the case of two substitutes. A reduction in the price of X would lead the consumer to move from point A in to a point such as B, where less of Y is consumed than at point A.

Figure 4-8



If goods X and Y were complements, a reduction in the price of X would have led the consumer to move from point A in Figure 4-8 to a point such as C, where more of Y is consumed than before. Please note that this would require a new set of indiffence curves (showing a different consumer's preferences), which is not shown here.

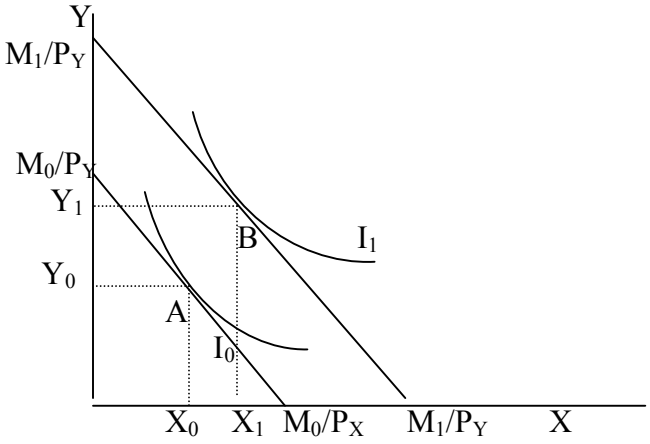
Price changes alter consumer incentives to buy different goods, thereby changing the mix of goods they purchase in equilibrium. The primary advantage of indifference curve analysis is that it allows a manager to see how price changes affect the mix of goods that consumers purchase in equilibrium. These changes may be the result of other firms initiating price changes, or due to a change in the frim's own strategy.

2.6 Income Changes

Indifference curve analysis also allows us to see how changes in income affect the mix of goods purchased by consumers. When the consumer's income changes, the consumer will respond by changing his or her optimal bundle of goods (X and Y). Normally, increases in income, all else remaining constant, will lead to increased demand for the products available to the consumer. We show such a situation in Figure 4-9, where the consumer's income has increased and that consumer subsequently purchases more of both products available. Note that at the initial prices and income level (M_0), and given the consumer's taste and preference pattern (indicated by the indifference curves), the consumer maximizes at point A.

Now suppose the consumer's income increases to M_1 so that his or her budget line shifts out in a *parallel* fashion. Clearly the consumer can now achieve a higher level of satisfaction than before. This particular consumer finds it in his or her interest to choose bundle B in Figure 4-9, where the indifference curve through point B is tangent to the new budget line.

Figure 4-9



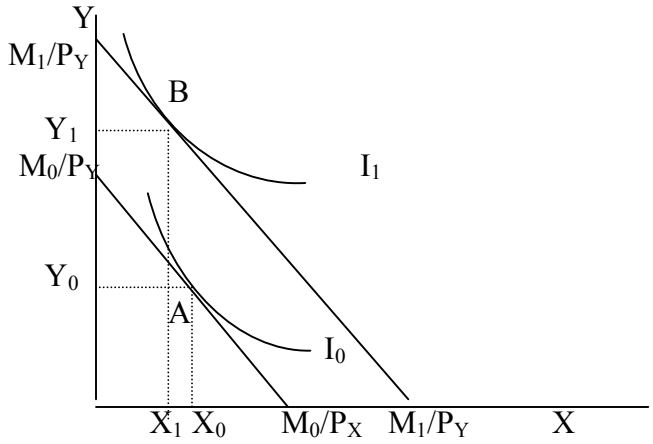
In this case, the quantity demanded of both products has increased as a result of the increase in the consumer's income. Since income and quantity demanded moved in the

same direction, we say that the income effect was positive for both goods. We call such products *normal* goods.

For some products the income effect is negative, meaning that if income increases, the quantity demanded of those products actually declines. This may happen to a product that is an *inferior* substitute for some superior product. As people become richer, they tend to switch away from inferior goods to superior substitutes. Conversely, as incomes fall (in a recession, for example), people tend to switch back to the inferior substitutes as they find themselves unable to afford the superior substitutes.

In Figure 4-10 we show the income effect for an inferior good. The product on the horizontal axis, travel by train, is an inferior good, and travel by air is a superior substitute for this particular consumer. Note that when the consumer's income increases and the budget line moves outward, the quantity demanded of travel by train actually decreases while travel by air increases. This consumer, now able to afford more air travel, substitutes away from the inferior good and in favour of the superior good. Looking at this from the opposite perspective, if the consumer's income were to fall from the higher level to the lower level, the consumer would reduce consumption of the superior good and increase consumption of the inferior good.

Figure 4-10



For any particular product at any particular point in time, some consumers may regard that product as a superior good while others may regard it as an inferior good. This difference in attitude toward a particular product stems from a difference in income levels and/or a different pattern of tastes and preferences.

2.7 The Income and Substitution Effects of a Price Change

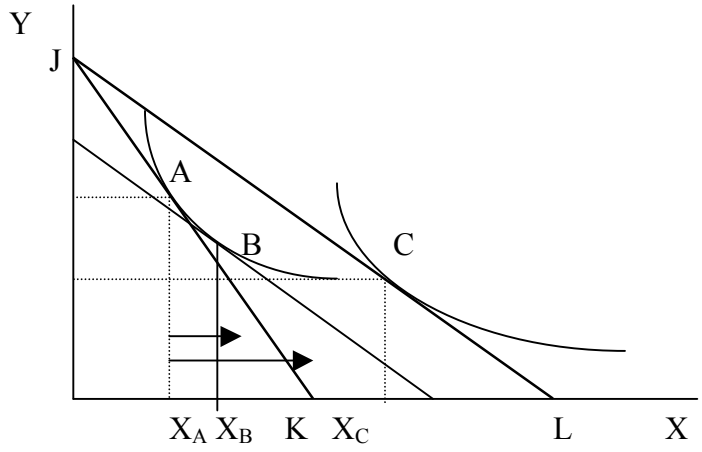
We can combine our analysis of price and income changes to gain a better understanding of the effect of a price change on consumer behaviour. The effect of the price change can be thought of as comprising two separate effects, known as the *income effect* and the *substitution effect.* When a product's price falls (and all other things remain unchanged),

the consumer experiences an increase in his or her real income. *Real income* is the purchasing power of money income. When the price of a particular good falls, the consumer can buy the same combination as before and have some money left over. This additional real income can then be spent, both on the product whose price has fallen and on other products. Normally, the income effect of a price reduction causes the consumer to buy a little more of the product in question, but if the product in question is an inferior substitute for some other product, the consumer may reduce purchases of the inferior product (whose price fell) in favour of the other product.

The second part of the price effect is the substitution effect. When the price of a product falls, consumers will tend to substitute in favour of that product because it is now cheaper relative to other substitute products that serve the same need. Thus, when the price of a product falls, *all else being kept constant,* the consumer would normally buy more of that product, first because his or her real income has increased, and second because he or she substitutes toward that product and away from other substitutes that are now relatively more expensive.

Suppose a consumer initially is in equilibrium at point A in Figure 4-11, along the budget line connecting points J and K. Suppose the price of good X decreases so that the budget line rotates counter-clockwise, and becomes the budget line connecting points J and L. The consumer now maximizes his or her satisfaction at point C. The movement from A to C is referred to as the *total effect*. The total effect, however, is composed of substitution and income effects. The substitution effect reflects a movement along an indifference curve, from A to B, thus isolating the effect of a relative price change on consumption: $X_B - X_A$. The income effect results from a parallel shift in the budget line; thus, it isolates the effect of increases of 'real income' on consumption and is represented by the movement from B to C: $X_C - X_B$. The total effect of a price decrease, which is what we observe in the marketplace, is the movement from A to C: $X_C - X_A$. Therefore, the total effect of a change in consumer behaviour results not only from the effect of a lower relative price of good X (the movement from A to B) but also from the increased real income of the consumer (the movement from B to C).

Figure 4-11



Learning Tip

An important consideration in running a firm is the impact of changes in income on the demand for the firm's product. Suppose you are the manager of a firm that sells a product that is a normal good and are considering expanding your product line to include another good. There are several things you may wish to consider in making your decision. Since your product is a normal good, you will sell more of it when the economy is booming (consumer incomes are high) than when times are tough (incomes are low).

2.7.1 Demonstration Problem

Suppose your product is a cyclical product, that is, sales vary directly with the economy. How can this information be useful to you when considering alternative products to include in your store?

Answer:

If you expand your offerings to include more normal goods, you will continue to have an operation that sells more during an economic boom than during a recession. But, if you include in your operation some inferior goods, the demand for these products will increase during bad economic times for normal goods. This is not to say that the optimal mix of products involves a 50-50 mix of normal and inferior goods; indeed, the optimal mix will depend on your own risk preference. Therefore, you may be better off diversifying, at least to a degree.

2.7.2 Demonstration Problem

Does running a gourmet food store involve a higher level of risk of going out of business than running a supermarket, or vice versa, when the economy heads for a recession?

Answer:

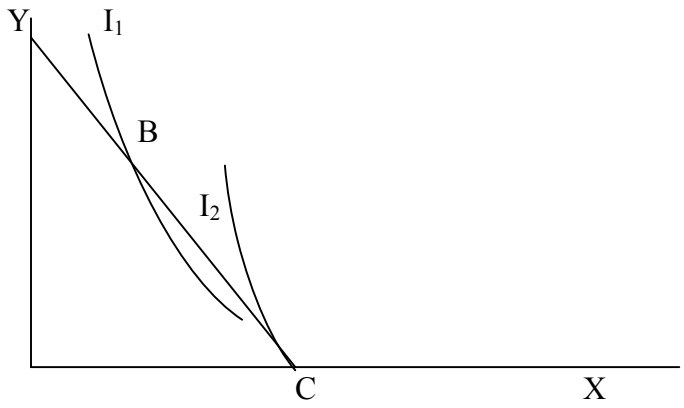
Note that gourmet shops sell almost exclusively normal goods, while supermarkets have a more 'balanced portfolio' of normal and inferior goods. This explains why, during recessions, many gourmet shops go out of business while supermarkets do not. This is the case because the outlet that sells low, or even perhaps negative, income elastic products tends to survive and in fact do well in a recession.

2.8 Corner Solutions

In all the examples considered so far, the optimal consumer basket has been interior, meaning that the consumer consumes positive amounts of both goods. In reality, though, a given consumer will not purchase positive amounts of all available goods. For example, not all consumers own an automobile or a house. Some consumers may not spend money on tobacco or alcohol. If the consumer cannot find any interior basket at which the budget line will be tangent to an indifference curve, then he or she may find an optimal basket at a corner point, that is, at a basket along an axis, where one of the goods is not purchased at all. If an optimum occurs at a corner point, the budget line may not be tangent to an indifference curve at the optimal basket.

Let's consider again our consumer who chooses between just two goods, food and clothing. If the consumer's indifference map is as illustrated in Figure 4-12, no indifference curve is tangent to his or her budget line. At any interior basket on the budget line, such as basket B, the slope of the indifference curve is steeper (more negative) than the slope of the budget line. This means that MRS > P_X/P_Y. In this case, our consumer is willing to exchange more Y for an additional unit of X than he or she has to, based on the market. This is true not only at basket B, but at all baskets on the budget line. The consumer would continue to substitute X for Y, moving along the budget line until he or she reaches the corner point basket C. At basket C the slope of the indifference curve I_2 is still steeper than the slope of the budget line, he or she would like to substitute more X for Y if that were possible. But no further substitution is possible because no more X can be purchased beyond basket C. Therefore the optimal choice for this consumer is basket C, because that basket gives the consumer the highest utility possible on the budget line.

Figure 4-12

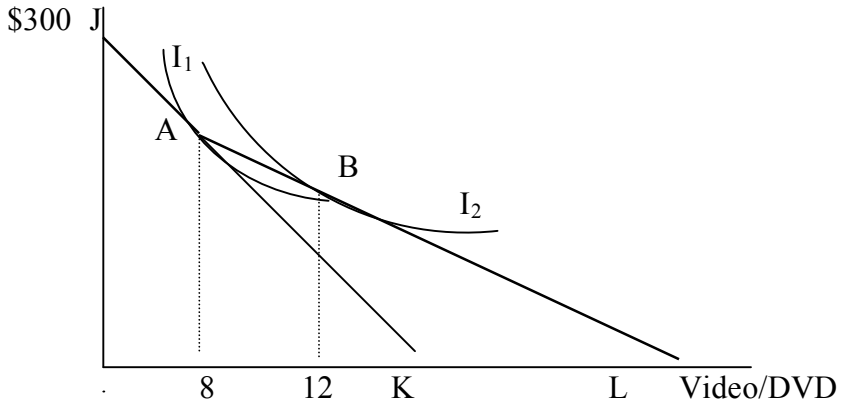


3 Application of Indifference Curves

3.1 Quantity Discounts

In many instances sellers offer consumers quantity discounts. We can use indifference curve analysis to understand how such discounts work. Here is an example. Suppose a video/DVD store sells its products according to the following formula. For the purchase of the first 8 videos/DVDs the seller charges a price of $15 for each video/DVD. However, a hefty discount is offered on purchase of *each additional* unit, $10 per video/DVD. Figure 4-13 shows how the quantity discounts affect the consumer's behaviour assuming that the buyer starts with a fixed budget of $300 to be spent on all goods including video/DVD.

Figure 4-13



As shown in Figure 4-13, in the absence of a quantity discount, each unit sells for $15 and the budget line facing the consumer is represented by JK. The slope of this line is (–15). Point J, the vertical intercept is a point where the buyer buys no video/DVD, and keeps his or her money for other products. The vertical axis hence can be regarded as a composite good (all other goods). Point K, on the other hand, is a point where all $300 is spent on video/DVD. The consumer, however, would choose point A, where 8 video/DVDs are purchased. Suppose the discount kicks in. The budget line is now will be represented by the line JAL, a broken line. The upper segment of the budget line (JA) has the slope of (15) whereas, the bottom segment (AL), has the slope of (–10), and the consumer will buy the total of 12 videos/DVDs, point B. The quantity discount scheme has induced the consumer to purchase additional 4 units.

3.2 Voucher versus Cash Subsidies

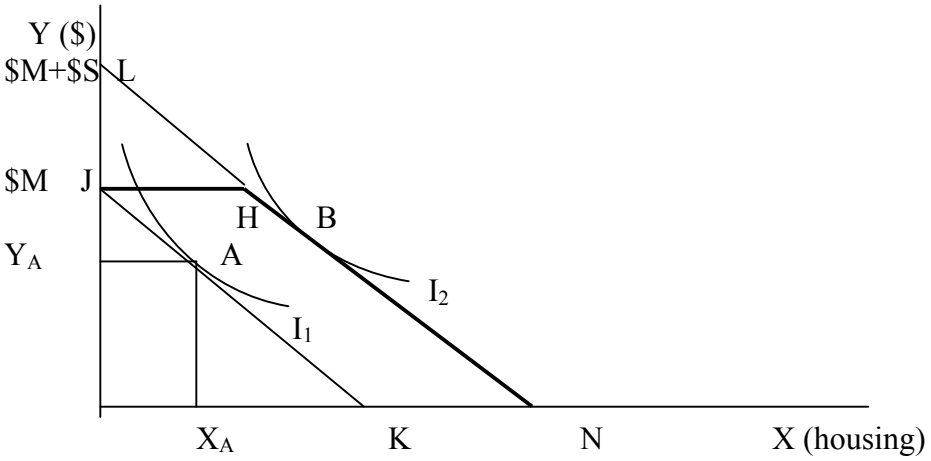
Governments in most countries often employ a wide range of publicly subsidized 'assisted housing' programs. One of the main objectives of all of these programs is to provide access to a primary human requirement for those most in need. For instance in Canada, the current annual cost of housing subsidies in Canada is estimated at over $5

billion. Increasingly, in the allocation of the limited stock of subsidized housing units, the focus is on those high-need, low-income families who have to spend over 30 per cent of their family income to rent adequate accommodation.

Suppose, the consumer (household's) preferences for housing and other goods are represented by the indifference curves as in Figure 4-14. Let the Y-axis denote the composite good (measured in dollars per month) and X represents the housing units. Assume that this household has a monthly budget of \$M and must pay P_X for each unit it rents. The budget line is represented by JK. Given the budget line and the indifference curves, equilibrium is found at point A with X_A units of housing. The consumer spends ($\$M - Y_A$) on housing and the rest, Y_A, on other goods.

Now suppose the government steps in to increase the housing consumption. There are two types of programs that might be implemented to increase the consumer's purchases of housing: (a) the subsidy as an additional amount of housing the family receives, *coupon (housing voucher)* or (b) the subsidy as an additional income, *cash subsidy*.

Figure 4-14

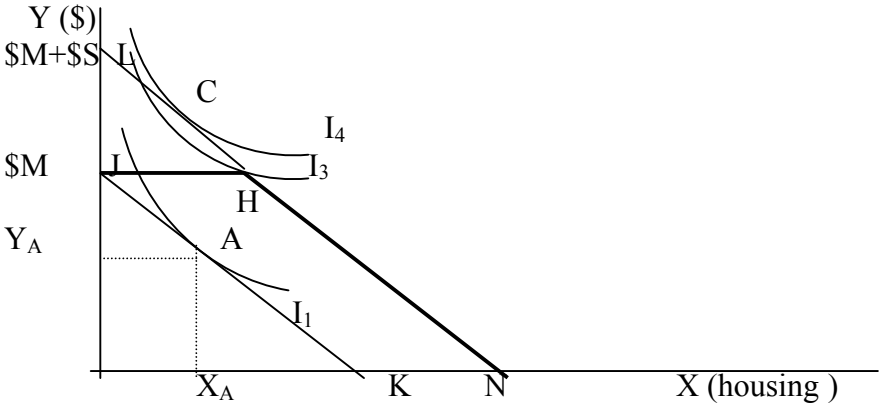


If the consumer receives a cash subsidy of \$S from the government, the budget line shifts from JK to LN. If the government gives the consumer a voucher also worth \$S that can only be spent on housing, the budget line will be represented by JHN. Because the consumer cannot apply the voucher to purchase other goods, the maximum he or she can spend on other goods is still \$M, whereas in the case of a cash (income) subsidy, the consumer is free to allocate whatever amount he or she wishes to allocate to housing and other goods. Assuming that the consumer's preferences are represented by the set of indifference curves I_1 and I_2, he or she will be indifferent between a cash subsidy and a voucher, and the *choice of the program* does not matter. The consumer's choice will be point B in either case.

However, if the consumer's preferences are represented by a different set of indifference curves, Figure 4-15, the choice of the program does matter. Under the policy of cash subsidy, the consumer will select the bundle at point C, on I_4, whereas under the voucher,

the consumer will choose the bundle at point H, on a lower indifference curve, I_3. Clearly, the consumer is worse off under the voucher scheme than the cash subsidy.

Figure 4-15



If such strong conclusions can be drawn from such a simple model, then why do governments spend billions of dollars to provide subsidized housing, instead of just making cash grants? One political reason is that earmarking the subsidy for housing guarantees that resources used will put roofs over people's heads rather than being spent in frivolous or illicit ways. An economic reason the likelihood of the consumer finding housing is less under cash subsidy than voucher. This is the case because the landlord may perceive the risk of non-payment to be greater when the entire rental payment is made by the tenants themselves than when a portion of it is guaranteed by the government.

4 Relationship Between Indifference Curve Analysis and Demand Curves

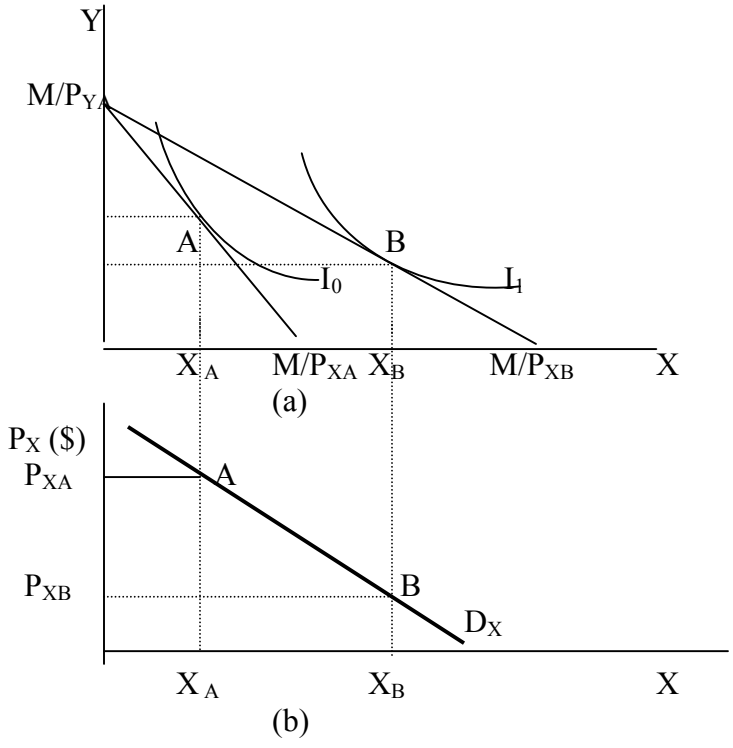
The indifference curve approach developed in this Block is, in fact, the basis for the demand functions we studied in Block 3. We have seen how the consumption patterns of an individual consumer depend on variables that include the prices of substitute goods, the prices of complementary goods, tastes (i.e., the shape of indifference curves), and income. We conclude by examining the link between indifference curve analysis and demand curves.

4.1 Individual Demand

Recall from Block 3 that a market demand curve is a relationship that tells how much of a good the market as a whole wants to purchase at various prices. Suppose we want to generate a demand schedule for a good, X, not for the market as a whole but for only a single consumer. Holding income, preferences, and the prices of all other goods constant, how will a change in the price of shelter affect the amount X the consumer buys? To answer this question, we begin with this consumer's indifference map, with X on the horizontal axis and the composite good Y on the vertical axis, Figure 4-16(a). The consumer initially is in equilibrium at point A, where income is fixed at M, and prices are P_{XA} and P_{YA}. But when the price of good X falls to the lower level, indicated by P_{XB} ($P_{XB} < P_{XA}$), the consumer reaches a new equilibrium at point B. The important thing to notice is that the only change that caused the consumer to move from A to B was a change in the price of good X; income and the price of good Y are held constant in the diagram. When the price of good X is P_{XA} the consumer consumes X_A units of good X; when the price falls to P_{XB}, the consumption of X increases to X_B.

The panel (b) in Figure 4-16 shows this relationship between the price of good X and the quantity consumed of good X. This consumer's demand curve for good X indicates that, holding other things constant, when the price of good X is P_{XA} the consumer will purchase X_A units of X; when the price of good X is P_{XB}, the consumer will purchase X_B units of X.

Figure 4-16



The line connecting point A to point B in panel (b) represents the demand curve for our consumer. An individual consumer's demand curve is like the market demand curve in that it tells the quantities the consumer will buy at various prices.

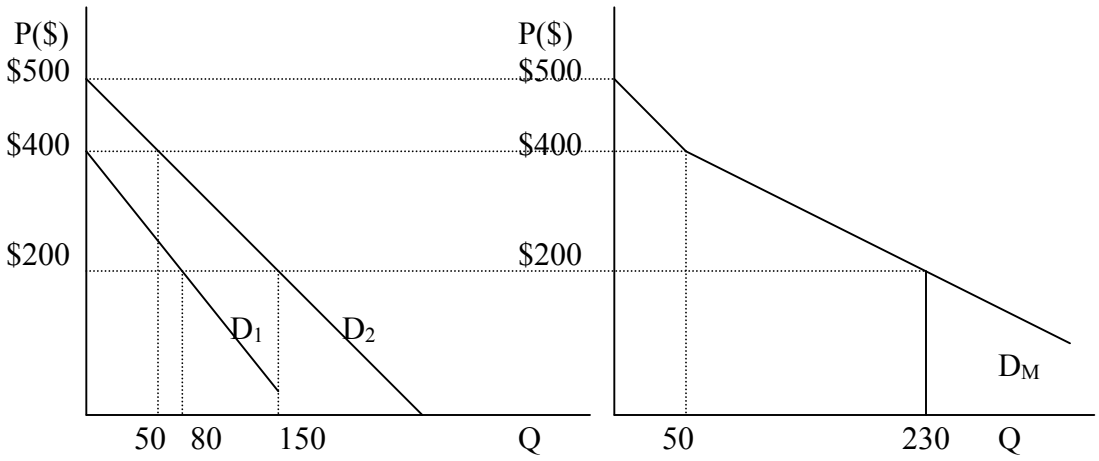
Learning Tip

Note that the $X_B - X_{A,}$ in Figure 4-16(b) consists of both the substitution as well as the income effect of the change in the price. Therefore, the typical demand curve obtained this way shows the total effect of a price change not the substitution effect per se.

4.2 Market Demand

The market demand curve is the horizontal summation of individual demand curves and indicates the total quantity all consumers in the market would purchase at each possible price. This concept is illustrated graphically in Figures 4-17(a) and 4-17(b). The curves D_1 and D_2 represent the individual demand curves of two hypothetical consumers: Individual 1 and Individual 2. When the price is \$500, Individual 1 and Individual 2 buy 0 units. Thus, at the market level, 0 units are sold when the price is \$500, and this is one point on the market demand curve (labeled D_M in Figure 4-17(b)). When the price is \$400, Individual 1 buys 50 units (point A) and Individual 2 buys 0 units (point B). Thus, at the market level 50 units are sold when the price is \$400, and this is another point (point A + B) on the market demand curve. When the price of good is \$200, Individual 1 buys 150 units and Individual 2 buys 80 units; thus, at the market level, 230 units are sold when the price is \$200.

Figure 4-17



(a) Individual demand (b) Market demand

5 The Attribute Approach to Consumer Choice

Our discussion of consumer behaviour so far has failed to explain why some customers opt to purchase brand A over brand B for the same general good or service. To analyze consumer choice criteria for competing brands, we can study the modern consumer theory known as *Attribute analysis*. The Attribute model of consumer behaviour holds that consumer's derive utility not from the products themselves but from the characteristics or attributes provided by the products. It is a product's characteristics, performance features, or attributes that create utility; thus, what causes a buyer to prefer one brand over another has to do with the different attributes of rival brands. For instance, what business travelers care about in their notebook computers are low weight, long battery life, and high computing power, rather than the logo on the case. There are trade-offs among these good things; what differentiates one notebook from another is where in this 'attribute space' or 'characteristics space' they are located.

Therefore, the major contribution of the Attribute model is its ability to explain that a consumer's preference for brand A over brand B is rooted in the fact that the consumer attaches more utility or satisfaction to some attributes than to others. Thus, the Attribute Approach greatly facilitates the explanation of consumer choice within groups of substitutes.

5.1 Depicting Products in Attribute Space

For the purposes of our graphical analysis here, consider the situation of buying a notebook computer assuming that a prospective buyer is familiar enough with its two most important attributes (other than price), i.e., 'low weight' and 'high computing power.' To demonstrate how a consumer might choose among products to maximize utility derived from the two attributes, suppose there are three notebook brands. After a careful evaluation, the buyer ranks each notebook on a scale of 1 to 10, with 1 being the lowest and 10 the highest, and comes up with the following information, Table 4-1 below. We shall assume that each product supplies both attributes in a particular ratio.

Table 4-1: Attributes and Prices of Three Brands of Notebook Computers

Notebook Brand	Notebook Price	Attribute Rating: Low Weight	Attribute Rating: High Computing Power	Ratio of Low Weight to High Computing Power
A	$2,500	9	4	2.25
B	2,800	7	6	1.16
C	3,200	5	9	0.55

In Figure 4-18, below, each notebook brand is shown in the attribute space as a ray drawn from the origin. Graphically, the slope of each ray is determined by the ratio of low weight to computing power, as listed in the last column in the Table. If the buyer buys notebook A, he or she will travel out along the steepest ray, absorbing the two attributes in the ratio 2.25. The other notebook computers are indicated by the lower rays because these offer the two attributes at lower ratios. If the traveller buys notebook B, he or she will travel along the second steepest ray with the slope of 1.16. Finally, for brand C, the ratio is 0.55.

What determines how far along each ray the buyer could go and how much of each product the buyer can purchase depends on the buyer's budget (budget constraint). In case of an indivisible purchase, such as purchase of an automobile or a notebook computer – the products that are, on the one hand, available only in discrete units and, on the other hand, the price of which are large in relation to the consumer's income – the consumer is expected to stop at purchasing just one unit of the product: one automobile and one notebook computer. This, of course, implicitly assumes the buyer's budget is large enough to permit the purchase of all three brands despite the price differences.

Following this logic, the *efficiency frontier*, which is the outer boundary of the attainable combinations of attributes, is obtained. It is called 'efficient' because only combinations on this frontier allow the consumer to maximize utility. This frontier consists of the points joining the limit points – points A, B, and C – on each product ray. Each point is found by first dividing the buyer's income by the price of the respective product in order to determine the number of the units of the product, and then multiplying the outcome by the attribute content of each unit. The point depicted along each ray shows the maximum intake of the two attributes that can be obtained by consuming each notebook.

Learning Tip

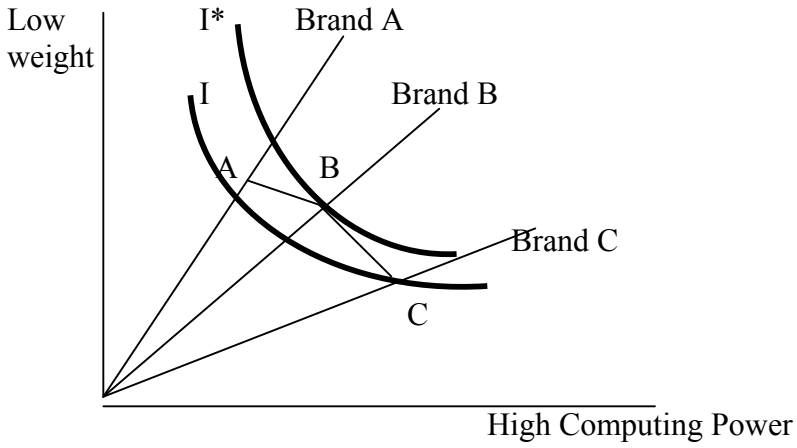
If our example involved a product like going to restaurants or going to the movies, more than '1 unit' might well be purchased over a short period of time. In this case, the distance on each that the buyer can travel is found by (1) dividing the consumer's allotted budget for the item by the price of each brand to arrive at how many units can be purchased and (2) calculating the total amounts of each attribute which can thus be obtained.

5.2 Utility Maximization from Attributes

Just as consumers can express preference or indifference between combinations of products, they can express preference of indifference between combinations of attributes. At any particular combination of 'light weight' and 'high computing power,' our buyer will be able to express a marginal rate of substitution between the two attributes: at any point, an extra unit of 'low weight' will be worth giving up some amount of 'high computing power.' These evaluations produce a set of indifference curves expressing the buyer's tastes and preferences at each possible attribute combination. Since our buyer cannot mix brands but rather needs to choose one notebook brand or another, he or she must be content with the constrained point: A, B, or C, Figure 4-18. Therefore, the efficient frontier is represented by the kinked line ABC. The buyer maximizes utility by choosing the notebook brand with the attribute combination on the highest indifference curve attainable, I*.

Of these options, point B allows the buyer to attain the highest indifference curve, I*.

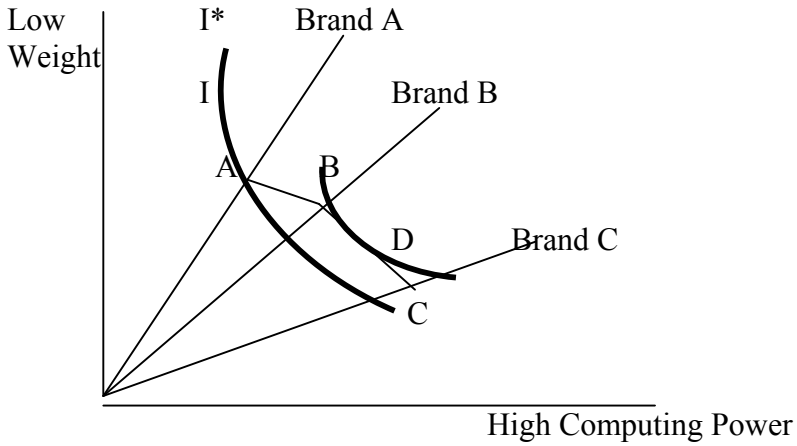
Figure 4-18



Learning Tip

If these products were mixable, the consumer would maximize utility by mixing brands to attain the highest indifference curve that is tangent to a flat section of the efficiency frontier, BC, to consume a combination of two goods (two restaurants or two movies), point D, instead of touching one of the corners (points A, B or C), in Figure 4-19.

Figure 4-19



6 Summary and Review

The utility approach to consumer behaviour is grounded upon psychological principles, and is best explained in terms of utility. Conceptually, utility can be measured in cardinal units, as in Block 2, or ordinal approach, as discussed in this Block, in which case the consumer is able to rank product in order of preference. Consumer's equilibrium is reached when the marginal utility per dollar is the same for all products, or when the slope of the indifference curve and the slope of the budget line are the same. The consumer's demand curve for a particular product is directly related to the marginal utility of that product. The downward sloping demand curve is due to the law of diminishing marginal utility. In general, the individual's demand curve consists of two effects, the substitution effect and the income effect. The income effect may be positive or negative depending on the taste and preferences of the consumer. Changes in consumer's income will change the quantity demanded at a given price level. The quantity demanded of normal goods will move in the same direction as income, whereas the quantity demanded of inferior goods will move in the opposite direction to income.

The Attribute Approach to consumer behaviour gives several valuable insights into consumer choice, which are not readily apparent using the product approach. This approach allows the entire range of substitutes available to the consumer to be depicted on the same graph. It is, therefore, able to explain why a consumer buys one brand of a product over another one. Furthermore, this approach can explain why a consumer will purchase combinations of substitute products.

7 Review Problems

1. A consumer has an income of $50. He can choose any combination of product X and product Y for consumption. If price of X (P_x) is $5 and price of Y ($P_y$) is $2.50. Draw the budget line of the consumer. Find the market rate of substitution. What is the maximum quantity of Y that he can purchase? If the income goes up to $75, then how will the budget line change? Find the maximum quantity of Y that he can purchase with the new budget.

2. A consumer can choose any combination of apples and oranges. A, B, and C are three different points on his indifference curve. If he moves from A to B he gives up 2 oranges for 1 apple. If he moves from B to C then he gives up 1 orange for 1 apple. Does the consumer's preferences comply with the diminishing marginal rate of substitution? Explain.

3. Consider the problem in question 2. If the price of apple is $0.50 per unit and the price of orange is $0.25 per unit, then with an income of $10 at which point (among points A, B, and C) will he maximize his welfare? Explain. If welfare is not maximized at any of these points, then explain how he should change his consumption.

4. Prices of food and clothing are currently $10 per unit and $25 per unit respectively. With an income of $500, find the market rate of substitution. Suppose that the consumer's preference shows that marginal rate of substitution at some point is 2 units food for 1 unit of clothing. Given that the consumer's indifference curve shows diminishing marginal rate of substitution, how should the consumer adjust his consumption to maximize utility given the budget constraint?

5. Suppose that the price of a hamburger is $1 and the price of a pizza is $2. With an income of $10, a consumer chooses 4 hamburgers and 3 pizzas. Graph the consumer's budget line and indicate his point of consumption. With his income rising to $15, the consumer chooses 6 pizzas and 3 hamburgers. Show this change in the graph. Is hamburger a normal good? Explain. What type of good is pizza? Explain.

6. A consumer faces the following prices of tea and coffee. The price of coffee is $1.50 per cup and the price of tea is $1 per cup. If she has an income of $15, then draw the opportunity set of consumption for the consumer. If price of coffee goes down to $1 per cup, how will you show the change in the graph? If tea and coffee were substitutes the how would you expect the ratio of tea and coffee consumption to change? Explain through a graph. Also explain the probable change if tea and coffee are complements.

7. ABC Telecom is offering a cell-phone package with the price of $1 per call, whereas the price per call with the landline is $0.50. With an income of $20, a consumer chooses to make 5 calls through her cell-phone from ABC and 30 calls through landline. Draw the budget line of the consumer. If ABC offers a promotion

that reduces the price to \$0.50 on all calls after the 10^{th} call, then how will the budget line change? How would the budget line have changed if instead of the promotion ABC offered \$0.50 per call on any amount of calls? Explain through a graph.

8. Suppose a consumer has a utility function of $U = X^{0.5} Y^{0.5}$, where X and Y are the products that she consumes. If price of X is \$5 and price of Y is \$10, find the demand functions for X and Y if her income is \$100.

9. Consider the situation described in question 8. What combination of X and Y will the consumer choose, if price of X goes up to \$10 per unit. Find the new combination that will be chosen by the consumer. Are X and Y substitutes or complements? Explain.

10. Jones has a utility function $U = S^2 C^3$. Where S stands for shoes and C stands for clothing. He has \$600 for spending on 'S' and 'C.' The nearest shop has the following price tags: shoes \$50, clothing \$30. Find the welfare maximizing quantities of 'S' and 'C' that he will choose. It takes \$50 for transportation to go to the discount shop where the price of shoes is \$40 and clothing \$20. What are the welfare maximizing quantities if Jones buys from the discount shop? Jones will prefer to buy from which shop? Explain.

11. A consumer has a utility function of $U = 10I^{1/2}$ where I is her income. She has an initial wealth of \$10,000 and she faces an uncertain situation in her business where she can earn \$5,000 or lose \$5,000 with equal probabilities. How much premium will she be willing to pay to avoid uncertainty and what is the amount that she will settle for with certainty?

12. Provide examples and draw graphs of the indifference curves showing:

    a. Perfect Substitutes

    b. Perfect Complements

13. Distinguish between a change in consumer perception and a change in consumer tastes.

8 Answer Key to Review Problems

1. Income = 50

   P_x =5, P_y = 2.50

   Market rate of substitution = –5/2.5 = –2. Maximum Y she can purchase = 20. The budget line shifts parallel to the right. Maximum Y she can purchase after the shift = 30

2. From A to B, MRS = 2/1 = 2

   From B to C, MRS = 1/1 = 1

   Therefore, the consumer's preferences comply with the diminishing MRS.

3. Market rate of substitution = 0.5/0.25 = 2, MRS at A = 2

   At A the consumer maximizes welfare.

4. Market rate of substitution = 25/10 = 2.5

   MRS = 2

   The consumer can improve welfare by increasing consumption of food and reducing consumption of clothing. Given that her preferences show diminishing MRS, with this change she can reach a point where her preference will be to give up 2.5 units of food for 1 unit of clothing, at which point MRS will be equal to the price ratio, maximizing her utility.

5. Hamburger is an inferior good. With income increasing from 10 to 15, consumption of burger reduced from 4 to 3. Pizza is a normal good. With income increasing from 10 to 15, consumption of pizza increased from 3 to 6.

6. With coffee price reducing, the budget line rotates pivoting at 15 cups of tea. If tea and coffee are substitutes, then with lower price of coffee, consumption of coffee should increase both from income and substitution effects. If they are complements, then consumption of both tea and coffee are likely to increase. In the former case, consumption of tea is likely to decrease. But if the income effect may more than offsets the substitution effect for tea, it will also lead to increased consumption of tea.

7. With $1.00 per cell phone call, the consumer chooses point A with 30 land line calls and 5 cell phone calls. With the promotion, the budget line is horizontal at 10 cell calls and 20 land calls up to the point 20 cell calls and 20 land calls. If the price is reduced for all calls, then the budget line rotates pivoting at 40 land calls. The benefit of such promotion is that some consumers may find it more satisfactory to make more calls through the cell-phone, and at the same time ABC Telecom

ensures that a high number of calls are made through cell-phones. If price reduced for all calls then the consumers may find it even more satisfying to reduce the number of cell-phone calls.

8. Marginal utility of X, $MU_x = 0.5X^{-0.5}Y^{0.5}$

Marginal utility of Y, $MU_y = 0.5X^{0.5}Y\text{-}0.5$

$MU_x/P_x = MU_y/P_y$

$(0.50 \times 5X^{-0.5}Y^{0.5})/5 = (0.5X^{0.5}Y–0.5)/10$

$Y/5 = X/10$, and $5X = 10Y$

The budget line, $5X + 10Y = 100$, $5X + 5X = 100$, and therefore, $X = 10$ and $Y = 5$

9. If price of X goes up to 10, then the revised budget line will be $10X + 10Y = 100$ or $X + Y = 10$. Also, $MU_x/P_x = MU_y/P_y$,

$(0.50.5X^{-0.5}Y^{0.5}) = (0.5X^{0.5}Y–0.5)$, since $X = Y$, $2X = 10$, and $X = 5$ and $Y = 5$.

10. Marginal utility of $S = 2SC^2$

Marginal utility of $C = 2S^2C$

$MU_s/P_s = MU_c/P_c$

$2SC^2/50 = 2S^2C/30$ …… $3C=5S$

The budget line, $50S + 30C = 600$, $C= 10$ and $S = 6$

If Jones buys from the discount store then the welfare maximizing condition is,

$2SC^2/40 = 2S^2C/20$, $C = 2S$.

The budget line, $40S + 20C = 600 – 50 = 550$. Substituting $C = 2S$, we get $C = 13.75$ and $S = 6.875$. Jones will prefer to buy from discount store, since more shoes and clothing will give him higher utility.

11. Wealth in good situation = 15,000

Wealth in bad situation = 5,000

$U(15000) = 10.(15000)^{0.5}=1000(1.5)^{0.5}$

$U(5000)=10(5000)^{0.5}=100(50)^{0.5}$

Expected utility $= 0.5[1000(1.5)^{0.5}+100(50)^{0.5}] = 965.93$

12. The indifference curves in the case of perfect substitutes are straight lines, whereas for the case perfect complements are right-angled (L-shaped).

13. Changes in perceptions affect the attribute ranking and hence the point along the ray form the origin in the attribute space, whereas, changes in the consumer's tastes affect the position of the indifference curves.

8 References and Suggested Further Reading

Baumol, W. J., *Business Behavior, Value and Growth***,** Rev. Ed. New York: Harcourt Brace Johanovich, 1967.

Douglas, J. E., *Managerial Economics: Analysis and Strategy*, Prentice Hall, 1992.

Lancaster, K, *Consumer Demand: A New Approach*, New York, Columbia University, 1971.

Pindyck, R. and Rubinfeld, D., *Microeconomics*, Fifth Edition, Prentice Hall, 2001.

D. Besanko, and R. Braeutigam, *Microeconomics: An Integrated Approach*, Wiley, 2002.

Commonwealth of Learning Executive MBA/MPA

E5 Managerial Economics

Block 5

Production Analysis

Course author: Farrokh Zandi, PhD
Professor, Schulich School of Business
York University
Toronto, Canada

Course designer: Douglas R. Cronk, M.A., ABD
Vancouver, BC, Canada

Copyright © Commonwealth of Learning, 2003

All rights reserved. No part of this course may be reproduced in any form by any means without prior permission in writing from:

The Commonwealth of Learning
1285 West Broadway
Suite 600
Vancouver, BC V6H 3X8
CANADA
e-mail: info@col.org

Contents

1 Block Five: Objectives and Introduction 1
1.1 Introduction 1
2 Production Analysis 2
2.1 The Production Function 2
2.2 The Short-Run Versus Long-Run 2
2.3 Production in the Short Run 3
3 Measures of Productivity 3
3.0.1 Demonstration Problem 4
3.1 The Law of Diminishing Returns 5
3.2 Total Product, Average Product, and Marginal Product 6
3.3 The Optimal Level of Employment of a Variable Input 10
3.3.1 Demonstration Problem 11
3.4 The Production Function of Multiple Variable Inputs (Isoquants and Isocosts) 11
3.5 Shapes of Isoquants 14
3.6 Cobb-Douglas Production Function 16
3.7 Isocosts 17
3.7.1 Demonstration Problem 18
3.8 Cost Minimization (Economic Efficiency) 18
3.8.1 Demonstration Problem 20
3.8.2 Demonstration Problem (Mathematically Challenging) 21
3.9 Input Substitution 21
3.10 Returns to Scale 22
3.11 Testing Production Function for Returns to Scale 25
4 Technological Changes 25
5 Summary and Review 28
6 Review Problems 28
7 Answer Key to Review Problems 30
8 References and Suggested Further Reading 31

1 Block Five: Objectives and Introduction

After working through Block Five of this course, you should be able to:

1. Outline the major decisions that firms must take.
2. Distinguish the short run from the long run.
3. Describe the relationship between average product, marginal product and total product.
4. Apply the Isoquant Approach to production.
5. Distinguish (diminishing) returns to factor from (diminishing) returns to scale.
6. Describe the rule of costs minimzation.

1.1 Introduction

This Block moves from the decisions of utility-maximizing households to examine the factors governing the behaviour of perfectly competitive profit-maximizing producers. It is a transition block that introduces many central topics, among which are productivity, costs, and economic profits. The most profitable way to employ the firm's resources to produce a given product requires an understanding of production function. Production function is simply an input-output relationship between one or more factors of production (input) and the good or service produced (output). This relationship is analyzed and quantified during a production study to determine the most economical combination of input resources to obtain a given level of output. A study of production functions is also fundamental to cost analysis. Once production function has been identified, its cost function can be derived from the production function. Hence the production affects the firms' status in its industry.

2 Production Analysis

Once demand for a given product or service has been determined, management decides the most profitable way to employ the firm's resources to produce that good or service. Such decisions involve an understanding of production functions.

A production function is simply an input-output relationship between one or more factors of production (input) and the good or service produced (output). This relationship is analyzed and quantified during a production study to determine the most economical combination of input resources to obtain a given level of output. Or conversely, it may involve the determination of the maximum output obtainable from a given level and mix of inputs. A production study may pertain not only to the production of goods, such as planes, computer games, scanners, ice cream, but also to the production of services, such as internet services, hospitality industry, or health care.

2.1 The Production Function

Production function is an equation, table, or graph showing the amount of output of a product that a firm can produce per period of time for each set of inputs. Both inputs and outputs are measured in physical rather than in monetary units. Technology is assumed to remain constant during the period of the analysis. Technology summarizes the feasible means (know-how) of converting raw inputs, such as steel, labour, and machinery, into an output such as an automobile.

A study of production functions is also fundamental to cost analysis. Once a firm's production function has been identified, its cost function can be derived from the production function, provided the market prices of the input factors are known. Hence, the production function strongly affects the firm's status in its industry, and the study of production functions is even more basic than the study of cost functions.

Like a demand function, a production function can be expressed as a schedule, graph, or an equation such as

$$Q = f(K, L) \tag{1}$$

where Q, a specific output, is a function of the input factors, K and L. Note that the general ideas presented here are valid for any two inputs, not just for labour and capital. It is important to remember that a production function relates to some given level of technology. If the technology changes through the upgrading of labour, materials, machinery, equipment, processes, or management, the production function changes accordingly.

2.2 The Short-Run Versus Long-Run

In production and cost theory, the distinction is made between the *short run,* in which the quantities of some inputs are variable while others are in fixed supply, and the *long run* in which all factors may be varied. Consequently, it is useful to classify the inputs on the

basis of whether or not they are variable in the short run. Labour is typically regarded as variable and capital as fixed in the short run.

It is important to understand that *the long run* does not refer to a long period of time. It is a peculiarity of the economists' jargon that the term has no direct connection with time at all, and that the firm is likely to be in a long-run situation for relatively short periods of time. When intending to change its scale of production, the firm must continue to operate in a short-run situation until its most-fixed factor becomes variable.

2.3 Production in the Short Run

As a manager, your job is to use the available production function efficiently. This effectively means that you must determine how much of each input to use to produce output. In the short run, some factors of production are fixed, and this limits your choices in making input decisions. For example, it takes several years for Honda Motor Company to build an assembly line. The level of capital is generally fixed in the short run. However, in the short run Honda can adjust its use of inputs such as labour and steel; such inputs are called variable factors of production.

Therefore, the short run by definition is that period of time during which at least one input is fixed. Variable inputs are those inputs whose quantities are directly related to the level of production.

Mathematically, this function is captured by the following relationship:

$$Q = f(\overline{K}, L) \qquad (2)$$

where, $\overline{K}$, refers to the fixed level of input of capital. The horizontal bar in this equation indicates that the input factors to its left are regarded as fixed in the production process under analysis, while the factor to its right is variable. The quantity of the output product, Q, is the result of combining a variable quantity of input factor, e.g., skilled labour, with fixed quantities of other input, e.g., buildings or equipment.

3 Measures of Productivity

The productivity of a factor of production refers to the amount of output that can be produced by that input, holding constant the input of all other factors of production. Obviously, an input, such as human resources, can be more productive if it works with modern mechanical and computer-assisted equipment and high-quality raw materials. Similarly, the plant or equipment can be more productive if it is being operated by highly skilled and well-trained workers. We use the phrase *the state of technology* when we refer to the quality of the resources involved in the production. Table 5-1 is a numerical representative of this function.

Table 5-1 Alternative Short-Run Production Functions

Row	Capital Inputs (machine month)	Labour Inputs (person month) (L)								
1										
2	(K)	0	1	2	3	4	5	6	7	8
3	1	0	15	35	77	112	130	145	155	162
4	2	0	22	48	80	125	161	188	208	223
5	3	0	35	72	120	180	255	310	358	387
6	4	0	500	102	162	230	305	385	444	493

Note that this table in fact represents four short-run production functions and not just one. Column 2 shows the amount of capital used per month. Accordingly, there are four possible *fixed* sizes of capital, varying from 1 to 4, $\overline{K} = 1$, $\overline{K} = 2$, $\overline{K} = 3$ and $\overline{K} = 4$, with 1 being the smallest and 4 the largest plant. Furthermore, each $\overline{K}$ is combined with the variable input of labour, varying from 0, column 3, to 8, the last column. The shaded area, rows 3, 4, 5, and 6, shows the level of output (total product). For example, row 3 represents a production function that combines 1 K with different Ls, 0 to 8, while row 4 represents a different production function that combines 2 units of K (a bigger plant) with labour, 0 to 8 units. Clearly, the manager has several plant options.

3.0.1 Demonstration Problem

Referring to Table 5-1, suppose the firm is currently producing 160 units of output and using a production function that employs 2 K. Now suppose that the demand for the firm's output increases permanently to, say 240, with a possibility of a further 10% increase in the following year. How many ways can this firm meet the market demand?

Answer:

The firm can produce 240 units of output using its current plant by stretching it almost to its limit. This will be done by combing 2K with 8L. Alternatively, the firm can produce the same output by expanding its plant (K) using a production function that combines either 3K (the next plant size) with 5L, or 4K with 4L. However, when we consider the possibility of the additional 10% increase in demand, the manager is left with only two choices, plant 3, $\overline{K} = 3$, and plant 4, $\overline{K} = 4$.

Learning Tip

Note that when there is expectation of future business growth, the manager may find it advantageous to choose a larger plant as opposed to a smaller plant to benefit from it in the long run.

Learning Tip

Note that in producing a given level of output, shown in Table 5-1, the firm is able to substitute capital for labour. This is known as the 'variable proportions' production function.

3.1 The Law of Diminishing Returns

Table 5-2, extracted from Table 5-1, shows that, in a given state of technology and keeping the input of capital constant, $\overline{K} = 2$, additional units of labour (the variable input) yield increasing output per unit of input up to a point. But eventually a point is reached beyond which further additions of labour yield diminishing returns per unit of input. In this illustration, the point of diminishing returns is reached at an input of 5 units of L. Put differently, the relative productivity of the marginal units of the variable inputs (L) initially increases (up to L = 4) and diminishes subsequently. Hence, the initial increasing returns to the variable input gives way to diminishing returns.

These, which are short-term phenomena, can be stated more broadly as the law of variable proportions. They state that as more and more of the variable factors are added to a given quantity of the fixed factors, the increment to output attributable to each of the additional units of the variable factor will increase at first, will later decrease, and will eventually become negative.

Table 5-2 The Law of Variable Proportions

Units of Input (L)	Units of Output (for $\overline{K} = 2$)	Increment to Output	Returns to the Variable Input
0	0	-	-
1	22	22	Increasing
2	48	26	Increasing
3	80	32	Increasing
4	125	45	Increasing
5	161	36	Diminishing
6	188	27	Diminishing
7	208	20	Diminishing
8	223	15	Diminishing

3.2 Total Product, Average Product, and Marginal Product

Total product (TP) is simply the maximum level of output that can be produced with a given amount of inputs, Column 2, in Table 5-2.

Average Product (AP_L), the most popular measure of productivity, is defined as total product divided by the quantity of the variable input (L) used, $AP_L = Q/L$.

Marginal Product (MP_L) is the change in total output attributable to the last unit of the input of labour, $MP_L = \frac{\Delta Q}{\Delta L}$.

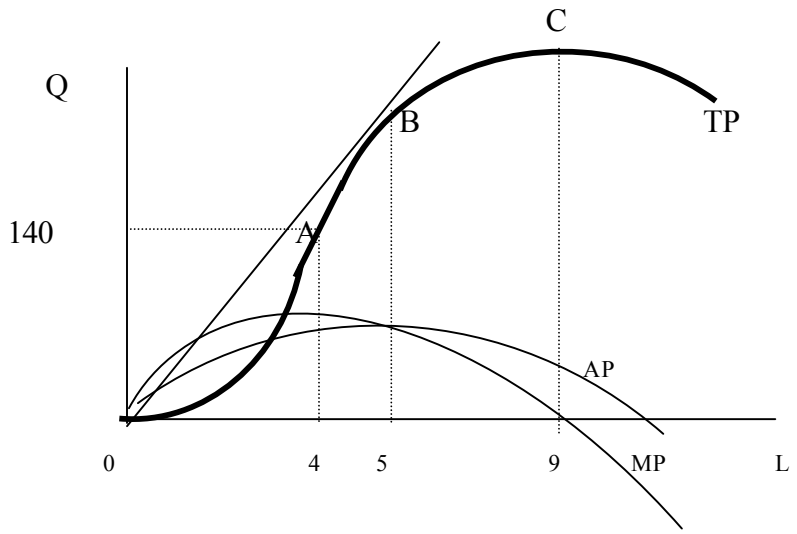
Note that the average and the marginal product concepts can also be defined for the fixed input of capital. Accordingly, AP_K, the average product of capital, is defined as total product divided by the quantity of the fixed input (K), whereas the marginal product of capital *(MP_K)* is the change in total output divided by the change in capital, $MP_K = \frac{\Delta Q}{\Delta L}$.

Figure 5-1 shows graphically the relationship among total product, marginal product, and average product. The first thing to notice about these curves is that as the use of labour increases between points 0 and A, the slope of the total product curve increases (becomes steeper). That is, output increases at an increasing rate. Since the slope of the TP curve is $MP_{L,}$, the marginal product curve will be increasing over this range. The range over which marginal product increases is known as the range of *increasing marginal returns.*

Figure 5-1 also shows that marginal product reaches its maximum at point A, where 4 units of labour are employed. This is indicated on the TP curve by the inflection point, point A. That is, the point at which the curve changes from concave upward to concave downward. As the usage of labour increases from the 4th through the 8th unit, total output increases, but at a decreasing rate. This is why marginal product declines between 4 and 8 units of labour but is still positive. The range over which marginal product is positive but declining is known as the range of *diminishing marginal returns* to the variable input.

Marginal product becomes negative when more than 9 units of labour are employed. After a point, using additional units of input actually reduces total product, which is what it means for marginal product to be negative. The range over which marginal product is negative is known as the range of *negative marginal returns.*

Figure 5-1



The relationship between marginal product and total product, exhibited in Figure 5-1, can be summarized as follows:

1. As long as the MP_L curve is rising, the total product curve increases at an increasing rate and is convex to the horizontal axis.

2. The quantity of input L at which the TP curve changes its curvature corresponds precisely to the point at which the MP_L curve peaks. This occurs at approximately 4 units of input, as shown by point A.

3. When the total-product curve reaches point C, its maximum MP_L is zero. Beyond this point, the marginal product is negative and the total product declines.

Learning Tip

The points discussed above can be summarized as follows:

Returns to the Variable Input (L)	**Marginal Product of the Variable Input (MP_L)**
Increasing returns	Positive and Rising
Decreasing returns	Positive and falling
Negative returns	Negative and falling

The law of variable proportions discussed above can be also discussed in the context of the average revenue. The average product of labour, AP_L, can be extracted from Table 5-2 as follows:

L	**AP_L**
0	0
1	22
2	24
3	26.67
4	31.25
5	32.2
6	31.33
7	29.7
8	27.87

AP_L is initially subject to the law of increasing returns, reaching a maximum at approximately 5 units of labour. Beyond that, it becomes subject to the law of diminishing returns. This relationship is also exhibited in Figure 5-1. To sum up:

1. The average product increases as variable input increases as long as the marginal product exceeds the average product.

2. When the marginal product is less than the average product, the average product decreases as variable input increases.

3. When the average product is at a maximum, the average product and marginal product are equal.

Learning Tip

It should be noted that even when the marginal product of labour slopes downward from its maximum point, the average product is still rising. It will continue to rise as long as the marginal product is greater than the average product. The average product will reach its maximum when the marginal product and the average product are equal ($MP_L = AP_L$). This is the point of maximum production efficiency for the variable input. Managers need to know this input level in order to make strategic decisions about production in the short run.

Figure 5-1 also illustrates the three stages of a typical production function.

Stage 1. This stage extends from zero input of the variable factor to the level of input where the average product is at a maximum. In this stage, the fixed factors are excessive relative to the variable input. Consequently, output can be increased by increasing the variable input relative to the fixed input.

Learning Tip

If a large department store were understaffed with salespersons, sales could be increased by hiring more salespersons, labour (the variable input), relative to floor space and fixtures (the fixed input). Throughout Stage 1, marginal product is greater than the average product.

Stage 2. This stage extends from the end of Stage 1 (the point where the marginal product and the average product are equal) to the point where the marginal product is zero and the total product is at a maximum, point C in Figure 5-1. Stage 2 is a rational stage in which relatively good balance has been achieved between the variable and fixed inputs.

Stage 3. In this stage, in which input is greater than 9 units, the variable-input factor is excessive relative to the fixed factors, the marginal product is negative, and the total product is falling. It is completely irrational to produce in this stage.

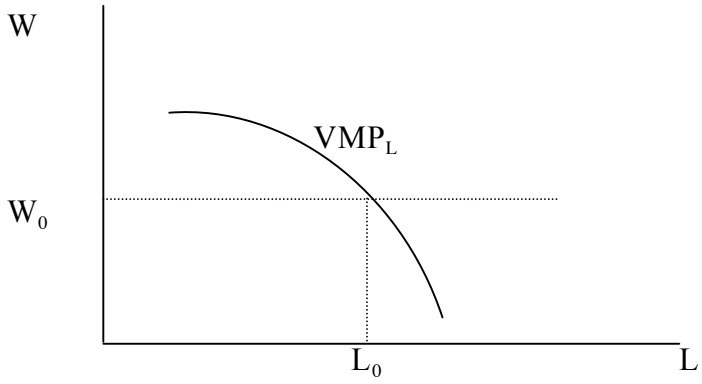
3.3 The Optimal Level of Employment of a Variable Input

To determine the optimum input-output relationship for maximum profit, we must shift the analysis from physical input-output relationships to economic relationships. The precise amount of the input variable needed to produce maximum profit depends upon the revenue and cost associated with hiring an additional unit of the variable input, where the cost side depends on the price of the input variable, while the revenue side is a function of marginal product of the input variable as well as the selling price of the output.

The additional revenue associated with hiring an extra unit of the input of labour equals the value of output produced by this additional hiring. Being referred to as the *value of marginal product (VMP)*, this is defined as MP_L x P, where $MP_L = \Delta Q/\Delta L$ and P is the price of the product sold. On the other hand, the additional cost of hiring is the wage rate, W. Therefore, optimality requires that:

$$MP_L \text{ x } P = W \qquad (3)$$

Figure 5-2



Recalling that the rational stage of production is stage 2, where MP_L is diminishing (Figure 5-2), the firm should hire labour up to a point where the value of its marginal product equals its marginal cost, L_0. Alternatively, equation (4) expresses the employment condition in real terms. The left hand side of this equation is the marginal physical product of labour, whereas the right hand side is the real wage:

$$MP_L = W/P \qquad (4)$$

Learning Tip

Note that the downward sloping VMP_L is also the demand for labour by the profit-maximizing firm. Since MP_L declines as more L is hired, VMP_L also declines. The firm continues adding to its work force as long as VMP_L exceeds W_0. In competitive environments, individual firms behave as *price takers* and therefore exert no influence on market prices. In this analysis, our individual price taking competitive firm takes P and W as given. Suppose initially $VMP_L > W$. Accordingly, the firm is not currently maximizing its profit and therefore should expand hiring. However, additional hiring will cause MP_L to decline, while P and W remaining constant. As the firm, in the process of hiring, gains additional profit, the left-hand side gradually diminishes and eventually equals the right-hand side.

3.3.1 Demonstration Problem

Suppose the production function is given by the following equation

$$Q = 20L - 5L^2$$

where L is defined in thousands of units of labour. What is the expression for the marginal product and average product of labour function? At what level of labour does TP reach its maximum? If P = \$5 per unit and W = \$2 per hour, what is the optimal level of employment of labour?

Answer:

$$MP_L = \frac{dQ}{dL} = 20 - 10L.$$

$$AP_L = \frac{Q}{L} = 20 - 5L$$

Note 'd' in the MP expression is the mathematical notation of derivative. The marginal product is the derivative of the total product with respect to quantity. That is, it shows the amount of change in output when the input of labour changes infinitesimally. TP is maximized when MP_L is set equal to zero: $20 - 10L = 0$, hence L = 2 (000).

V MP_L= W, \$5.(20 –10L) = \$2, and L = 1.96 (000) = 1960

3.4 The Production Function of Multiple Variable Inputs (Isoquants and Isocosts)

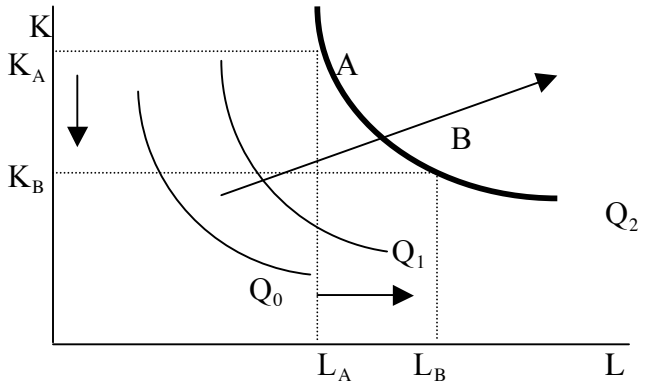
The firm's production problem is to choose the optimal levels of all inputs, not just a single input, that enter the production process such that output is maximized for a given

budget constraint. Therefore, our next task is to examine the optimal choice of capital and labour in the long run, when both inputs are free to vary. In the presence of multiple variables of production, there exist various combinations of inputs that enable the manager to produce the same level of output.

The basic tools for understanding how alternative inputs can be used to produce output are *isoquants* and *isocosts*. An *isoquant* is a line joining combinations of inputs (K, L) that generate the same level of output. The word *isoquant* comes from the Greek word *iso* meaning equal and the Latin word *quantus* meaning quantity. As with indifference curves, there will be a family of isoquant curves, with higher curves preferred to lower curves. Similarly, the curves do not cross, and they are convex to the origin. Figure 5-3 depicts a typical set of isoquants. Because input bundles A and B both lie on the same isoquant, each will produce the same level of output, namely Q_2 units. Input mix A implies a more capital-intensive plant than the input mix B. As more of both inputs are used, a higher isoquant is obtained. Thus as we move in the northeast direction in the figure, each new isoquant is associated with higher and higher levels of output,

$$Q_2 > Q_1 > Q_0.$$

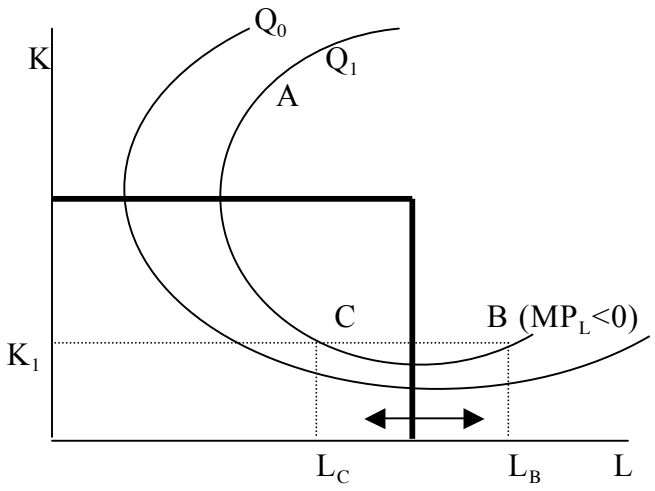
Figure 5-3



Two points need to be appreciated.

1. Unlike indifference curves, isoquant curves are not negatively sloped throughout their length. Isoquants may bend back, as shown in Figure 5-4, at both relatively high and relatively low ratios of capital to labour. The reason for this difference is that in consumer theory we assume that marginal utility would never be negative, whereas in producer theory we make no such restrictive assumption concerning marginal products.

Figure 5-4



To understand this better, we should distinguish between technical efficiency, whereby there is no wasted resources, and economic efficiency, whereby there is no waste of expenditure. Points A and B on the isoquant Q_1, in Figure 5-4, are technically inefficient points. For example, the firm at point B is using too much labour relative to point C. By moving from B to C, the firm, while maintaining its output level (Q_1), can save ($L_B - L_C$) in labour. Note that in reducing the input of labour, employment of K remains unchanged. Similarly, point A is also technically inefficient. The two perpendicular (bold) lines drawn against the vertical and horizontal axis, separate the space into the economic and uneconomic region. Points above and the to the right of these lines (outside the box) such as B and A, are in the uneconomic region, while points below and to the left of these two lines (inside the box), such as C, are in economic region.

Learning Tip

When the marginal product of labour is negative, which happens when the capital-labour ratio is low, the slope of the isoquant must be positive. Hence, the isoquant bends upward at the right-hand end. Conversely, where the marginal product of capital is negative, at relatively high capital-labour ratios, the slope of the isoquant must take a positive value, and thus the curve bends back at the left-hand. In between, where both marginal products are positive, the slope must be negative.

2. Furthermore, isoquants are convex. The reason isoquants are typically drawn with a convex shape is that inputs such as capital and labour are not perfectly substitutable. In Figure 5-3, for example, if we start at point A and begin substituting labour for capital, it takes increasing amounts of labour to replace each unit of capital that is taken away. The slope of an isoquant at any point can,

therefore, be expressed as the ratio of the amount of capital that can be subtracted from the production process to the amount of labour that is added to the production process, such that output level remains constant. This ratio, known as the *marginal rate of substitution (MRTS)* is the rate at which labour and capital can substitute for each other:

$$MRTS = -\frac{\Delta K}{\Delta L} \tag{5}$$

where ΔK is the decrement to capital that will just allow output to remain unchanged given a one-unit increment, ΔL, to the labour input.

Learning Tip

Just as the slope of an indifference curve is equal to the negative of the ratio of the marginal utilities, the slope of an isoquant is equal to the negative of the ratio of the marginal products. To see a more precise connection between MRTS and diminishing marginal return, we need to develop mathematical argument. We can begin by noting that when we change the quantity of labour by ΔL units and the quantity of capital by ΔK units of capital, the change in output that results from this substitution will be as follows:

ΔQ = change in output from change in quantity of capital + change in output from change in Quantity of labour

$$\Delta Q = (\Delta K \text{ x } MP_K) + (\Delta L \text{ x } MP_L).$$

Since changes in output caused by a movement along an isoquant is zero:

$$\Delta Q = 0 = (\Delta K \text{ x } MP_K) + (\Delta L \text{ x } MP_L).$$

Rearranging this equation we obtain:

$$\frac{\Delta K}{\Delta L} = -\frac{MP_L}{MP_K} = \text{MRTS} \tag{6}$$

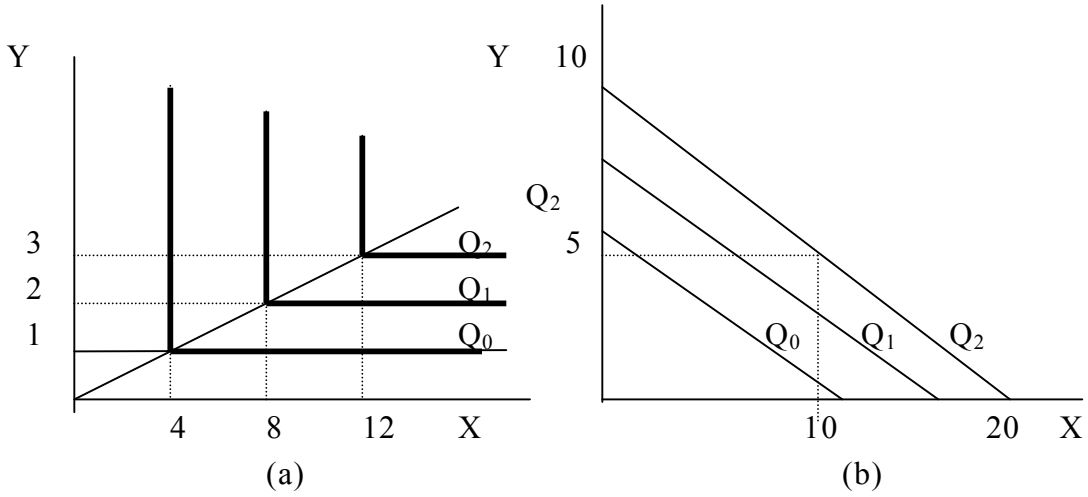
Specifically, with capital on the vertical axis, the slope of an isoquant is equal to the negative of the marginal product of labour over the marginal product of capital.

3.5 Shapes of Isoquants

The possibility of substitution of input L for input K and the degree of substitution between them determine the shape and slope of the isoquant. In addition to the normal shape of isoquants illustrated above, there are peculiar shapes associated with perfect

substitutes and perfect complements, as shown in Figure 5-5. Panel (a) shows the right-angled isoquants resulting from two inputs that are perfectly complementary; i.e., input of a single variable by itself will not produce any output. To obtain output, both Y and X must be input in a fixed ratio. For example, the manufacturing of a car requires an input of one body casting (chassis) and four wheels, a ratio that never changes. Therefore, if two castings are input, eight wheels must also be input to obtain two units of output, because the two inputs are complementary. Note that in panel (a), Y = body casting, and X = wheels.

Figure 5-5



Panel (b) shows the isoquants resulting from two inputs that are perfect substitutes for one another. Suppose that a firm is choosing between two types of computers to store company data. One has a high-capacity hard drive that can store 10 gigabytes of data, while the other has a low-capacity hard drive that can store 5 gigabytes of data. If it needs to store 100 gigabytes of data, it could either purchase 10 high-capacity computers and no low-capacity computers (the vertical intercept), or it could purchase no high-capacity computers and 20 low-capacity computers (the horizontal intercept). Or it could purchase 5 high-capacity computers and 10 Low-capacity computers. Note that in panel (b), Y = high-performance computers, while X = low-performance computers.

For the fixed-proportions production function (perfect complements), the expression for the production function is:

$$Q = \min (4W, \text{casting}) \qquad (7)$$

Learning Tip

Note that adding more wheels to a body casting gives us no more cars, neither does adding a body cast to four wheels. The line connecting the corners of the L-shaped isoquants in panel (a) of Figure 5.5 shows the fixed proportion (combinations) of the body cast and wheels, i.e., there is only one possible input combination: one body cast per 4 wheels.

On the other hand, the production function for the linear (perfect substitutes) isoquants is represented by

$$Q = 10H + 5L \tag{8}$$

For this production function, the slope is constant and the MRST does not change as we move along the curve.

Learning Tip

The above analysis suggests that a firm's ability to substitute among inputs depends on the curvature of its isoquants. Specifically, when the production function offers limited input substitution opportunity, the MRTS changes substantially as we move along an isoquant. In this case, the isoquants are nearly L-shaped, as in panel (a) of Figure 5.5. When the production function offers abundant input substitution opportunities, the MRTS changes gradually as we move along an isoquant. In this case, the isoquants are nearly straight lines, as in (b).

3.6 Cobb-Douglas Production Function

This is an intermediate between a linear production function and a fixed proportions production function. This production function is known as the *Cobb-Douglas* production function, and it is given by the formula

$$Q = AK^a L^b \tag{9}$$

where A, a, and b are positive constants. A is the scale variable, where a and b (the exponents) are the elasticity of output with respect to capital and labour respectively.

In the Cobb-Douglas production function, capital and labour can be substituted for each other. Unlike a fixed proportions production, capital and labour can be used in variable proportions. Unlike a linear production function, though, the rate at which labour can be

substituted for capital is not constant as you move along an isoquant. Note, however, that this type of production function is not the only one that allows substitutability between L and K. It is, however, one that has several friendly mathematical properties.

3.7 Isocosts

An isocost line is a locus of combinations of inputs that require the same total expenditure. Let us express the firm's expenditure on inputs as

$$TC = K \times P_K + L \times P_L \tag{10}$$

where TC is the total dollar expenditure; P_K and P_L are the unit prices of capital and labour, respectively; and *K* and L are the number of physical units of capital and labour that are to be employed in the production process. This can be rearranged to appear as

$$K = \frac{TC}{P_K} - \frac{P_L}{P_K} L \tag{11}$$

in which form it is perhaps more recognizable as a linear equation explaining K in terms of L and three known values (TC, P_L, and P_K). It can therefore be plotted in the same space as the isoquant curves. The intercept of the isocost line on the capital axis occurs when $L = 0$, and this intercept value of *K* is simply the total expenditure divided by the price of the capital units, resulting in a particular physical quantity of capital units, TC/P_K. As we purchase units of labour, it is evident that our purchase of capital units, for the same expenditure (isocost) level, is drawn down in the ratio of the price of labour to the price of capital.

Figure 5-6

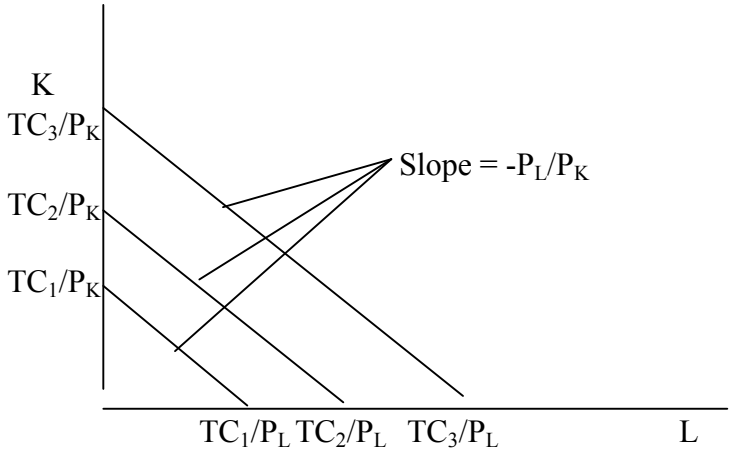


Figure 5-6 shows the graph of isocost lines for a variety of different total cost levels, TC_1, TC_2, and TC_3, where $TC_3 > TC_2 > TC_1$. In general, there are an infinite number of isocost lines, one corresponding to every possible level of total cost.

Learning Tip

Note that isocosts associated with different levels of costs are parallel to each other. This is so because input prices (P_L, P_K) are kept constant. However, when input prices change, the line will rotate. If for example, P_L declines, the horizontal intercept gets larger and the line rotates counter clockwise.

3.7.1 Demonstration Problem

Suppose due to a general slowdown in the economy, the price of all factors of production, labour and capital, decline by the same percentage. How would this change impact the isocost line?

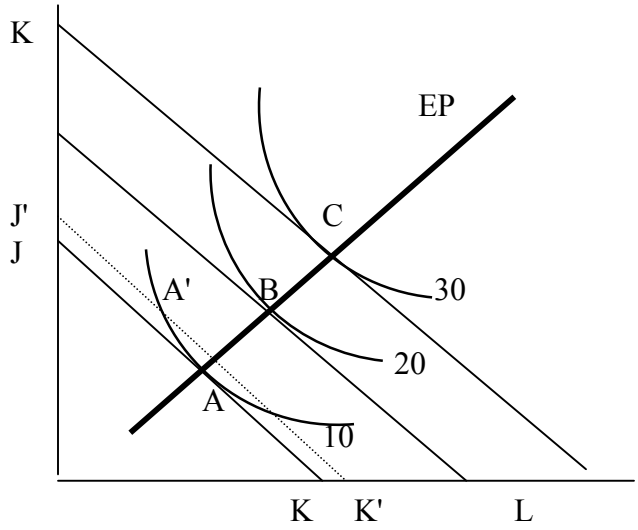
Answer:

A proportionate decrease in the price of both factors leaves the slope of the isocost line, P_L/P_K, intact while shifting it to the right in a parallel fashion.

3.8 Cost Minimization (Economic Efficiency)

For each output level, there will be a minimum-cost combination of the factors necessary to produce that output level. In Figure 5-7, we show three isoquant curves representing 10, 20, and 30 units of output. The cost of producing 10 units is minimized at point A, since any other capital/labour combination producing 10 units, such as at A', will lie to the right of the isocost line *JK* and would thus require a larger total expenditure to purchase, J'K'. Recall that the intercepts on the capital and labour axes are equal to TC/P_K, where P_K and P_L are presumed to remain unchanged. Hence, larger total expenditures are represented by higher intercept points and higher isocost lines. Similarly, the output level 20 is produced at the least-cost by the input combination represented by point B, as 30 units of output is produced at the least-cost by the input combination at point C.

Figure 5-7



Input cost is minimized for a given output level, where the isoquant representing that output level is just tangent to the lowest attainable isocost line. Equivalently, output is maximized for a given input expenditure level where the isocost line representing that expenditure level is just tangent to the highest attainable isoquant. At the point of tangency between the isoquant and the isocost curve, the slopes of these curves are the same. That is, optimality requires that the rate at which the firm can technically substitute labour for capital equals the rate that the market allows it to.

$$MRTS = \frac{MP_L}{MP_K} = \frac{P_L}{P_K} \tag{11}$$

Rearranging terms, we have

$$\frac{MP_L}{P_L} = \frac{MP_K}{P_K} \tag{12}$$

This alternative arrangement holds that cost minimization or output maximization requires that the ratios of the marginal products to prices of all inputs be equal. This is

useful in that this way the rule of optimization can be generalized to a larger number of inputs, 1,2, 3,…..

$$\frac{MP_1}{P_1} = \frac{MP_2}{P_2} = \frac{MP_3}{P_3} = ... \quad (13)$$

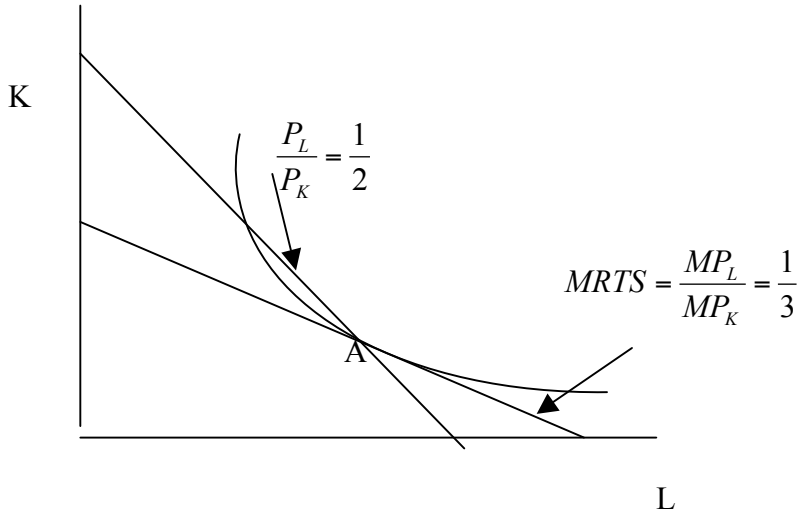
3.8.1 Demonstration Problem

Suppose $MP_L = 1$, $MP_K = 3$, whereas P_L and P_K are respectively \$1 and \$2. Is the firm that faces these factors optimizing information?

Answer:

Optimality requires $\frac{MP_L}{MP_K} = \frac{P_L}{P_K}$. Plugging in: $\frac{MP_L}{MP_K} = \frac{1}{3}$, whereas $\frac{P_L}{P_K} = \frac{1}{2}$. Therefore, the rate by which the firm is able to substitute labour for capital, 1/3, is lower than the rate the market allows it to, 1/2. As such, the firm is not minimizing its costs. Specifically, the firm is using too much labour and too little capital. In terms of a diagram, Figure 5-8, below, portrays this situation. Note that, at point A, $MRTS = \frac{MP_L}{MP_K} = 1/3$ is the slope of the isoquant, and $\frac{P_L}{P_K} = 1/2$ is the slope of the isocost line. The firm needs to substitute K for labour. Doing so will increase MP_L and decreases MP_K, and hence will increase MRTS toward 1/2. The process of substitution stops when the MRTS equals the price ratio (1/2).

Figure 5-8



3.8.2 Demonstration Problem (Mathematically Challenging)

Suppose the production is represented by a Cobb-Douglas production function, $Q = K^{0.5}L^{0.5}$. Also, suppose that P_K and P_L are \$10 and \$5, respectively.

a. Determine the optimal combination of inputs.

b. If the manager has \$500 to spend on labour and capital, how many units of K and L should the firm acquire?

Answer:

a. First we need to take the (partial) derivative of Q to find the marginal products of labour and capital: $MP_K = (0.5)K^{-0.5}L^{0.5}$ and $MP_L = (0.5)K^{0.5}L^{-0.5}$. Hence, $\frac{MP_L}{MP_K} = \frac{0.5K^{0.5}L^{-0.5}}{0.5K^{-0.5}L^{0.5}} = \frac{K}{L}$. Setting this equal to $\frac{P_L}{P_K} = \frac{1}{2}$, we have: $\frac{K}{L} = \frac{1}{2}$. Meaning that for every one unit of K, the firm needs to hire two units of L in order to minimize its costs, L = 2K.

b. The isocost line is represented by $TC = P_K xK + P_L xL = 10.K + 5.L = \$500$. Substituting in for either L or K: \$500 = 10(K) + 5(2K) = 20K, hence K = 25, L = 50.

In Figure 5-7, the line that connects A, B, and C is referred to as the *expansion path (EP)*. The expansion path, or more precisely, the *long run expansion path*, is a locus of the tangency points between various isoquant and isocosts that shows the least-cost combinations of labour and capital a firm would choose as it expanded its output level. This assumes constant price of labour and capital and constant state of technology.

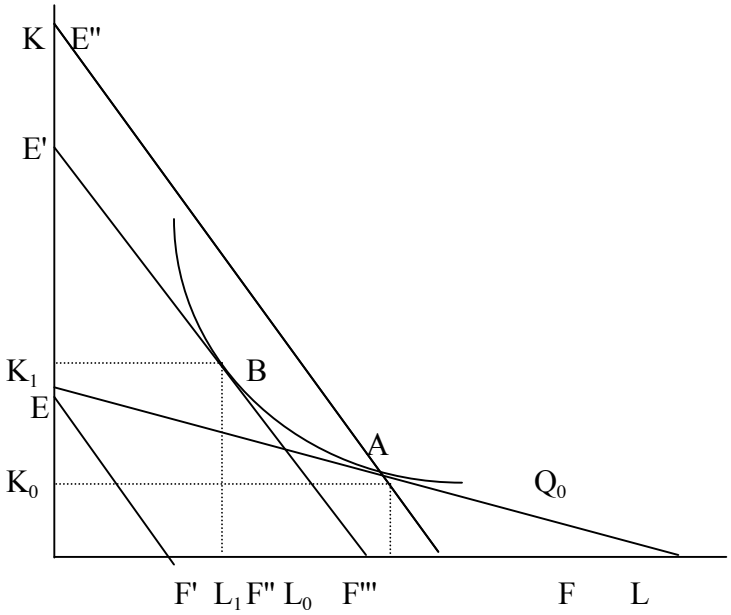
3.9 Input Substitution

Economic efficiency depends on the relative factor prices. If the price of one factor changes, all else being kept constant, the profit-maximizing firm will attempt to substitute away from the factor that has become relatively more expensive and in favour of the factor that has become relatively less expensive. Suppose the initial situation in Figure 5-9, is point A, where output level Q_0 is being produced economically efficiently by a combination of K_0 and L_0.

Now suppose that labour prices rise, for example, because of a new agreement with the labour union. This causes an increase in the cost of labour and the isocost line rotates down (clockwise) from EF to EF′. If the firm wishes to maintain its current expenditure, it cannot produce Q_0. Alternatively, if the firm wishes to maintain its output level at Q_0 to hold its market share, it will need to spend more money on the inputs.

In the long run, the firm will substitute capital for labour, i.e., the firm will increase its plant size to K_1 and decrease its labour input to L_1, point B. The increased input price ratio, P_L/P_K has caused the production process to become relatively more capital intensive. Note that E'F'', the new isocost line that is tangent to the isoquant Q_0 at point B, is parallel to EF'.

Figure 5-9



Learning Tip

In the short run, however, the firm is constrained to the plant size indicated by K_0. This constraint requires that the firm, wishing to produce Q_0, continue to hire more labour than is optimal ($L_0 > L_1$), causing the cost of production to exceed that of the long run, E''F''', in Figure 5-9.

3.10 Returns to Scale

Figure 5-10 illustrates three sets of isoquants for different production processes. In each panel, the points on the ray from the origin show the proportion by which output increases when both inputs are increased proportionately. The distance between the successive isoquants brought about by a proportionate increase in inputs is a reflection of

how output responds to changes in inputs. For example doubling all inputs means doubling the distance from the origin along the ray. We refer to changing all inputs by the same proportion as a change in *scale.* What is the effect on output of a change in scale?

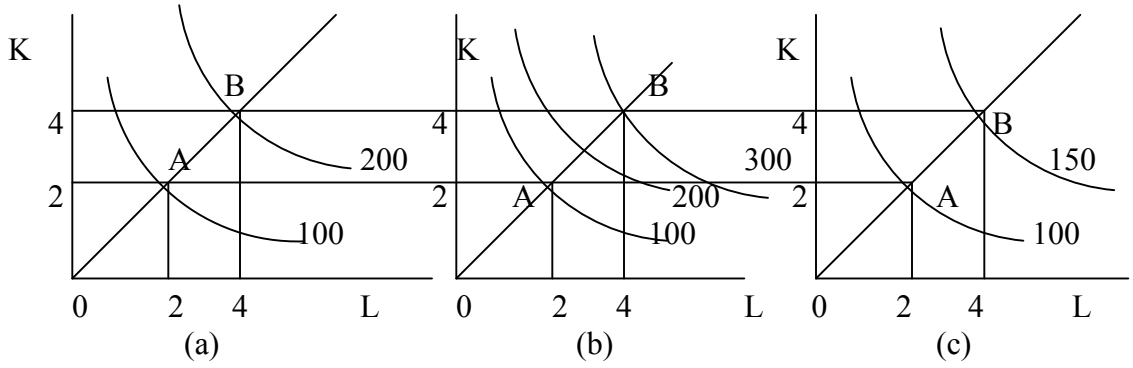
In panel (a), we find that doubling inputs leads to a doubling in output. That is, the move from point A to point B doubles the distance from the origin and also doubles output from 100 to 200. The property of output increasing proportionately to the scale of inputs is called *constant return to scale*, because the amount of output per bundle of inputs is always constant.

A production process exhibits constant returns to scale if increasing all inputs proportionately increases output in equal proportion.

Unlike the proportionately spaced isoquants in panel (a) of Figure 5-10, panel (b) has isoquants that bunch closer together and panel (c) has Isoquants that spread out as the distance from the origin along a ray increases. In panel (b), we find that doubling inputs leads to a more than doubling in output (tripling). That is, the move from point A to point B triples output and hence less than doubles the distance between 100 and 200 from the origin. The property of output increasing more than proportionately to the scale of inputs is called *increasing return to scale*, because the amount of output per bundle of inputs increases with the scale.

Panel (c) shows the case of a production function that in which output increases less than proportionately to scale. That is, the move from point A to point B *more than* doubles the distance between 100 and 200 from the origin and also less than doubles output (50% increase). The property of output increasing less proportionately to the scale of inputs is called *decreasing returns to scale,* because the amount of output per bundle of inputs falls as scale increases.

Figure 5-10



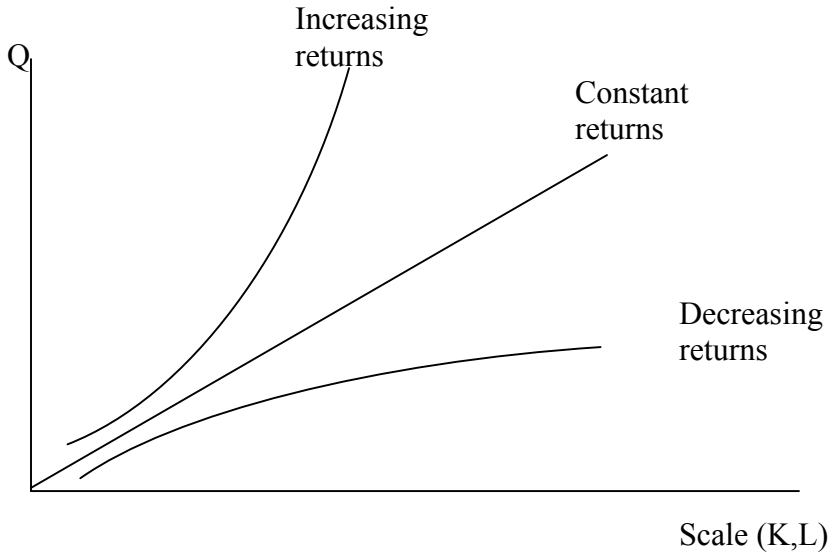
Why might some production processes exhibit constant returns to scale and others exhibit increasing or decreasing returns? One answer is that a larger scale of operation often

allows workers and managers to specialize in different tasks. Henry Ford's assembly line, for example, could produce more cars than the same workers and equipment could do if production had to be done one car at a time in separate garages, with each worker designing, machining, assembling, and finishing the cars. On the other hand, large scale can also introduce inefficiencies, often associated with breakdown of coordination and critical information flows.

From a policy perspective, increasing returns to scale seem to imply that large firms are desirable and should be encouraged, because they are able to produce more efficiently than small ones and thus should be able to sell their product at a lower price. Yet, as we will see later, large firms often have greater market power that would allow them to charge high prices, and so large firms are sometimes broken into smaller, competing firms to encourage lower prices. Alternatively, as in the case of regional electric utilities, large firms may be regulated by government bodies.

Another way to describe the effects of different returns to scale is to graph output against scale. Assuming that we have fixed a particular ratio of inputs (that is, we have limited ourselves to a particular ray from the origin), scale is measured on a graph like Figure 5-10 as the distance from the origin. Plotting scale versus output leads to Figure 5-11. The product-to-scale curve for decreasing returns to scale is concave with respect to the bottom of the graph. For increasing returns to scale, the curve is convex; and for constant returns to scale, it is a ray from the origin. The positioning of the curves in this Figure has no significance, as each type of curve will be positioned according to the equation that represents the expansion path.

Figure 5-11



3.11 Testing Production Function for Returns to Scale

If the production function is known, it can be analyzed algebraically for returns to scale. Suppose we have a Cobb-Douglas production function $Q = AK^a L^b$. This is a general expression assuming no prior knowledge of returns to scale. Let us double the inputs of K and L

$$Q' = A(2K)^a (2L)^b$$

or

$$Q' = (2)^{a+b}(AK^a L^b) = (2)^{a+b}.Q$$

Hence, if (a + b) = 1, Q' = 2Q, which is the case of constant returns to scale. If (a + b) >1, Q' > 2Q, which is the case of increasing returns to scale, and if (a + b) <1, Q' < 2Q, which is the case of decreasing returns to scale.

4 Technological Changes

So far, we have treated the firm's production function as fixed; that is, it remains stationary over time. But as knowledge in the economy evolves and as firms acquire know-how through experience and investment in research and development, a firm's production function will change. The notion of *technological progress* captures the idea that production functions can shift over time. In particular, technological progress refers to a situation in which a firm can achieve more output from a given combination of inputs, or equivalently, the same amount of output from less inputs.

We can classify technological progress into three categories: neutral technological progress, labour-saving technological progress, and capital-saving technological progress. Figure 5-12 illustrates *neutral technological progress*. In this case, an isoquant corresponding to a given level of output (10 units) shifts inward (indicating that lesser amounts of labour and capital are needed to produce a given output), but the isoquants shift so as to leave MRTS unchanged along any ray (e.g., OA) from the origin. Under neutral technological progress, in effect, the firm's entire map of isoquants is simply relabelled, each one now corresponding to a higher level of output, but the isoquants themselves retain the same shape.

Figure 5-12

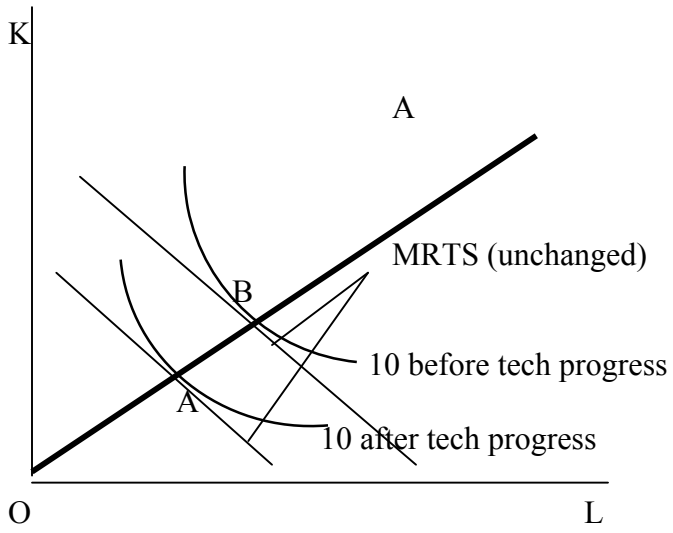


Figure 5-13 illustrates *labour-saving technological progress.* As before, the isoquant corresponding to a given level of output shifts inward, but now along any ray from the origin, the isoquant becomes flatter, indicating that the MRTS is now less than it was before. You should recall that MRTS = MP_L/MP_K, so the fact that the MRTS decreases implies that under this form of technological progress the marginal product of capital increases more rapidly than the marginal product of labour. This form of technological progress would arise when technical advances in capital equipment, robotics, or computer increase the marginal productivity of capital relative to the marginal productivity of labour.

Figure 5-13

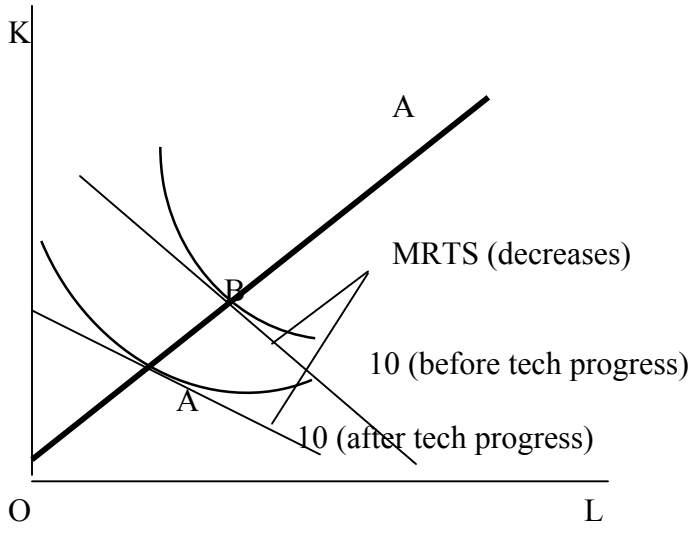
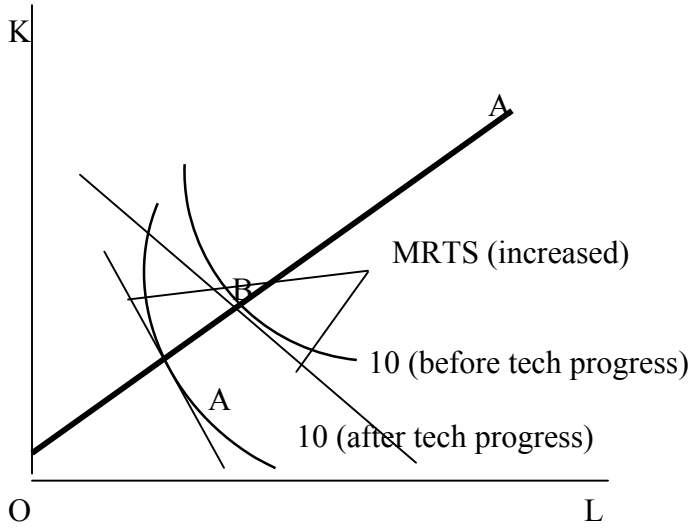


Figure 5-14 depicts *capital-saving technological progress*. Here, as an isoquant shifts inward, MRTS increases, indicating that the marginal product of labour is increasing more rapidly than the marginal product of capital. This form of technological progress would arise if, for example, the education or skill level of the firm's actual (and potential) work force rose, increasing the marginal productivity of labour relative to the marginal product of capital.

Figure 5-14



5 Summary and Review

Production of any goods requires input of at least two basic factors. In the short run, there has to be at least one fixed factor of production and one variable. This results in the law of variable proportions: diminishing and increasing. In the process of production, as the input of the variable factor (labour) increases, the marginal product of labour first increases (increasing returns), and then decreases (diminishing returns).

An isoquant shows the various combinations of two inputs that can be used to produce a specific level of output. Its slope, the 'marginal rate of technical substitution (*MRTS*),' is equal to the ratio of the marginal products of the two inputs. In order to minimize production costs or maximize profit, the firm must produce where an isoquant is tangent to an isocost.

Returns to scale is a long-range planning concept that enables management to determine the effect of increasing all inputs by the same proportion. 'Constant,' 'increasing,' and 'decreasing' returns to scale refer to the situation where output changes, respectively, by the same, by a larger, and by a smaller proportion than inputs. Increasing returns to scale arise because of specialization and division of labour and from using specialized machinery. Decreasing returns to scale arise primarily because as the scale of operation increases, it becomes more and more difficult to manage the firm and coordinate its operations and divisions effectively. In the real world, most industries seem to exhibit near-constant returns to scale.

6 Review Problems

1. The manager of a plant calculated the cost at different output levels. The result is in the table below:

Units of Labour	Total Product (Units)
0	850
10	1,700
20	3,500
30	6,900
40	10,000
50	11,500
60	12,600
70	11,550
80	10,400

   Find the average product and nmarginal product of labour for different levels of labour input.

2. A firm uses only labour to produce to output Q. The production function is:

   $Q = 10 \times L^{1/2}$

   If labour costs $ 25.00 per hour, then find the firm's total cost, average cost, and marginal cost functions.

3. The owner of H.J. Car Wash believes that the relationship between the number of car washed and labour input is

   $$Q = -0.8 + 4.5L - 0.3L^2$$

   where Q is the number of cars washed per hour, and L is the number of people employed. His firm receives $5 for each car washed, and the wage rate for each person employed is $4.50.

   a. How many people should he employ to maximize profits?

   b. What would be the firm's hourly profit?

4. Cost of labour is $100 per unit and capital is $200 per unit. If the marginal product of labour is 1,000 and marginal product of capital of is 2,500, then is it optimal to use the combination of 10 labour and 30 capital? If not, then how should the owner revise the combination?

5. A firm makes optimal production decision based on its technology and prices of inputs. The inputs that are used are labour (L) and capital (K). Suppose labour cost is $20 per unit and capital cost is $10 per unit. What is the MRTS for production of 500 units? If prices of inputs remain the same, will the MRTS change for production of 1,000 units? If the firm employs 100 units of labour and 100 units of capital with the technology $K^{0.5}L^{0.5}$, then is the firm optimizing its production decision?

6. A firm faces Cobb-Douglas technology for production as given below:

   $Q = K^{0.75}L^{0.25}$

   Here, K is capital and L is labour. In the short run K is fixed. What should be the wage rate in term K and L? If wage is $5.00, K is 100 and L is 625, then what is the price that the firm should charge to maximize profit?

7. A software developer finds that development of its 'Accpack' accounting software follows the function $10K^{3/4}L^{1/4}$, where K and L are capital and labour inputs respectively. Find the marginal products of capital and labour. Also find the marginal rate of technical substitution. If capital costs $5.00 per unit and wage is $10.00 per unit, then what is the condition for cost minimization? If the owner currently employs 10,000 units of capital and 10,000 units of labour to produce 100,000 units of Accpack, is the firm following cost optimization rule? What should be the optimum combination of K and L to produce 100,000 units of Accpack?

7 Answer Key to Review Problems

1. The manager of a plant calculated the cost at different output levels.

Labour	Total Product	Average Product	Marginal Product
0	850	100	0
10	1,700	170	850
20	3,500	175	1,800
30	6,900	230	3,400
40	10,000	250	3,100
50	11,500	230	1,500
60	12,600	210	1,100
70	11,550	165	-1,050
80	10,400	130	-1,150

2. Marginal Product of Labour = $500 \times L^{-1/2}$. In equilibrium, Value of Marginal Product of Labour = Wage, Since P =1, $1 \times 500 \times L^{-1/2} = 25$, therefore,

   $500 \times L^{-1/2} = 25$, $L^{1/2} = 20$, $L = 400$, and Quantity = $1000(400)^{1/2} = 20{,}000$.

3. H.J. Car Wash:

   a. Value of marginal product = wage. Hence, $\$5 \times (4.5 - 0.6L) = \$4.50$, therefore, $L = 6$

   b. Profit = TR – TC. TR = P x Q = $\$5 \times (-0.8 + 4.5L - 0.3L^2) = 5 \times (-.8 + 4.5 \times 6 - 0.3(6)^2) = \$77$. TC = 6 x \$4.50 = \$27.0. Therefore, profit equals \$50 per hour.

4. Wage (w) = \$100, Rent (r) = \$200

   MPL = 1,000, MPK = 2,500

   MPK/r = 2,500/200 = 12.5 > MPL/w = 1,000/100 = 10

   Therefore more capital (or less labour) needs to be employed until MPK/r = MPL/w.

5. In equilibrium (when optimal decisions are made) MRTS = Input Cost Ratio = 10/20 = 0.5. MRTS at an output level of 500 should not differ from MRTS at an output level of 1,000 if the input costs do not change.

   Technology = $K^{0.5}L^{0.5,}$ MPK = $0.5K^{-0.5}L^{0.5}$ and MPL = $0.5K^{0.5}L^{-0.5}$

   Therefore, MPK/MPL = L/K. When optimal decisions are made, L/K = 0.5, and L = 0.5K

   If 100 L and 100 K are used to produce any level of output, since L/K = 1 > 0.5.

6. $Q = K^{0.75}L^{0.25}$

   Value of Marginal Product of Labour, VMPL = $0.25K^{0.75}L^{-0.75}$ x P = Wage

   Wage = 5.00, hence, K = 10,000, L = 625.

   $0.25K^{0.75}L^{-0.75}.P = 5$, $0.25(10000)^{0.75}(625)^{-0.75}.P = 5$, and therefore, P = 2.5

7. MPK = $7.5K^{-1/4}L^{1/4}$ …… (i)

   MPL = $2.5K^{3/4}L^{-3/4}$ …… (ii)

   Since MPK / MPL = w/r, dividing (i) by (ii), we get

   $(7.5K^{-1/4}L^{1/4})/(2.5K^{3/4}L^{-3/4}) = 5/10$, 6L = K

   It is not optimal since L = K, and the optimality condition as derived above is 6L = K.

   Quantity to produce = 100,000, $10(6L)^{3/4}L^{1/4} = 100{,}000$, $L^{3/4}L^{1/4} = 10{,}000/6^{3/4}$ L = 2,608.

8 References and Suggested Further Reading

Besanko, D. and Braeutigam, R. *Microeconomics: An Integrated Approach*, Wiley, 2002

Baye, M., *Managerial Economics and Business Strategy*, McGraw Hill, Irwin, 2002.

Douglas, J. E., *Managerial Economics: Analysis and Strategy*, Prentice Hall, 1992.

Pindyck, R. and Rubinfeld, D., *Microeconomics*, Fifth Edition, Prentice Hall, 2001.

Salvatore, D., *Managerial Economics in a Global Economy*, McGraw Hill, 1993.

Commonwealth of Learning Executive MBA/MPA

E5 Managerial Economics

Block 6

Cost Theory

Course author: Farrokh Zandi, PhD
Professor, Schulich School of Business
York University
Toronto, Canada

Course designer: Douglas R. Cronk, M.A., ABD
Vancouver, BC, Canada

Copyright © Commonwealth of Learning, 2003
All rights reserved. No part of this course may be reproduced in any form by any means without prior permission in writing from:

The Commonwealth of Learning
1285 West Broadway
Suite 600
Vancouver, BC V6H 3X8
CANADA
e-mail: info@col.org

Contents

1 Block Six: Objectives and Introduction 1
1.1 Introduction 1
2 Short Run Costs 1
2.1 The Total Variable Cost Curve 2
2.2 Average Variable and Marginal Costs 3
2.2.1 Demonstration Problem 1 6
2.3 Fixed and Sunk Costs 6
2.4 The Short-Run Supply Decision 7
3 Long Run Costs 8
3.1 Minimum Efficient Scale 12
3.1.1 Demonstration Problem 2 13
3.2 Economies and Diseconomies of Scale 13
3.3 Reasons for Economies of Scale 15
3.4 Reasons for Diseconomies of Scale 15
3.5 Economies of Scope 15
3.5.1 Demonstration Problem 3 17
4 The Nature of Costs 17
4.1 Economic Versus Accounting Costs 17
4.2 Direct and Indirect Costs 18
4.3 Explicit and Implicit Costs 18
4.3.1 Demonstration Problem 4 18
4.4 Opportunity Costs and Accounting Costs 19
4.5 Economic versus Accounting Profits 19
4.5.1 Demonstration Problem 5 20
4.5.2 Demonstration Problem 6 20
4.6 Normal and Pure Profits 21
4.7 Incremental Cost Analysis 22
4.8 Relevant Costs and Irrelevant Costs 22
4.8.1 Demonstration Problem 7 23
5 Summary and Review 23
6 Review Problems 24
7 Answer Key to Review Problems 26
8 References and Suggested Further Reading 28

1 Block Six: Objectives and Introduction

After working through Block Six of this course, you should be able to:

1. Examine the short run and long run cost curves.
2. Analyze the nature of costs and their relevance in managerial decision making.
3. Explain the importance of the long-run average cost curve.
4. Discuss economies and diseconomies of scale and analyze the factors behind them.
5. Explain the phenomenon of economies of scope vis a vis economies of scale.

1.1 Introduction

Firms incur costs when they buy inputs to produce the goods and services that they plan to sell. In this block we examine the link between a firm's production process and its total cost. The firm's cost functions are derived from the optimal input combinations examined in the last unit and show the minimum cost of producing various levels of output. In this chapter, we focus on the nature of costs of production consisting of explicit and implicit costs, relevant or opportunity costs, and incremental costs. We, subsequently, examine plant size and economies of scale and scope.

2 Short Run Costs

In the short run, just as businesses use fixed and variable inputs, they face corresponding fixed and variable costs. Fixed costs, or total fixed costs, (TFC), do not change when a business changes its quantity of output, since these costs relate to fixed inputs such as machinery and land. Variable costs, or total variable costs, (TVC), in contrast, relate to variable inputs, which change when a business adjusts the quantity produced. The most important variable costs are wages and payments for materials used in production, whereas the typical fixed cost is the cost of machinery. Total cost (TC) is the sum of all inputs, both fixed and variable, and is found by adding fixed and variable costs at each quantity of output.

Costs and production are two sides of the same coin. A firm's total cost reflects its production function, whereas the firm's supply curve, discussed in the earlier Blocks, is a reflection of its costs relationships.

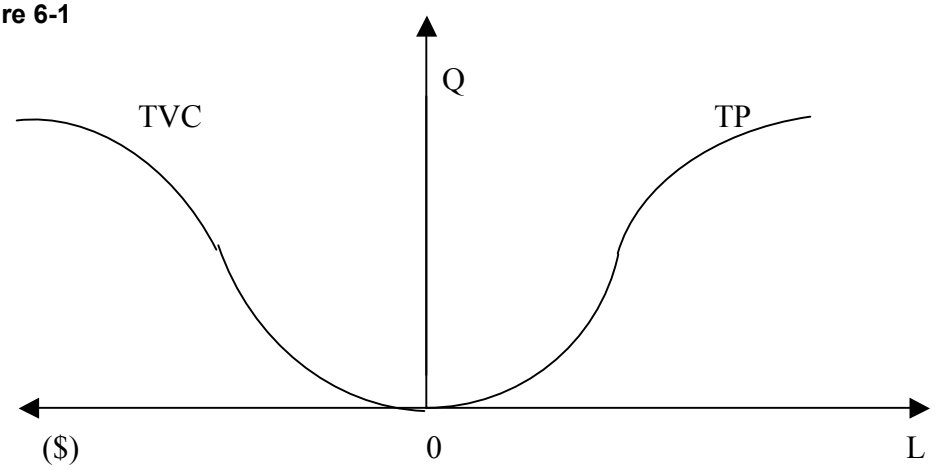
2.1 The Total Variable Cost Curve

The Total Variable Cost (TVC) curve can be derived from the TP curve simply by multiplying the level of variable inputs by the cost per unit of those inputs and by plotting these cost data against the output level. Suppose that the variable factor unit costs \$5 each. Using the data from the total product curve, shown in Table 5-2 in Block 5 partially reproduced below, and multiplying each unit of the variable factor by \$5, we can plot the cost of the variable input against the output, Figure 6-1.

Table 6-1 The Law of Variable Proportions

Units of Input (L)	Units of Output (for $\overline{K} = 2$)
0	0
1	22
2	48
3	80
4	125
5	161
6	188
7	208
8	223

Figure 6-1



Note that the scale on the left-hand side of the horizontal axis is five times that on the right-hand side, so that the TVC curve is a mirror image of the TP curve.

A firm's total cost is the sum of fixed and variable costs.

2.2 Average Variable and Marginal Costs

Average variable cost (AVC) is equal to TVC divided by output at every level of Q. That is,

$$AVC = \frac{TVC}{Q} \tag{1}$$

Average fixed cost (AFC) is equal to TFC divided by output at every level of Q. That is,

$$AFC = \frac{TFC}{Q} \tag{2}$$

Average total cost (ATC) is equal to TC (TVC + TFC) divided by output at every level of Q. That is,

$$ATC = \frac{TC}{Q} = \frac{TFC + TVC}{Q} = AFC + AVC \tag{3}$$

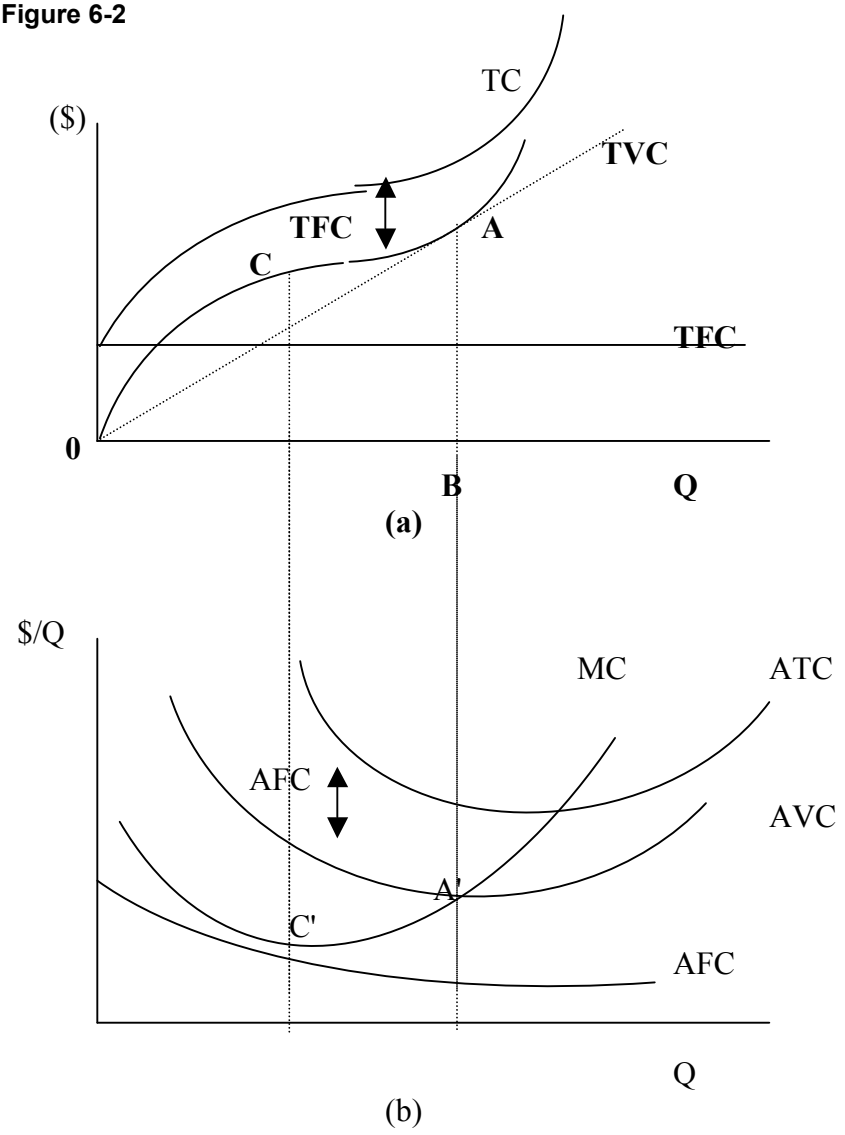
Marginal cost (MC) is equal to the change in total costs caused by a one-unit change in output:

$$MC = \frac{\Delta TC}{\Delta Q} = \frac{\Delta TVC}{\Delta Q} \tag{4}$$

Note that the fixed cost by definition does not change with the level of output.

In panel (a) in Figure 6-2, we show the TVC curve tipped on its side (rotated 90 degrees to the right), as well as the TC and TFC curve. Note that the difference between TC and TVC is the Fixed costs (TFC). In panel (b), the average and the marginal curves are shown.

Figure 6-2



Note that the AVC for each level is equal to the ratio of the vertical distance from the quantity axis to the TVC curve. Average variable cost at point *A* on the TVC curve, panel (b) Figure 6-2, is equal to the slope of OA (AB/OB), and is shown as point *A'* on the AVC curve, where MC = AVC, below.

It can be seen that the slopes of rays from the origin joining points on the TVC curve become progressively flatter as we begin to move up the TVC curve away from the origin. Thus the AVC, which is equal to the value of these slopes, must fall over this range. MC is, on the other hand, equal to the slope of the TVC curve at each output level. If we were to put tangents against the TVC curve at every output level, we would see that the slopes of these tangents would fall, at first, up to the point of inflection on the TVC curve, point C in panel (a) and would then rise. The point of inflection occurs where the TVC curve changes from convexity to concavity (when viewed from above). The slope of the TVC becomes progressively flatter until point C is reached.

Learning Tip

When the MC curve lies below the AVC curve, the latter is falling. In effect, the lower marginal cost is pulling down the average cost. When the MC curve lies above the AVC curve, the latter must be rising, being pulled up by the marginal costs. It follows that when the MC crosses the AVC curve, the AVC must be at its minimum value. This is evident in Figure 6-2. At this point, MC = AVC, since both are given by the slope of the tangent to the TVC curve at that point, (points A and A').

Learning Tip

Note that marginal costs are inversely related to the marginal product of the variable factors. When marginal productivity of labour is falling, marginal cost of output is rising, and vice versa. But if the efficiency of the marginal units of the variable factor is constant, output can be produced at a constant level of marginal cost. These relationships can be confirmed by another look at Figure 6-1. Notice that when the TP curve is concave from above, exhibiting increasing marginal productivity of the variable factor(s), the TVC curve is indicating decreasing marginal costs per unit of output. Alternatively, when diminishing returns set in for the variable factor(s) after the inflection point, the TVC curve begins to increase at an increasing rate, and marginal costs are increasing over this range.

Mathematically, this can be presented as:

$$AVC = \frac{TVC}{Q} = \frac{P_L x L}{Q} = \frac{P_L}{\frac{Q}{L}} = \frac{P_L}{AP_L}, (5)$$

$$MC = \frac{\Delta TVC}{\Delta Q} = \frac{\Delta (P_L x L)}{\Delta Q} = \frac{P_L x \Delta L}{\Delta Q} = \frac{P_L}{\frac{\Delta Q}{\Delta L}} = \frac{P_L}{MP_L} (6)$$

In many production processes the fixed factors are indivisible, in the sense that we must utilize all of the fixed factors or none at all. In case of a steel mill, for example, the blast furnace is either on or off. Therefore, it must be very inefficient to produce low volumes. In cases like this, over a wide range of output, the higher the level of output, the lower the average cost. Beyond certain point, however, average cost begins to rise.

In other production processes, however, the plant, or some parts of it, are divisible in the sense that the variable inputs are not obliged to work with the entire plant, but may instead work with a more efficient subset of the plant. In these instances, the firm may be able to work with a reasonably constant input ratio over a wide range of output. The upshot being that, in the former, the average cost curve will be U shaped, as in Figure 6-2, whereas in the latter case, it will be closer to a flat line.

In practice, cost functions may take many forms, but the cubic cost function is the most frequently used function, closely approximating any cost function.

2.2.1 Demonstration Problem

The total cost function for a car rental company is

$$TC = 100Q - 3Q^2 + 0.1Q^3.$$

1. What is the equation expressing the AVC curve?
2. What is the equation expressing the MC equation?
3. Determine the level of output that will minimize the AVC curve.
4. Determine the level of output that will minimize the MC curve.

Answer:

1. AVC = TVC/Q. Since TFC is zero, all costs are variable in this case. Hence,

$$AVC = 100 - 3Q + 0.1Q^2$$

2. $MC = 100 - 6Q + 0.3\ Q^2$

3/4. The minimum point of a curve is found by setting its derivate equal to zero:

$$\frac{d(AVC)}{dQ} = -3 + 0.2Q = 0$$

$$\frac{d(MC)}{dQ} = -6 + 0.6Q = 0$$

Therefore, AVC reaches a minimum at Q = (3/0.2) = 15, and MC at Q = (6/0.6) = 10.

2.3 Fixed and Sunk Costs

We now make an important distinction between fixed costs and sunk costs. Recall that a fixed cost is a cost that does not change when output changes. A related concept, called *sunk* cost, is a cost that is lost forever once it has been paid. To be concrete, imagine that you are the manager of a local air commuter company and have paid $100,000 to lease a 70-seater plane for six months. This expense reflects a fixed cost to your firm. The cost is $100,000 regardless of whether you use the plane to carry passenger, use it half empty or use it fully loaded. How much of this $100,000 is a sunk cost depends on the terms of your lease. If the lease does not permit you to recoup any of the $100,000 if you should decide to break your lease in the interim, the entire $100,000 is a sunk cost. You have already incurred the cost, and there is nothing you can do to change it. If the lease states that you will be refunded 50% in the event you do choose to break your lease within the

first month, then only $50,000 of the $100,000 in fixed costs is a sunk cost. Sunk costs are thus the amount of these fixed costs that cannot be recouped.

Learning Tip

Since sunk costs are lost forever once they have been paid, they are irrelevant to decision making. A decision maker should ignore sunk costs to maximize profits or minimize losses. However, notice that, while sunk costs are irrelevant in making your decision, they do affect your calculation of total profits.

2.4 The Short-Run Supply Decision

In the short run the firm must incur its fixed costs, since it cannot, by definition, change the input of these factors of production. The firm's variable costs, on the other hand, are discretionary, in the sense that the firm may decide whether or not to incur these costs, since the input of variable factors can be varied all the way back to zero. What induces a firm to incur costs? A firm will set up a plant and incur fixed and variable costs to produce a particular product because it expects to make a profit.

Once committed to its present size of plant, however, the fixed costs of that plant are obligatory in the short run. That is, the firm must incur them because it cannot vary them in the short run, by definition. But variable costs are *discretionary* and should be incurred only if the firm expects revenues at least to cover these variable costs. Suppose that in a particular situation price exceeds average variable cost but is less than the average total cost, meaning that the firm is making a loss. If the firm continues production when P is greater than AVC but less than ATC, there is, for each unit sold, some excess of revenue over variable costs that contributes toward the fixed costs. Therefore, it reduces its losses by continuing to produce when ATC > P > AVC, rather than ceasing production. If it were to cease production, its loss would equal the entire costs, rather than just part of fixed costs.

Learning Tip

The firm should incur variable costs and hence supply the product market whenever price exceeds the firm's average variable cost level. If price exceeds the ATC level, the firm makes a profit. If price falls below ATC, the firm makes a loss, but it can minimize its loss by staying in production as long as P > AVC. If price slips below AVC, the firm would minimize losses by ceasing production and until either the price rises above AVC again or until it can liquidate its fixed costs or terminate the associated fixed costs. The latter option implies a long-run situation, of course.

3 Long Run Costs

As noted earlier, cost behaviour in the short run is an *operating* concept. At any given time, now or in the future, the firm will operate an existing plant that incurs a set of fixed costs. Operational decisions making is, therefore, concerned with minimizing variable costs. In contrast, cost behaviour in the long run is a *planning* concept. It is based on the promise that all inputs can be varied and that for each possible output level, there is a least-cost combination of inputs. Therefore, it is possible to construct a plant of optimum size for any desired level of production.

Because of the flexibility afforded in the long run, the firm must begin the planning process by first establishing the technology and production methods that are or will be available for the firm's expansion. The firm is then free to choose the size of the plant.

In Block 5 we defined the long run as the time period during which all inputs are variable. Thus, all costs are variable in the long run (i.e., the firm faces no fixed costs). The length of time of the long run depends on the industry. In some service industries, such as dry-cleaning, the period of the long run may be only a few months or weeks. For others that are very capital intensive, such as the construction of a new electricity-generating plant, it may be many years. It all depends on the length of time required for the firm to be able to vary all inputs.

The firm's long-run total cost (LTC) curve is derived from the firm's expansion path and shows the minimum long-run total costs of producing various levels of output. The firm's long-run average and marginal cost curves can then derived from the long-run total cost curve, below.

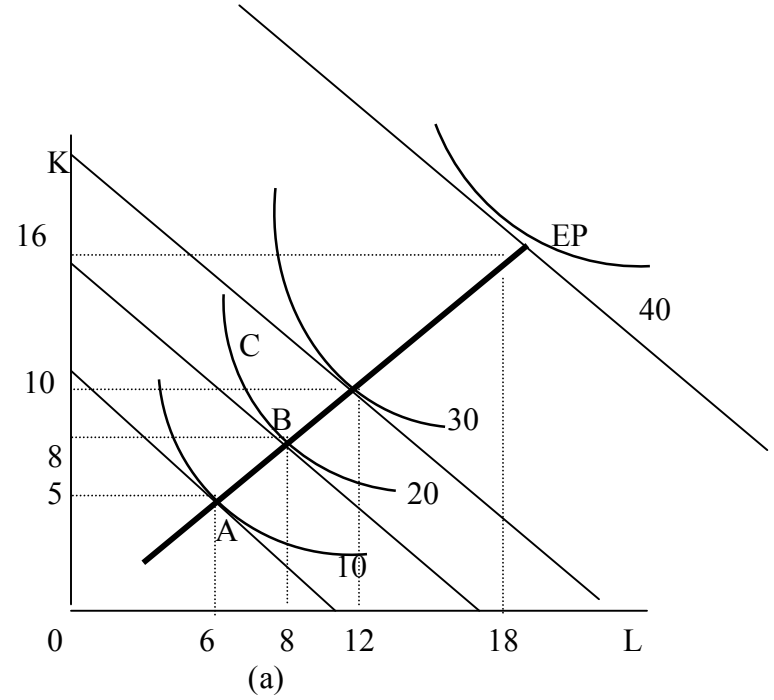
Panel (a) of Figure 6-3 shows the expansion path of the firm. As explained in Block 5, the expansion path shows the optimal input combinations to produce various levels of output. For example, point A shows that in order to produce 10 units of output (10Q), the firm uses 6 units of labour (6L) and 5 units of capital (5K). If the wage of labour (P_L) is \$5 per unit and the rental price of capital (P_K) is also \$5 per unit, the minimum total cost of producing 10Q is

$$TC = P_L \times L + P_K \times K$$

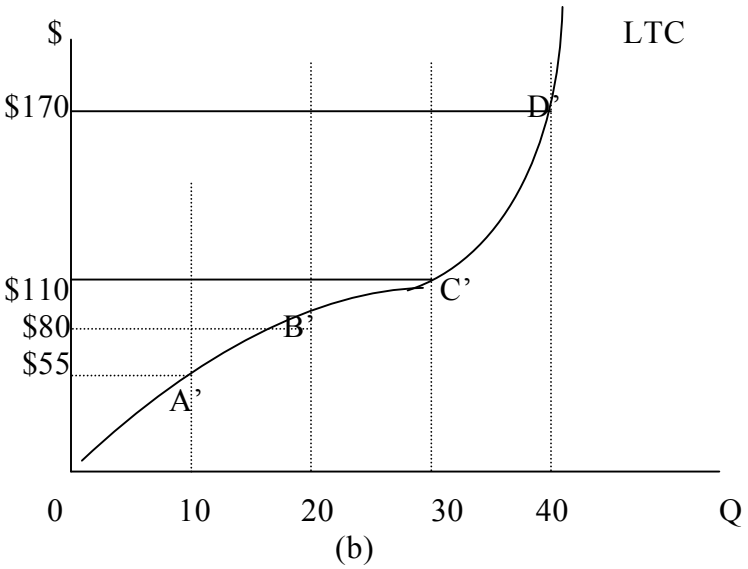
$$TC = (\$5)\ (6L) + (\$5)\ (5K) = \$55.$$

This is shown as point A' in panel (b), where the vertical axis measures total costs and the horizontal axis measures output. From point B on the expansion path in panel (a), we get point B' (\$80) on LTC curve in panel (b). Other points on the LTC curve are obtained in a similar fashion.

Figure 6-3



(a)



(b)

Learning Tip

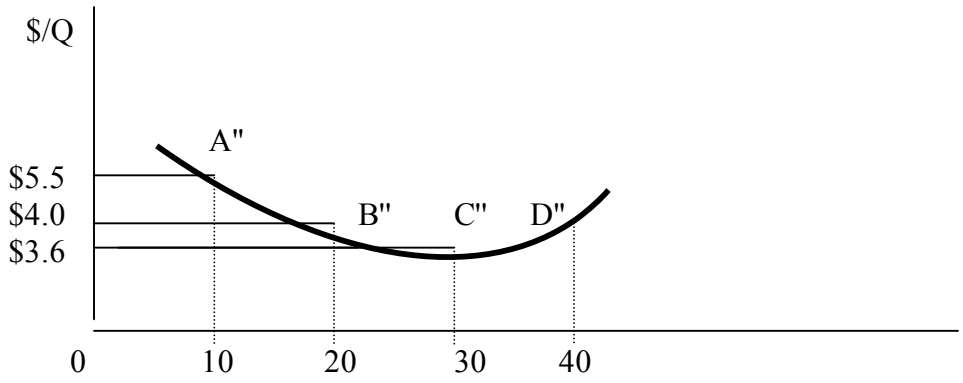
Note that the LTC curve starts from the point of origin because there are no fixed costs in the long run.

From the LTC curve we can derive the firm's long-run average cost (LAC) curve. LAC is equal to LTC divided by Q. That is,

$$LAC = \frac{LTC}{Q} \tag{5}$$

For example, the LAC to produce 10Q is obtained by dividing the LTC of $55 (point A' on the LTC curve in panel (b) of Figure 6-3) by l0. This is the slope of a ray from the origin to point A' on the LTC curve and is plotted as point A" in Figure 6-4 below. Other points on the LAC curve are similarly obtained. Note that the slope of a ray from the origin to the LTC curve, in Figure 6-3, declines up to point C' and then rises. Thus, the LAC curve, in Figure 6-4, declines up to point C" (4Q) and rises thereafter.

Figure 6-4



From the LTC curve we can also derive the long-run marginal cost (LMC) curve. This measures the change in LTC per unit change in output,

$$LMC = \frac{\Delta LTC}{\Delta Q} \tag{6}$$

The *long-run average cost* (LAC) curve shows the least cost of production for each level when all inputs may be varied. It is the locus of points from various ATC curves which allow each output level to be produced at lowest cost, given the ability to change plant size (that is, to vary continuously the input of capital).

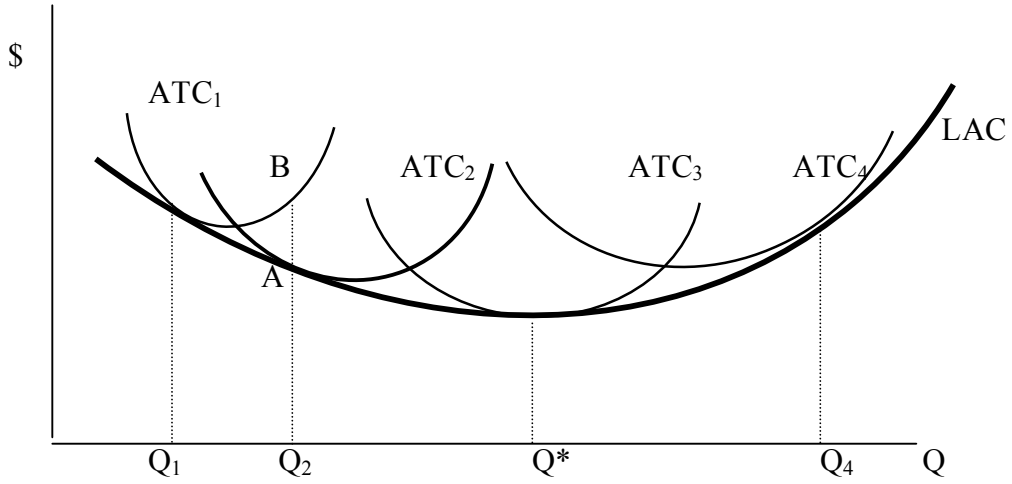
To show this we can proceed by finding a series of ATC curves, one for every level of fixed inputs. Each level of capital (fixed factor) input will give rise to a TP curve, from which we can derive a TVC curve and ultimately obtain the appropriate ATC curve, as we did earlier. This procedure would give us a series of ATC curves, each with a slightly larger capital input level as we move from left to right. As shown in Figure 6-5, the long-run average cost (LAC) curve is the 'envelope curve' of all these short-run curves. A corollary of this is that for any point on the LAC curve there is an ATC curve lying tangent at that point.

The U shape of LAC curve implies that expansion of plant size – each ATC representing a different size with ATC_4 associated with the largest and ATC_1 the smallest plant – results in lower and lower per unit costs until the optimum plant size is reached. This occurs at Q* units of output, corresponding to plant 3, where the LAC curve is at its lowest point. Thereafter, further increase in production will require successively larger plants and successively higher unit costs.

As shown in Figure 6-4, LAC is tangent to only one point on each ATC curve. At the optimal plant size, the tangency occurs at the lowest point on both the short-run and long-run average cost curves, shown by ATC_3. For all other plant sizes, the tangency point occurs: (1) to the left of the minimum-cost point on all short-run curves that are to the left of the optimum curve (ATC_3) and (2) to the right of the minimum-cost point on all short-run curves that are to the right of the optimum curve. Therefore, for outputs less than Q*, it is more economical to operate at less than capacity (minimum point of the short run curve), underutilizing a plant that is slightly larger than necessary. Therefore, it would be cheaper to produce output Q_2 with a plant designated by ATC_2 than with one represented by ATC_1.

Conversely, at outputs beyond the optimum level Q*, it is more economical to operate at a somewhat higher level than capacity, overutilizing a plant with a capacity that is slightly smaller than the required production. Therefore, it is cheaper to produce Q_4 with plant ATC_4 than with ATC_3.

Figure 6-5



Learning Tip

It is important to keep in mind that while the U shape of the short-run average cost (ATC) curve is based on the operation of the law of diminishing returns (resulting from the existence of fixed inputs in the short run), the U shape of the LAC curve depends on increasing, constant, and decreasing returns to scale, respectively, as will be explained below.

Learning Tip

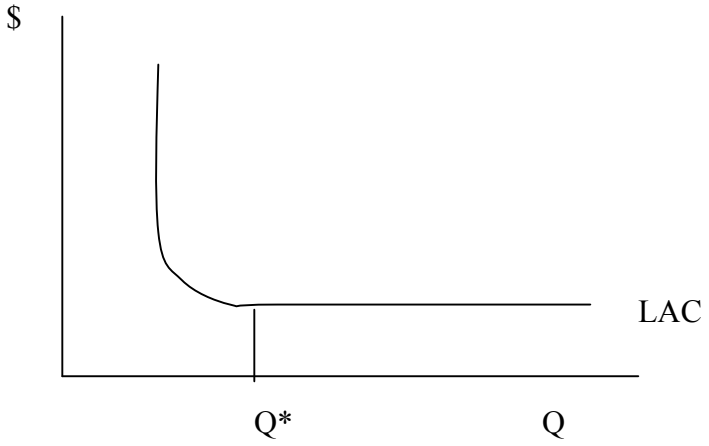
The LAC curve may not be a smooth line for many firms in the real world, since the smoothness depends on the availability of a large number of different plant sizes, with the sizes of these plants varying by relatively small increments from the smallest available to the largest available. In many industries there are only a few alternative sizes of plant, each one significantly different in size from the next largest and the next smallest.

3.1 Minimum Efficient Scale

The shape of the long-run average cost curves of production plants is important not only because it governs cost-efficient plant size but also because it indicates the number of firms that will emerge in an industry. In Figure 6-6, we see that Q* is the level of output that first minimizes the long-run average cost of the plant. This is the minimum size or scale of the plant that is cost efficient, when cost efficient means lowest ATC. Hence we can refer to this plant as the *minimum efficient scale* (MES).

There is evidence that in many industries, the LAC faced by a firm curve may not be U-shaped, as portrayed above, instead it may be L-shaped, as shown in Figure 6-6. Here, once the minimum LAC is reached at a plant with the capacity to produce Q*, successively larger plants can maintain that same minimum LAC. This means that there is a wide range of plant sizes in which there is no particular cost advantage to any one size.

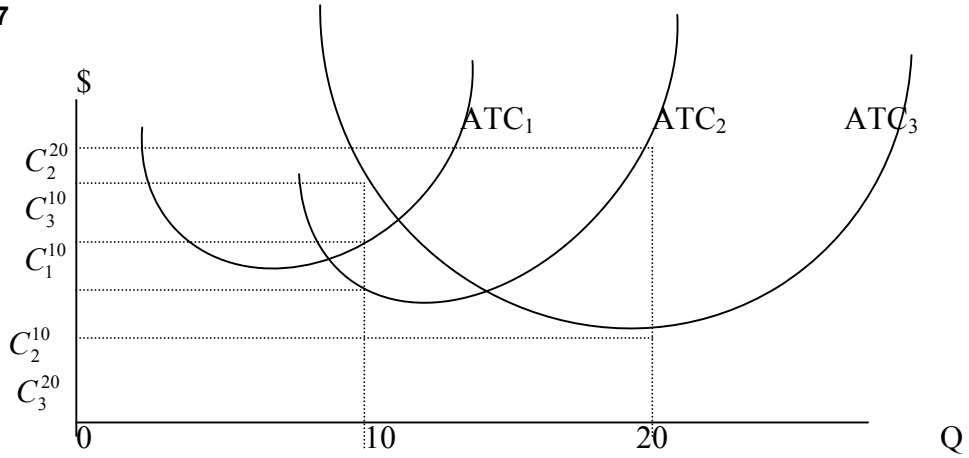
Figure 6-6



3.1.1 Demonstration Problem

Suppose that there are three short-run average cost curves as shown in Figure 6-7. Each curve is associated with a different plant size. Which plant can produce 10 units of output most efficiently? Which one can produce 20 units of output most efficiently?

Figure 6-7



Answer:

The smallest plant is ATC_1, the largest plant is ATC_3, and the medium-size plant is ATC_2. If the firm wished to produce 10 units of output, it would choose the medium-size plant at the per unit cost of C_2^{10}, where $C_2^{10} < C_1^{10} < C_3^{10}$. If the firm wishes to produce 20 units, the lowest per unit cost would belong to plant 3, C_3^{20}.

3.2 Economies and Diseconomies of Scale

The term *economies of scale* is defined as the decrease in the unit cost of production as a firm increases all its inputs of production. This phenomenon is illustrated in Figure 6-5, above.

In this figure, we saw that the average total cost curve, labelled 'Plant 1,' represented a certain amount of capacity. 'Plant 2' represented a greater production capacity because it is positioned to the right of Plant 1. In addition, it is located on a lower level than Plant 1, signifying that over a certain range of output, the larger plant is able to produce greater amounts of output at a lower average cost than the smaller one. 'Plant 3' represented still a larger plant than plant 2, and the lowest cost attained at Q*. This implies that between 0 and Q*, an expansion of output allows the firm to produce at a lower long-run average cost. When there are economies of scale, increasing the size of the operation decreases the minimum average cost.

Sometimes 'economies of scale' is used interchangeably with the term *increasing returns to scale.* Increasing returns to scale is a long-term phenomenon indicating that the firm's output grows at a rate that is faster than the growth rates of its inputs. For example, a 100 per cent increase in inputs results in more than 100 per cent increase in output, say 200 per cent. In this case, as the firm expands, the per-unit cost drops. The following are the main causes of economies of scale.

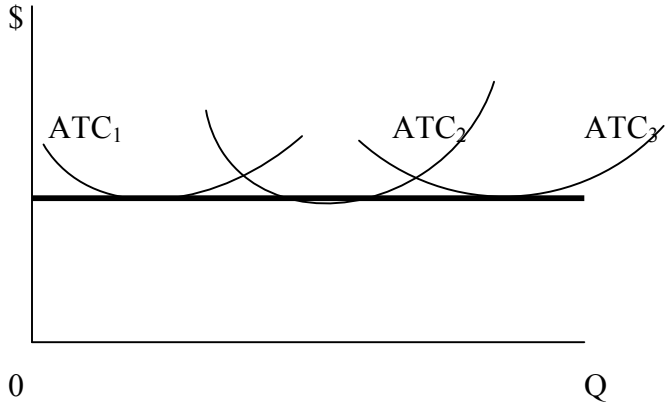
Learning Tip

With economies of scale, output rises more rapidly than the total cost of inputs, so that average cost falls as the scale of production expands.

With the final plant expansion, from ATC_3 to ATC_4, the company's short-run average cost curve not only shifts to the right but also up. This shift reflects *diseconomies of scale*. Since the plant's output is rising less rapidly than the total cost of input costs, the average cost curve rises as the production of automobiles continues to increase.

Constant returns to scale usually result when making more of an item requires repeating exactly the same tasks used to produce previous units of output. In objective terms, constant returns to scale occur in a business that expands inputs a given percentage and sees output rise by the same percentage. In terms of Figure 6-8, this implies that the technology in an industry allows a firm to produce different levels of output at the same minimum average cost.

Figure 6-8



3.3 Reasons for Economies of Scale

Economies of scale, also synonymously used as increasing returns to scale or decreasing costs, arise because of technological and financial reasons. At the technological level, economies of scale arise because as the scale of operation increases, a greater division of labour and specialization can take place, and more specialized and productive machinery can be used. Specifically, with a large-scale operation, each worker can be assigned to perform a repetitive task rather than numerous different ones. This results in increased proficiency and the avoidance of the time lost in moving from one machine to another. At higher scales of operation, more specialized and productive machinery can also be used. For example, using a conveyor belt to unload a small truck may not be justified, but it greatly increases efficiency in unloading a whole train or ship. Furthermore, some physical properties of equipment and machinery also lead to increasing returns to scale. For example, doubling the diameter of a pipeline more than doubles the flow without doubling costs, doubling the weight of a ship more than doubles its capacity to transport cargo without doubling costs, and so on. Thus, per-unit costs decline. Firms also need fewer supervisors, fewer spare parts, and smaller inventories per unit of output as the scale of operation increases.

Besides the above technological reasons for increasing returns to scale or decreasing costs, there are financial reasons that arise as the size of the firm increases. Because of bulk purchases, larger firms are more likely to receive quantity discounts in purchasing raw materials and other intermediate (i.e., semi-processed) inputs than smaller firms. Large firms can usually sell bonds and stocks more favourably and receive bank loans at lower interest rates than smaller firms. Large firms can also achieve economies of scale or decreasing costs in advertising and other promotional efforts. For all these technological and financial reasons, the LAC curve of a firm is likely to decline as the firm expands and becomes larger. Furthermore, the indivisible nature of many types of capital equipment may also result in economies of scale. In situations like this, the average cost (cost per unit) gets to decline as the volume of output produced increases.

3.4 Reasons for Diseconomies of Scale

Diseconomy of scale, also used synonymously as decreasing returns, or increasing costs, arises primarily because as the scale of operation increases, it becomes ever more difficult to manage the firm effectively and coordinate the various operations and divisions of the firm. The number of meetings, paperwork, and telephone bills, increase more than proportionately to the increase in the scale of operation and it becomes increasingly difficult for top management to ensure their directives and guidelines are properly carried out by their subordinates. Thus, efficiency decreases and costs per unit tend to rise. Other reasons for diseconomy of scale are disproportionate rise in transportation costs and input market imperfections (e.g., wage rates driven up).

3.5 Economies of Scope

Up until now, our analysis of the production process has focussed on situations where the firm produces a single output. There are also numerous examples of firms that produce multiple outputs. Honda Motors Co. produces both cars and motorcycles (and many

varieties of each); Hewlett Packard produces many different types of computers and printers. While our analysis for the case of a firm that produces a single output also applies to a multi-product firm, the latter raises some additional issues. This section will highlight these concepts.

This way, it is possible for managers to identify cost-saving phenomenon known as *economies of scope*. This occurs when it is possible to produce two or more products together at a lower per-unit cost than for each product separately. A key factor in this form of cost savings is the sharing of a company's fixed cost by multiple products.

Another way that a company can utilize economies of scope is to produce goods or services that require similar skills and experience. For example, when PepsiCo expanded into the snack and fast-food business, it was able to utilize its background in one type of fast-moving consumer item (soft drinks) to another (chips, tacos, and fried chicken). The product development, channels of distribution, and marketing know-how are very similar in these two product groups.

In this section, we will assume that the cost function for a multi-product (in this case a two-product) firm is given by $TC(Q_1, Q_2)$, where Q_1, is the number of units produced of product 1 and Q_2, is the number of units produced of product 2. Therefore, the *multi-product cost function,* which itself is derived from a *multi-product production function*, assumes technical efficiency.

In general, economies of scope are present when the joint output of a single firm is greater than the output that can be achieved by two different firms each producing a single product (with equivalent production inputs allocated between the two firms).

Economies of scope is present when

$$TC(Q_1,0) + TC(0,Q_2) > TC(Q_1,Q_2) \tag{7}$$

Assume that the general expression for a multi-product function is captured by a quadratic function as follows:

$$TC(Q_{1,}Q_2) = a + bQ_1Q_2 + Q_1^2 + Q_2^2 \tag{8}$$

where the joint cost of producing both outputs is captured by (bQ_1Q_2). To determine if the economies of scope is present, we first plug in equation (8) to obtain $TC(Q_{1,}0) = a + Q_1^2$ and $TC(0,Q_2) = a + Q_2^2$, and then plug in the results in equation (7):

$$a + Q_1^2 + a + Q_2^2 > a + bQ_1Q_2 + Q_1^2 + Q_2^2$$

or

$$a - bQ_1Q_2 > 0 \tag{9}$$

Therefore, given that a, Q_1, and Q_2 are all positive – a being the fixed cost and Q_1 and Q_2 being the level of production – economies of scope are present if $b < 0$.

3.5.1 Demonstration Problem

Does this production function exhibit economies of scale?

$$TC(Q_1, Q_2) = 175 \square 0.55Q_1Q_2 + 0.2Q_1^2 + 0.2Q_2^2$$

Answer:

Yes, because the coefficient to Q_1Q_2, which is (– 0.55), is negative and therefore, $a + b\ Q_1Q_2 = 175 + 0.55\ Q_1Q_2 > 0$.

4 The Nature of Costs

To determine what a cost is, you must begin with the firm's objective. Let's find the seemingly obvious answer to what costs are by focusing on your diapers business. It is conceivable that you started your firm because of an altruistic desire to provide nearby families of infants with diapers. More likely, however, you started your business to make money. Economists normally assume that the goal of a firm is to maximize profit, and they find that this assumption works well in most cases.

What is a firm's profit? The amount that the firm receives for the sale of its output (diapers) is called its total revenue. The amount that your firm pays to buy inputs (fabric, absorbent filler, workers, sewing machines, etc.) is called its total cost. You get to keep any revenue that is not needed to cover costs. We define profit as a firm's total revenue minus its total cost. That is,

$$\text{Profit} = \text{Total revenue} - \text{Total cost} \qquad (10)$$

Your objective is to make your firm's profit as large as possible. To see how a firm goes about maximizing profit, we must consider fully how to measure its total revenue and its total cost. Total revenue is the easy part: it equals the quantity of output the firm produces times the price at which it sells its output. The measurement of your firm's total cost, however, is more challenging.

4.1 Economic Versus Accounting Costs

The data for decision making with respect to costs typically come not from economists but from accountants. In most cases these data are adequate and appropriate, but in some cases, since they were derived for different purposes, they are less suitable for direct insertion into economic decision-making procedures. We shall examine several different economic and accounting cost concepts and the relationships between them.

4.2 Direct and Indirect Costs

In the business firm some costs are incurred that can be directly attributed to the production of a particular unit of a given product. The use of raw materials, labour inputs, and machine time involved in the production of each unit can usually be determined. On the other hand, the cost of fuel for heating, electricity, office and administrative expenses, depreciation of plant and buildings, and other items cannot easily and accurately be separated and attributed to individual units of production (except on an arbitrary basis). Accountants speak of the *direct* costs per unit when referring to the separable costs of the first category and of *indirect*, or *overhead*, costs when referring to the joint costs of the second category.

Note that direct and indirect costs do not coincide exactly with the economist's variable cost and fixed cost categories. The criterion used by the economist to divide cost into either fixed or variable is whether or not the cost varies with the level of output, whereas the criterion used by the accountant is whether or not the cost is separable with respect to the production of individual output units. To bring the accounting costs into line with the economic concepts, we must find that part of the indirect or overhead costs that varies with the output level.

4.3 Explicit and Implicit Costs

The accounting process is predominantly concerned with explicit costs. These are costs that actually involve transfer of funds from the firm to another party that had previously supplied some materials or services. These are out-of-pocket expenses in the current time period, since they are an actual cash outflow in payment for resources. Other cost items, however, are implicit costs, in the sense that they do not involve an actual cash outflow in the current time period.

An easy way of remembering the difference between explicit costs and implicit costs is as follows. *Explicit costs* are generally associated with factors that are not owned by the firm. The opportunity cost of those factors, which are not owned by a firm, is simply the price that firm has to pay for them. Each cost requires direct payment of money by firms. Implicit costs, by contrast, are associated with factors that are owned by the firm. When a firm owns machinery, for example, it does not normally have to pay out money to use that machinery. *Implicit costs* are equal to what the factors could earn for the firm in some alternative use, either within the firm or rented out to some other firm.

4.3.1 Demonstration Problem

Suppose you give up the opportunity to earn money as a manager of a local posh restaurant in favour of running your own little pizzeria. In preparing an income statement, how would your accountant take your business decision into consideration differently from that prepared by an economist?

Answer:

Your accountant would not count this as a cost of your pizzeria. Because no money flows out of the business to pay for this cost, it never shows up on the accountant's financial statements. An economist, however, would count the forgone income as a cost because it would affect the decisions that you make in your pizzeria business. For example, if your wage as a manager of the local posh restaurant rises from $200 to $300 per day, you might decide that running your own business is too costly and choose to shut down the factory in order to become a full-time manager.

4.4 Opportunity Costs and Accounting Costs

As indicated before, there is a fundamental difference between accounting (absolute) cost and opportunity (alternative) cost. *Accounting cost* is the historical outlay of funds for wages and salaries, raw materials, rent, utilities, interest, and so forth. Accounting costs also include estimated periodic reductions in asset valuations, such as depreciation, amortization, and depletion expense.

Opportunity cost is the cost of forgoing certain opportunities or alternatives in favour of pursuing others. Opportunity cost exists because all resources are scarce. A resource that is used for one purpose cannot be used for something else at the same time. For example, consider a firm that chooses to invest $100,000 in a building instead of buying a $100,000 government Treasury Bill. The interest that could have been earned on the Bills is a forgone opportunity and therefore an opportunity cost. In the firm's cost analysis, this opportunity cost should be added to the building's explicit cost of $100,000.

4.5 Economic versus Accounting Profits

When most people hear the word profit, they think of accounting profits. *Accounting profit* is the total amount of money taken in from sales (total revenue, or price times quantity sold) minus the dollar cost of producing goods or services. Accounting profits are what show up on the firm's income statement and are typically reported to the manager by the firm's accounting department.

A more general way to define profits is in terms of what economists refer to as economic profits. *Economic profits* are the difference between the total revenue and the total opportunity cost of producing the firm's goods or services. The opportunity cost of using a resource includes both the explicit (or accounting) cost of the resource and the implicit cost of giving up the next-best alternative use of the resource. The opportunity cost of producing a good or service generally is higher than accounting costs because it includes both the dollar value of costs (explicit, or accounting, costs) and any implicit costs.

Implicit costs are very hard to measure and therefore are often overlooked by managers. Effective managers, however, continually seek out data from other sources to identify and

quantify implicit costs. Managers of large firms can use sources within the company, including the firm's finance, marketing, and/or legal departments, to obtain data about the implicit costs of decisions. In other instances managers must collect data on their own.

Learning Tip

The fact that Accounting profit is different from Economic profit is not to say that either the accountant's or economist's view of profit is incorrect; each is designed for a different purpose. The accountant's purpose is to find, once the capital has been invested in a particular pursuit, the return to the owners of that capital. The economist's purpose is to ensure that all resources are employed in their most efficient uses. The existence of economic profit confirms this.

4.5.1 Demonstration Problem

What does it cost you to read a book?

Answer:

The answer is that the price you paid the bookstore for a book is an explicit (or accounting) cost, while the implicit cost is the value of what you are giving up by reading the book. You could be studying some other subject or watching TV, and each of these alternatives has some value to you. These alternatives constitute your implicit cost of reading this book; you are giving up this alternative to read a book. The opportunity cost of reading a book is the value of the 'next best' alternative.

4.5.2 Demonstration Problem

Suppose you own a building in New York that you use to run a small pizzeria. Food supplies are your only accounting costs. At the end of the year, your accountant informs you these costs were $20,000 and that your revenues were $100,000. What is your accounting profit, and how different would that be from your economic profit?

Answer:

Your accountant would report to you that your profits are $80,000.

However, these accounting profits overstate your economic profits, because the costs include only accounting costs. First, the costs do not include the time you spent running the business. Had you not run the business, you could have worked for someone else, and this fact would be an economic cost not accounted for in accounting profits. To be concrete, suppose you could have worked for someone

for $30,000. Your opportunity cost of time would have been $30,000 for the year. Thus, $30,000 of your accounting profits are not profits at all but one of the implicit costs of running the pizzeria.

Second, accounting costs do not account for the fact that, had you not run the pizzeria, you could have rented the building to someone else. If the rental the building is $100,000 per year, you gave up this amount to run your own business. Thus, the costs of running the pizzeria include not only the costs of supplies ($20,000) but the $30,000 you could have earned in some other business plus $100,000 you could have earned in renting the building to someone else. The economic cost of running the pizzeria is $150,000 – the amount you gave up for your business. Considering the revenue of $100,000, you actually lost $50,000 by running the pizzeria.

4.6 Normal and Pure Profits

Normal profits are earned when total revenues equal total costs, if total costs are calculated to reflect the opportunity costs of *all* services provided. If revenues just equal these costs, then all factors are earning the same in that particular employment as they could earn elsewhere. If revenues exceed these costs, we say that the firm is earning a *pure*, or economic, profit. Remembering that the owners of the firm are the effective suppliers of the services of the land and buildings mentioned, you will see that an economic profit means that the owners of the firm are earning more profit than they could by investing their capital elsewhere. The accounting profit must be adjusted for the opportunity cost of the owned resources – that is, for what the firm would pay for the services of those resources if they were purchased or hired – before the alternative investment possibilities can be assessed. Accounting profit will exceed economic profit if some implicit opportunity costs are not subtracted from revenues.

Learning Tip

A normal profit (when TR = TC) does not mean *no* profit. Since total costs in the economic sense include the opportunity cost of all resources used, the return on capital invested is included as a cost, rather than counted as a residual in the accounting sense. Normal profit means a *sufficient return* on the owner's investment in the firm, sufficient to prevent him or her from liquidating this investment and investing it in the next best alternative investment, since the return on the next best alternative investment opportunity is included as an economic cost of production. Normal profit, therefore, means as much profit as the owners could get elsewhere.

4.7 Incremental Cost Analysis

Incremental costs are those costs that will be incurred as the result of a decision. Incremental costs are measured by the change in total costs that results from a particular decision. Incremental costs may therefore be either fixed or variable, since a new decision may require purchase of additional capital facilities plus extra labour and materials.

Learning Tip

Note that incremental costs are not identical with marginal costs. As defined above, marginal costs are the change in total cost for a one-unit change in the output level. Incremental costs, on the other hand, are the aggregate change in costs that results from a decision. This decision may involve a change in the output level of 20 or of 2,000 units, or it may not involve a change in the output level at all. For example, the decision may be whether or not to introduce a new technology of producing the same output level. Knowledge of marginal costs, however, may be very important for the calculation of the incremental costs.

4.8 Relevant Costs and Irrelevant Costs

The *relevant costs* for decision-making purposes are those costs that will be incurred as a result of the decision being considered. The relevant costs are, therefore, the incremental costs. Put differently, when a decision has to be made in which cost is a factor, only those costs that will change as a result of the decision are relevant. On the other hand, costs that have been incurred already and costs that will be incurred in the future regardless of the present decision are *irrelevant* costs as far as the current decision problem is concerned. Costs that do not change as a result of the decision are called sunk costs. Since they are not affected by the decision, sunk costs are irrelevant.

For example, a decision to add a second shift workers would increase relevant variable costs, such as costs of wages, materials, and supplies, but it would not change the cost already sunk into land, buildings, machinery and other fixed assets used in production.

Learning Tip

It must be emphasized that incremental cost is a *short-run* decision-making concept based on the existence of sunk costs that can be ignored as irrelevant. In the long run all costs must be recovered or the business will fail.

4.8.1 Demonstration Problem

Laila, an accounting major at Midtown College, is on a tight budget, so she has kept careful records on the cost of driving her second-hand automobile. Her records show that when depreciation, interest on her investment (opportunity cost), license fees, parking fee, and insurance are added to her direct operating costs for gas, oil, tires, and maintenance the total cost comes to $0.50 per kilometre if she drives 10,000 kilometres per year.

Laila wants to drive home for summer holidays, which is a 1000-kilometer round trip. She can't afford the trip unless she takes other students along to share the cost. She notes that the total cost of the trip at $0.50 per mile would be 1,000 x $.50 = $500. This would mean that the pro rata cost for herself and four riders would be $100 each, a sum that neither she nor her equally penniless classmates can afford.

Fortunately, she learned about incremental costs during a course in cost accounting and recognizes that her problem is one of short-run decision making. Her objective is merely to cover the added operating cost of making a 1000-kilometer round trip for which only the direct operating costs for gas, oil, tires, and maintenance are relevant. According to Laila's record, these costs amount to 12 cents ($0.12) per kilometre.

Answer:

Thus 1,000 kilometres of driving will increase Laila's total cost by 1,000 x $.12 = $120, not $500. If Laila can find four riders willing to pay $30 each, she will get her own transportation for free.

5 Summary and Review

In the short run, we have fixed and variable cost costs. The shape of the total variable cost curve follows from the law of diminishing returns. The long-run total cost curve is derived from the expansion path and shows the minimum total cost of producing various levels of output when the firm can build any desired scale of plant. A U-shaped long-run average cost curve is based on the assumption that economies of scale prevail at small levels of output and diseconomies of scale prevail at larger levels of output.

The firm's choice of plant size depends on its desired level of output. If demand is predictable with certainty, the plant size chosen is the one that allows the production of the quantity demanded at minimum average cost. If demand is uncertain, the plant size chosen is one that minimizes the expected value of average cost, given the probability distribution formed by the meshing of the demand probabilities and the known cost curves at each output level. Choosing plant size on this basis typically means that the firm

builds in some excess capacity, which benefits the firm in case of market growth. In this discussion, we also distinguished the relevant from the irrelevant costs, and, economic profit from accounting profits. Finally, we studied the notion of economies of scope.

6 Review Problems

1. The manager of a plant calculated the cost at different output level. The result is in the table below:

Output	Total Cost	Fixed Cost
0	100	100
1	200	100
2	290	100
3	370	100
4	440	100
5	500	100
6	550	100
7	600	100
8	660	100
9	730	100
10	810	100

   Find the variable cost, average total cost, and marginal cost for the different levels of output.

2. The cost function of a firm is as follows:

   $TC(Q) = 100Q^3 – 30Q^2 + 10Q + 500$

   Find the marginal cost, average cost, average variable cost functions. If the firm produces 10 units of product then what will be the marginal cost, average cost, and total cost?

3. The production manager of a computer manufacturing company finds that the cost of production follows the function:

   $TC(Q) = 500 + 50Q – 10Q^2 + 25Q^3$

Find the average cost and the marginal cost functions. At which level of output will the marginal cost be minimum? Find the average total cost and the marginal cost at this level of output.

4. A company's long run cost function is as follows:

   $C(Q) = 100,000 + 25,000Q$

   Is the long run marginal cost constant? Does long run marginal cost depend on output level? If not, then will the average cost of producing 500 units and 1,000 units be the same? Does economies of scale exist in the long run?

5. Cost of labour is $100 per unit and capital is $200 per unit. If the marginal product of labour is 1,000 and marginal product of capital of is 2,500, then is it optimal to use the combination of 10 labour and 30 capital? If not, then how should the owner revise the combination?

6. A newly floated company requires an initial investment outlay of $1 million in land and machinery. Should this cost be considered as sunk cost? Explain why or why not? How is sunk cost relevant to profit maximization?

7. A fast food outlet sells sandwiches (Q_1) and chips (Q_2). The manager estimates that the cost of producing Q and Q exhibits a function:

   $C(Q,Q) = 200 + 5Q + 10Q – 2QQ$

   Will the outlet be more cost efficient if it were to specialize in any of the two products? Is there any cost-complementarity in production?

7 Answer Key to Review Problems

1. The manager of a plant calculated the cost at different output level.

Output	Total Cost	Fixed Cost	Variable Cost	Average Total Cost	Marginal Cost
0	100	100	0	–	
1	200	100	100	200.0	100.0
2	290	100	190	145.0	45.0
3	370	100	270	123.3	26.7
4	440	100	340	110.0	17.5
5	500	100	400	100.0	12.0
6	550	100	450	91.7	8.3
7	600	100	500	85.7	7.1
8	660	100	560	82.5	7.5
9	730	100	630	81.1	7.8
10	810	100	710	81.0	8.0

2. Total Cost = $100Q^3 – 30Q^2 + 10Q + 500$

   Marginal Cost = $300Q^2 – 60Q + 10$

   Average Cost = $100Q^2 – 30Q + 10 + 500/Q$

   At Q = 10, TC = $100(10)^3 – 30(10)^2 +10(30) + 500 = 97,600$

   MC = $300(10)^2 – 60(10) + 10 = 29,410$

   ATC = $100(10)^2 –30(10) + 10 + 500/10 = 9,760$

3. Average Cost = $500/Q + 50 – 10Q + 25Q^2$

   Marginal Cost = $50 – 10Q + 75Q^2$

   Marginal Cost will be minimum where the first derivative is zero.

$$\frac{dMC(Q)}{dQ} = 0 \Rightarrow 150Q - 10 = 0 \Rightarrow Q = 15$$

Marginal cost will be minimum when quantity is 15. Marginal cost at this point = $50 - 10(15) + 75(15)^2 = 16{,}775$. Average cost at this point = 5,558.33.

Average cost is minimum when $\frac{d\left[C(Q)/Q\right]}{dQ} = 0$

But, $\frac{d\left[C(Q)/Q\right]}{dQ} = \frac{MC(Q)}{Q} - \frac{C(Q)}{Q^2} = \frac{1}{Q}\left(MC(Q) - AC(Q)\right)$

The above expression will be zero only when MC(Q) = AC(Q). We have seen earlier that this is not true at the minimum MC point.

4. Long run marginal cost = 25,000 (constant). It does not depend on output.

   Average Cost = 100,000/Q + 25,000 (dependent on output). Therefore Average Cost will not be the same for different levels of output.

   Average Cost when Q = 500 is [100000/500 + 25000] = 25,200

   Average Cost when Q = 1000 is [100000/1000 + 25000] = 25,100

   No. Economies of scale do not exist in the long run since the marginal cost is constant in the long run.

5. Wage w = 100

   Rent r = 200

   Marginal Product of Labour (MPL) = 1,000

   Marginal Product of Capital (MPK) = 2,500

   MPK/r = 2,500/200 = 12.5 > MPL/w = 1,000/100 = 10

   Therefore, more capital (or less labour) needs to be employed until MPK/r = MPL/w.

6. Whether the total or part of the initial investment outlay is to be considered as sunk cost depends on how value can be salvaged by disposing the land and machinery. If we consider that there will be no salvage value or in other words one hundred percent depreciation then the total investment should be considered as sunk cost. Otherwise, there will always be some part that can be recovered. The amount that cannot be recovered is the 'sunk cost.'

7. $MC_1 = 5 - 2Q_1$

$MC_2 = 10 - 2Q_2$

Therefore cost complementarities exist, since increasing Q_2 will reduce MC of 1, and also increasing Q_1 will reduce MC of 2.

$C(Q_1,0) + C(0,Q_2) - C(Q_1,Q_2) = 200 + 5Q_1 + 200 + 10Q_2 - 200 - 5Q_1 - 10Q_2 + 2Q_1Q_2 = 200 + 2Q_1Q_2 > 0$

Therefore economies of scope exists, and thus specialization in one product will not be beneficial.

8 References and Suggested Further Reading

Anthony, R.N., 'What Should Cost Mean?' *Harvard Business Review*, May-June 1970.

McElroy, F. W. 'Return to Scale of Cost Curves: Comment,' *Southern Economic Journal*, 1970.

Besanko, D. and Braeutigam, R., *Microeconomics: An Integrated Approach*, Wiley, 2002

Baye, M., *Managerial Economics and Business Strategy*, McGraw Hill, Irwin, 2002.

Douglas, J. E., *Managerial Economics: Analysis and Strategy*, Prentice Hall, 1992.

Pindyck, R. and Rubinfeld, D., *Microeconomics***,** Fifth Edition, Prentice Hall, 2001.

Salvatore, D., *Managerial Economics in a Global Economy*, McGraw Hill, 1993.

Commonwealth of Learning Executive MBA/MPA

E5 Managerial Economics

Block 7

Nature of Industry

Course author: Farrokh Zandi, PhD
Professor, Schulich School of Business
York University
Toronto, Canada

Course designer: Douglas R. Cronk, M.A., ABD
Vancouver, BC, Canada

Copyright © Commonwealth of Learning, 2003
All rights reserved. No part of this course may be reproduced in any form by any means without prior permission in writing from:

The Commonwealth of Learning
1285 West Broadway
Suite 600
Vancouver, BC V6H 3X8
CANADA
e-mail: info@col.org

Contents

1 Block Seven: Objectives and Introduction 1
1.1 Introduction 1
2 Different Perspectives on Competition 1
2.1 Chamberlin's Resource-Based View 1
2.2 Schumpeter's Technology Based View 2
2.3 The Industrial Organization Model 2
3 Market Structure 3
3.1 Measures of Industry Concentration 4
3.1.1 Concentration Ratios 4
3.1.2 Demonstration Problem 5
3.1.3 Herfindahl-Hirschman Index 5
3.1.4 Demonstration Problem 6
3.1.5 Demonstration Problem 6
3.2 Limitations of Concentration Measures 7
4 Market Performance 9
4.1 Economic Profits or Rates of Return on Investment 9
4.2 The Lerner Index 9
4.2.1 Demonstration Problem 10
5 Summary and Review 10
6 Review Problems 11
7 Answer Key to Review Problems 12
8 References and Suggested Further Reading 13

1 Block Seven: Objectives and Introduction

After working through Block Seven of this course, you should be able to:

1. Explain various theories of competition.
2. Discuss the key elements of industrial organization's view of competition.
3. Distinguish alternative measures of industry concentration.
4. Analyze the elements of industry performance.

1.1 Introduction

General environmental analysis is useful for answering broad corporate strategy questions. It helps companies to choose the industries in which they should operate. Such general analysis, however, serves only as a starting point for strategy formulation. Formulating competitive strategies at the business unit level requires more detailed analysis of industry competition. Competitive strategies should be based on a detailed understanding of industry structure. The structure of competition and determinants of profitability in an industry are fundamental to formulating good competitive strategies. Note that they can help managers determine opportunities and threats facing their business. Then, competitive strategies can be formulated to exploit these opportunities and avoid threats.

The industry is a part of the general economic environment. It refers to a set of firms producing the same, or functionally similar, products or services and vying for the same customers. An industry can include both domestic and foreign firms. For example, the automobile industry includes firms such as GM, Ford, Toyota, Honda, etc.

2 Different Perspectives on Competition

Competition is a complex idea that can be studied from many different perspectives. The current comprehensive views of competition, known as Industrial Organization views, succeed the two traditional views of competition: Chamberlin's view, which is a resource based view, and Schumpeterian view, a technology based view.

2.1 Chamberlin's Resource-Based View

In the 1930s, Edward Chamberlin and his colleagues began systematic study of economic competition between firms. They viewed competition as a resource-based phenomenon. It represented attempts by firms to use their unique resources to create lasting performance benefits. The resources used by firms included finances, technology, patents, trademarks, brand awareness, marketing knowledge, and administration. For example, John D.

Rockefeller's Standard Oil (now Exxon) gained a competitive edge by buying up oil rights. It bought exploration and extraction leases all across the United States and abroad. Once it had captured this unique resource, it developed exploration, drilling, refining, and marketing facilities. Soon it became the largest American oil company.

Firms in every industry have some unique and some overlapping resources. These resources explain different types of market environments, which will be discussed in Block 8: monopolistic, oligopolistic, monopolistic competition, and perfect competition. Industries often evolve through these structures as firms respond to or create profit opportunities.

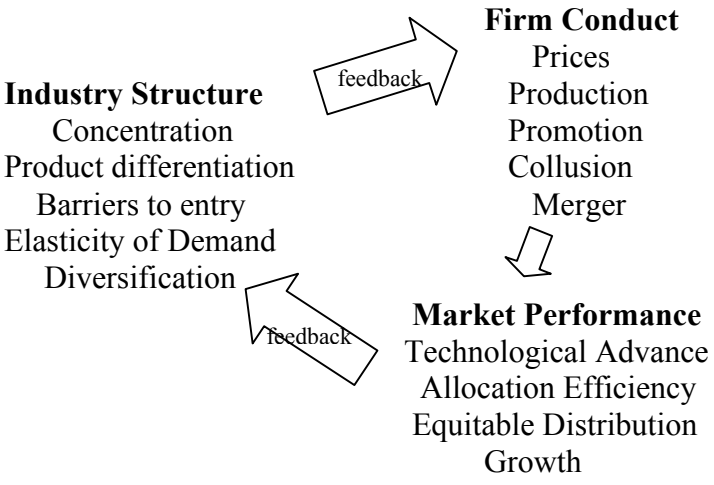
2.2 Schumpeter's Technology Based View

In contrast to the resource-centred view of competition was the work of Joseph Schumpeter in the late 1940s and early 1950s. In explaining the historical process of economic development in Western industrial economies, Schumpeter focused on the role of the entrepreneur in creating technological revolutions and product-market shifts. These factors, he argued, were at the heart of capitalistic competition and economic change. Revolutionary innovations in products, markets, and technologies gave firms temporary competitive advantages. These revolutions, which he coined as a 'creative destruction' process, could completely displace all currently competitive firms. Other factors, such as price, marketing strategies, and financial resources, were less important.

2.3 The Industrial Organization Model

Neither of the two views of competition introduced above, however, captures the complexity of today's global business environment. A third view of competition, shown in Figure 7-1, that combines and extends the elements of the two is the industrial organization model of competition. This theory is advanced by Bain in late 1950s.

Figure 7-1



The original model of industrial organization competition argued that financial returns in an industry are a function of two variables: the competitive structure of the industry and the conduct of firms within the industry. Several economic variables determine the structure of an industry. They include the number and relative size of competitors, degree of product differentiation, barriers to entry into the industry, and the elasticity of demand. These structural forces jointly determine the average level of industry profits and how the industry profits and how members behave, and in the past decade, it has served as the basis for examining competitive structures for strategy formulation.

3 Market Structure

How would you measure the market power of a firm or its exercise in a market? How would you measure the impact of different factors on the market power of a firm or its exercise in a market?

The traditional Structure-Conduct-Performance (SCP) approach relies on accounting data regarding profits and costs to measure market power. For many decades, economists have conducted *(SCP)* studies that concentrate on the second question, which concerns the relationship between market performance and market structure. Market *performance* is the success of a market in producing benefits for consumers. For example, a market is performing well if prices are near the marginal cost of production. Market structure affects market performance through the *conduct or behaviour* of the firms.

Market *structure* consists of those factors that determine the competitiveness of a market including factors such as the number of firms that compete in a market, the relative size of the firms, technological and cost conditions, demand conditions, and the ease with which firms can enter or exit the industry. Different industries have different structures, and those structures affect the decisions the prudent manager will make. The following subsections provide an overview of the major structural variables that affect managerial decisions.

The SCP approach assumes that there is a stable, *causal* relationship between the structure of an industry, firm conduct, and market performance. This stable relationship follows a line of causality that runs from structure through conduct to performance. Since conduct was thought to be difficult, if not impossible, to observe directly, the focus of the approach is on identifying structure variables that are observable and measurable and that are linked with market power. If a stable relationship is established between structural variables and market power, then the SCP implication is that this structural variable facilitates the exercise of market power. The basic idea is to establish relationships between structural variables and market performance that generalize, or hold, across industries. The aspect of market performance that has attracted almost exclusive interest is the exercise of market power. As discussed in Figure 7-1, the measures of market structure are industry concentration, demand elasticity, barriers to entry, diversification, and product differentiation.

3.1 Measures of Industry Concentration

In this section, we will focus on the key measure of market structure, i.e., *industry concentration*, that affects managerial decisions. The issue is one of the number and the size distribution of firms within an industry. That is, are there many small firms within an industry or only a few large firms? This question is important because, as we will see in later Blocks, the optimal decisions of a manager who faces little competition from other firms in the industry will differ from those of a manager who works in an industry in which there are many firms.

Here, we examine two popular measures of industry concentration.

3.1.1 Concentration Ratios

Concentration ratios measures how much the total output of an industry is produced by the largest firms in the industry

$$CRn = S_1 + S_2 + S_3 + \ldots\ldots + S_n = \sum S_i \quad (1)$$

$i = 1, 2, .. n$ with $n < N$

where $S_1, S_2, S_3, \ldots S_n$ are output of firm 1, 2, 3… n, as a percentage of the industry output, and N is the number of firms in the industry.

By far, the most common variable used to measure the market structure of an industry is the four-firm concentration ratio, CR4, which is the share of industry sales accounted for by the four largest firms.

$$CR_4 = S_1 + S_2 + S_3 + S_4 \quad (2)$$

Another measure of concentration is the eight-firm concentration ratio, CR8, which is the share of industry sales accounted for by the eight largest firms.

$$CR_8 = S_1 + S_2 + S_3 + S_4 + S_5 + S_6 + S_7 + S_8 \quad (3)$$

It is, of course, arbitrary to focus attention on the top four firms in defining concentration ratios.

Learning Tip

Concentration ratios do not adjust for variation in firm size. For example, the CR4 for an industry when there are five firms of equal size is 0.80. A market with six sellers where the distribution of market shares (in percentages) is 50, 10, 10, 10, 10 ,10, would also have a CR4 of 0.80.

3.1.2 Demonstration Problem

Eight firms compete in a market to sell detergent. The total sales of all firms selling the product are $100 million. In ranking firms' sales from the highest to lowest, we find the top four firms' sales to be $25 million, $20 million, $15 million, and $12 million, respectively. Calculate the four-firm concentration ratio in the market for detergent.

Answer:

The sales of the four largest firms are

$$S_1 + S_2 + S_3 + S_4 = 25/100 + 20/100 + 15/100 + 12/100 = 0.72.$$

This means that the four largest firms in the industry account for 72 per cent of total industry output.

When an industry is composed of a very large number of firms, each of which is very small, the four-firm concentration ratio is close to zero. When four or fewer firms produce all of an industry's output, the four-firm concentration ratio is 1. The closer the four-firm concentration ratio is to zero, the less concentrated is the industry, the closer the ratio to 1, the more concentrated is the industry.

3.1.3 Herfindahl-Hirschman Index

An alternative measure of concentration is the *Herfindahl-Hirschman Index*, HHI, which equals the sum of the squared market shares of in the industry, multiplied by 10,000.

$$HHI = 10{,}000(S^2_1 + S^2_2 + S^2_3 + \ldots S^2_n) = 10{,}000\sum S_i^2 \quad (4)$$

where $i = 1, 2, 3 \ldots N$ = number of firms in an industry

By squaring the market shares before adding them up, the index weighs firms with high market shares more heavily. The value of the HHI lies between 0 and 10,000. The HHI = 10,000 when there is one firm, and the HHI = 0 when there are numerous infinitesimal firms. Therefore, the more concentrated the industry, the closer the index is to 1. In the case of equal-sized firms, the larger the HHI, the fewer the number of firms in the industry. In this case, the value of the HHI index is 1/N. If the market share distribution is not uniform, the HHI > 1/N and it grows with the increase in inequality between the firms in the industry.

Learning Tip

The HHI is based on the market shares of all firms in an industry and therefore is more complete than concentration ratios that focus only on shares of a small number of large firms.

3.1.4 Demonstration Problem

Suppose an industry has three firms with market shares of 50, 30, and 20 per cent. What is the HHI?

Answer:

The sum of squared market shares:

$S_1^2 + S_2^2 + S_3^2 = (50)^2 + (30)^2 + (20)^2 = 2{,}500 + 900 + 400 = 3{,}800.$

Learning Tip

Note that changes in HHI arise from both a change in the absolute number of firms and the size distribution of firms. A wider distribution of the firm sizes relative to the average firm size in the industry, the larger the squared of firm sizes, and hence the larger the HHI.

3.1.5 Demonstration Problem

Calculate the HHI for the following cases (a) there are two firms where each firm's share is $S_1 = 1/2$; $S_2 = 1/2$, (b) there are three firms where each firm's share is $S_1 = 1/3$; $S_2 = 1/3$; $S_3 = 1/3$, (c) there are three firms where each firm's share is $S_1 = 1/2$; $S_2 = 1/2$; $S_3 = 1/2$.

Answer:

a. $HHI = S_1 + S_2 = (1/2)^2 + (1/2)^2 = 1/2$

b. $HHI = S_1 + S_2 + S_3 = (1/3)^2 + (1/3)^2 + (1/3)^2 = 1/3$

c. $HHI = S_1 + S_2 + S_3 = (1/2)^2 + (1/4)^2 + (1/2)^2 = 3/8$

When as in (a) and (b), N are equal-sized firms, then the value of the HHI is 1/N, respectively. In case (c) the industry is about as concentrated as an industry with 2.6 equal-sized firms.

3.2 Limitations of Concentration Measures

Unfortunately, concentration measures have serious problems, of which the most important problem is the fact that many factors influence seller concentration measures. For example, profitability affects the degree of concentration in an industry by affecting entry. On the other hand, the SCP requires that the causality run from the structure to performance, i.e., the structure to affect performance (profits) and not vice versa. In order for this to be the case, however, the market structure has to be exogenous. Endogeneity means that the structure is determined before profitability and that profitability does not affect structure. The problem is that most commonly used measures of market structure are not exogenous.

Another problem is that concentration ratios, published by the government for specific industries and products definitions, do not necessarily coincide with relevant economic markets. Concentration measures are often based on national statistics. If the geographic extent of the market is local because transportation costs are very high, national concentration statistics may misleadingly indicate markets are less concentrated than is true.

Similarly, concentration measures are often biased because they ignore imports and exports. For example, the earlier evidence for the U.S. put the four-firm concentration ratio for the U.S. automobile industry at 92 per cent. This figure indicated a very concentrated industry. However, it ignored the imports of Italian, Korean, Japanese, and German cars, which were over 25 per cent of total sales in the United States.

Finally, concentration ratios reveals very little about conditions of entry or product differentiation within an industry.

The following table shows the estimates of the concentration measures for selected Canadian industries.

Table 7-1 Four-Firm Concentration Ratios and Herfindahl-Hirschman Indexes for Selected Canadian Manufacturing Industries

Industry	CR4	HHI (Value Shipped)
Logging	1.4	219
Gold, quartz mines	75.0	1874
Uranium Mines	(n/a)	2708
Vegetable Oil Mills	70.9	1499
Breweries	99.00	3177
Wineries	72.0	1543
Men's Clothing factories	20.6	225
Sawmills and planning mills	19.2	181
Furniture re-upholstery and repair	8.1	45
Asphalt roofing manufacturers	96.3	2458
Hardware, tool and cutlery manufacturers	12.9	106
Machine shops	6.4	42
Motor vehicle manufacturers	93.7	3865
Shipbuilding and repair	56.4	1066
Manufacturers of major appliances	77.0	1751
Jewellery and silverware industry	53.2	990
Signs and display manufacturers	15.9	141

Source: *Industrial organization and concentration in the manufacturing, mining and logging industries,* Statistics Canada, 1980.

Notice that there is a significant variation among industries in the degree of concentration. In Table 7-1, the largest four-firm concentration ratio belongs to the Brewery industry with 99 per cent, whereas the smallest four-firm concentration ratio, 1.4 per cent, belongs to Logging industry. Also, note that the industry with the largest four-firm concentration ratio does not equally correspond the largest HHI. For example, while the Brewer industry outranks other industries in regard to the four-firm concentration ratio, it falls short of the Motor Vehicle industry in regard to the HHI; 3,865 for Motor Vehicle versus 3,177 for Breweries. This, however, would be the case if the firms in the Motor Vehicle industry were individually larger than those in the Brewery industry. The logic behind this is squaring the market shares before adding them up giving rise to the index weighing firms with high market shares more heavily.

4 Market Performance

SCP studies typically measure market performance by using one of the following two profitability variables.

4.1 Economic Profits or Rates of Return on Investment

Economic profits are the difference between revenues and the opportunity cost of all inputs. In the long run, economic profits are an indicator of market power. In competitive markets, economic profits are eliminated by entry. Monopoly profits can only persist in the long run if there are entry barriers that provide the firm with market power. However, profits are an imperfect indicator since a firm may have market power, but not earn economic profits. And in the short run, price-taking firms could earn economic profits.

A rate of return on investment is the ratio of earnings or income to investment. If economic profits are positive, then the firm's rate of return will be greater than the competitive rate of return. Economic profits and excess rates of return both indicate profitability in excess of opportunity cost.

4.2 The Lerner Index

To avoid the problems associated with calculating rates of return, many economists use a different measure of performance, the Lerner Index, or the Price-Cost Margin:

$$L = (P - MC)/P \tag{5}$$

where P is price and MC is marginal cost. Thus, the Lerner Index measures the difference between price and marginal cost as a fraction of the price of the product.

When a firm sets its price equal to the marginal cost of production, the Lerner Index is zero; consumers pay a price for the product that exactly equals the cost to the firm of producing another unit of the good. When a firm charges a price that is higher than marginal cost, the Lerner Index takes on a value greater than zero, with the maximum possible value being unity. The Lerner Index therefore provides a measure of how much firms in an industry mark up their prices over marginal cost.

The higher the Lerner Index, the greater is the firm's mark-up. In industries in which firms rigorously compete for consumer sales by attempting to charge the lowest price in the market, the Lerner Index is close to zero. When firms do not rigorously compete for consumers through price competition, the Lerner Index is closer to 1.

A mathematical derivation helps see the way price elasticity and the Lerner Index are related.

Recall the useful relationship between MR and P for a firm that is maximizing its profit:

$$MR = P(1 - 1/E_P)$$

and

$$MR = MC.$$

Plugging for MR, and rearranging:

$$(P - MC)/P = 1/E_P \quad (6)$$

This relationship provides a rule of thumb for pricing. The relationship says that this mark-up should equal the inverse of the elasticity of demand. Accordingly, the more price elastic the demand curve faced by an imperfectly competitive firm, the smaller the mark-up.

4.2.1 Demonstration Problem

A firm in the tourist industry has a marginal cost of $300 and charges a price of $500. What are the Lerner Index and the price elasticity of demand?

Answer:

The Lerner index is: $(P - MC)/P = (500 - 300)/300 = 2/3 = 0.66$.

The elasticity is: $1/E_P = 0.66$, and $E_P = 1/0.66 = 1.33$.

5 Summary and Review

The most comprehensive theory of competition is the Industrial Organization theory. Three elements of this theory are concentration, performance, and conduct. The SCP approach assumes that there is a stable, causal relationship between the structure of an industry, firm conduct, and market performance running from structure through conduct to performance. If a stable relationship is established between structural variables and market power, then the SCP implication is that this structural variable facilitates the exercise of market power.

Several concentration measures are employed to determine the structure of an industry, such as concentration ratios, CR4 and CR8, as well as Hefindahl-Hirschman index. Market performance is measured by the rate of return on investment as well as by the Lerner Index.

6 Review Problems

1. Briefly explain why understanding market structure is important in managerial decision making. Provide a list of relevant factors that determine market structure.

2. Explain the following measures of industry concentration – the four-firm concentration ratios and the Herfindahl-Hirschman Index.

3. There are ten firms in an industry. The largest five firms have reported sales of $5 billion, $4.2 billion, $2.5 billion, $1.1 billion, and $1 billion. The rest of the firms have reported sales of $1 billion in aggregate. Find the four-firm concentration ratios and the Herfindahl-Hirschman Index of the industry.

4. 'The lower the Lerner Index, the higher the level of competition' – explain whether the statement is true, false, or uncertain.

5. A firm has a Lerner Index of 0.60. Find the mark-up factor. If the price of the product is $100, what is the firm's marginal cost?

6. Discuss the limitations of the measures of the concentration indices.

7. Illustrate how concentration ratios do not adjust, as the Herfindahl-Hirschman index does, for variation in firm size?

8. Calculate the CR_4 and the HHI for an industry when there are five firms of equal size.

7 Answer Key to Review Problems

1. Understanding market structure is important in order to make better decision in firm management. By analyzing market structure, a manager can systematically and methodically generate relevant information that leads to more efficient decision making (e.g. output level, advertising level, decision to enter a market, etc.).

   Relevant Factors: Industry Concentration, Technological Image, Market Demand, Entry Barrier.

2. Four-Firm Concentration Ratio = market Share of Largest Four Firms = (Sales of Largest Four Firms/Total Industry Sales)

   Herfindahl-Hirshman Index (HHI) =Sum of Squared Market Shares of All Firms x 10,000 =10,000. $\sum_{i=1}^{n} s_i^2$,Where, s_i = Share of company 'i' ; n = total number of companies

3. Four-firm concentration ratio = (5 + 4.2 + 2.5 + 1.1)/(5 + 4.2 + 2.5 + 1.1 + 1 + 1) = 13.8/15.8

   $HHI = (5/15.8)^2 + (4.2/15.8)^2 + (2.5/15.8)^2 + (1.1/15.8)^2 + (1/15.8)^2 + 5(0.2/15.8)^2$

   $\Rightarrow HHI = 0.246.$

4. P: Price, MC: Marginal Cost, L: Lerner Index

   L = (P – MC)/P

   In a competitive model. P = MC, therefore L = 0

   In a monopolistic model, MR = MC and P >MC, therefore L >0. The higher the difference between P and MC in the monopoly case, the greater the market power of the monopolist. Thus the lower the LI, the more competitive is the market.

5. Mark-up factor 3.5, (1/1–L) = 3.5 or, 1–L = 1/3.5, or L = 2.5/3.5. Therefore, 2.5/3.5 = (P–100)/P, or P = 350

   Mark-up factor = 1/(1–L) = 1/0.4 = 2.5, L = (P–MC)/P, or 0.6 = (100–MC)/100, or MC = 40.

6. Limitations:

   - non inclusion of foreign firms
   - lack of information on geographical concentration
   - definitional issues for industry and product classes

7. Since the HHI is based on the squared of the firms' market share, the larger the firm the greater their weight influences the index. Suppose that there are 4 firms in an industry with equal market shares, 0.25. The 4-firm concentration ratio in this case is 1.00. The HHI in this case is 4 x 10,000 $(0.25)^2$ = 2,500. Now suppose the firms' shares are altered as follows: 0.10, 0.20, 0.30, and 0.40. The 4-firm concentration ratio is still equal to 1.00, whereas the HHI is now equal to $10{,}000(0.10^2 + 0.20^2 + 0.3^2 + 0.4^2) = 3{,}000$. Therefore, the size distribution matters to the HHI but not to the 4-firm ratio.

8. See answer to question 7.

8 References and Suggested Further Reading

Baumol, W. J., *Business Behaviour, Value and Growth*, Rev. Ed. New York: Harcourt Brace Johanovich, 1967.

Besanko, D. and Braeutigam, R., *Microeconomics: An Integrated Approach*, Wiley, 2002

Baye, M., *Managerial Economics and Business Strategy*, McGraw Hill, Irwin, 2002.

Douglas, J. E., *Managerial Economics: Analysis and Strategy*, Prentice Hall, 1992.

Pindyck, R. and Rubinfeld, D., *Microeconomics*, Fifth Edition, Prentice Hall, 2001

Robinson, J., *Economics of Imperfect Competition*, London, McMillan Inc. 1933.

Stigler, G.J., 'The Kinky Oligopoly Demand Curve and the Rigid Prices,' *Journal of Political Economy*, 1947.

Commonwealth of Learning Executive MBA/MPA

E5 Managerial Economics

Block 8

Market Structure

Course author: Farrokh Zandi, PhD
Professor, Schulich School of Business
York University
Toronto, Canada

Course designer: Douglas R. Cronk, M.A., ABD
Vancouver, BC, Canada

Copyright © Commonwealth of Learning, 2003
All rights reserved. No part of this course may be reproduced in any form by any means without prior permission in writing from:

The Commonwealth of Learning
1285 West Broadway
Suite 600
Vancouver, BC V6H 3X8
CANADA
e-mail: info@col.org

Contents

1 Block Eight: Objectives and Introduction 1
1.1 Introduction 1
2 Perfect Competition 2
2.1 Single Firm Versus the Market 3
2.1.1 Demonstration Problem 5
2.2 Short-Run Supply Decision 5
2.2.1 Demonstration Problem 6
2.2.2 Demonstration Problem 7
2.2.3 Demonstration Problem 9
2.3 Long Run 9
2.4 Efficiency 11
2.4.1 Consumer Surplus 11
2.4.2 Producer Surplus 12
2.4.3 Deadweight Loss 12
3 Monopoly 12
3.0.1 Demonstration Problem 14
3.0.2 Demonstration Problem 15
3.1 Welfare Loss of Monopolies 16
3.2 The Multi-Plant Monopoly 17
3.2.1 Demonstration Problem 18
4 Monopolistic Competition 19
4.1 Short-Run Price and Output Determination 20
4.2 Long-Run Price and Output Determination 21
4.2.1 Demonstration Problem 22
5 Oligopoly 23
5.1 The Kinked Demand Curve Model of Oligopoly 23
5.1.1 Demonstration Problem 25
5.2 Non-Co-operative Oligopoly 26
5.3 Nash Equilibrium 27
5.4 Quantity Competition 27
5.4.1 The Cournot Model 27
5.4.2 Collusion 30
5.5 The Stackelberg Model 31
5.6 The Bertrand Model (Price Competition) 32
6 Summary and Review 34
7 Review Problems 34

8 Answer Key to Review Problems 36
9 References and Suggested Further Reading 39

1 Block Eight: Objectives and Introduction

After working through Block Eight of this course, you should be able to:

1. Explain different forms of market structure and their underlying assumptions.
2. Analyze the significance of barriers to entry and the relative size of the firms in an industry in shaping the structure of the industry.
3. Explain why monopolies may, in some circumstances, be undesirable from the perspective of the society.
4. Compare the short and long-run equlibrium situations under perfect and monopolistic competition.
5. Outline various models of oligopoly.
6. Explain why oligopolistic market structures cannot be explined via one single model.
7. Explain the nature of the game implicit in the theory of oligopoly.

1.1 Introduction

In the previous Block, we examined the nature of industries and saw that industries differ with respect to their underlying structures, conduct, and performances. In this Block, we characterize the optimal price, output, and advertising decisions of managers operating in environments of (1) perfect competition, (2) monopoly, (3) monopolistic competition, and (4) oligopoly.

It is logical to start with the simplest case: a situation where managerial decisions have no perceptible impact on the market price. Thus, in the first section of this Block we will analyze output decisions of managers operating in perfectly competitive markets. In subsequent sections, we will examine output decisions by firms that have market power: monopoly and monopolistic competition, and oligopoly.

There are two key assumptions used in the economic theory of firms you should review before looking at pricing and output decision-making in the four types of markets:

- The firm's primary objective is the short-run maximization of profit. This may not be, however, the case for oligopoly, where time horizons typically extend beyond the short run. High short-run profits may induce the entry of new competitors to cause a more competitive market for the firms later in the planning period.
- The opportunity cost of producing a particular good or service is included in the cost of doing business – economic costs.

In analyzing a firm's pursuit of short-run profit, the economic theory of the firm posits that its managers must address three basic questions:

1. Should our company be in this business? That is, should it be selling this particular product at all?

2. If so, how much should we produce?

3. And if we are able to set the price, what price should we charge?

Firms operating in perfectly competitive markets cannot set their own price. Therefore, this question does not apply to them.

2 Perfect Competition

Perfect competition provides a benchmark against which the behaviour of other markets is judged. In this chapter, we stress five key points:

1. Competition has many desirable properties.

2. Free entry and exit is a crucial factor in determining whether a market is competitive and efficient.

3. Welfare is maximized under competition.

4. The desirability of competition is reduced in the presence of externalities such as pollution.

5. Even if some of the necessary conditions for perfect competition do not hold, markets can come close to achieving the desirable properties of perfect competition.

Even though perfect competition is rarely, if ever, encountered in the real world, we study the perfect competition model because it provides an ideal framework against which we compare other models and markets.

We define perfect competition as a market outcome in which all firms produce homogeneous, perfectly divisible output, and face no barriers to entry or exit; producers and consumers have full information, incur no transaction costs, and are price takers; and there are no externalities. That is, the main assumptions of perfect competition are:

1. *Homogeneous Good.* All firms sell an identical product. Consumers view the products of various firms as the same and hence are indifferent between them.

2. *Perfect Information.* Buyers and sellers have all relevant information about the market, including the price and quality of the product.

3. *Price Taking.* Buyers and sellers cannot individually influence the price at which the product can be purchased or sold. Price is determined by the market, so each buyer and seller takes the price as given.

4. *No Transaction Costs.* Neither buyers nor sellers incur costs or fees to participate in the market.

5. *No Externalities.* Each firm bears the full costs of its production process. That is, the firm does not impose externalities – uncompensated costs – on others. For example, pollution produced by a firm is an externality because the firm does not recompense the victims.

6. *Free Entry and Exit.* Firms can enter and exit the market quickly at any time without having to incur special expense. That is, firms do not face barriers to entry or exit.

7. *Perfect Divisibility of Output.* Firms can produce and consumers can buy a small fraction of a unit of output. As a result, the amount of output demanded or supplied varies continuously with price. This technical assumption avoids problems caused by large discrete changes in either supply or demand in response to small price changes.

Some economists also assume that a perfectly competitive market has a large number of buyers and sellers. If there are many firms, no one firm can charge a price above the market price without losing all its customers, so the firm views the price at which it can sell as beyond its control. Similarly, consumers cannot find a firm willing to sell below the market price, so consumers must view the market price as beyond their control.

However, even if there are relatively few firms in a market, no firm could raise its price above the market price without losing all its customers if another firm could quickly enter the market and underprice it. Thus, because we assume firms and consumers are price takers and that there is free entry and exit, we do not also have to assume that there are a large number of firms. Competitive markets typically have a large number of firms and consumers, but industries can have all the properties of perfect competition even though there are few firms in those industries.

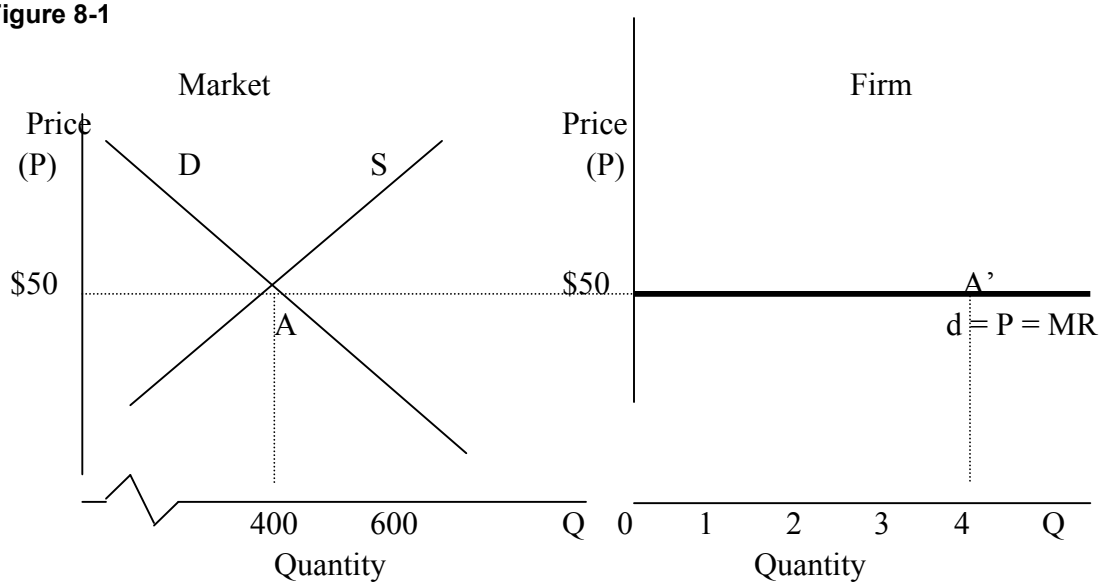
2.1 Single Firm Versus the Market

Under perfect competition, the price is determined by interaction of many suppliers and demanders. The firm can sell its products only at this price, no less and no more. If the firm chooses to charge a higher price, given that its rivals are selling the same product, it will lose all its customers. The firm should not want to sell its product at a lower price, either; this strategy would have made sense, if the firm had some the capacity and the size to meet the needs of the customers lured from other firms, but it does not. These perfectly competitive firms are small; selling for less is not practical. Therefore, a perfectly competitive firm faces a demand curve that is horizontal at the market equilibrium price. Put differently, the demand is *perfectly elastic*.

Given that the market price is determined at the interaction of the market demand curve and the market supply curve, the perfectly competitive firm is a *price taker*. That is the perfectly competitive firm takes the price of the product as given and has no perceptible effect on that price by varying its own level of output and sales product.

In Figure 8-1, D is the market demand curve, and S is the market supply curve of the product. Equilibrium price is P = $50 and is determined at point A, at the intersection of D and S. At this point, the quantity demanded equals quantity supplied, 400.

Figure 8-1



The equilibrium price and quantity can be determined algebraically by setting the demand and supply functions equal to each other and solving for the equilibrium price. Then substituting the result into the demand or supply functions and solving for Q, we get the equilibrium quantity.

$$\text{Demand: } Q_D = 900 - 10P \qquad (1)$$

$$\text{Supply: } Q_S = 150 + 5P \qquad (2)$$

Setting Q_D equal to Q_S, and solving for P, we have

$$Q_D = Q_S$$

$$900 - 10P = 150 + 5P$$

$$750 = 15P \text{ , } P = \$50$$

Substituting into either function and solving for Q

$$Q_D = 900 - 10(\$50) = 400$$

$$Q_S = 150 + 5(\$50) = 400$$

2.1.1 Demonstration Problem

Assume that in the example above, the industry (market) consists of 100 identical firms. How much would each firm produce? Depict the point of production for a representative firm.

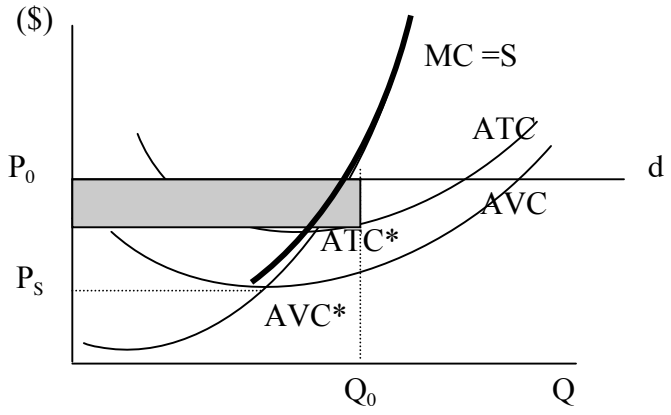
Answer:

It is clear that each firm will produce (1/100) of the industry's output. Since the industry as whole produces 400 units, the share of each firm is 4 units. The price charged by each firm is $50 along the horizontal line (d) and quantity produced is 4, point A' in Figure 8-1.

2.2 Short-Run Supply Decision

Let us first examine the incentives of a typical firm. Suppose a firm has the short-run cost curves in Figure 8-2 and faces a market price of P_0. How much should it produce? Indeed, should it produce anything at all?

Figure 8-2



The objective of any firm, including a competitive firm, is to maximize its profits (or, equivalently, minimize its losses). The competitive firm's profits, π, are

$$\pi = TR(PQ) - TC(Q) \qquad (3)$$

where P is price, Q is output, and TR and TC are total revenue and total costs, respectively.

A firm produces only if doing so is more profitable than not producing. It produces only if the revenues from producing exceed avoidable costs: the costs that are not incurred if a firm ceases production. The revenues earned in excess of avoidable cost are called *quasi-rents,* which are the payments above the minimum amount necessary to keep a firm operating in the short run. For simplicity, assume that all fixed costs are sunk. An example of a sunk cost in Block 6 is that a firm is not refunded its incorporation fee if it ceases operation. In this case, avoidable costs are the same as variable costs. Thus, the rule for deciding whether to remain in business is: produce and sell only if revenues are at least as great as total variable cost. Equivalently, the firm should produce and sell at price P only if it equals or exceeds average variable cost (AVC).

A shown in Figure 8-2, the minimum point of average cost (the lowest point on the ATC curve), ATC*, is greater than minimum AVC, AVC*, in the short run, because average costs are average variable costs plus average fixed costs. Thus, a firm finds it more profitable to produce than to shut down if price is below minimum average cost, $P < ATC^*$, but above minimum average variable cost, $P > AVC^*$. It is more profitable to produce and earn some revenue in excess of variable cost than to shut down and earn no revenue (which can help offset the fixed costs). That is, the firm chooses to produce though it is losing money when all costs are considered.

2.2.1 Demonstration Problem

Consider an example to clarify this apparent contradiction. Suppose a firm's fixed cost is $300 and sunk. Its marginal cost (MC) is constant at $20 at quantities less than 150 units, while at more than 150 units, MC rises drastically. If the price is $20, i.e., P = MC = $20, the firm will produce and sell 150 units. The firm just covers production cost and makes no contribution to the $300 fixed cost; it loses $300. What would be the effect of marginal costs of $19 and $21?

Answer:

If the price is $19, the firm is better off not producing at all, because it loses an additional $1 for every unit it produces and would lose $450 ($300 in fixed cost plus $150 in variable cost) if it produced 150 units. It is better to shut down and lose only $300 than to produce and suffer greater losses.

If the price is $21, by producing 150 units, the firm now more than covers variable cost: it earns $150 above variable cost. It still loses money overall (–300 + 150 = –150) because of the fixed cost of $300, but it is better to lose $150 than $300.

Learning Tip

The point of the example is that the decision to produce or not is *independent* of the fixed sunk cost. If fixed costs are sunk (incurred whether the firm produces or not), they should be ignored in deciding whether to produce.

The price at which a firm ceases production is the *shutdown point*, which is P_S in Figure 8-2. That is, if price exceeds AVC*, the firm operates along its MC curve. The firm's *supply curve* reflects the quantity that a firm is willing to supply at any given price. The competitive firm's supply curve is the portion of the MC curve above AVC*, the shutdown point.

2.2.2 Demonstration Problem

If a firm suffers losses in the short run (the period in which costs are sunk), should it continue to operate and remain unprofitable in the future?

Answer:

No. In the long run, a firm that is losing money will not reinvest – it will not continue to sink costs. Short-run losses are a signal that the firm should not invest further to replace plant and equipment. In the long run, a rational firm shuts down if it expects to have losses in each period forever. It prefers to cease production rather than invest in new facilities or maintenance and lose even more.

When a firm loses in the short run, its revenues are below the long-run opportunity cost of its resources. Because opportunity cost includes a normal profit, a firm that is making a loss may not literally be paying out more money than it receives; it is simply earning less than it could have earned had it invested its (already) sunk costs elsewhere.

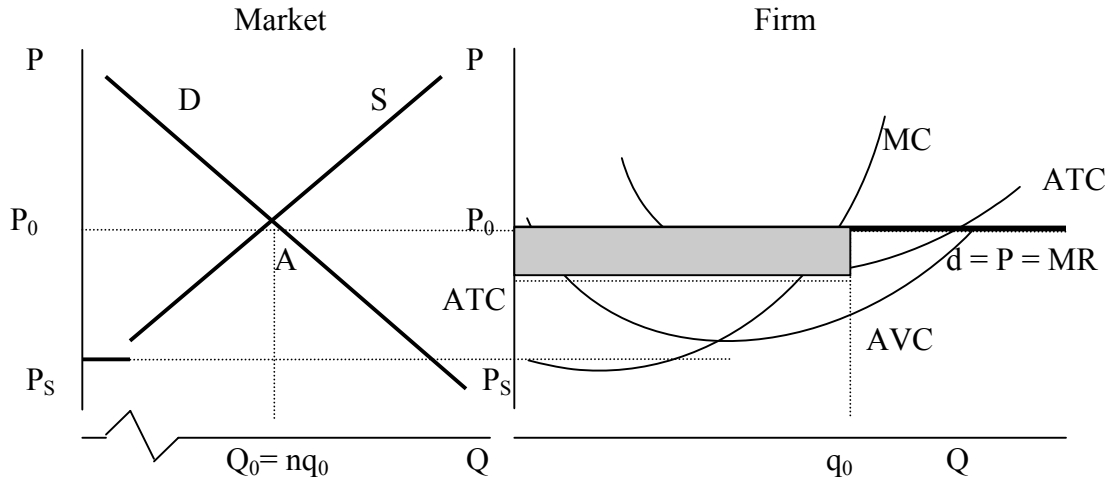
Learning Tip

If fixed costs are not sunk, the shutdown decision depends on whether revenues exceed *avoidable* costs. If some costs are avoidable, a price equal to AVC* is not high enough to prevent the firm from shutting down. Accordingly, the price at which shutdown occurs is above average variable cost and closer to average, the greater the proportion of fixed costs that are avoidable. If there are no sunk costs (all costs avoidable), the shutdown point coincides with the minimum point on the ATC (ATC*). Therefore, if it has no sunk costs, a firm shuts down before it incurs economic losses.

As we have already seen, the short-run profit maximization is attained by P = MC. Suppose there are *n* identical firms and that all fixed costs are sunk in the short run. The short-run market supply curve, S in Figure 8-3, is the horizontal sum of the supply curves of each firm (the upper portion of the firm's marginal cost curve above the minimum of the AVC curve). The horizontal portion of the market supply curve reflects (1) that no output is forthcoming if price falls below the shutdown point and (2) that at a price slightly above the shutdown point, all firms produce.

The interaction of the market demand curve and the short-run market supply curve determines the *short-run equilibrium price*, P_0 and quantity, $Q_0 = nq_0$, where q is the volume of output produced by an individual firm, which produces 1/n of the market output.

Figure 8-3



Learning Tip

Note that marginal revenue (MR) is the derivative of the total revenue function. For a perfectly competitive firm, total revenues is TR = PQ, where P is the market equilibrium price. Thus, MR = dTR/dQ = P.

The shaded rectangle in Figure 8-3 represents the maximum profits of the firm. To see this, note that the area of the shaded rectangle is given by its base (q_0) times the height [$P_0 - ATC(q_0)$]. Recall that $ATC(q_0) = TC(q_0)/q_0$; that is, average total cost is total cost divided by output. Intuitively, [$P_0 - ATC(q_0)$] represents the profits per unit produced. When this is multiplied by the profit-maximizing level of output (q_0), the result is the amount of total profits earned by the firm.

2.2.3 Demonstration Problem

Suppose the cost function of for a firm is given by $TC(Q) = 10 + 2Q^2$

If the firm sells output in a perfectly competitive market and other firms in the industry sell output at a price of $40, what price should the manager of this firm put on the product? What level of output should be produced to maximize profits? How much profit will be earned?

Answer:

The derivate of the total cost function is the marginal cost function:

$$d(TC)/dQ = MC(Q) = 2(2Q) = 4Q$$

Since the firm competes in a perfectly competitive market, it must charge the same price other firms charge; thus, the manager should price the product at $40. To find the profit-maximizing output, we must equate price with marginal cost $40 = 4Q$, so the profit-maximizing level of output is 10 units. The maximum profits are thus:

$$\text{Profit} = (40)(10) – [5 + 2(10)^2] = 400 – 205 = \$195$$

2.3 Long Run

In the short run, a firm decides whether to produce or not and also if the decision is to produce, how much to produce. In the long run, the firm faces less constraint. They can indeed adjust their levels of capital so that they can enter this market. Remember, long run is a time frame in which an existing firm can change the size of its fixed input (plant) – there are no fixed inputs – or to decide to leave the industry to avoid losses. By a similar token, in the long run, new firms can enter the industry in pursuit of profits.

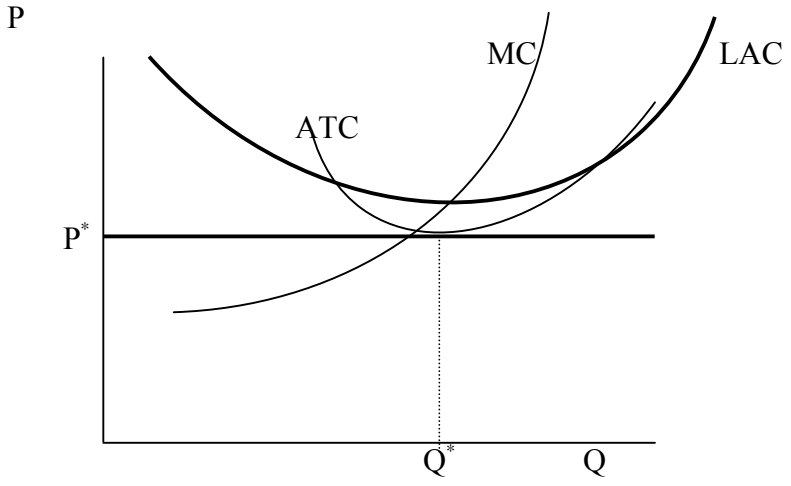
As indicated earlier, in the short run, a firm might face three different possibilities: making an economic profit, taking an economic loss, or breaking even. In the long run, however, a perfectly competitive firm, despite its short-run situation, cannot either run an economic profit or an economic loss. Indeed, the only outcome in the long run for a perfectly competitive firm is one in which the firm makes normal profit.

An industry in which firms are making an economic profit in the short run adjusts in two different ways. First, the number of firms in the industry increases. Second, the existing firms expand to take advantage of the profit. *Entry* by new firms is a rational decision by investors who respond to market incentives (signals) prompted by an economic profit. By the same token, economic profit also serves as the incentive (signal) to the existing firms to expand their existing operation.

The presence of short-run losses prompts a different set of rational responses. First, the number of firms in the industry decreases as some existing firms *exit*. Second, the existing firms contract the scale of their business.

The process just described continues until ultimately the market price is such that all firms in the market earn zero economic profits. Moreover, the best level of output is the one at which price equals the minimum point of ATC (short-run average cost curve) as well as the lowest point of the long-run average cost curve (LAC). This is shown in Figure 8-4. At the price of P^* each firm receives just enough to cover the average costs of production (ATC) because in the long run there is no distinction between fixed and variable costs, and economic profits are zero.

Figure 8-4



Learning Tip

In the long run, perfectly competitive firms produce a level of output such that P = MC, and also P = minimum of ATC. The first condition is the optimality condition, which needs to hold in the short run as well as in the long run. The second condition is strictly a long-run condition that needs to hold in light of free entry and exit implying not only that firms are earning zero economic profit (normal profits) but also that all economies of scale have been exhausted. There is no way to produce the output at a lower average cost of production. This implies that at Q^*, LAC = ATC.

It is important to remember the distinction we made in Block 6 between economic profits and accounting profits. The fact that a firm in a perfectly competitive industry earns zero economic profits in the long run does not mean that accounting profits are zero; rather, zero economic profits implies that accounting profits are just high enough to offset any implicit costs of production. The firm earns no more, and no less, than it could earn by using the resources in some other capacity. This is why firms continue to produce in the long run even though their economic profits are zero.

2.4 Efficiency

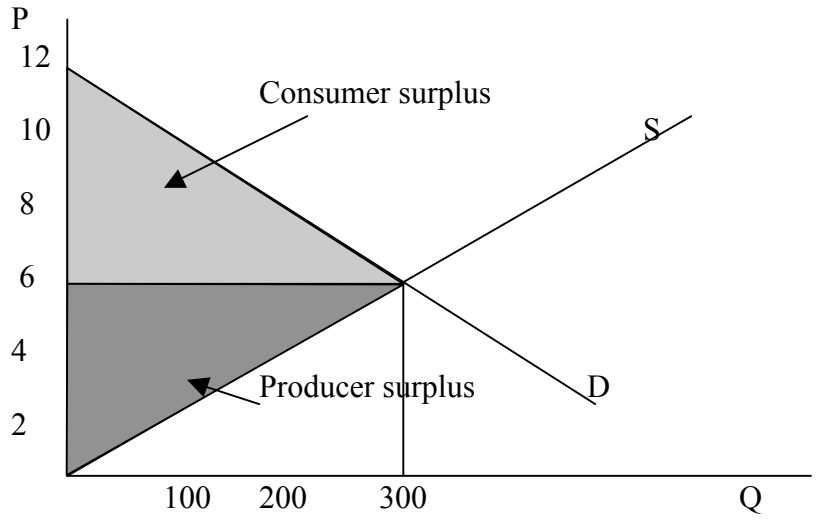
We now describe a common measure of welfare showing that competition maximizes this measure of welfare for any given distribution of income, and illustrate that departures from competition lower welfare.

2.4.1 Consumer Surplus

Typically, consumers value the goods they purchase above what they actually pay for them. *Consumer surplus* is the amount above the price paid that a consumer would willingly spend, if necessary, to consume the units purchased.

The demand curve reflects the value that consumers place on consuming alternative units of a good. For example, the demand curve in Figure 8-5 indicates that consumers would pay \$10 for 100 units of the good, \$8 for 200 units, and \$6 for 300 units. In the competitive equilibrium, consumers pay \$6 for 300 units. They would, however, be willing to pay \$4 more for the first 100 units, \$2 more for the first 200 units, and no extra amount for 300 units. The total consumer surplus is the area below the demand curve and above the equilibrium price of \$6 up to the equilibrium quantity of 300 units. This area equals \$900 (= [\$12 – \$6] x 300/2). The amount that consumers did actually pay is 300 x \$6 = \$1,800. Therefore, the consumer surplus is 50 per cent of what is actually paid.

Figure 8-5



2.4.2 Producer Surplus

Similarly, firms may receive more for the goods they sell than it costs them to produce those goods. *Producer surplus* is the largest amount that can be subtracted from a supplier's revenues and yet the supplier would still willingly produce the product.

We can use information from the supply curve to calculate firms' producer surplus. The marginal cost of producing output represents the firm's supply curve. For example in Figure 8-5, it costs firms $2 to produce 100 units, $4 to produce 200 units, and $6 to produce 300 units. The producer surplus is the area above the supply curve and below the price line. Therefore, the producer surplus is the area equal to $900. That is, the firm would be willing to depend on an extra $900 than going without any sale.

One common measure of welfare from a market is the sum of consumer and producer surplus. It is the amount that consumers and producers are willing to pay to exchange the equilibrium quantity of output.

2.4.3 Deadweight Loss

The cost to society of a market not operating efficiently is called deadweight loss. It is the welfare loss – the sum of the consumer surplus and producer surplus lost – from a deviation from the competitive equilibrium. We will discuss the application of this concept below.

3 Monopoly

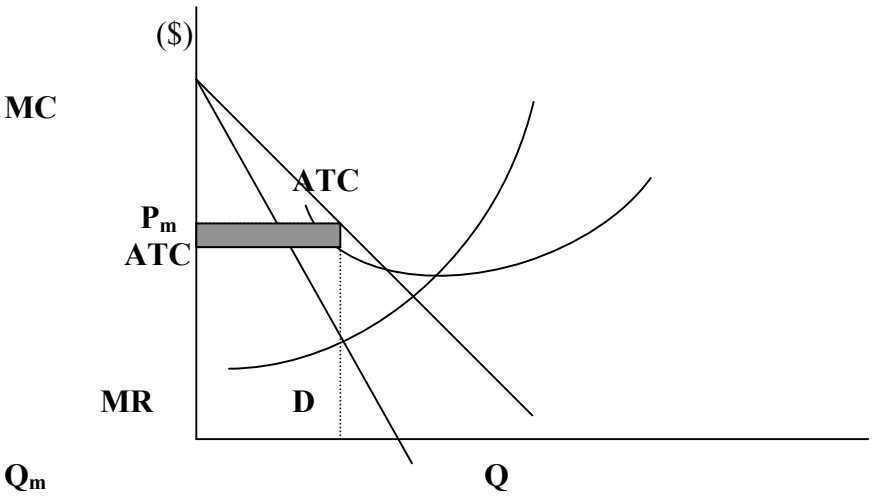
In the previous section, we characterized the optimal output decisions of firms that are small relative to the total market. In this context, small means the firms have no control over the prices they charge for the product. In this section, we will consider the opposite extreme: monopoly. Monopoly refers to a situation where a single firm serves an entire market for a good for which there are no close substitutes. As such, monopolists can safely ignore the reaction of their rivals, because there are no close rivals.

Monopolies persist over time if the entry of new firms is prevented by *barriers to entry*. Barriers to entry are additional costs that a potential entrant must incur before gaining entry to a market. These costs may arise because of economies of scale, the best location, possession of the necessary information, the right to produce the product (such as a patent or a government mandate), the existing firms controlling an essential raw material, and so on. Consumer loyalty to the existing firm also poses a barrier to a potential entrant, since the new firm would have to spend substantial amounts on advertising to offset it. In some cases, the entrant cost is infinite (where it is impossible to acquire the necessary input or permission to produce).

When one thinks of a monopoly, one usually envisions a very large firm. This needn't be the case, however; the relevant consideration is whether there are other firms selling close substitutes for the good in a given market. For example, a gas station located in a small town that is several hundred kilometers from another gas station is a monopolist in that town. In a large town, there are usually many gas stations, and the market for gasoline is not characterized by monopoly.

The profit-maximizing monopolist will wish to expand output until marginal costs rise to equal marginal revenues. Since this firm faces the entire market demand, the firm's demand curve and the market demand curve are one and the same. Thus, the market marginal revenue curve is the firm's marginal revenue curve. We saw in previous Blocks that the marginal revenue curve associated with a negatively sloping linear demand curve has the same vertical intercept on the graph and twice the slope of the demand curve. In Figure 8-6, we show the market demand curve (D) faced by the monopolist and the corresponding marginal revenue (MR) curve. Superimposed on these are the cost curves of the monopolist – the short-run average cost (ATC) and marginal cost (MC) curves. The profit-maximizing monopolist produces up to the point where marginal cost per unit rises to meet the falling marginal revenues. This point occurs at output level Q_m. Notice that for every unit to the right of Q_m, MC > MR, it will therefore not be produced. Conversely, for every unit to the left of Q_m, MC < MR, it therefore will be produced and sold. The firm's profits can be visualized as the rectangle (P_m – ATC) x Q_m.

Figure 8-6



Learning Tip

In summary, the monopolist is restricted by consumers to choose only those price-quantity combinations that lie along the market demand curve. The monopolist can choose a price or a quantity, but not both. The monopolist can sell higher quantities only by lowering the price. If the price is too high, consumers may choose to buy nothing at all.

Monopolists in the real world usually do have some peripheral competition from distant and partial substitutes. The post office experiences competition from the telephone company and from such delivery services as local or international courier companies for some of the services it provides. The electricity company competes with the gas company in some households. In all these cases, however, the extent of substitution of these products for the monopolist's product is quite small.

3.0.1 Demonstration Problem

In the discussion above, what economic measure can determine the extent of substitution between products?

Answer:

The answer is the cross elasticity of demand between the monopolist's product and its distant substitutes.

Learning Tip

Note that the reason why MR is less P for a monopolist can be explained as follows. Suppose a monopolist sells 5 units of output at a price of $8 per unit, for a total revenue of $40. What happens to revenue if the monopolist produces the 6th unit of output? Total revenue increases by less than $8. To see why, note that the monopolist can sell one more unit of output only by lowering price, say, from $8 to, say, $7 per unit. But, the price reduction necessary to sell the 6^{th} unit lowers the price received on all first 5 units from $8 to $7. The total revenue associated with 6 units of output is therefore $42. The change in revenue due to producing one more unit is therefore $2 (= 42 – 40), which is less than the price charged for the product.

Mathematically, the monopolist should produce the output, Q_m, such that marginal revenue equals marginal cost:

$$MR(Q_m) = MC(Q_m) \quad (4)$$

Given the level of output, Q_m, that maximizes profits, the monopoly price is the price on the demand curve corresponding to the Q_m units produced: $P_m = P$ (when $Q = Q_m$).

3.0.2 Demonstration Problem

Suppose the inverse demand function for a monopolist's product is given by

$$P = 60 - Q$$

And the cost function is given by

$$TC\ (Q) = 20 + 2Q$$

Determine the profit-maximizing price and quantity and the maximum profits.

Answer:

As we know, MR has the same intercept as the demand curve but a slope that is twice as large as that in the demand curve

$$MR = 60 - 2Q$$

and $MC = 2$

Next, we set MR = MC to find the profit-maximizing level of output:

$$60 - 2Q = 2$$

$$2Q = 58$$

Solving for Q yields the profit-maximizing output of $Q_m = 29$ units. We find the profit-maximizing price by plugging in for Q_m in the demand function:

$$P = 60 - (29) = \$31$$

Thus, the profit-maximizing price is \$31 per unit. Finally, profits are given by the difference between revenues and costs:

$$\text{Profits} = ((P_m)(Q_m)) - TC(Qm)$$

$$= (31)\ (29) - [20 + 2(29)]$$

$$= \$821$$

Learning Tip

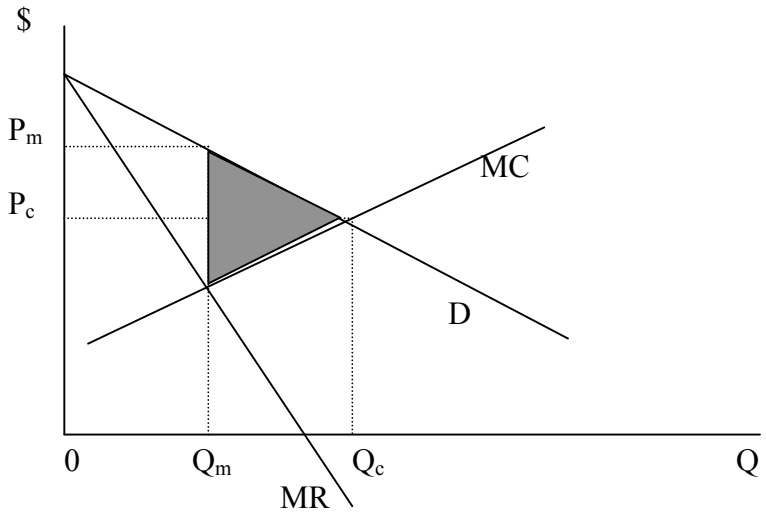
Recall that a supply determines how much will be produced at a given price. Since perfectly competitive firms determine how much output to produce based on price (P = MC), supply curves exist in perfectly competitive markets. In contrast, a monopolist determines how much to produce based on marginal revenue, which is less than price (P > MR = MC). As a consequence, there is no supply curve in markets served by a monopolist.

3.1 Welfare Loss of Monopolies

Often, the flip side of the power that a monopolist enjoys is viewed as a cost to the society, *social cost*. As we discussed earlier, the cost to the society of a market not operating efficiently is called deadweight loss.

In comparison to perfectly competitive industries, monopolies tend to create a deadweight loss. This is the case since a monopolist charges too high a price and produces too little output, compared to the socially optimal level of price and output obtained under perfect competition. Under perfect competition, the price paid by consumers reflects the marginal cost – which in absence of externalities is the true cost to the society – while under monopoly, the price is marked up above marginal cost. Therefore, monopolies, in comparison, operate on a higher point on the demand curve, where the price is higher and output is lower. As illustrated in Figure 8-7, P_m and Q_m correspond to the monopoly price and quantity, i.e., where MR = MC, whereas P_c and Q_c are the price and quantity under perfect competition where P = MC.

Figure 8-7



The deadweight loss, or the loss to the society, is shown by the shaded area, which is partly the loss to consumers and partly to producers.

3.2 The Multi-Plant Monopoly

Up until this point, we have assumed that the monopolist produces output at a single location. In many instances, however, many firms operate more than one plant. They may have built two or more smaller plants to avoid diseconomies of plant size that would occur if they tried to serve their market from a single, much larger, plant. A second reason for multiplant operation is relatively high transportation costs for the finished goods. If the market is spread over a wide geographical area, the firm will find it more economical to operate two or more plants in two or more locations if the savings in transportation costs more than outweigh the economies of plant size foregone. Third, the firm may operate two or more plants as a result of a merger or takeover, and the costs of scrapping both plants and building another larger one more than offset the economies of plant size foregone.

The multi-plant monopolist, unlike a simple monopolist, has to do two things:

1. It must determine its profit-maximizing price and total output level, and

2. at the same time it must determine the cost-minimizing allocation of the total output among its two or more plants.

The standard profit-maximizing rule, that marginal costs should equal marginal revenue, remains applicable, but our interpretation of the marginal cost curve must reflect the fact that marginal costs now derive from two or more production sources. The relevant marginal cost curve must show the marginal cost for each incremental unit produced when the firm is free to choose the plant in which that incremental unit should be produced. Clearly, the firm will always nominate the plant that can produce the incremental unit at the lowest additional (or marginal) cost.

For simplicity, suppose the firm operates just two plants, plant A with the total cost function $TC_A(Q_A)$ and plant B represented by the cost function $TC_B(Q_B)$, where ($Q_A + Q_B = Q$ = total output of the monopolist). The demand for the monopolist's product is typically given by $P = f(Q) = f(Q_A + Q_B)$. Note that consumers do not care which plant their purchased products come from.

It follows that, at any total output level (Q), the marginal cost from each plant should be equal, since if this were not true, the last unit might be taken away from the plant with the higher marginal cost and be produced instead by the plant with the lower marginal cost, and thus reduce the cost to the firm as a whole. That is,

$$MC_A(Q_A) = MC_B(Q_B)$$

The firm's profits will be maximized when

$$MR(Q) = MC_A(Q_A)$$

$$MR(Q) = MC_B(Q_B) \qquad (5)$$

Learning Tip

Note that marginal revenue in equation (5) is found by taking the derivative of the market demand function with respect to *total* output (Q), whereas marginal costs are found by taking the derivative of each cost function with respect to its own *individual* output, Q_A, or Q_B.

3.2.1 Demonstration Problem

Suppose a monopolist faces the following demand and cost curves.

Demand: $P = 100 - (Q)$, hence

$MR = 100 - 2Q = 100 - 2\,(Q_A + Q_B)$

and costs:

$$TC_A = 2Q_A^2, hence$$
$$MC_A = 4Q_A$$
$$TC_B = Q_B^2, hence$$
$$MC_B = 2Q_B$$

Determine (a) the level of output produced by each plant, (b) the total level of output, and (c) the price charged.

Answer:

$MR(Q) = MC_A$

$MR(Q) = MC_B$

Plugging in

$100 - (Q_A + Q_B) = 4Q_A$

$100 - (Q_A + Q_B) = 2Q_B$

From the first equation:

$Q_B = 100 - 5Q_A$

Substituting this into the second equation:

$100 - (Q_A + 100 - 5Q_A) = 2(100 - 5Q_A)$

$Q_A = 18.25$

$Q_B = 100 - 5(18.25) = 9.75$

$P = 100 - (28) = \$72.$

Learning Tip

The above analysis of the multi-plant firm applies to any market situation where a firm envisages a negatively sloping *ceteris paribus* demand curve. It is thus appropriate for monopoly, monopolistic competition, or oligopoly situations. An important issue for the manager of such a multi-plant monopoly is the determination of how much output to produce at each plant.

4 Monopolistic Competition

In monopolistic markets, the firms must choose a price knowing that the consumer has many close substitutes to choose from. If the price is too high, in view of the consumer's perception of the value of the differentiating features of the firm's products, the consumer will purchase a competing firm's product instead. Thus, the monopolistic competitor must expect a relatively elastic demand response to changes in its price level. Yet at the same time, it expects to change price without causing any other firm to retaliate and, consequently, without causing a change in the general price level in the market. This is possible because the firm is one of many firms, and it expects the impact of its actions to be spread imperceptibly over all the other firms, giving no one firm any sufficient reason to react to the initial firm's price change.

Monopolistic competition is so called because it has elements of both monopoly and pure competition. The firm has a significant amount of monopolistic pricing and pure competition. The firm has a significant amount of the monopolist's pricing power by virtue of the differentiation of its product. It can change price up or down without experiencing the extreme response of pure competition. For price increases it will suffer a loss of sales, but this loss in not total, as it would be for the pure competitor. Like a monopoly, the monopolistic competitor can adjust the price upward or downward to the level which maximizes its profits. But like the pure competitor, the monopolistic competitor has many rivals in the short run, compounded by the free entry of new firms in the long run.

The following sums up the conditions that prevail in a monopolistically competitive industry: (a) Many buyers and sellers, (b) differentiate products, and (c) rather easy entry to and exit form the industry by firms in the long run.

In reality, monopolistically competitive markets are found where a large number of vendors gather to sell similar products to a gathering of potential buyers. The weekly fruit and vegetable market in some communities may be characterized as monopolistic competition.

4.1 Short-Run Price and Output Determination

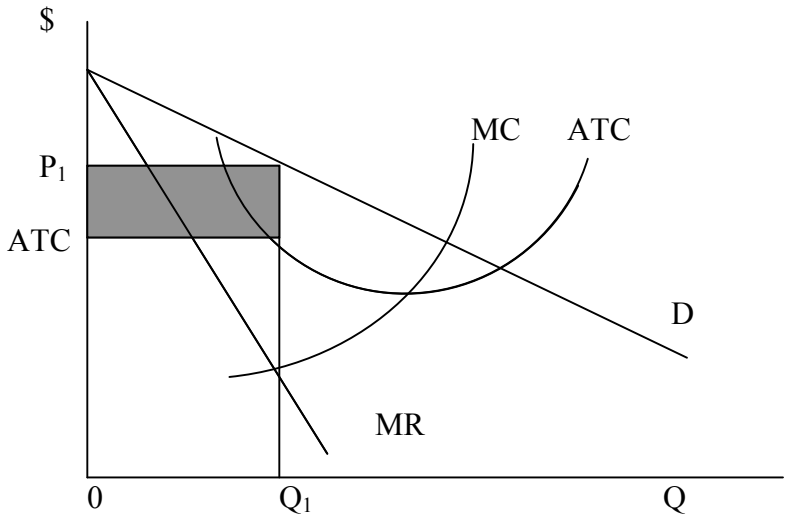
Figure 8-8 shows the equilibrium situation for a monopolistically competitive firm. The demand curve faced by this firm is negatively sloping because firms in this industry are price makers selling differentiated products. The monopolistically competitive firm maximizes its profits at the price and output level where marginal revenue equals marginal costs. Since the firm produces a differentiated product, the demand curve it faces is negatively sloped. That is the firm is a price maker to an extent. However, the demand is highly price elastic since there are many close substitutes for the product.

Learning Tip

The price elasticity of demand is greater and the demand curve is flatter the smaller is the degree of product differentiation.

The best level of output and price is found when the firm sets its MR = MC, where the short-run economic profit is shown by the shaded are.

Figure 8-8



Note that a monopolistically competitive firm can make a positive economic profit, or a negative economic profit (loss), or zero economic profit in the short run. In the above case in Figure 8-8, the firm makes a positive economic profit since P > ATC at the production point, Q_1.

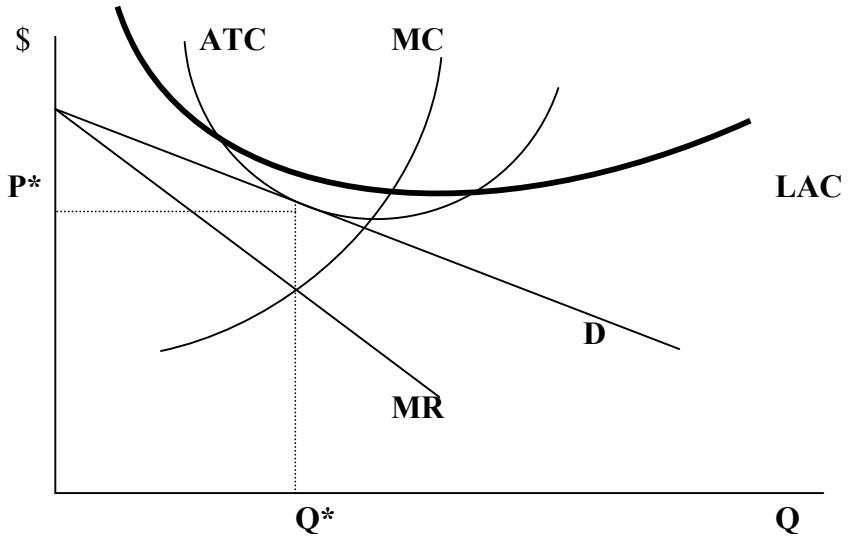
4.2 Long-Run Price and Output Determination

Because there is free entry, if firms earn short-run profits in a monopolistically competitive industry, additional firms will enter the industry in the long run to capture some of those profits. Similarly, if existing firms incur losses, in the long run firms will exit the industry.

An industry in which monopolistically competitive firms are making an economic profit in the short run adjusts in two different ways. First, the number of firms in the industry increases. Second, the existing firms expand to take advantage of the profit. Entry by new firms is a rational decision by investors who respond to market incentives (signals) prompted by an economic profit. By the same token, economic profit also serves as the incentive (signal) to the existing firms to expand their existing operation. On the other hand, the presence of short-run losses prompts a different set of rational responses. In this case, on the one hand, the number of firms in the industry decreases as some existing firms exit, and on the other hand, the existing firms contract the scale of their business.

The process just described continues until ultimately the market price is such that all firms in the market earn zero economic profits. The best level of output in the long run is, therefore, the one at which price equals ATC (short-run average cost curve) as well as LAC (the long-run average cost curve). This is shown in Figure 8-9. At the price of P^* each firm receives just enough to cover the average costs of production (ATC) because in the long run there is no distinction between fixed and variable costs, and economic profits are zero.

Figure 8-9



In the long run, monopolistically competitive firms produce at Q*, where two conditions are satisfied: (1) P > MC = MR, and (2) P = ATC > minimum of average costs. The first condition is the requirement for profit maximization that is the objective of the firm in the short run as well as in the long run. However, unlike perfect competition, the point of

optimality occurs to the left of the minimum of ATC. Therefore, P > MC implying that monopolistically competitive firms produce less output than is socially desirable. The second condition ensures that firms earn zero economic profits just as firms in perfectly competitive markets do. Even though the firms have some control over price, competition among them leads to a situation where no firm earns more than its opportunity cost of producing.

Learning Tip

There are two important differences between a monopolistically competitive market and a market serviced by a monopolist. First, while a monopolistically competitive firm faces a downward-sloping demand for its product, there are other firms in the industry that sell similar products. Second, in a monopolistically competitive industry, there are no barriers to entry. As we will see later, this implies that firms will enter the market if existing firms earn positive economic profits.

4.2.1 Demonstration Problem

You are the manager of a monopolistically competitive firm, and your demand and cost functions are given by $P = 10 – 1/2\ Q$ and $TC(Q) = 100 – 20Q + Q^2$.

a. Find the profit-maximizing price and output, and the maximum profit.

b. What are the adjustments in the long run?

Answer:

a. Setting MR = MC to obtain Q

MR (= 10 – Q) = MC (= –20 + 2Q), hence, Q = 10 and P = \$5. Profit = TR – TC = 10(\$5) – (100 –20(10) + 10^2) = \$50.

b. In the long run, new firms enter causing the existing firms' market share to diminish; the demand curve shifts to the left, shifting the MR curve with it. Economic profit disappears and each firm will make only normal profit, where MR = MC coincides with P = ATC.

5 Oligopoly

Oligopolies are markets in which there are only a few sellers of a homogenous or differentiated product. In the real world, the great majority of market situations are oligopolies. Prominent examples are the automobile, steel, aluminum, chemical, breakfast cereals, cigarettes, and many other industries. The homogenous products include steel, whereas differentiated products include breakfast cereals and automobiles.

Oligopolies are mutually dependent because there are only a few of them sharing a particular market, and any one firm's sales gain, resulting from a price reduction, for example, will be accompanied by sales losses for its rivals. The rival firms must be expected to react to their loss of sales, probably by also reducing price to win back their lost customers. Clearly, such myopic price-cutting could precipitate a 'price war' which could inflict heavy financial losses on all firms. Conversely, a price increase that is ignored by rivals may lead to a substantial loss of market share and profitability, since at least some of its customers will switch to a rival firm's product.

When General Motors, for example, introduces price rebates in the sale of its automobiles, Ford and Daimler Chrysler are likely to follow with their own rebates. Or when Honda makes an aggressive move to increase production of its Accord automobiles, Toyota and Ford might react by cutting their production of Camry and Taurus, respectively. Therefore, oligopolists should attempt to predict the probable reaction of rivals to its strategic actions and decide whether or not an action is likely to be worthwhile in the final analysis before taking that action.

As a general rule, oligopolists can compete on the basis of price, or non-price, such as product differentiation, quantity, advertising, service, etc. However, since price competition can be destructive, there is a tendency for oligopolists not to engage in price competition.

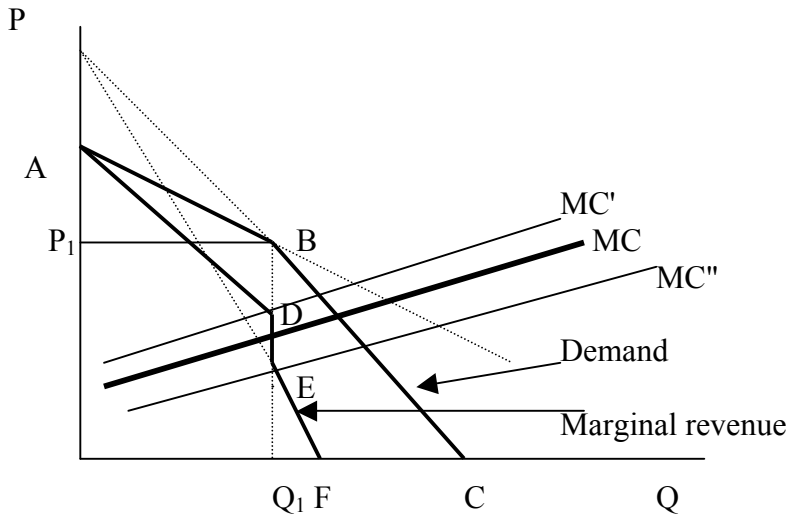
5.1 The Kinked Demand Curve Model of Oligopoly

We shall first examine the *kinked demand curve model.* Also known as Sweezy Oligopoly, this model is characterized by (a) few firms serving the market, (b) differentiated products, and (c) presence of barriers to entry. Furthermore, this model assumes that the firm's actions in regard to a price change may invite an asymmetrical reaction by other firms. For price increases, the firm will very likely lose most of its customers as there will be no reaction from rivals, since the rival firms will be content to sit back and receive the extra customers who switch away from the firm raising its price. However, price reductions by the firm will not increase its market share since its rivals will likely match the price reduction to maintain their own share of the market.

In conclusion, other firms do not react to this firm's price increases, the firm's demand curve above the current price level will be relatively elastic, while below the current price level, the demand curve will be relatively inelastic since the firm's market share does not change as other firms match its price reduction. Therefore, the demand curve faced by the firm has a kink at the current level of price.

In Figure 8-10, we show a firm's current price and output levels as P_1 and Q_1. For prices above P_1, the firm envisages the relative elastic demand curve shown by the line AB. For prices below P_1, it envisages the relatively inelastic demand curve shown by the line BC. The demand cure facing the firm is, therefore, ABC, being kinked at the current price level. The marginal revenue curve corresponding to this demand curve will have two separate sections. The upper section of the marginal revenue curve, shown as AD, relates to the upper section of the demand curve and therefore shares the same intercept and has twice the slope of the line AB. The lower section, EF, relates to the lower section of the demand curve and is positioned such that it has twice the slope of the line BC, and if extended up to the price axis would share its intercept point with the line BC similarly extended. Note that the kink in the demand curve, B, causes the DE discontinuity in the marginal revenue. The best level of output of the oligopolist with marginal cost curve MC is Q_1, found at the intersection of MC with the vertical section of MR. The oligopolist then charges the price of P_1, given by point B (at the kink on the demand curve).

Figure 8-10



Learning Tip

You will note that there is a vertical discontinuity in the marginal revenue curve, shown as the gap DE in Figure 8-10. Given the foregoing, it is apparent that the length of this gap depends on the relative slopes of the two segments of the demand curve, which in turn is related to the elasticity of demand under the two alternative scenarios of not reacting and matching behaviour.

An important implication of the Sweezy model is that there is a range (DE) over which changes in marginal cost do not affect the profit-maximizing level of output. That is, the marginal cost curve can rise to MC' or fall to MC" within this range without inducing the

oligopolist to change the prevailing price of P_1 and sales of Q_1 units (as long as P > AVC). Only when the marginal cost curve shifts above the MC' or below MC" will the oligopolist be induced to change its optimum combination of price and quantity.

Learning Tip

Provided marginal costs remain in a given range, the Sweezy model holds that firms have an incentive *not* to change their pricing behaviour. This is based on the assumption that rivals will match price cuts not price increases. Firms in this model do not want to change their prices because of the effect of price changes on the behaviour of other firms in the market.

Learning Tip

The Kinked demand curve model is not a complete theory of price determination, because it is unable to tell us how the firm arrives at the initial price and output levels. However, it is useful in that it shows us how strategic interactions among firms can have a profound impact on pricing decisions.

5.1.1 Demonstration Problem

Safe Ride Products markets its products in an oligopoly in which all makers of similar products are keenly aware of what each other is doing. Safe Ride Products perceives the demand function for its new road-emergency kit to be a kinked line with two different slopes.

Above the kink, $D_1 : Q_1 = 85 - P_1$

while below the kink, $D_2 : Q_2 = 32.5 - 0.25P_2$

Where, Q = Output in thousands of units, P = Price in dollars.

The firm's total cost, TC, is

$TC = 375 + 25Q + O.6Q^2$

a. Using the kinked-demand-curve model, what is the firm's output and price at the kink?

b. Are this output, price, and profit optimal?

c. Graph this problem.

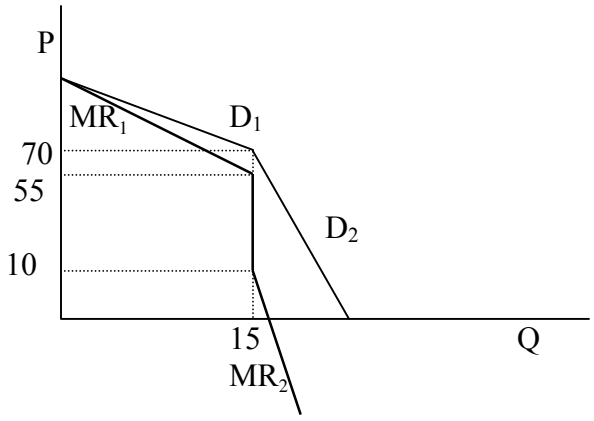
Answer:

a. The kink occurs at the intersection of the two differently sloped demand curves. But we first must invert the demand equations: $P_1 = 85 - Q_1$, and $P_2 = 130 - 4Q_2$. Now setting $P_1 = P_2$, and treating Q_1 and Q_2 as the same quantity: $85 - Q = 130 - 4Q$, $Q = 15$, and hence $P_1 = P_2 = \$70$.

Profit = $[(70(15)] - [375 - 25(15) - .6(15)^2] = \$165(000)$

b. Optimality requires that MR = MC. The upper point on the discontinued vertical MR curve is obtained by first plugging into MR_1, $MR_1 = 85 - 2Q = 85 - 2(15) = \$55$, whereas the lower point on this discontinued segment of the MR curve is found by plugging into the $MR_2 = 130 - 8Q = 130 - 8(15) = \$10$. Since at the current level of output (15), $MC = 25 + 1.2Q = 43$ falls between 10 and 55, the price, and output found above are those of the optimal situation.

c.



5.2 Non-Co-operative Oligopoly

Most work on oligopoly theory has concentrated on single-period. Such models are appropriate for markets that last for only brief periods of time, so that rival firms compete once, but never again. In such models, complex, long-run strategies and reputations for hard-nosed competition are irrelevant.

The three best-known single-period oligopoly models that use the concept of a Nash equilibrium, Cournot, Bertrand, and Stackelberg models, are discussed below. Cournot and Stackelberg are models of quantity competition, whereas Bertrand is a model of price competition.

5.3 Nash Equilibrium

John F. Nash (1951) defined the most widely used equilibrium concept. A set of strategies is called a **Nash equilibrium** if, holding the strategies of all other firms constant, no firm can obtain a higher payoff (profit) by choosing a different strategy. Thus, in a Nash equilibrium, no firm *wants* to change its strategy.

In the Cournot and Stackelberg models, firms' strategies concern setting quantities. In the Bertrand model, firms set prices. The Nash equilibrium concept is useful when strategies include setting advertising or other variables in addition to output or price.

5.4 Quantity Competition

5.4.1 The Cournot Model

The French mathematician Augustin Cournot presented the first – and probably still the most widely used – model of non-co-operative oligopoly in 1838. Cournot assumed that each firm acts independently and attempts to maximize its profits by choosing its output. The discussion focuses on the *duopoly*, or two-firm case.

Consider a Cournot duopoly for microprocessors characterized by the following conditions:

- *No entry*: There are two firms and no entry by other firms is possible (these firms own the only available technology in the market).
- *Homogeneity*: The firms produce identical (homogenous) microprocessors, so the sum of their outputs equals industry output: $Q = Q_1 + Q_2$, where Firm 1 produces Q_1 and Firm 2 produces Q_2.
- *Single period*: This market and the two firms only exist for one period.
- There are *a few buyers* in their market.
- Each firm believes rivals *will hold their output* constant.

To highlight the implications of Cournot oligopoly, suppose each duopoly must make an output decision, and each firms believes that its rival will hold output constant as it changes its own output. To determine its own output level, Firm 1 will equate marginal revenue with its marginal cost. Notice that since this is a duopoly, Firm 1's marginal revenue is affected by Firm 2's output level. In particular, the greater the output of Firm 2, the lower the market price and thus the lower is Firm 1's marginal revenue. This means that the profit-maximizing level of output for Firm 1 depends on Firm 2's output level and the profit-maximizing level of output for Firm 2 depends on Firm 1's output. These interrelationships are referred to as reaction functions.

A *reaction function* defines the profit-maximizing level of output for a firm for given (expected) output levels of the other firm. It tells us a firm's best response to the output level of a rival firm.

What strategy should Firm 1 use to choose its output level? The answer depends on its belief about Firm 2's behaviour. If Firm 1 believes that Firm 2 will sell Q_2 processors, it can determine Q_1 that will maximize its profit. Firm 1 can sell all but Q_2 units of the amount demanded by the market; that is, it faces the *residual demand curve*, Q_1 = market demand $-Q_2$, which is the market demand curve minus the expected output of Firm 2, Q_2.

Note that Firm 1 has a monopoly over those consumers whose demands are not met by Firm 2. We can summarize the relationship between Firm 1's profit-maximizing quantity and Firm 2's quantity in an equation,

$$Q_1 = R_1(Q_2) \tag{7}$$

which is the *best-response function* (or reaction function) showing the best (highest profit) action (output) by a firm given its beliefs about the action its rival takes.

Similarly, Firm 2's best-response function is derived in a similar way. Since the firms are identical (same costs, identical products), Firm 2's reaction function is the mirror image of Firm 1's:

$$Q_2 = R_2(Q_1) \tag{8}$$

Accordingly, Firm 2's choice of output depends on the output it expects Firm 1 to produce.

Suppose the market demand in a Cournot model is represented by

$$P = 100 - Q \tag{9}$$

where $Q = Q_1 + Q_2$. Each firm has a constant marginal cost, $MC_1 = MC_2 = \$10$ and no fixed costs. Each firm can produce enough output to meet the entire market's demand. Producers set their output with an eye on the market price. But, since the market price depends on total output produced by both firms, how much Firm 1 produces depends on how much it expects Firm 2 to produce. Put differently, Cournot equilibrium is the situation where neither firms has an incentive to change its output given the output of the other firm.

We can determine the reaction function for each firm as follows. To maximize profit each firm sets marginal revenue equal to marginal cost.

$$\begin{aligned} &TR_1 = (100 - Q_1 - Q_2)xQ_1 = 100Q_1 - Q_1^2 - Q_1Q_2 \\ &and \\ &MR_1 = 100 - 2Q_1 - Q_2 \end{aligned} \tag{10}$$

Now setting MR_1 equal to MC_1 (=10) and solving for Q_1:

$$100 - 2Q_1 - Q_2 = 10$$

Firm 1's reaction function (R_1):

$$Q_1 = 45 - 0.5Q_2 \tag{11}$$

The same calculation applies to Firm 2:

$$TR_2 = (100 - Q_1 - Q_2)xQ_2 = 100Q_2 - Q_2^2 - Q_1Q_2$$
$$and \tag{12}$$
$$MR_2 = 100 - 2Q_2 - Q_1$$

Setting MR_2 equal to MC_2 (=10) and solving for Q_2:

$$100 - 2Q_2 - Q_1 = 10$$

Firm 2's reaction function (R_2):

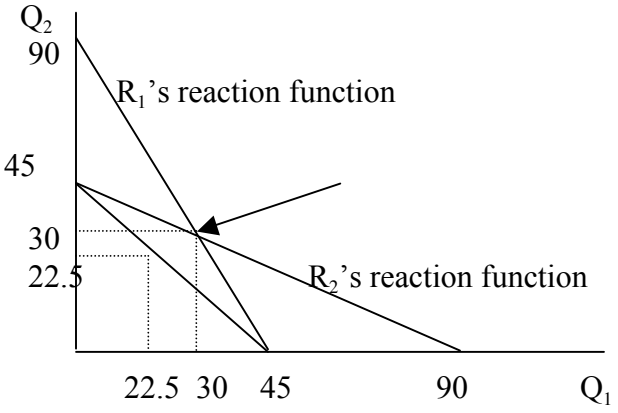
$$Q_2 = 45 - 0.5Q_1 \tag{13}$$

The equilibrium output levels are the values for Q_1 and Q_2 that are at the intersection of the reactions functions.

$$Q_1 = Q_2 = 30$$
$$P = \$40,$$
$$\pi_1 = \pi_2 = \$900$$

If each firm believes the other will produce 30 units and if each firm produces 30 units, neither firm wants to change its output. A firm is unwilling to produce at a point not on its reaction function because doing so would result in a lower profit. The only point where both firms are on their best-response functions is at the intersection of the best-response functions. A non-intersection point cannot be an equilibrium; an equilibrium point is one in which neither firm wants to change its behaviour. In the Cournot equilibrium, total market output is 30 + 30 = 60 processors and the price is \$40, Figure 8-11.

Figure 8-11



To understand why reaction functions are shaped as they are, let us highlight a few points in Figure 8-11. First, it is clear from Firm's 1 reaction function, equation (11), that if Firm 2 produced zero units of output, the profit-maximizing level of output for Firm 1 would be 45 (on horizontal axis). At this point ($Q_1 = 45$ and $Q_2 = 0$), Firm 1 is a monopolist. Similarly, it is clear from Firm's 2 reaction function, equation (13), that if Firm 1 produced zero units of output, the profit-maximizing level of output for Firm 2 would be 45 (on vertical axis). At this point ($Q_2 = 45$ and $Q_1 = 0$), Firm 1 is a monopolist.

In a model in which firms only choose output levels, any output levels at which no firm believes it can increase its profits by increasing or decreasing its output are, by definition, a Cournot equilibrium, and no combination of outputs could be an equilibrium except a Cournot equilibrium. Thus, where firms must choose output levels independently, the Cournot equilibrium is not only realistic, it is the only plausible equilibrium.

Because the Cournot equilibrium is a special case of the Nash equilibrium where firms have strategies over quantities, it is often referred to as a *Cournot-Nash equilibrium* or a *Nash-in-quantities equilibrium*.

5.4.2 Collusion

The firms are worse off and the consumers are better off at the Cournot equilibrium than if firms act collusively as a cartel (monopoly). Firms can benefit at the expense of consumers by agreeing to restrict output or, equivalently, charge higher prices. Such an act by firms is known as *collusion.*

If the two firms decide to collude and antitrust laws do not prevent them from doing so, they would set their output to maximize *total* profit, behaving like a monopolist and presumably split the profit evenly. This is done by choosing total output Q so that marginal revenue equals marginal cost, which equals 10 in this example.

$$TR = P \times Q = (100 - Q) \times Q = 100Q - Q^2$$

Marginal revenue is therefore

MR = 100 – 2Q = (MC = \$10)

Q = 45, and hence $Q_1 = Q_2 = 22.5$

P = \$55

Each firm earns profit of \$1,012.50.

As expected, under collusive mechanism, which is a co-operative scenario, both firms produce less, charge a higher price, and naturally earn more profit than in the Cournot equilibrium.

5.5 The Stackelberg Model

So far, we have assumed that our two duopolists make their output decisions at the same time. In the Stackelberg oligopoly, firms may set their decisions in sequence. Specifically, one firm is assumed to make an output decision before the other firm does. The former is referred to as the *leader* and the latter as the *follower.* The implication is that the follower takes the leader's output as given, and chooses its own output that maximizes profit. That is, the follower behaves in a Cournot fashion. The leader, however, observes the follower's output in setting a level of output that maximizes its own profit. Other conditions required for the Stackelberg Oligopoly are similar to those of Cournot's.

Let's start with Firm 2. Since Firm 2 is the follower and sets it output after the leader, its profit maximizing level of output is determined by its reaction function:

$$R_2: Q_2 = 45 - 0.5Q_1$$

Firm 1 is the leader and goes first, *first mover,* anticipating Firm 2's move. That is, the leader will choose a point on Firms 2's reaction function that maximizes its own profit.

$$TR_1 = (100 - Q_1 - Q_2)xQ_1 = 100Q_1 - Q_1^2 - Q_1Q_2$$

and

$$TR_1 = 100Q_1 - Q_1^2 - Q_1(45 - 0.5Q_1) = 55Q_1 - 0.5Q_1^2$$

hence

$$MR_1 = 55 - Q_1 = MC_1 = 10, therefore$$

$$Q_1 = 45, and Q_2 = 22.5.$$

Therefore, the leader produces twice as much as the follower. Given that both charge the same price for their homogenous products, we can conclude that it pays to move first, *first mover advantage*.

P = 100 – (45 + 22.5) = \$32.5

Profits earned by Firm 1 is \$1,012.50 and by Firm 2 is \$506.25.

5.6 The Bertrand Model (Price Competition)

Cournot's work was well ahead of its time. The first major challenge to his book came in 1883, 45 years after it was published. In this critique, Joseph Bertrand argued that it is hard to see who sets prices in oligopolistic markets if the firms do not set them. Cournot, by having firms choose output rather than price, fails to explain explicitly the mechanism by which prices are determined (but, for that matter, so does a competitive model).

In Bertrand's model, firms set prices rather than output. If consumers have complete information and realize that firms produce identical products, they buy the one with the lowest price. In a Bertrand model, since each firm believes its rival's price is fixed by a slight price cut, the firm is able to capture all its rival's business. In the Bertrand equilibrium discussed below, firms make zero profits and no firm can increase its profits by raising or lowering its price, which, when it exists, is equivalent to the social optimum (competitive equilibrium) discussed above.

To illustrate the Bertrand equilibrium, let us make the same assumptions as in the Cournot example: no entry, homogenous products, and single period. We also assume the same demand and cost relationships as above. The only important change is that firms now set prices rather than quantities. Each firm is willing to sell as much quantity as is demanded at the price it sets.

To explain more precisely, because the products are identical, all consumers will purchase from the firm charging the lowest price. For concreteness, suppose Firm 1 charges the profit-maximizing price. However, by slightly undercutting this price, Firm 2 would capture the entire market and make positive profits, while Firm 1 would sell nothing. Therefore, Firm 1 would retaliate against Firm 2 by undercutting Firm 2's lower price, thus recapturing the entire market.

When would this 'price war' end? When each firm charged a price that equals marginal cost, $P_1 = P_2 = MC$. Given the price of the other firm, neither firm would choose to lower its price, for then its price would be below marginal cost and would make a loss. Also, no firm would want to raise its price, for then it would earn nothing. In short, Bertrand oligopoly and homogeneous products lead to a situation where each firm charges marginal cost and economic profits are zero. Note that the duapolists in this case would be acting as perfectly competitive firms.

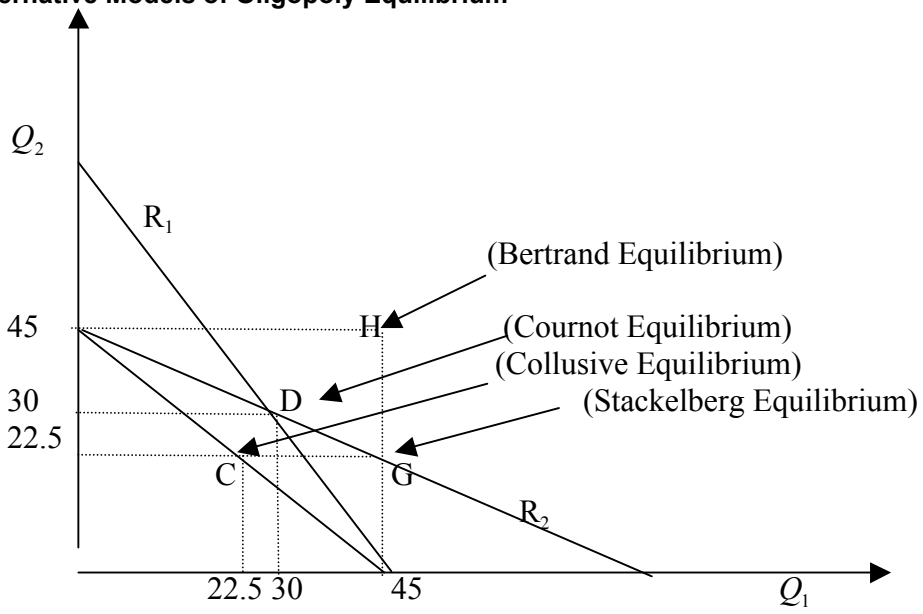
Therefore, with the given demand and cost functions,

$$P_1 = P_2 = MC = \$10$$
$$100 - (Q_1 + Q_2) = 10,$$
therefore,
$$Q_1 + Q_2 = 90, \textit{and}$$
$$Q_1 = Q_2 = 45$$

Each firm will earn zero economic profit.

The following Figure captures all these oligopoly models.

Figure 8-12 Alternative Models of Oligopoly Equilibrium



Points D, C, G, and H are, respectively, representative of the Cournot, Collusion, Stackelberg, and Bertrand equilibrium. Note that in all four cases, we have assumed homogenous products.

Learning Tip

Firms under the collusive mechanism (as in cartels) make the most profit and produce the least amount of output. Firms under Bertrand Oligopoly produce the most and charge the least. Therefore, from the perspective of producers, collusive equilibrium (co-operative mechanism), if allowed and workable, is the best scenario, while from the perspective of consumers, Bertrand equilibrium presents the best scenario.

Note that in practice, firms tend not to trust each other and, hence, are more likely to behave as Cournot Oligopolies rather than joint profit maximizing (collusive) firms.

6 Summary and Review

The process by which price and quantity are determined in the real world is a function of the market structure within which firms compete against each other. Four different market structures are identified: perfect competition, monopolistic competition, monopoly, and oligopoly. The key factors that differentiate these markets are the type of product, number and the nature of buyers and sellers, conditions of entry and exit to and from the industry, and the presence of economies of scale.

Under perfect and monopolistic market conditions, there are numerous firms competing in the industry selling products that are homogenous, in the former, or somewhat differentiated, in the latter. Furthermore, market entry and exit are free from all barriers. Under pure monopoly, there is only one seller of a product that has no substitute. Monopolies arise from patents, control of raw materials, etc. Since barriers to entry are so high, the monopolies can maintain their profit in the long run.

Oligopolies are characterised by a few sellers selling either differentiated or homogenous products. The main feature of oligopolistic market structure is interdependency amongst the firms.

7 Review Problems

1. Suppose a manager of a firm is risk neutral. Explain how the pricing strategy by the manager will differ from the perfectly competitive situation, where the price is set on the basis of equating marginal cost and marginal revenue (note: marginal revenue for a firm in competitive setting is the price itself).

2. A cellular phone provider estimates the following pricing situation for the coming year:

   a. Boom (30% probability) – price per set $40.00

   b. Unchanged (50% probability) – price per set $30.00

   c. Recession (20% probability) – price per set $20.00

   The cost function for the firm is $C(Q) = 10 – 5Q^2$. What should be the optimal output decision if the firm maximizes expected profit? Calculate the expected profit of the firm.

3. How does a monopolistically competitive firm use the measure of elasticity in pricing decisions? If a firm measures the elasticity of demand to be –2.5 and has a cost function $C(Q) = 50 + 4Q^2$, then how much should the firm charge the consumers for its products?

4. In a Cournot duopoly situation, the firms have the same cost structure of $C(Q) = 48 + 2.5Q$. If the elasticity of demand is –1.8, then calculate the optimum price in this situation.

5. For a firm maximizing its profits in a perfectly competitive situation, find the conditions of profit maximization in terms of price and marginal cost. Why will the firm not sell a quantity where the price is greater or less than the marginal cost?

6. If a firm has a cost function $C(Q) = 100 + 4Q^2 – 2Q$ and faces a price of \$150 on a perfectly competitive world, then what is the level of output that it should produce? Find the profit at this level of output.

7. If a firm earns positive economic profit in the short run in a competitive scenario, what is likely to happen in the long run? Explain the long run equilibrium conditions.

8. A monopolist always operated in the region where the elasticity of demand is greater than one – explain why.

9. What is the condition of profit maximization for a monopolist? Provide an intuitive explanation of the profit maximization rule.

10. If a monopolist faces an inverse demand function $P = 60 – 20Q$ and if the firm's cost structure follows the function $C(Q) = 5Q^2 + 10Q$ then what will the optimal level of output be? Find the profit for the firm at this level of output.

11. A monopolist has two separate markets under its coverage. The inverse demand functions for the firm's products are as follows:

    a. market A: $P_a = 20 – 10Q_a$

    b. market B: $P_b = 15 – 15Q_b$

    Should P_a and P_b be the same, if the monopolist exercises its power in both the markets? Find the values of P_a and P_b in this case, assuming that MC =0, and there are no fixed costs.

12. Suppose there are only two firms in an industry producing homogeneous products and facing the inverse demand function $P = 100 – 5Q$ (Q: Total output; Q_1: Firm 1's output; Q_2: Firm 2's output). Marginal costs of both the firms are equal to 4. Find the (a) Cournot equilibrium and (b) Stackelberg equilibrium.

13. In the situation described in question (10), what should be the Bertrand and the collusive equilibria?

8 Answer Key to Review Problems

1. Under an uncertain situation the manager of a firm makes an expectation about the marginal revenue, based on the expected price and demand, and equates the expected marginal revenue with marginal cost. Thus the profit-maximizing rule is similar to market models without uncertainty. If E[MR] is expected marginal revenue and MC is marginal cost, then the rule is E[MR] = MC.

2. $E[P] = 40(0.3) + 30(0.5) + 20(0.2) = 31$

   $MC = 10Q$

   Setting E[P] = MC, we get $Q = 3.1$

   $E[\text{Profit}] = 31(3.1) - 10 - 5(3.1)^2 = 38.05$

3. $MR = P[1+1/E_f] = MC$

   $MC = 4$, $E_f = 2.5$

   Therefore, $P = 6.67$.

4. $P = [\ nE_m/(1+nE_m)\ ]\ MC$

   $MC = 2.5$ $E_m = -1.8$, $n = 2$, $P = 3.46$.

5. To maximize profit, the firm sets MR = MC. However, since in perfect competition, the firm faces a horizontal demand curve, MR = P, therefore, P = MR = MC. When P > MC, the firm is in a position to increase its total profit by producing yet additional units of output, which otherwise would remain unexploited. Similarly, when P < MC, the firms has produced too much. By cutting back on its production, the firm can cut its loss, or increase its profit.

6. P = MC. Plugging in: $150 = 8Q – 2, and hence Q = 9. Profit = TR (19 x $150) – TC ($100 + 4 \times 19^2 - 2 \times 19$) = $2850 – $1506 = $1344.

7. Firms will enter the industry to earn above normal returns. This causes the existing firm's share of output to diminish, and hence the price they can charge to fall. Assuming that factor costs do not change in response to the entry, cost curves will not be affected. The price will keep falling until profit per unit equals zero, at which point, the entry should stop, long run.

8. A monopolist always operated in the region where the elasticity of demand is greater than one.

$$MR = P\left(1 - \frac{1}{|E|}\right)$$

$MR > 0, P > 0$

$\Rightarrow 1 - \frac{1}{|E|} > 0$

$\Rightarrow |E| > 1$

Therefore, a monopolist always operates in the elastic region.

9. MR=MC is the profit maximizing condition. If Q is at a level where MR <MC, then revenue generated from the last unit is not able to offset the cost of the last unit. Therefore profit is not maximized at this level. On the other hand, if MR >MC, then revenue from producing another unit will be greater than the cost of that unit. Therefore profit can be increased. Therefore MR = MC is the profit maximizing condition.

10. Monopolist: optimum level of output and profit at that level.

$P = 60 - 20Q$

$\Rightarrow R = 60Q - 20Q^2$

$\Rightarrow MR = 60 - 40Q$

$MC = 10Q + 10$

Therefore at MR = MC,

$10Q + 10 = 60 - 40Q$

$\Rightarrow 50Q = 50$

$\Rightarrow Q = 1$

P (at Q = 1) = 40

Cost (at Q = 1) = 15

Profit (at Q = 1) = 25

11. A monopolist has two separate markets.

$MR_a = MC = 0$, $20 - 20Q_a = 0$. Hence, $Q_a = 1$, $P_b = \$10$, Profit = \$10.

$MR_b = MC = 0$, $15 - 30Q_b = 0$. Hence, $Q_b = 0.5$, $P_b = \$7.5$, Profit = \$3.75.

12. Two firms in an industry producing homogeneous products.

$P = 100 - 5Q = 100 - 5(Q_1+Q_2)$

Revenue 1 = $100Q_1 - 5Q_1^2 - 5Q_1Q_2$

Marginal Revenue 1 = $100 - 10Q_1 - 5Q_2$

Similarly, Marginal Revenue 2 = $100 - 10Q_2 - 5Q_1$

MR1 = MC

Or, $100 - 10Q_1 - 5Q_2 = 50$ …… $Q_1 = 5 - 0.5Q_2$ (reaction function of 1)

Similarly, $Q_2 = 5 - 0.5Q_1$ (reaction function of 2)

Cournot Equilibrium is at the intersection of these two curves,

$$Q_1 = 5 - 0.5Q_2 = 5 - 0.5(5 - 0.5Q_1)$$
$$\Rightarrow Q_1 = 5 - 2.5 + 0.25Q_1$$
$$\Rightarrow Q_1 = 3.33$$
$$\Rightarrow Q_2 = 3.33$$
$$\Rightarrow P = 100 - 5(Q_1 + Q_2) = 66.67$$

Stackelberg Equilibrium,

$$Q_2 = 5 - 0.5Q_1$$

$$\text{Profit}(1) = \Pi(1) = PQ_1 - Cost = [100 - 5(Q_1 + Q_2)]Q_1 - 50Q_1$$
$$\Rightarrow \Pi(1) = 25Q_1 - 2.5Q_1^2$$
$$\Rightarrow \frac{\partial \Pi(1)}{\partial Q_1} = 25 - 5Q_1 = 0$$
$$\Rightarrow Q_1 = 5 \Rightarrow Q_2 = 2.5 \Rightarrow P = 62.5$$

13. Bertrand and the collusive equilibria.

Bertrand Equilibrium, P = MC = 50 = 100 –5Q or Q = 10

Both firms will share the quantity, given same structure, and therefore each firm will sell 5 units.

Collusion Equilibrium

MC of both firms = 50

MR of market = $\frac{\partial(PQ)}{\partial Q} = \frac{\partial}{\partial Q}(100 - 5Q)Q = 100 - 10Q$

Setting MR=MC, we get 100-10Q = 50 …… Q = 5

Both firms will share the equilibrium quantity (5) and each will sell 2.5 units.

Price in the collusion equilibrium = 100 – 5(5) = 75.

9 References and Suggested Further Reading

Baumol, W. J., *Business Behavior, Value and Growth*, Rev. Ed. New York: Harcourt Brace Johanovich, 1967.

Baye, M., *Managerial Economics and Business Strategy*, McGraw Hill, Irwin, 2002.

Besanko, D. and Braeutigam, R., *Microeconomics: An Integrated Approach*, Wiley, 2002.

Douglas, J. E., *Managerial Economics: Analysis and Strategy*, Prentice Hall, 1992.

Pindyck, R. and Rubinfeld, D., *Microeconomics*, Fifth Edition, Prentice Hall, 2001

Robinson, J., *Economics of Imperfect Competition*, London, McMillan, 1933.

Stigler, G.J., 'The Kinky Oligopoly Demand Curve and the Rigid Prices,' *Journal of Political Economy*, 1947.

Commonwealth of Learning Executive MBA/MPA

E5 Managerial Economics

Block 9

Pricing Practices

Course author: Farrokh Zandi, PhD
Professor, Schulich School of Business
York University
Toronto, Canada

Course designer: Douglas R. Cronk, M.A., ABD
Vancouver, BC, Canada

Copyright © Commonwealth of Learning, 2003
All rights reserved. No part of this course may be reproduced in any form by any means without prior permission in writing from:

The Commonwealth of Learning
1285 West Broadway
Suite 600
Vancouver, BC V6H 3X8
CANADA
e-mail: info@col.org

Contents

1 Block Nine: Objectives and Introduction 1
1.1 Introduction 1
2 Cost-Plus Pricing 2
2.0.1 Demonstration Problem 4
3 Extracting Consumer Surplus 4
3.1 Price Discrimination 4
3.1.1 First-Degree Price Discrimination 5
3.1.2 Demonstration Problem 6
3.1.3 Second-Degree Price Discrimination 7
3.1.4 Third-Degree Price Discrimination 8
3.1.5 Demonstration Problem 10
3.1.6 Demonstration Problem 12
4 Two-Part Tariffs 12
5 Pricing Product Bundles 14
5.0.1 Demonstration Problem 15
5.1 Decentralization and Transfer Pricing 18
5.2 Transfer Pricing with No External Market for the Intermediate Product 18
5.2.1 Demonstration Problem 20
5.3 Transfer Pricing with a Perfectly Competitive External Market 20
5.3.1 Demonstration Problem 22
5.4 Transfer Pricing with an Imperfectly Competitive External Market 22
6 Summary and Review 24
7 Review Problems 24
8 Answer Key to Review Problems 25
9 References and Suggested Further Reading 28

1 Block Nine: Objectives and Introduction

After working through Block Nine of this course, you should be able to:

1. Explain the objectives of pricing strategies.
2. Explain how an imperfectly competitive firm is able to apply various methods of pricing strategy.
3. Explain the conditions under which a firm may be able to successfully price discriminate.
4. Analyze the underlying strategy of the two-part pricing and the bundling.
5. Discuss the significance of transfer pricing.
6. Explain how the firm's transfer-pricing strategy changes with the underlying market conditions.

1.1 Introduction

In previous blocks, we saw that firms in order to maximize profits produce where marginal revenue (MR) equals marginal cost (MC) and then charge the price indicated on the demand curve they face. This is true under all types of market structures, except when firms in oligopolistic markets may pursue other objectives, such as sales revenue maximization in the long run. Throughout our discussions, however, we employed a framework that assumed firms would (a) have perfect information about the buyers and sellers, and hence about demand and costs functions, or (b) sell in only one market, or (c) produce only one product. None of these assumptions, however, is generally true for most firms today. That is, most firms produce more than one product, sell products in more than one market, are organized (at least large corporations) into a number of decentralized profit centres, and have only a general rather than a precise knowledge of the demand and cost curves they face. As a result, our discussion of the pricing decision presented in the previous block must be expanded to take into consideration actual pricing practices.

In this Block, we examine the firm's pricing under imperfect knowledge, cost-plus pricing that approximates the MR = MC rule, price discrimination, price strategies that aim to capture consumer surplus such as bundling, two-part pricing (pricing of multiple products), and transfer pricing or the pricing of (intermediate) products transferred between the firm's divisions.

2 Cost-Plus Pricing

The most widely used method of pricing is known as *cost-plus pricing*. It is a procedure whereby the price is determined by adding a fixed mark-up of some kind to the cost of producing or acquiring the product. Thus development of a cost-plus price requires two basic steps: determination of the relevant cost and determination of what the 'plus' should be.

The formula for the mark-up percentage is

$$Markup = \frac{\text{Price} - Cost}{Cost} \tag{1}$$

For example, if the cost (average variable cost = AVC) of a product that sells for \$12 is \$10, the mark-up on cost would be

$$= \frac{12 - 10}{10} = 0.2 = 20\%$$

Suppose the percentage of mark-up is denoted by X. Rearranging equation (1) yields

$$\begin{aligned} &X = \frac{P - AVC}{AVC}, \\ &or \\ &P - AVC = X.AVC \\ &and \\ &P = AVC + X.AVC \\ &or \\ &P = AVC(1 + X) \end{aligned} \tag{2}$$

This mark-up, in dollar terms, is the per-unit contribution to overheads and profits, and hence choosing the size of the mark-up amounts to choosing the contribution margin. As a matter of practice, the mark-up should be large enough so that, at the actual level of sales volume, the total contributions to overheads and profits actually cover the overheads and allow a profit to be made.

The size of the mark-up is constrained, however, by the willingness of consumers to pay the higher prices associated with a higher mark-up. Even a firm that has a monopoly over a certain product must acknowledge that at higher prices consumers in aggregate will buy fewer units. Oligopolists must take into account the relative prices of competitors. An individual firm will lose sales to rivals if its price is significantly above the prices of its rivals and if consumers do not think the item is worth the extra money.

Mark-up pricing is often thought to be simply cost based, but it is evident that the amount by which price can be marked up is highly dependent on the demand conditions facing the firm. When asked what factors determine the size of the mark-up percentage, businesspeople often respond that they choose the mark-up with an eye to 'what the market will bear.' These statements carry an implicit message about the price elasticity of demand facing the firms. Therefore, the size of the mark-up is both cost as well as demand based, contrary to the naive view that mark-up pricing depends on cost alone.

At this juncture, two questions arise:

1. Does the practice of mark-up pricing ignore the marginalist principles of pricing?
2. And if not, is there any way to determine if cost-plus pricing is maximizing?

The answer to the first question is NO; it does not ignore marginalist principles. To the second question, the answer is YES, if we are able to determine the price elasticity of the product.

First, let us remember that

$$MR = P(1 - 1/E_P) \tag{3}$$

where E_P is the absolute value of the price elasticity of demand. When profit is maximized, MR = MC. Therefore, we can substitute MC for MR in equation (3)

$$MC = P(1 - 1/E_P) \tag{4}$$

If, we assume, for simplicity, that the cost function is linear (AVC = MC), we will have

$$AVC = P\,(1 - 1/E_P) = P(E_P - 1)/E_P]$$

Isolating P

$$P = AVC\,[E_P/(E_P - 1)]$$

or

$$P - AVC = AVC[(E_P/(E_P\text{-}1)] - AVC$$

and therefore,

$$\frac{P - AVC}{AVC} = \frac{1}{E_P - 1} \tag{5}$$

Equation (5) indicates that the percentage of mark-up varies with the price elasticity of demand.

From this analysis we can draw the following conclusions:

1. Different products with different elasticities should have different percentages of mark-up even if they have the same marginal cost.

2. The more price-elastic the product, the smaller should be the rate of mark-up.

3. If the mark-up rate is chosen correctly, it will give exactly the profit-maximizing price. For example, for a product that has a price elasticity of 3 (in absolute value), a mark-up of 50 per cent ($\frac{1}{3-1} = \frac{1}{2}$) would be consistent with the marginalist principles. And for a product with a price elasticity of demand of 2, a mark-up of 100 per cent ($\frac{1}{2-1} = \frac{1}{1}$) would be consistent with the profit maximization principle.

2.0.1 Demonstration Problem

Suppose a retail sport store charges $30 for a jersey and the cost to the firm is $10. The firm has reasons to believe that the price elasticity of demand for its product at that price is 2 (in absolute value). Is this price optimal?

Answer:

The percentage of actual mark-up is

$$X = (30 - 10)/10 = 2 = 200\%$$

The optimal (profit maximizing) mark up, however, is

$$1/(Ep - 1) = 1/(2 - 1) = 100\%$$

Since the optimal mark-up rate of 100 per cent is less than the actual mark-up of 200%, the actual (current) mark-up and the resulting price are too high.

3 Extracting Consumer Surplus

3.1 Price Discrimination

Price discrimination describes in general a method that can be used by some sellers to tailor their prices to the specific purchasing situations or circumstances of their buyers. Specifically, it is defined as the practice by a seller of charging different prices to the same buyer at different times or to different buyers for the same good or service at the same time, without corresponding differences in cost. For analytical purposes, it is convenient to distinguish among three degrees of differential pricing.

Learning Tip

Note that if a firm's practice of charging a higher price during certain hours of day reflects the higher cost of serving the market during those hours, this practice does not constitute price discrimination. The differential pricing *not* associated with the underlying cost differences would be considered price discrimination.

3.1.1 First-Degree Price Discrimination

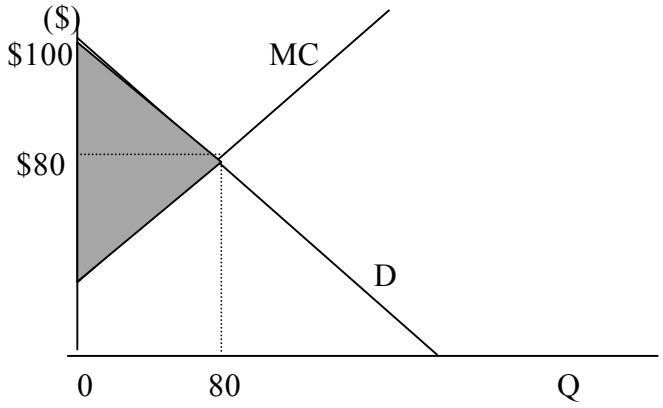
In differential pricing of the first degree, the seller charges the same buyer a different price for each unit bought, thereby extracting the consumer's maximum willingness to pay *reservation price*. By shading the price down to the buyer for each additional unit purchased, the seller obtains larger total revenue than if the same price per unit were charged for all units bought. Unfortunately for managers, first-degree price discrimination (also called perfect price discrimination) is extremely difficult to implement because it requires the firm to know precisely the maximum price each consumer is willing and able to pay for alternative quantities of the firm's product.

Nonetheless, some service-related businesses, including car dealers, mechanics, and lawyers, successfully practice a form of first-degree price discrimination. For example, when a firm sells a product at an auction, it is attempting to get consumers to bid up the price so that the consumer with the highest reservation price purchases the good. Also, most car dealers post sticker prices on cars that are well above the dealer's actual marginal cost, but offer 'discounts' to customers on a case-by-case basis. The best salespersons are able to size up customers to determine the minimum discount necessary to get them to drive away with the car. In this way they are able to charge different prices to different consumers depending on each consumer's willingness and ability to pay. This practice permits them to sell more cars and to earn higher profits than they would if they charged the same price to all consumers. Similarly, most professionals also charge rates for their services that vary, depending on their assessment of customers' willingness and ability to pay.

Figure 9-1 shows how first-degree price discrimination works. Each point on the market demand curve reflects the maximum price that consumers would be willing to pay for each incremental unit of the output, i.e., reservation price. Consumers start out with 0 units of the good, and the firm can sell the first incremental unit for $100. Since the demand curve slopes downward, the maximum price the firm can charge for each additional unit declines, ultimately to $80 at an output of 80 units. The first unit goes to the customer with the highest reservation price, $100. The firm would sell the first unit for $100 and captures all of the consumer surplus. The second unit is sold to the second unit to the person with the second highest reservation price of $99, and similarly, the third unit goes for $98, etc. This way, the seller charges each customer his or her highest reservation price for that unit, which is the height of the demand curve. The difference between each point on the demand curve and the firm's marginal cost represents the

profits earned on each incremental unit sold. Thus, the shaded area between the demand curve and the firm's marginal cost curve reflects the firm's total profit when it charges each consumer the maximum price he or she will pay for small increments of output between 100 and 80 units. This strategy allows the firm to earn the maximum possible profits. Notice that consumers receive no consumer surplus on the 5 units they purchase: the firm extracts all surplus under first-degree price discrimination. This practice is also referred to as *perfect price* discrimination.

Figure 9-1



3.1.2 Demonstration Problem

Suppose a perfectly discriminating monopolist faces the following demand curve, P = 30 – Q. The monopolist's marginal cost is MC = 10, and there are no fixed costs.

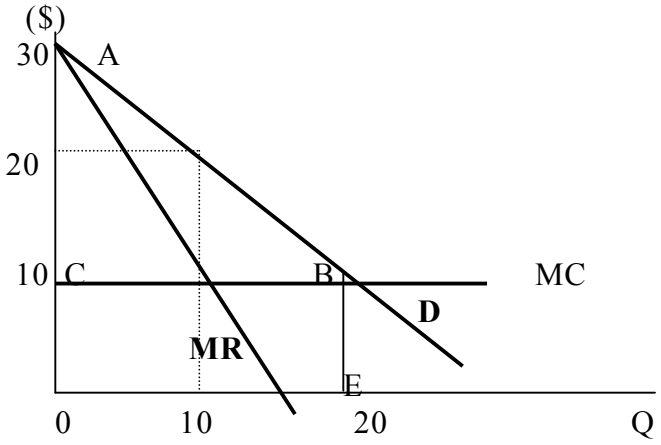
a. What are the profit maximizing price and quantity under first-degree (perfect) price discrimination? What is the gain to the monopolist?

b. What would be the price-maximizing price and quantity under simple (non-discriminating) solution?

Answer:

a. A first-degree price discriminating monopolist maximizes profit by setting P = MC, 30 – Q = 10, hence producing Q = 20, and charging the marginal Price (price of the last unit) of $10. At this point, the firm's total revenue is represented by the area ABC in Figure 9-2:

Figure 9-2



The numerical value of this area is $200. The producer receives the area of the triangle ABC ($200) plus the area of the rectangle OCBE (P x Q = 20 x 10 = $200), equal to $400, of which the cost of production is the rectangle OCBE. The net profit is, therefore, $200.

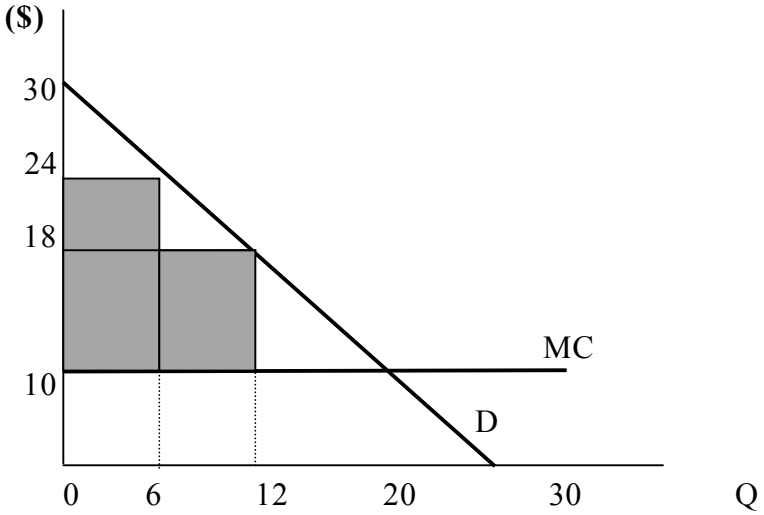
b. If the firm cannot price discriminate and instead charges a single price for all units sold, it will produce where MR = MC. MR = 30 – 2Q = MC =10, and Q = 10. Total revenue, in this case, equals $200 (10 x $20) and total cost equals $100 (10 x $10). Therefore, the profit is $100.

3.1.3 Second-Degree Price Discrimination

More practical and common is second-degree price discrimination. Differential pricing of the second degree, more commonly known as volume discounting or quantity discounting, involves the same underlying principle as first-degree differential pricing, except that the seller charges different prices for blocks of units instead of for individual units. By doing so, the firm will extract part, but not all, of the consumers surplus. By charging a higher price for smaller quantities, the seller receives higher total revenue than if a single price were charged. This practice is very common in the electric utility industry, where firms typically charge a higher rate on the first hundred kilowatt hours of electricity used than on subsequent units. Thus, the firm charges different prices to different consumers, but does not need to know specific characteristics of individual consumers.

For example, suppose that the firm of Figure 9-2 sets the price of $24 per unit on the first 6 units of the product and the price of $18 per unit on the next batch or block of 6 units of the product, Figure 9-3.

Figure 9-3



The total revenue of the firm would then be *$144* ($24 x 6) from the first batch of 6 units of the product and *$108* ($18 x 6) from the second batch or block of 6 units, for the overall total revenue of *$252 and profit of $132* (TR – TC = $252 – $10 x 12), as compared to *$160* with first-degree price discrimination and *$100* without any price discrimination). Thus, consumers end up with some consumer surplus, which means that second-degree price discrimination yields lower profits for the firm than it would have earned if it were able to perfectly price discriminate. Nonetheless, profits are still higher than they would have been if the firm had used the simple strategy of charging the same price for all units sold. In effect, consumers purchasing small quantities (or alternatively, those having higher marginal valuations) pay higher prices than those who purchase in bulk.

3.1.4 Third-Degree Price Discrimination

The final type of price discrimination is commonly practiced by firms that recognize that the demand for their product differs systematically across consumers in different demographic groups. In these instances firms can profit by charging different groups of consumers different prices for the same product, a strategy referred to as *third-degree price discrimination.*

Differential pricing of the third degree occurs when the seller segregates buyers according to income, geographic location, individual tastes, kinds of uses for the product, or other criteria, and charges different prices to each group or market despite equivalent costs in serving them. Thus, as long as the demand elasticities among different buyers are unequal, it will be profitable to the seller to group the buyers into separate classes according to elasticity, and charge each class a separate price. This is what is referred to, more generally, as *market segmentation*, that is, the carving up of a total market into subgroups from the standpoint of pricing.

There are many examples of third-degree price discrimination. One of these is provided by electrical power companies, which usually charge higher rates to residential than to commercial users of electricity. The reason for this is that the price elasticity of demand for electricity is higher for the latter than for the former because the latter could generate their own electricity if its price rose above the cost of building and running their own power plants. This choice is generally not available to households. Other examples of third-degree price discrimination are the higher air fares charged to airlines to business travellers than to vacationers, the higher price charged by telephone companies during business hours than at other times, and the higher prices charged for many services to all customers, except children and the aged.

To practice price discrimination effectively, two conditions must be satisfied:

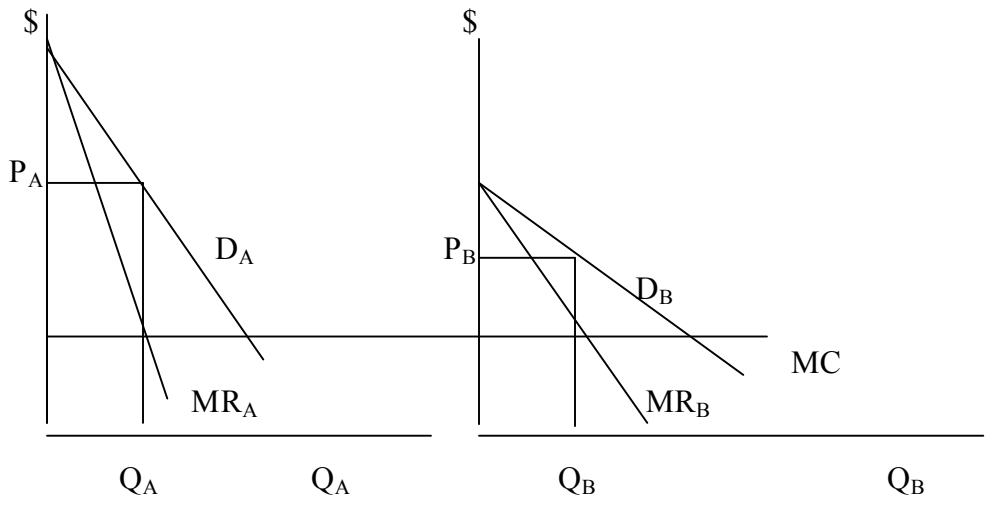
1. Market Segmentation. The seller must be able to partition (segment) the total market by segregating buyers into groups or submarkets according to elasticity. Profits can be enhanced by charging a different price in each submarket.

2. Market Sealing. The seller must be able to prevent – or natural circumstances must exist that will prevent – any significant resale of goods from a lower-priced submarket to a higher one. Any resale (leakage) by buyers between submarkets tends to neutralize the effect of different prices and narrow the effective price structure toward a single price to all buyers.

To see how the third-degree price discrimination enhances profits, we will consider a firm with market power that can charge two different prices to two groups of consumers and the marginal revenues of selling to group A and group B are MR_A and MR_B, respectively. The basic profit-maximizing rule is to produce output such that marginal revenue is equal to marginal cost. This principle is still valid, but the presence of two marginal revenue functions introduces some ambiguity. The condition for maximum profit using price discrimination is that the marginal revenue be the same in all markets and equal to marginal cost. That is, the additional revenue gained from selling one more unit in market A shall be equal to the additional revenue obtained by selling one more unit in market B. If the MRs are not equal, the firm can increase its revenue and profit by selling more to the market with the higher MR. The concept is illustrated in Figure 9-4.

Figure 9-4 illustrates that the customers are willing to pay a higher price than those in market B, perhaps because there is less competition in market A. Less competition implies higher mark-up. In contrast, in market B the lower price represents lower mark-up, which is the symptom of a weaker market power, which in turn can be a function a greater competition in that market.

Figure 9-4



3.1.5 Demonstration Problem

Suppose an airline company faces two different groups of clients, business people and vacationers. The demand equation representing each group is given as follows:

Business passengers: $P_A = 20 - 0.2Q_A$

Vacationers : $P_B = 10 - 0.05Q_B$

Also assume that this firm faces a cost function represented by:

$TC(Q) = 4Q_T$, where $Q_T = Q_A + Q_B$

a. Show how this airline company can increase its profit by engaging in a third-degree price discrimination. What price will the airline company charge and what output will it produce in each market?

b. If for some reason the airline company is unable to practice price discrimination and, therefore, it charges a single price, what will be the price and output?

Answer:

a. First we should find the MRs and MC. $MR_A = 20 - 0.4Q_A$, and

$MR_B = 10 - 0.1Q_B$, and $MC = 4$. Next, we set $MR_A = MC$, and $MR_B = MC$ to obtain Q_A and Q_B as follows:

$20 - 0.4Q_A = 4$, yielding $Q_A = 40$

$10 - 0.1Q_B = 4$, yielding $Q_B = 6$, for the total of $Q_T = Q_A + Q_B = 100$

Plugging for Q_A and Q_B in their own respective demand equations, we have: $P_A = \$12$, and $P_B = \$7$. Profit equals $TR_A + TR_B - TC = \$12 \times 40 + \$7 \times 60 - 4 \times 70 = \$620$.

b. In absence of price discrimination, the firm will sell its tickets at the same price in both markets ($P = P_A = P_B$). The total market demand faced by the airline company is: $Q_T = Q_A + Q_B$. We, therefore, need to invert the demand equations first and then add them:

$Q_A = 100 - 5P_A$, $Q_B = 200 - 20P_B$

Hence $Q_T = 300 - 25P$, so that $P = 12 - 0.04Q_T$. Setting MR =MC

$12 - 0.08\ Q_T = 4$

yields $Q_T = 100$, $P = \$8$. Profit equals = TR – TC = 100 x \$8 – (100 x \$4) = \$400, which is less than that earned under price discrimination practice.

In the example of the airline company, above, we noticed that the airfare in the business travellers market was set higher (full fare) than that in the vacationers market (excursion fare). This is so, because business travellers' willingness to pay is typically greater than that of the vacation travellers, due to the fact that the time and the duration of their business trips are less flexible. There are simply not as many good alternatives available to business travellers in regard to the mode of transport and the time of their trips as there are to excursion travellers. The availability of substitutes, however, is reflected in the price elasticity of demand.

Recall the relationship between MR and P. Applying this to both markets, we have:

$$MR_A = P_A (1 - 1/E_{PA})$$

$$MR_B = P_B (1 - 1/E_{PB})$$

Since in equilibrium, $MR_A = MR_B$

$$P_A (1 - 1/E_{PA}) = P_B (1 - 1/E_{PB})$$

Hence

$$\frac{P_A}{P_B} = \frac{(1 - 1/E_{PB})}{(1 - 1/E_{PA})} \qquad (6)$$

Based on equation (6), one can see that the airline company tends to charge full fare in market A, where the price elasticity is lower (more inelastic), and charge an excursion (lower) price in market B, where the elasticity is greater.

Learning Tip

Note that the relationship between the elasticity and the price is an inverse one. Since the demand by group A is less elastic than by group B, $E_{PA} < E_{PB}$, and hence the ratio of P_A/P_B will be greater.

3.1.6 Demonstration Problem

Suppose that the airline company, above, faces the following price elasticities of demand: $E_{PA} = 1.2$ and $E_{PB} = 2$. What is the ratio of the airfare (full fare) in market A to the fare (excursion fare) in market B?

Answer:

$$\frac{P_A}{P_B} = \frac{(1 - 1/2)}{1 - 1/1.2} = 3$$

4 Two-Part Tariffs

Another form of pricing strategy, which is a variation of second-degree price discrimination, is two-part tariffs (pricing).

With a two-part tariff, the firm charges a consumer a fixed fee (the first tariff) plus a per-unit charge (the second tariff) for the right to buy as many units of the good as the consumer wants.

Learning Tip

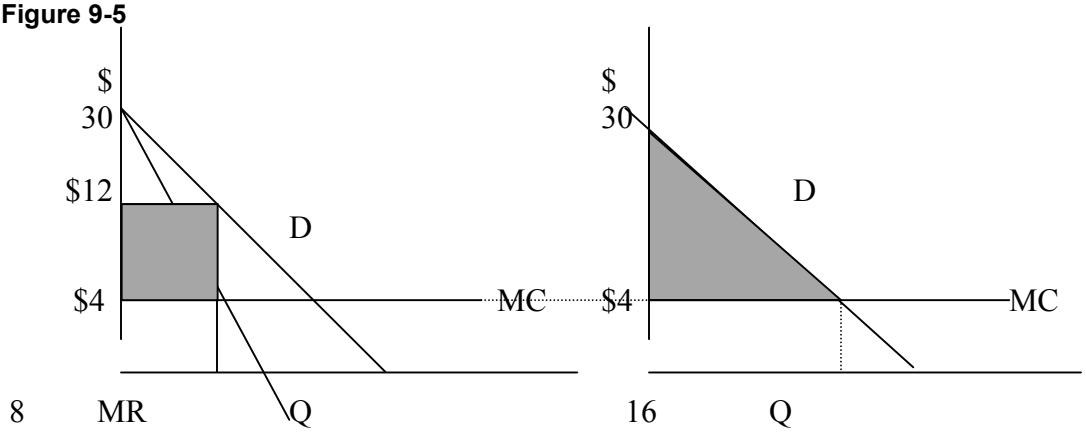
Because of the fixed fee, consumers pay more per unit if they buy a small number of goods than if they buy a larger number.

This strategy is usually used by car rental companies to enhance profits. They typically charge a per-day fee and a price per kilometre driven. Fairs usually charge an entrance fee and a price for each ride.

To profit from two-part tariffs, a firm must have market power, know how demand differs across customers or with the quantity that a single customer buys, and successfully prevent re-sales.

We now examine how a firm uses a two-part tariff to extract consumer surplus. Suppose that a monopoly has a constant marginal and average cost of $4 and no fixed cost, and consumers are represented by the demand curve equation P = 20 – Q, Figure 9-5. If the monopolist adopts a policy of charging a single price to all customers, its profit maximizing-price and quantity would be found by setting its MR = MC. That yields a price of $12 per unit and output of 8 units, respectively.

Figure 9-5



In this case, the profit equals to TR ($12 x 8) – TC ($4 x 8) = $64. If on the other hand, the monopolist engages in a two-part pricing, the outcome would be very much like the first-degree price discrimination, which allows a firm to extract all consumer surplus from consumers. Accordingly, the monopolist would charge a price equal to its MC = $4 and would sell 16 units that will allow it to break even on each unit sold. However, in doing so, it can also charge a lump-sum fee of up to $208, which is the potential consumer surplus, the shaded triangle under the demand curve and above the price line.

Learning Tip

With two-part pricing, all profits are derived from the lump-sum (fixed) fee. By setting a low price that equals MC, the seller can maximize the size of the consumer, and thereby can set the highest possible fixed fee.

5 Pricing Product Bundles

Product bundling is the practice of selling two or more products together as a package deal for a single price. Firms that sell two or more goods may use bundling to raise profits. Bundling allows firms that cannot directly price discriminate to charge customers different prices.

A series of examples will indicate how widespread product bundling is. Computers are often bundled with monitors and with software and sold as a package deal. Restaurants offer fixed menus which include soup, main course, dessert, and coffee for a single price. Automobiles offer luxury or sports packages that must be sold in conjunction with the basic vehicle. Professional sports teams and symphony orchestras offer season tickets that bundle together a variety of events for a single price. And cable television companies bundle channels and sell them as a package rather than selling them individually.

Put simply, the firm has a profit incentive to bundle products together when doing so allows the extraction of a greater degree of consumer surplus from the potential customers. In general, it may be optimal to offer the products both separately and in the bundle, since some consumers will only want one of the products and would not be willing to pay the bundle price. Offering both the bundle and the separate products is known as 'mixed' bundling, as opposed to 'pure' bundling, where the products are only available as a package deal.

In practice the seller will be unable to determine each buyer's reservation price and will simply assume that buyers have a range of reservation prices, such that there is a negatively sloping demand curve for both the products separately and collectively (the bundle). The seller will typically increase profit (over that available from pricing the products separately) by raising the prices of each product *if sold separately* and offering the bundle as a package deal at a price which is less than the sum of the prices of the components of the bundle. Thus buyers who only want one or some subset of the products in the bundle must pay more than they otherwise would have, and other buyers who would not have purchased all the products separately end up buying the bundle because it is the cheapest way to get the products that they do want. Raising the price of individual products will cause the firm to lose some sales, but the gain in sales resulting from the availability of the bundle typically outweighs that loss, such that the overall sales and profit are higher than if the products were priced separately.

An application of bundling is *quantity discounts.* The theory of bundling explains why some firms offer a given product in several different sizes (such as small, medium, large, and jumbo bottles of Coke), and why some consumers buy only the small size and others pay more to buy larger sizes (but at a reduced price per unit, such as per ounce). The larger sizes can be viewed as bundles, or multiples, of the smallest size of the same product, and the buyer is given a discount for purchasing in quantity. Similarly, when a product is priced at, say, $3 each, or two for $5, the buyer is essentially getting a discount on the second unit of the product if he or she chooses to buy that additional unit. Why are some people induced to buy in quantity while others are not? A person will demand the

extra units of the product bundled together in the larger size if the incremental cost to the consumer is less than the consumer's reservation price for those extra units.

5.0.1 Demonstration Problem

A consumer might be willing to pay a maximum of $1 for a 250 ml can of cola, if she is very thirsty. The store price is, let us say, 80 cents. Since the asking price is less than the buyer's reservation price, the seller will make the sale. But suppose the seller also has a 500 ml can of the same cola for $1.20. Will the consumer buy that one instead?

Answer:

Suppose that the consumer's reservation price on the 500 ml size is $1.50, indicating she expects 50 cents worth of extra utility (consumer surplus) from the additional 250 mls. Since the additional 250 mls will cost her only 40 cents more, she will buy the larger size. Now, suppose the store also has a 750 ml size, priced at $1.80. The consumer's reservation price for the 750 ml size is, say, $1.75, indicating her willingness to pay only another 25 cents for the additional 250 mls. Since the additional quantity would cost her 60 cents, however, she chooses the 500 ml size. The seller collects $1.20 from this customer, compared to only 80 cents that would have been collected if the cola had been priced uniformly per millilitre (ml) regardless of bottle size. Assuming that the seller's marginal cost of the extra 250 mls is always less than the price premium attached to the larger sizes, the seller will have increased both sales and profit. Of course, more thirsty customers, or someone buying the family groceries, are likely to buy the 750 mls size, because their reservation prices are more likely to exceed the price asked.

A more elaborate discussion of bundling can be made by working through the following illustration.

Suppose, a tour company in the hotel and hospitality business identifies three groups of customers for an exclusive destination:

Group 1: 1/3 of the customers will use a package of hotel and airfare

Group 2: 1/3 of the customers will use only airfare

Group 3: 1/3 of customers will use only hotels.

The marginal cost of the airfare is $400 and the marginal cost of the hotel is $250. The maximum price each type of consumer will pay is as follows:

Table 9-1

Commodity Bundling	Reservation Price (Maximum Willingness to Pay)	
	Airfare	**Hotel**
Group 1	$800	$500
Group 2	$800	$400
Group 3	$500	$600

First let us see how much profit the firm can earn if it does not bundle the fare and the hotel.

Optimal price of airfare: What price should the tour company set for the airfare?

If it sets the airfare at $500, all three groups of customers will buy and the firm's profit will be (TR) – (TC) = (3 x $500) – (3 x $400) = $300. At the price of $800, the firm will sell airfare only to Group 1 and 2, and will earn (2 x $800) – (2 x $400) = $800. Therefore, it should set the price of airfare at $800.

Optimal Price of hotel: What price should the tour company set for the hotel?

If it sets the hotel at $400, all three groups of customers will buy and the firm's profit will be (TR) – (TC) = (3 x $400) – (3 x $250) = $450. At the price of $500, the firm will sell hotel only to Group 1 and 3, and will earn (2 x $500) – (2 x $250) = $500. At the price of $600, the firm will sell hotel only Group 3, and will earn (1 x $600) – (1 x $250) = $350 Therefore, the tour company should set the price of hotel at $500.

The best the firm can do without bundling is to set the price of airfare at $800 and the price of hotel at $500. It will then earn a total profit of $1,300.

Now consider the option to bundle the airfare and the hotel, selling two components in a single package.

Optimal price of package: What price should the tour company set for the package?

In order to answer this, we first need to summarize the customers' reservation price for the package. Building on the information in Table 1:

Table 9-2

Commodity Bundling	Reservation Price (Maximum Willingness to Pay)		
	Airfare	**Hotel**	**Value of Bundle**
Group 1	$800	$500	$1,300
Group 2	$800	$400	$1,200
Group 3	$500	$600	$1,100

If the bundle is sold at $1,100, all three groups would buy the package and the firm will earn (3 x $1,100 – 3 x $650) = $1,350. If the bundle is sold at $1,200, only Group 1 and 2 will buy the package and the profit will be (2 x $1,200 – 2 x $650) = $1,100. And if the tour company sets the price at $1,300, its profit from the sale of the package to only Group 1, will be $650. Therefore, the optimal bundle price is $1,100. Note that the bundling has increased profits from $1,300 without bundling to $1,350 with bundling.

Learning Tip

The firm should consider offering its products in bundles because bundling can increase the firm's profits. Bundling includes the packaging together of different products, the offering of the same product in different-sized containers, and the offering of quantity discounts for multiple units of the same product.

Learning Tip

Note that the key as to why bundling the two products increased profits in the above illustration is the negative correlation between customers' demands. As you can see in Table 9-1, the preferences of Group 2 and Group 3 are inversely related. Group 2 is willing to pay more for the airfare, whereas Group 3 is willing to pay more for the hotel. By bundling them, the tour company is offering an incentive to the customers to take both products, while they may not have done so otherwise.

5.1 Decentralization and Transfer Pricing

Many large firms are organized into divisions or profit centres to facilitate the efficient operation of the firm. The decentralization of decision making in this way is considered to have a beneficial impact on the firm's overall efficiency and profitability, since each division manager is judged by the profit performance of his or her division or profit centre. A problem can arise, however, when one division or profit centre supplies a component product (or intermediate product) to another division or profit centre that uses this intermediate product as the basis for the firm's finished product, which is sold to consumers. This transaction is essentially internal to the firm and does not take place in the market for that intermediate product, if indeed there is a market for it. Since the transfer of such products involves a transaction between autonomous units within the firm, it becomes necessary to establish a *transfer price* at which the product may be sold by one division and purchased by the other.

If the transfer price is set at a relatively high level, the supplying division will make more profits and the buying division less, presuming there is competition of some sort in the market for the finished product. If the transfer price is set at a relatively low level, the opposite will prevail. The firm as a whole will wish to set the transfer price at a level which serves to maximize the profit of the firm as a whole, rather than at some arbitrary or negotiated level that may not best serve the firm's objective.

We shall analyze the transfer pricing problem under three scenarios, first considering the case where there is no external market for the intermediate product, then considering the existence of a perfectly competitive external market for the intermediate product, and finally considering the existence of an imperfectly competitive market for the intermediate product.

5.2 Transfer Pricing with No External Market for the Intermediate Product

The marginalist rule, that marginal cost equals marginal revenue, determines the profit-maximizing output and price level for the firm as a whole. But if the firm has two divisions, for example, its overall marginal cost at any output level will be the sum of the marginal costs in its two divisions. Suppose a firm has two divisions, P (production) and M (marketing). The intermediate product is produced in Division P, and then packaged and marketed by Division M. We suppose that each division has an upward-sloping marginal cost curve, as shown in Figure 9-6, where

MC_P = Division P's marginal cost of producing components for Division M,

MC_M = Division M's marginal cost of completing the firm's product,

MC_F = The firm's marginal cost, which is the vertical summation of MC_P+ MC_M,

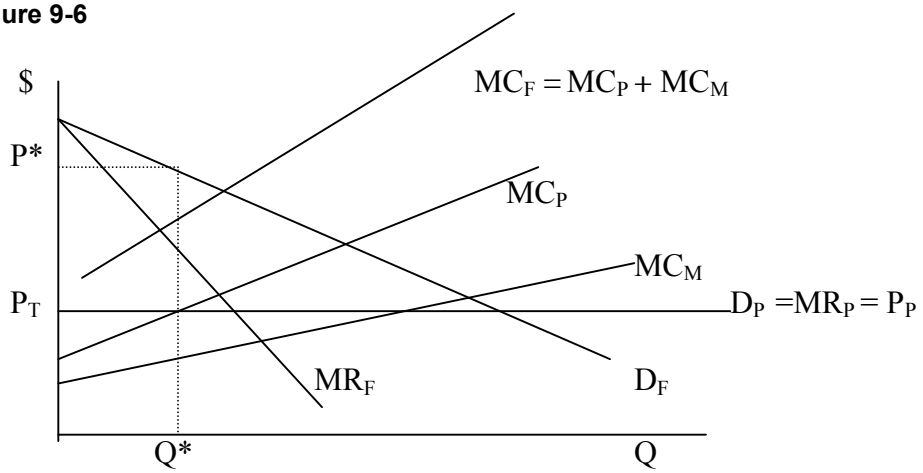
D_F = Demand for the firm's product,

MR_F = The firm's marginal revenue,

P_T = The transfer price.

Since there is no external market for Division P's product, its only customer is Division M. On the other hand, Division M's only source of the product is Division P. Hence, the production by Division P precisely equals to the demand by Division M. Therefore, Given this information, the firm maximizing price, output, and the transfer price are P* level Q*, and P_T, respectively.

Figure 9-6



What transfer price should be established to induce Division P to produce exactly the profit-maximizing output of the intermediate product? Similarly, what transfer price will induce Division M to set the market price at the level P* such that the firm's profits are maximized? In Figure 9-6, we show the optimal transfer price as P_T, chosen to equate Division P's marginal cost at output level Q*. As a result, Division P faces a horizontal demand curve (D_A) at the transfer price – it can sell as much as it wants to at that price.

However, it will only want to produce Q* units, since its marginal cost MC_P equals its marginal revenue ($P_A = MR_A$) at that output level. Thus setting the transfer price at the level P_T induces Division P to produce and supply exactly the optimal amount.

From the viewpoint of division M also, the market price P* and output/sales level Q* are consistent with profit maximization. At any other price and output combination in the market for the finished product, Division M's net additional revenue (net marginal revenue) would either exceed or fall short of the transfer price (its effective marginal cost of the intermediate product), and its divisional profit would not be maximized.

Thus establishing the transfer price at the level P_T provides the appropriate incentives for both divisions to produce and sell at the firm's profit-maximizing output level Q*. Each division maximizes its profits by producing Q* units, and the overall firm's profits are also maximized. Any other transfer price may have higher profits for either Division P or M, but would have lower profit for the firm as a whole, not to mention a shortage or surplus of the intermediate product.

5.2.1 Demonstration Problem

Suppose a vertically integrated firm has two divisions. The P (production) Division manufactures a patented product that S Division uses in a customized finished product that it assembles and installs. Demand and cost functions have been estimated as follows:

Demand for the firm's product is: $P_F = 500 - 0.01P_F$. Marginal cost in P Division: $MC_P = 10 + 0.001Q_P$, and marginal cost in S Division: $MC_S = 100 + 0.005Q_S$.

a. What are the firm's profit maximizing price and output?

b. What is the transfer price and what is the level of output produced by Division P? How many units of the customized finished product will be produced?

Answer:

a. Optimality requires that the firm set $MR_F = MC_F$. From the demand curve we obtain $MR_F = 500 - 0.02Q_F$, then we obtain MC_F by adding the two (divisional) marginal costs curves vertically

$MC_F = MP_F + MCs = (10 + 0.001Q_P) + (100 + 0.005Qs)$. Since $Q_P = Qs = Q_F$ at the profit-maximizing level of output, $MC_F = 110 + 0.006Q_F$. At the point of optimality,

$MR_F = MC_F$

$500 - 0.02Q_F = 110 + 0.006\ Q_F$,

$0.026Q, = 390$

$Q_F = 15{,}000$, and $P_F = \$350$.

b. The transfer price is the Production Division's marginal cost at a production level of 15,000 units:

$MC_P = 10 + 0.001Q_P = 10 + 0.001\ (15{,}000) = \$25$ per unit.

5.3 Transfer Pricing with a Perfectly Competitive External Market

Now suppose that the intermediate product is available from other suppliers as well as Division P. Moreover, Division P's product is identical to that of other suppliers. There will be a market price for the intermediate product determined by the forces of supply and demand in that external market. In this case the transfer price must be set equal to the market price for the intermediate product. If it is set higher, Division M will prefer to buy from the external market rather than from Division P. If the transfer price is set below the external market price, Division P will prefer to sell to the external market rather than to Division M. If, at the transfer price (equal to the market price), division P wants to

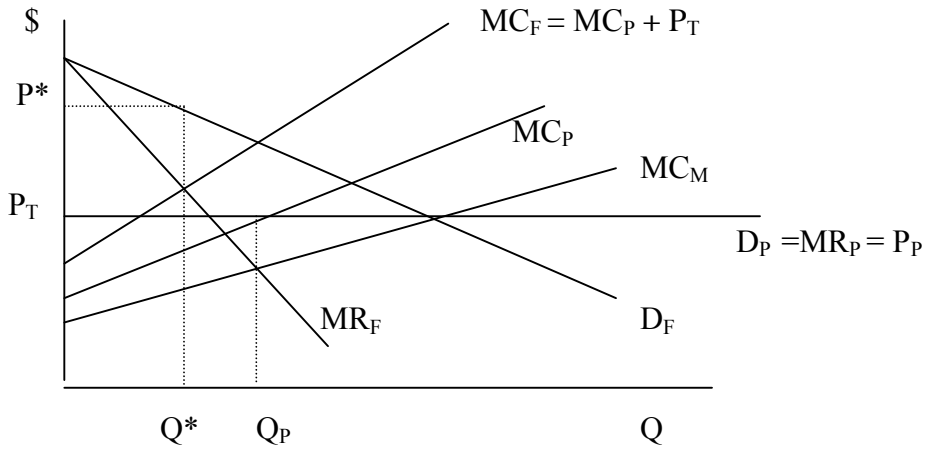
produce more than Division M wishes to purchase, it can sell the balance in the external market. Conversely, if Division M wants to purchase more than Division P wishes to produce, it can buy the remainder in the external market.

For the case of perfectly competitive external market, we shall illustrate the derivation of a transfer price with the aid of figure 9-7. The Figure conveys the following information:

> D_F = Demand curve for the firm's product, which is assembled and sold by Division M,
>
> D_P = Demand curve for Division P's product. Since the product is sold in a perfectly competitive external market, D_P is a horizontal line,
>
> MC_P = The marginal cost of selling Division P's product,
>
> MC_M = Marginal cost to Division M of assembling and marketing the firm's product, excluding the purchases from Division P or the external market,
>
> MC_F = Marginal cost of producing the firm's product, including purchases from Division P or the external market, i.e., $MC_F = MC_P + P_T$. Since Division P will establish production at a level where $MR_P = MC_P$, then $MC_T = MC_M + MC_P$.

As illustrated in Figure 9-7, the transfer price, P_T, is the external market price, since the external market is perfectly competitive. Note that Division P will operate at its profit-maximizing level of production, Q_P, such that $MC_P = MR_P$. It will, however, sell Q* units to Division M, and, thus, it will sell the remainder, shown by the distance $Q_P - Q^*$, in the external market for the intermediate product. Had the competitive external market price been lower, such that $Q_P < Q^*$, Division M would have purchased some from Division P and the remainder from the external market. Clearly, Division M's profit maximizing output is found by setting $MR_F = MC_T$. This is the quantity of Division P's product that Division M will buy.

Figure 9-7



5.3.1 Demonstration Problem

Suppose that demand, and cost curves for a vertically integrated firm with two divisions P and M are given as follows**:**

$P_F = 18 - 0.1Q_F$

$MC_P = 1 + 0.1Q_P$

$MC_M = 1 + 0.1Q_M$

Also assume the perfectly competitive external price for the intermediate product that Division P produces is $6. Determine the best level of output of the intermediate product for Division P. How much of this will be sold to the Marketing Division and how much to external buyers? What is the transfer price? What is the optimal level of price and output from the perspective of the firm?

Answer:

The profit-maximizing level of price and output for the production division:

$MC_P = 1 + 0.1Q_P = \$6 = P_T$ and hence $Q_P = 50$

The profit-maximizing level of price and output for the Marketing Division:

$MC_F = MC_M + P_T = 0.1Q_M + 6$

Then

$MC_F\ (= 0.1Q_M + 6) = MR_F\ (= 18 - 0.2Q_M)$

Hence,

$Q_M = 40$, and $P_M = \$14$.

Therefore, the Production Division produces 50 units, selling 40 units to Division M, and the rest, 10 units, in the external market.

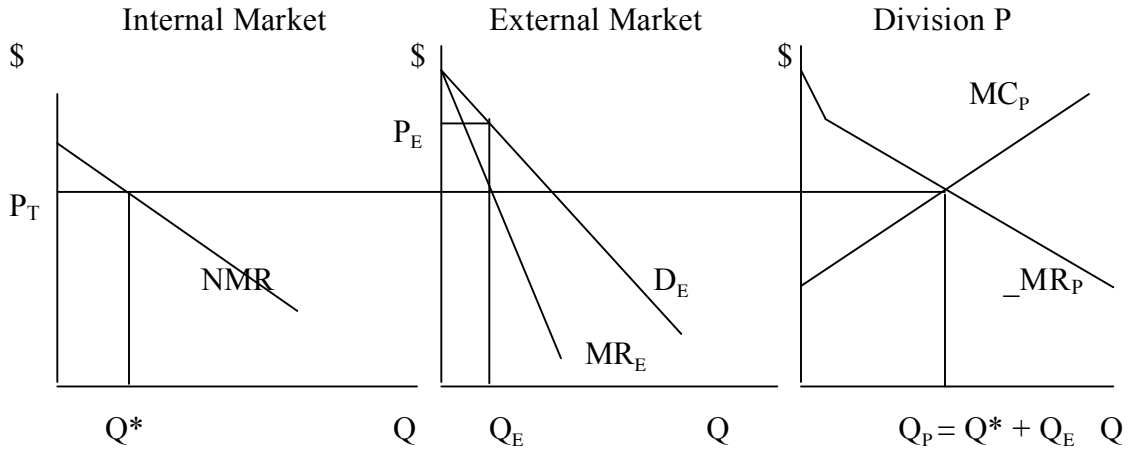
5.4 Transfer Pricing with an Imperfectly Competitive External Market

Now we suppose that the intermediate products are not identical across firms, and thus each supplier to the external market faces a downward-sloping demand curve in that market. In the left-hand panel of Figure 9-8, we show the net marginal revenue, NMR, curve for (the marketing) Division M of the firm. In the Division M market, the relevant marginal revenue is the net marginal revenue derived by subtracting the cost of the component purchased from Division P from the marginal revenue derived from demand for Division M's product (which is the firm's product as assembled and marketed by Division M). That is, $NMR_M = MR_F - MC_P$ .The cost of the component purchased from Division P is just Division P's marginal cost. In the middle panel, we show the demand

curve, D_E, and the marginal revenue curve, MR_E, of the imperfect external market for Division P's intermediate product.

Division P again has two markets for its output – it can sell to either Division M or it can sell to the external market, E. However, since the elasticities of demand are different in these two markets this case, Division P will find it profitable to practice price discrimination between the two markets. As shown in the right panel, Division P's marginal revenue is the horizontal summation of the total division's marginal revenue in both markets, i.e., $\Sigma MR_P = MR_E + NMR_M$, where Σ (the Greek letter sigma) stands for the sum. The intersection of ΣMR_P and MC_P determines Division P's optimal output level is Q_P units, which must be allocated between the internal and external market. The optimal transfer price is $P_T = MC_P$.

Figure 9-8



At the transfer price of P_T dollars, Division M will purchase Q_P units of Division P's product. In the external market, the optimum sales level is Q_E units, which will clear the market at a price of P_E dollars, which is more than the transfer price, P_T. Setting the transfer price equal to Division P's marginal cost insures that Division M will demand from Division P a quantity that will maximize profit for the firm as a whole.

6 Summary and Review

A firm with market power can influence the price in the market. Firms with market power can use many strategies to increase their profits. One common practice is price discrimination, by which the firm charges different prices to different consumers at the same time or to the consumers at different times, for reasons not associated with costs. To be able to implement this successfully, however, the firm requires information about its customers and their willingness to pay. Bundling is another strategy by which firms may be able to enhance their profit. This is a form of tie-in sales that requires customers to purchase goods in a package. Bundling increases profits when customers have different demands and tastes. Transfer pricing refers to the determination of the price of the intermediate products sold by one semiautonomous division of a firm to another semiautonomous division of the same enterprise. Appropriate transfer pricing is essential in determining the optimal output of each division and the firm as a whole.

7 Review Problems

1. A monopolist uses 'block-pricing' strategy to extract all consumer surpluses. If the inverse demand function faced by the firm is $P = 10 – 0.5Q$ then what bundle should the firm offer to consumers? (Assume that the firm produces at constant marginal cost of $4.00). How much surplus is extracted in this situation?

2. A fast food restaurant estimated that there are four categories of consumers who visit the pace with different valuation of burgers and pies, as stated below:

Category of Consumer	Valuation of Burger	Valuation of Pie
I	$3.50	$1.00
II	$3.00	$2.00
III	$2.50	$3.00
IV	$2.00	$4.00

   The cost of making a burger is $0.50, which is the same as the cost of making a pie. If the consumers are differentiable then what should be the optimal pricing strategy by the firm. If the owner offers a package deal of $5.00 for a burger and a pie, what will be the profit for the owner (assume there is only one individual per category of consumer).

3. The local telephone company has monopoly power over the market. A researcher who has been employed by the firm submitted a report that shows significant difference in demand for international calls in the night compared to the number of calls in the day time. How should the company decide on the price for international calls to maximize profit?

4. Why is transfer pricing important? Explain the transfer price formula given that a firm has one upstream and one downstream division. If the cost function of these two divisions are $C_u(Q) = 5Q + 10Q^2$ and $C_d(Q) = 50Q$ respectively, then find the profit maximizing level of output and profit for the firm. Assume that the inverse demand function is $P = 100 – 2Q$.

5. Explain the relevance of the following pricing strategies for a monopolist:

   a. first degree price discrimination

   b. second degree price discrimination

   c. third degree price discrimination.

6. A monopolist faces an inverse demand function of $P = 10 – 0.2Q$. The marginal cost of the firm is $6.00. If the firm uses standard monopoly pricing policy find the equilibrium price, output and profit. With a 'two-part' pricing strategy what should the firm do with regard to price setting?

7. The demand for ice-cream in summer follows the function: $P = 10 – 0.005Q$, and in winter the function is $P = 8 – 0.01Q$. If the cost function of the firm is $C(Q) = 0.002Q + 5$, then how should the firm maximize its profit?

8. A water supply company has monopolistic power over the market. The firm charges different rates per unit of water depending on the level of usage by the consumers. The price chart says $0.10 per unit for use up to 300 gallons, $0.12 for use of 300-500 gallons and $0.15 for use of over 500 gallons. What sort of pricing strategy does the water supplier use and why?

8 Answer Key to Review Problems

1. If P = MC then $10 – 0.5Q = 4$, therefore, $Q = 12$ and $P = 4$

   If $Q = 0$, then $P = 10$

   Consumer Surplus = $0.5(10 – 4)12 = 36$

   Bundle, 12 units for $36 + (4 \times 12) = 84$

2. If differentiable then charge different consumers as per their valuation.

   If not, then $5.00 will be attractive only to groups II, III, & IV.

   Cost of Burger = 0.5 per unit x 5 = 1.5

   Cost of Pie = 0.5 per unit x 5 = 1.5

   Total Cost = $3.00

Total Revenue = \$5.00 x 3 = \$15.00

Profit = \$12.00

3. Determine two demand curves for daytime and night time

   a. Equate MR of daytime with MC and decide on quantity and price for daytime only.

   b. Equate MR of night time with MC and decide on quantity and price for night time only.

   c. Given two different demand functions, the MR's will be different and therefore both quantity and price will differ from daytime to night time.

4. Transfer pricing is important since large firms with different divisions producing parts of products at different places consider these divisions as independent cost centres. Therefore, costs need to be optimized for all divisions, otherwise this could result in a loss for the firm. The rule is to equate net marginal cost of downstream with upstream MC to find an optimal level of output.

   Marginal Cost of Upstream = $MC_u = 5+20Q$

   Marginal Cost of Downstream = $MC_d = 50$

   Revenue function of Downstream = $PQ = 100Q – 2Q^2$

   Marginal Revenue of Downstream = $MR_d = 100 – 4Q$

   Net Marginal Revenue of Downstream = $MR_d – MC_d = 50 – 4Q$

   Setting Net Marginal Revenue of Downstream = MC_u, we get

   $5 + 20Q = 50 – 4Q$ or $Q = 45/24 = 1.875$ units.

5. A monopolist, by virtue of being the only seller has market power to set prices. Market power allows the monopolist to charge arbitrary prices for profit maximization. First-, second-, and third-degree price discriminations are various strategies to extract whole or part of the consumer surplus over and above the producer surplus that the monopolist would generate by following the profit-maximizing rule only.

   a. First-Degree Price Discrimination: This strategy is applicable only if the monopolist can differentiate the consumers individually, and is fully aware of the demand function. The monopolist extracts the entire consumer surplus in this way.

b. Second-Degree Price Discrimination: The monopolist puts consumers in different price brackets. Prices are usually declining with higher level of purchase. For example, if the monopolist charges one price for 0-10 units, another price for 11-20 units, another price for 21-30 units and so on, then the firm is able to extract some consumer surplus without having to know about all consumers individually.

c. Third-Degree Price Discrimination: This strategy is relevant when there are markets that are not overlapping. The firm sets prices according to individual market demand curves, setting marginal cost equal to the marginal revenue from that particular market.

6. $P = 10 - 0.2Q$ [if $Q = 0$ then $P = 10$]

$\text{Revenue } R = 10Q - 0.2Q^2$

$\text{Marginal Revenue } MR = 10 - 0.4Q$

Setting MR = MC

$10 - 0.4Q = 6$

$Q = 10$

$P = 8$

$\text{Profit} = (10 \times 8) - (8 \times 6) = 32$

$\text{Consumer Surplus} = 0.5\ (10 - 8) \times 8 = 8$

Two part pricing, buy 10 for $[32 + (10 \times 8)] = 112$

$\text{Profit} = 112 - 48 = 64$

7. The firm can maximize profit by using third-degree price discrimination.

Summer demand: $P_s = 10 - 0.005Q$

Revenue function $R_s = 10Q - 0.005Q^2$

Marginal Revenue $MR_s = 10 - 0.01\ Q$

Setting MR = MC

$10 - 0.01Q = 0.002Q + 5$

or $Q = 416.67$ and $P = 7.92$

Winter demand: $P_w = 10 – 0.01Q$

Revenue function $R_w = 10Q – 0.01Q^2$

Marginal Revenue $MR_w = 10 –0.02\ Q$

Setting MR = MC

$10 – 0.02Q = 0.002Q + 5$

or $Q = 227.27$ and $P = 7.27$

In summer the firm should sell 416.67 units at a price of $7.92 per unit and in winter it should sell 227.27 units to maximize profits at a price of $7.27 per unit.

8. The water supplier uses second-degree price discrimination as a strategy. This allows the firm to charge different prices for different levels of usage. This results in extraction of consumer surplus from the consumers who consume lower amounts of water. The firm does not need to know beforehand the demand function for all consumers, but earns more profit compared to the usual profit maximizing scheme (by setting marginal revenue equal to marginal cost and selling all units at the same price).

9 References and Suggested Further Reading

Baye, M., *Managerial Economics and Business Strategy*, McGraw Hill, Irwin, 2002.

Besanko, D. and Braeutigam, R., *Microeconomics: An Integrated Approach*, Wiley, 2002

Douglas, J. E., *Managerial Economics: Analysis and Strategy*, Prentice Hall, 1992.

Pindyck, R. and Rubinfeld, D., *Microeconomics*, Fifth Edition, Prentice Hall, 2001.

Salvatore, D., *Managerial Economics in a Global Economy*, McGraw Hill, 1993.

Commonwealth of Learning Executive MBA/MPA

E5 Managerial Economics

Block 10

An Overview of Strategic Business Concepts for Managers

Course author: Farrokh Zandi, PhD
Professor, Schulich School of Business
York University
Toronto, Canada

Course designer: Douglas R. Cronk, M.A., ABD
Vancouver, BC, Canada

Copyright © Commonwealth of Learning, 2003
All rights reserved. No part of this course may be reproduced in any form by any means without prior permission in writing from:

The Commonwealth of Learning
1285 West Broadway
Suite 600
Vancouver, BC V6H 3X8
CANADA
e-mail: info@col.org

Contents

1 Block Ten: Objectives and Introduction 1
1.1 Introduction 1
2 Strategy Definitions 2
2.1 Strategy According to George Steiner 2
2.2 Strategy According to Henry Mintzberg 3
2.3 Strategy According to Treacy and Wiersema 4
3 Strategy at Different Levels of a Business 5
3.1 How Strategy is Managed – Strategic Management 5
3.2 Strategic Analysis 6
3.2.1 Introduction to PEST analysis 6
3.2.2 Porter's Five Forces Model 7
3.2.3 Threat of New Entrants 10
3.2.4 Threat of Substitutes 10
3.2.5 Bargaining Power of Buyers 11
3.2.6 Intensity of Rivalry 11
3.4 SWOT Analysis 12
3.4.1 The Key Distinction: Internal and External Issues 12
3.4.2 Areas to Consider 13
3.5 Ansoff Product/Market Matrix 13
3.5.1 Market Penetration 14
3.5.2 Market Development 14
3.5.3 Product Development 15
3.5.4 Diversification 15
4 Strategic Choice and Implementation 15
5 Game Theory – A Basic Definition, Prisoner's Dilemma, and Key Concepts 16
5.1 The Game Matrix, Equilibrium, Dominant Strategies, and Nash Equilibrium 18
6 Summary and Review 20
7 Review Problems and Answer Key 20
8 References and Suggested Further Reading 21

1 Block Ten: Objectives and Introduction

After working through Block Ten of this course, you should be able to:

1. Understand what is meant by 'business strategy,' both in the context of institutional definitions and specific strategy models.

2. Understand what elements are necessary in order to conduct a strategic analysis of an industry, a company, or a business opportunity.

3. Use the concepts to evaluate alternatives and make decisions based on the information you have available to you regarding an industry or business opportunity.

4. Understand the basic principles of game theory, to be extrapolated and put towards understanding how one firm's action affects, and is affected by, all other firms in that industry.

5. Combine the concepts from all ten blocks to create an 'analytical tool kit' for application in any business situation.

1.1 Introduction

Up to this point, the previous blocks have focused on providing the tools for consistent, rigorous analysis of a business situation to allow managers to make informed, substantiated decisions. However, in as much as the understanding of economics and finance drive business decisions, there is also the more creative side to rounding out decisions – understanding what your competitors are going to do, or understanding general trends in the industry and how to stay ahead of them. This component, the 'art' of business, can be summed up in one word: strategy.

The study of strategy for business situations has parallels to strategies for battle in war, or legal battles in a court case. Just as high-profile prosecutors would do everything in their power to understand the context of the case, the players involved, the motives of each player, and the evidence available, so must managers understand not just the issue at hand, but also its context, their competitors, and whether they have enough information to make a good decision. Business strategy ties together the analytical rigour of managerial economics to the art of understanding the customer and customer trends, or drivers of consolidation in an industry, or how to make your company achieve the goals you set for it.

Understanding strategy and how to implement it is essential to running a successful organization, or completing a project for a superior. No court case is ever won by an ill-prepared, poorly organized legal team. Similarly, all successful business ventures rely on being highly knowledgeable both in the task at hand, and the context of that task.

Johnson and Scholes (*Exploring Corporate Strategy*) define strategy as follows:

> Strategy is the direction and scope of an organization over the long-term: which achieves advantage for the organization through its configuration of resources within a challenging environment, to meet the needs of markets and to fulfil stakeholder expectations.

In other words, strategy is about:

- Where is the business trying to get to in the long-term? (Direction)
- Which markets should a business compete in and what kind of activities are involved in such markets? (Markets, scope)
- How can the business perform better than the competition in those markets? (Advantage)
- What resources (skills, assets, finance, relationships, technical competence, and facilities) are required in order to be able to compete? (Resources)
- What external, environmental factors affect the businesses' ability to compete? (Environment)
- What are the values and expectations of those who have power in and around the business? (Stakeholders)

2 Strategy Definitions

There are countless books available in the marketplace that disseminate what great minds have conceived strategy to be. Inherently, they touch on many of the topics outlined above – markets available, competitive advantages regarding specific firms and competitors, and resources available to them. While there is no single book or definition that provides the 'magic bullet' answer in creating a business strategy, the following definitions encompass the most accepted views of strategy.

2.1 Strategy According to George Steiner

George Steiner, a professor of management and one of the founders of *The California Management Review*, is generally considered a key figure in the origins and development of strategic planning. His book, *Strategic Planning*, is close to being a bible on the subject. Yet, Steiner does not bother to define strategy except in the notes at the end of

his book. There, he notes that strategy entered the management literature as a way of referring to what one did to counter a competitor's actual or predicted moves. Steiner also points out in his notes that there is very little agreement as to the meaning of strategy in the business world. Some of the definitions then in use to which he pointed include the following:

1. Strategy is that which top management does that is of great importance to the organization.

2. Strategy refers to basic directional decisions, that is, to purposes and missions.

3. Strategy consists of the important actions necessary to realize these directions.

4. Strategy answers the question: What should the organization be doing?

5. Strategy answers the question: What are the ends we seek and how should we achieve them?

Steiner was writing in 1979, at roughly the mid-point of the rise of strategic planning. Perhaps the confusion surrounding strategy contributed to the demise of strategic planning in the late 1980s. The rise and subsequent fall of strategic planning brings us to Henry Mintzberg.

2.2 Strategy According to Henry Mintzberg

Henry Mintzberg, in his 1994 book, *The Rise and Fall of Strategic Planning*, points out that people use 'strategy' in several different ways, the most common being these four:

1. Strategy is a *plan*, a 'how,' a means of getting from here to there.

2. Strategy is a *pattern* in actions over time; for example, a company that regularly markets very expensive products is using a 'high end' strategy.

3. Strategy is *position*; that is, it reflects decisions to offer particular products or services in particular markets.

4. Strategy is *perspective*, that is, vision and direction.

Mintzberg argues that strategy emerges over time as intentions collide with and accommodate a changing reality. Thus, one might start with a *perspective* and conclude that it calls for a certain *position*, which is to be achieved by way of a carefully crafted *plan*, with the eventual outcome and strategy reflected in a *pattern* evident in decisions and actions over time.

2.3 Strategy According to Treacy and Wiersema

The notion of restricting the basis on which strategy might be formulated has been carried one step farther by Michael Treacy and Fred Wiersema, authors of *The Discipline of Market Leaders*. In the 1993 *Harvard Business Review* article that presaged their 1994 book, Treacy and Wiersema assert that companies achieve leadership positions by narrowing, not broadening their business focus. Treacy and Wiersema identify three 'value-disciplines' that can serve as the basis for strategy: operational excellence, customer intimacy, and product leadership. As with driving forces, only one of these value disciplines can serve as the basis for strategy. Treacy and Wiersema's three value disciplines are briefly defined below:

Table 10-1 Treacy and Wiersema's three value disciplines

Operational Excellence	Strategy is predicated on the production and delivery of products and services. The objective is to lead the industry in terms of price and convenience.
Customer Intimacy	Strategy is predicated on tailoring and shaping products and services to fit an increasingly fine definition of the customer. The objective is long-term customer loyalty and long-term customer profitability.
Product Leadership	Strategy is predicated on producing a continuous stream of state-of-the-art products and services. The objective is the quick commercialization of new ideas.

Each of the three value disciplines suggests different requirements. Operational Excellence implies world-class marketing, manufacturing, and distribution processes. Customer Intimacy suggests staying close to the customer and entails long-term relationships. Product Leadership clearly hinges on market-focused R&D as well as organizational nimbleness and agility.

3 Strategy at Different Levels of a Business

Strategies exist at several levels in any organization – ranging from the overall business (or group of businesses) through to individuals working in it.

Corporate Strategy – is concerned with the overall purpose and scope of the business to meet stakeholder expectations. This is a crucial level since it is heavily influenced by investors in the business and acts to guide strategic decision-making throughout the business. Corporate strategy is often stated explicitly in a 'mission statement.'

Business Unit Strategy – is concerned more with how a business competes successfully in a particular market. It concerns strategic decisions about choice of products, meeting needs of customers, gaining advantage over competitors, exploiting or creating new opportunities, etc.

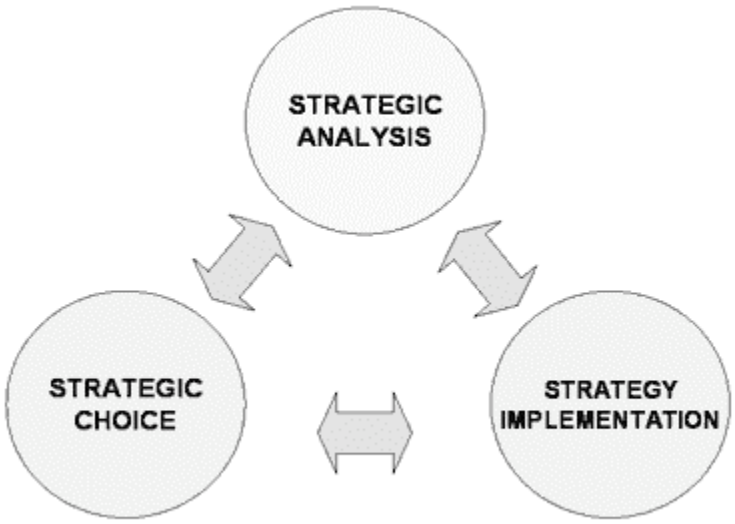
Operational Strategy – is concerned with how each part of the business is organized to deliver the corporate and business-unit level strategic direction. Operational strategy therefore focuses on issues of resources, processes, people, and so-on.

3.1 How Strategy is Managed – Strategic Management

In its broadest sense, strategic management is about taking 'strategic decisions' – decisions that answer the questions above.

In practice, a thorough strategic management process has three main components, shown in the figure below.

Figure 10-1 Strategic Management Process – Main Components



3.2 Strategic Analysis

This is all about the analyzing the strength of businesses' position and understanding the important external factors that may influence that position. There are many processes, models and theories regarding strategy, and this block provides only an overview of the main ones. The process of Strategic Analysis can be assisted by a number of tools, including:

PEST Analysis – a technique for understanding the 'environment' in which a business operates.

Five Forces Analysis – a technique for identifying the forces that affect the level of competition in an industry.

SWOT Analysis – a useful summary technique for summarizing the key issues arising from an assessment of a business's 'internal' position and 'external' environmental influences.

Ansoff Product/Market Matrix – A basic analytic tool to help a company understand how its growth strategy in specific markets can correlate to existing or potential products.

3.2.1 Introduction to PEST analysis

PEST analysis is concerned with the environmental influences on a business.

The acronym stands for the Political, Economic, Social and Technological issues that could affect the strategic development of a business.

Identifying PEST influences is a useful way of summarizing the external environment in which a business operates. However, it must be followed up by consideration of how a business should respond to these influences.

The table below lists some possible factors that could indicate important environmental influences for a business under the PEST headings:

Table 10-2 PEST factors

Political/Legal	Economic	Social	Technological
Environmental regulation and protection	Economic growth (overall; by industry sector)	Income distribution (change in distribution of disposable income	Government spending on research
Taxation (corporate; consumer)	Monetary policy (interest rates)	Demographics (age structure of the population; gender; family size and composition; changing nature of occupations)	Government and industry focus on technological effort
International trade regulation	Government spending (overall level; specific spending priorities)	Labour/social mobility	New discoveries and development
Consumer protection	Policy towards unemployment (minimum wage, unemployment benefits, grants)	Lifestyle changes (e.g. Home working, single households)	Speed of technology transfer
Employment law	Taxation (impact on consumer disposable income, incentives to invest in capital equipment, corporation tax rates)	Attitudes to work and leisure	Rates of technological obsolescence
Government organization/ attitude	Exchange rates (effects on demand by overseas customers; effect on cost of imported components)	Education	Energy use and costs
Competition regulation	Inflation (effect on costs and selling prices)	Fashions and fads	Changes in material sciences
	Stage of the business cycle (effect on short-term business performance)	Health & welfare	Impact of changes in Information technology
	Economic 'mood' – consumer confidence	Living conditions (housing, amenities, pollution)	Internet!

3.2.2 Porter's Five Forces Model – Analyzing Competitive Industry Structure

In his book, *Competitive Strategy* (Free Press: 1980), Michael Porter identifies three fundamental competitive strategies and lays out the required skills and resources, organizational elements and risks associated with each strategy. The table below is a shorthand way of referring to what Porter has to say.

Table 10-3 Porter's Fundamental Strategies

Competitive Strategy	Required Skills and Resources	Organizational Elements	Associated Risks
Overall Cost Leadership	Sustained capital investment and access to capital.	Tight cost control.	Technological change that nullifies past investments or learning.
	Process engineering skills.	Frequent, detailed reports.	Low-cost learning by industry newcomers or followers through imitation, or through their ability to invest in state-of-the-art facilities.
	Intensive supervision of labour.	Structured organization and responsibilities.	Inability to see required product or marketing change because of the attention placed on cost.
	Products designed for ease of manufacture.	Incentives based on meeting strict quantitative targets.	Inflation in costs that narrow the firm's ability to maintain enough of a price differential to offset competitors' brand images or other approaches to differentiation.
Differentiation	Strong marketing abilities.	Strong coordination among functions in R&D, product development, and marketing.	The cost differential between low-cost competitors and the differentiated firm becomes too great for differentiation to hold brand loyalty. Buyers thus sacrifice some of the features, services, or image possessed by the differentiated firm for large cost savings.
	Product engineering with creative flair.	Subject measurement and incentives instead of quantitative measures.	
	Strong capability in basic research.		
	Corporate reputation for quality or technological leadership.	Amenities to attract highly skilled labour, scientists, or creative people.	Buyers need for the differentiating factor falls. This can occur as buyers become more sophisticated.
	Long tradition in the industry or unique combination of skills drawn from other businesses.		Imitation narrows perceived differentiation, a common occurrence as industries mature.
	Strong co-operation from channels.		

Focus	Combination of the above policies directed at the particular strategic target.	Combination of the above policies directed at the particular strategic target.	The cost differential between broad-range competitors and the focussed firm widens to eliminate the cost advantages of serving a narrow target or to offset the differentiation achieved by focus.
			The Differences in desired products or services between the strategic target and the market as a whole narrows. Competitors find submarkets within the strategic target and outfocus the focuser.

Porter is considered the undisputed guru of competitive strategy. In the same book he identifies five forces that drive competition within an industry:

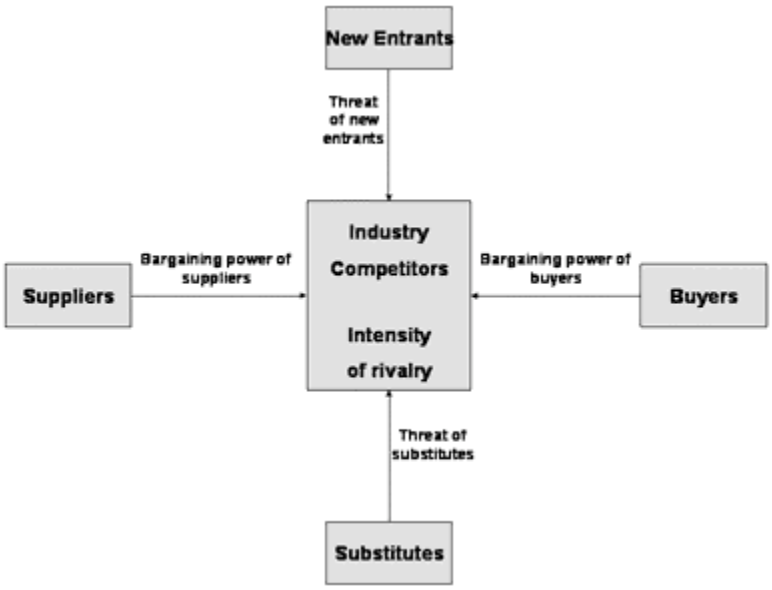
1. The threat of entry by new competitors.
2. The intensity of rivalry among existing competitors.
3. Pressure from substitute products.
4. The bargaining power of buyers.
5. The bargaining power of suppliers.

One obvious application of all this is to would-be entrants and the problem of entering new markets. Another is to the current competitors and the ongoing task of staying competitive in markets where they already operate.

Perhaps the most important thing to keep in mind is the inverse relationship between profit margins or returns and the intensity of competition: as the intensity of competition goes up, margins and returns are driven down. This can require changes in competitive strategy to remain in an industry and, under some circumstances, it can occasion the decision to exit a business or an industry.

The diagram in Figure 2 is a handy way of depicting the five forces Porter identifies and providing some elaborating detail. However, the diagram is no substitute for reading what Porter has to say.

Figure 10-2



3.2.3 Threat of New Entrants

New entrants to an industry can raise the level of competition, thereby reducing its attractiveness. The threat of new entrants largely depends on the barriers to entry. High entry barriers exist in some industries (e.g., shipbuilding) whereas other industries are very easy to enter (e.g., estate agency, restaurants). Key barriers to entry include:

- Economies of scale
- Capital/investment requirements
- Customer switching costs
- Access to industry distribution channels
- The likelihood of retaliation from existing industry players.

3.2.4 Threat of Substitutes

The presence of substitute products can lower industry attractiveness and profitability because they limit price levels. The threat of substitute products depends on:

- Buyers' willingness to substitute
- The relative price and performance of substitutes
- The costs of switching to substitutes.

- Bargaining Power of Suppliers (Suppliers are the businesses that supply materials and other products into the industry)

The cost of items bought from suppliers (e.g. raw materials, components) can have a significant impact on a company's profitability. If suppliers have high bargaining power over a company, then in theory the company's industry is less attractive. The bargaining power of suppliers will be high when:

- There are many buyers and few dominant suppliers
- There are undifferentiated, highly valued products
- Suppliers threaten to integrate forward into the industry (e.g., brand manufacturers threatening to set up their own retail outlets)
- Buyers do not threaten to integrate backwards into supply
- The industry is not a key customer group to the suppliers

3.2.5 Bargaining Power of Buyers

Buyers are the people/organizations who create demand in an industry.

The bargaining power of buyers is greater when:

- There are few dominant buyers and many sellers in the industry
- Products are standardized
- Buyers threaten to integrate backward into the industry
- Suppliers do not threaten to integrate forward into the buyer's industry
- The industry is not a key supplying group for buyers.

3.2.6 Intensity of Rivalry

The intensity of rivalry between competitors in an industry will depend on:

- The structure of competition – for example, rivalry is more intense where there are many small or equally sized competitors; rivalry is less when an industry has a clear market leader.
- The structure of industry costs – for example, industries with **high fixed costs** encourage competitors to fill unused capacity by price cutting.

- Degree of differentiation – industries where products are commodities (e.g., steel, coal) have greater rivalry; industries where competitors can differentiate their products have less rivalry.

- Switching costs – rivalry is reduced where buyers have high switching costs, i.e., there is a significant cost associated with the decision to buy a product from an alternative supplier.

- Strategic objectives – when competitors are pursuing aggressive growth strategies, rivalry is more intense. Where competitors are 'milking' profits in a mature industry, the degree of rivalry is less.

- Exit barriers – when barriers to leaving an industry are high (e.g., the cost of closing down factories) – then competitors tend to exhibit greater rivalry.

3.4 SWOT Analysis

SWOT is an abbreviation for Strengths, Weaknesses, Opportunities, and Threats and is an important tool for auditing the overall strategic position of a business and its environment.

Once key strategic issues have been identified, they feed into business objectives, particularly marketing objectives. SWOT analysis can be used in conjunction with other tools for audit and analysis, such as the PEST analysis and Porter's Five-Forces analysis. It is also a very popular tool with business and marketing students because it is quick and easy to learn.

3.4.1 The Key Distinction: Internal and External Issues

Strengths and weaknesses are internal factors. For example, strength could be your specialist marketing expertise. A weakness could be the lack of a new product.

Opportunities and threats are external factors. For example, an opportunity could be a developing distribution channel such as the Internet, or changing consumer lifestyles that potentially increase demand for a company's products. A threat could be a new competitor in an important existing market or a technological change that makes existing products potentially obsolete.

It is worth pointing out that SWOT analysis can be very subjective – two people rarely come up with the same version of a SWOT analysis even when given the same information about the same business and its environment. Accordingly, SWOT analysis is best used as a guide and not a prescription. Adding and weighting criteria to each factor increases the validity of the analysis.

3.4.2 Areas to Consider

Some of the key areas to consider when identifying and evaluating Strengths, Weaknesses, Opportunities, and Threats are listed in the example SWOT analysis below.

Table 10-4 SWOT Analysis

	Positive	Negative
	Strengths	**Weaknesses**
Internal Factors	Technological skills Leading Brands Distribution Channels Customer Loyalty/Relationship Production Quality Scale Management	Absence of important skills Weak brands Poor access to distribution Low customer retention Unreliable product/service Sub-scale Management
	Opportunities	**Threats**
External Factors	Changing customer tastes Liberalization of geographic markets Technological advances Changes in government politics Lower personal taxes Change in population age structure New distribution channels	Changing customer tastes Closing of geographic markets Technological advances Changes in government politics Tax increases Change in population age structure New distribution channels

3.5 Ansoff Product/Market Matrix

The Ansoff Growth matrix is a tool that helps businesses decide their product and market growth strategy.

Ansoff's product/market growth matrix suggests that the attempts of a business to grow depend on whether it markets new or existing products in new or existing markets.

The output from the Ansoff product/market matrix is a series of suggested growth strategies that set the direction for the business strategy. These are described below:

Table 10-5 Ansoff Product/Market Matrix

	Existing Products	New Products
Existing Markets	Market Penetration	Product Development
New Markets	Market Development	Diversification

3.5.1 Market Penetration

Market penetration is the name given to a growth strategy where the business focuses on selling existing products into existing markets.

Market penetration seeks to achieve four main objectives:

1. Maintain or increase the market share of current products – this can be achieved by a combination of competitive pricing strategies, advertising, sales promotion and perhaps more resources dedicated to personal selling.

2. Secure dominance of growth markets.

3. Restructure a mature market by driving out competitors; this would require a much more aggressive promotional campaign, supported by a pricing strategy designed to make the market unattractive for competitors.

4. Increase usage by existing customers – for example by introducing loyalty schemes.

A market penetration marketing strategy is very much about 'business as usual.' The business is focusing on markets and products it knows well. It is likely to have good information on competitors and on customer needs. It is unlikely, therefore, that this strategy will require much investment in new market research.

3.5.2 Market Development

Market development is the name given to a growth strategy where the business seeks to sell its existing products into new markets.

There are many possible ways of approaching this strategy, including:

- New geographical markets; for example, exporting the product to a new country

- New product dimensions or packaging, for example

- New distribution channels
- Different pricing policies to attract different customers or create new market segments.

3.5.3 Product Development

Product development is the name given to a growth strategy where a business aims to introduce new products into existing markets. This strategy may require the development of new competencies and requires the business to develop modified products which can appeal to existing markets.

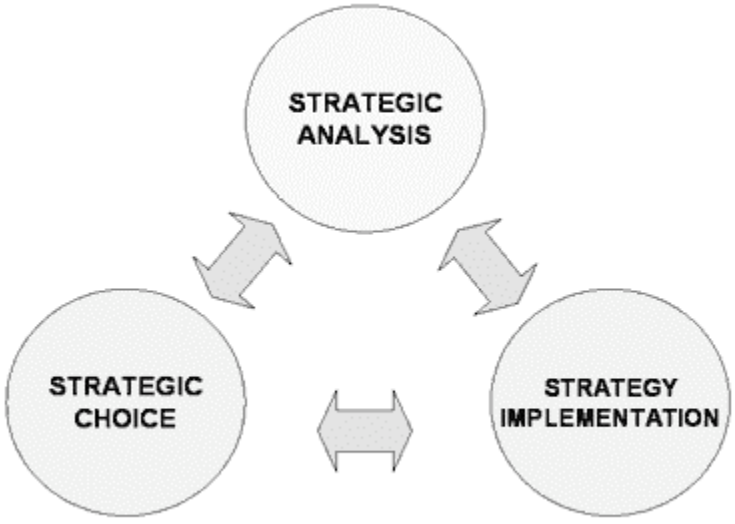
3.5.4 Diversification

Diversification is the name given to the growth strategy where a business markets new products in new markets.

This strategy is inherently more risky because the business is moving into markets in which it has little or no experience.

For a business to adopt a diversification strategy, therefore, it must have a clear idea about what it expects to gain from the strategy and an honest assessment of the risks.

4 Strategic Choice and Implementation



Recall the above diagram from earlier in this chapter. It is important to note that the analytic tools described in brief above are of no use if the manager cannot apply them, develop alternatives from the information they produce, pick a recommendation he or she

feels meets the goals of the project, department or company, and implements it. The application of any of the models above is fairly straightforward – using the models, compartmentalizing the information at hand to create an understanding of an industry, the company, or its opportunities. But how do you choose the best one?

The answer to this question relies on understanding your strengths and weaknesses (from the SWOT model), and how the company can best outweigh the weaknesses with the strengths to become more competitive. In other words, how do you leverage what you do best, your competitive advantages, to create a sustainable market or competitive position – this is the real question asked in choosing a recommended strategy. There are many components that play into this – the company's financial situation, the human resources available inside the company, and the skill set inherent in the company as a whole. Ultimately, choosing your strategy must fit with both what the company is able to do and what the company wants to do – the models above are simple guidelines to help you understand more completely what the details are to those components.

However, what is often more challenging to determine is what your competition is doing, and how your decisions affect your competitors' choices. This is the focus of the next section, Game Theory.

5 Game Theory – A Basic Definition, Prisoner's Dilemma, and Key Concepts

An excellent introductory definition to Game Theory comes from the *Economist*:

> Stripped to its essentials, game theory is a tool for understanding how decisions affect each other. Until the theory came along, economists assumed that firms could ignore the effects of their behaviour on the actions of rivals, which was fine when competition was perfect or a monopolist held sway, but was otherwise misleading. Game theorists argue that firms can learn from game players: no card player plans his strategy without thinking about how other players are planning theirs.
>
> *Economist*, June 15, 1996

Building on this definition, then, Game Theory analyzes strategic interactions in which the outcome of one's choices depends upon the choices of others. Perhaps the easiest-to-understand, and most famous, example of Game Theory is what is known as the Prisoner's Dilemma, and can be illustrated as follows.

Sasha and Richard are two hard-bitten criminals who have recently been caught for a crime that they committed together. They are being held in separate cells, and each has been offered the opportunity to confess to the crime. Now, Sasha and Richard have no special feeling for each other, they just want to do as little time as possible. The prosecuting attorney visits Sasha and says he's willing to make a deal with her. He's also offering the same deal to Richard, and she is aware of that.

He says, ‘We have some circumstantial evidence on both of you, and if neither of you tells me anything we can still get both of you a year in jail, the way things stand right now. But, if you confess, and admit your partner was with you, then we’ll let you off scott free because you were helpful, and he’ll get three years in jail. Of course, if *he* confesses and you don’t, then *you’re* the one who gets the three years and he walks free. Now, if you both confess, then we’ve got you both dead to rights and you both get two years in jail.’

The crown attorney leaves and goes back to his office, after telling Sasha he’ll be back in half an hour to get her answer. Here’s the dilemma: she can either co-operate with her partner and not confess, or try to leave Richard in the dust by confessing. If she confesses, she does either no time (if Richard doesn’t confess) or does two years (if Richard confesses, as well). If she doesn’t confess – if she co-operates with Richard – she either does one year in jail (assuming Richard doesn’t confess), or does three years (if Richard confesses and she doesn’t).

The kicker, of course, is that Richard is thinking exactly the same thing. This, then, is the dilemma. By doing what is best for themselves – by being selfish – they will do worse than by following a different, and apparently illogical, plan. Both will confess, assuming the other will, and therefore get two years in jail. Ideally, neither of them would confess, and get only one year in jail. However, the risk that the other criminal will confess in order to hedge his or her bets is too high.

This illustration of the Prisoner’s Dilemma highlights a number of key parts of Game Theory important for any analysis:

- The Players – Identify who the decision makers affecting the game are.
- Strategies – Understand all the feasible actions that can be taken in this game.
- Payoffs – Recognize the objectives each player has.
- Incentives – Understand the motives behind what the decision makers will do.

This example also helps the flesh out some of the basic rules of Game Theory:

- Time Frame – all games happen in a time frame, so identify it! Are the moves simultaneous or sequential?
- Nature of the Conflict – is it in the player’s interests to conflict or co-operate? Furthermore, will the players interact once, or repeatedly?
- Nature of Interaction – What is driving the decisions of the players?
- Information – Identify the information, and you’ve identified another aspect of the strategies involved.

And, in any game, you want to keep the following in mind:

- Are the threats by the parties credible?
- Can co-operation be sustained?
- Are there informational advantages out there?
- Assume that your opponent is as rational as you are – namely, that he or she is going to maximize his or her payoffs, and are perfect calculators of those payoffs.
- Assume that your opponent has access to the same common knowledge that you do.

5.1 The Game Matrix, Equilibrium, Dominant Strategies, and Nash Equilibrium

So, Game Theory is the description of strategic interaction between mutually aware players. You are self-interested and selfish, and so is everyone else.

The most useful way to develop an understanding of the Prisoner's Dilemma is if we quantify it. To do this, we use what is termed a game matrix, and it looks like this:

Table 10-6 Prisoner's Dilemma

		Richard	
		Co-operate	Confess
Sasha	Co-operate	(1,1)	(3,0)
	Confess	(0,3)	(2,2)

If Sasha and Richard co-operate with each other, the result can be read as (1, 1) for (Sasha, Richard). As we have noted before, the theory behind the Prisoner's Dilemma predicts that Sasha and Richard will sell each other out, as they are selfish and are following their own self-interest.

This, however, illustrates the central concept to Game Theory – the equilibrium – which is also a central component to the study of economics. The difference is that this particular example illustrates what is called a dominant strategy equilibrium – that is, the incentive of each player to confess does not depend on how the other player plays. Dominant strategy is the most persuasive notion of equilibrium known to game theorists.

However, obviously, there are games without Dominant Strategies. David K. Levine, a professor of Economics at UCLA, proposes the following example to contrast with the Prisoner's Dilemma.

Consider the following game matrix:

Table 10-7 Battle of the Sexes Game Matrix

		Wife	
		opera	ballgame
Husband	opera	1,2	0,0
	ballgame	0,0	2,1

This is known as the *Battle of the Sexes* game. The story goes that a husband and wife must agree on how to spend the evening. The husband prefers to go to the ballgame (2 instead of 1), and the wife to the opera (also 2 instead of 1). However, they prefer agreement to disagreement, so if they disagree both get 0. This game does not admit a dominant strategy equilibrium. If the husband thinks the wife's strategy is to choose the opera, his *best response* is to choose opera rather than ballgame (1 instead of 0). Conversely, if he thinks the wife's strategy is to choose the ballgame, his best response is ballgame (2 instead of 0). While in the prisoner's dilemma the best response does not depend on what the other player is thought to be doing, in the battle of the sexes the best response depends entirely on what the other player is thought to be doing. This is sometime called a *coordination game* to reflect the fact that each player wants to coordinate with the other player.

For games without dominant strategies the equilibrium notion most widely used by game theorists is that of *Nash equilibrium*. In a Nash equilibrium, each player plays a best response, and correctly anticipates that her opponent will do the same. The battle of the sexes game has two Nash equilibria: both go to the opera, or both go to the ball game; if each expects the other to go to the opera (ballgame), the best response is to go to the opera (ballgame). By way of contrast, one going to the opera and one to the ballgame is not a Nash equilibrium, since each correctly anticipates that the other is doing the opposite, neither one is a playing a best response.

Game Theory is an essential component of business strategy, as it can be applied to the cornerstone of any decision in business. It teaches managers about strategy – commitment versus flexibility, decision making in uncertain conditions, and gives them the opportunity to look forward and reason backwards. It can, in the words of Tom Copland, Director of corporate finance for McKinsey and Company, 'explain why oligopolies tend to be unprofitable, the cycle of over capacity and overbuilding, and the tendency to execute real options earlier than optimal.'

Game Theory should never be applied in a vacuum – there is no substitute for experience and learning on the job – but it can certainly provide a framework within which that experience can be applied at the theoretical level before any action is taken. Good managers will be able to look at their strategic alternatives and choose amongst them based not only on internal competitive advantage, but also with an understanding of how their choices are affected by, and will affect, the choices of their competitors.

6 Summary and Review

Strategy combines three key elements: analysis, strategic choice, and strategic implementation. This block focuses on helping the student understand how to properly analyse a situation to understand its strategic significance, using the following models:

PEST Analysis – a technique for understanding the 'environment' in which a business operates, this analysis helps the manager understand the Political, Environmental, Social, and Technological influences on a given industry or business opportunity.

Five Forces Analysis – a model that helps managers understand the five key forces influencing an industry, namely Supplier Power, Buyer Power, the threat of new Entrants, the Substitute products available to the market, and the internal rivalry in the industry. By completing this analysis, a manager will have a strong idea of trends and pressures at work that help define the industry, including what the industry drivers are and how concentrated the industry is.

SWOT Analysis – This is an internal/external analysis model that focuses on the Strengths and Weaknesses of a company, and compares them to the Opportunities and Threats the lie in the industry. Once completed, this model helps a manager understand how a company's strengths best match up to take advantage of the opportunities in the marketplace, despite the weaknesses of the company and the threats to, and in, the industry.

Ansoff Product/Market Matrix – A basic analytic tool to help a company understand how its growth strategy in specific markets can correlate to existing or potential products, including market penetration, market development, product development, and product diversification.

Developing strategic alternatives, choosing and implementing a strategy, is done only after as much information has been gathered as possible. Strategic implementation relies on what resources, both capital and human, are available to the company given the constraints of the situation it faces.

7 Review Problems and Answer Key

There are no review problems for this block.

It is only through application of these models that students will thoroughly understand their guiding principles and their uses. Therefore, this section is intended to prepare the student, in combination with the other nine blocks, to focus on the cases at the end of this course. Each case must be reviewed based on the merits of the information at hand, and the students must decide which model(s) will best be used in analyzing a given case. It will be through this analysis that the student understands what information is available, what information is missing, and whether strategic alternatives exist given the information available.

8 References and Suggested Further Reading

Andrew, Kenneth. 1980. *The Concept of Corporate Strategy*. Dow-Jones Irwin.

de Witt, Bob. 2000. *Strategy – Process, Content, Context*. ITP.

Mintzberg, Henry. 1994. *The Rise and Fall of Strategic Planning*. Basic Books.

Porter, Michael. 1986. *Competitive Strategy*. Free Press.

Robert, Michel. 1993. *Strategy: Pure and Simple*. McGraw-Hill.

Steiner, George. 1979. *Strategic Planning*. Free Press.

Treacy, Michael and Fred Wiersema. 1994. *The Discipline of Market Leaders*. Addison-Wesley.