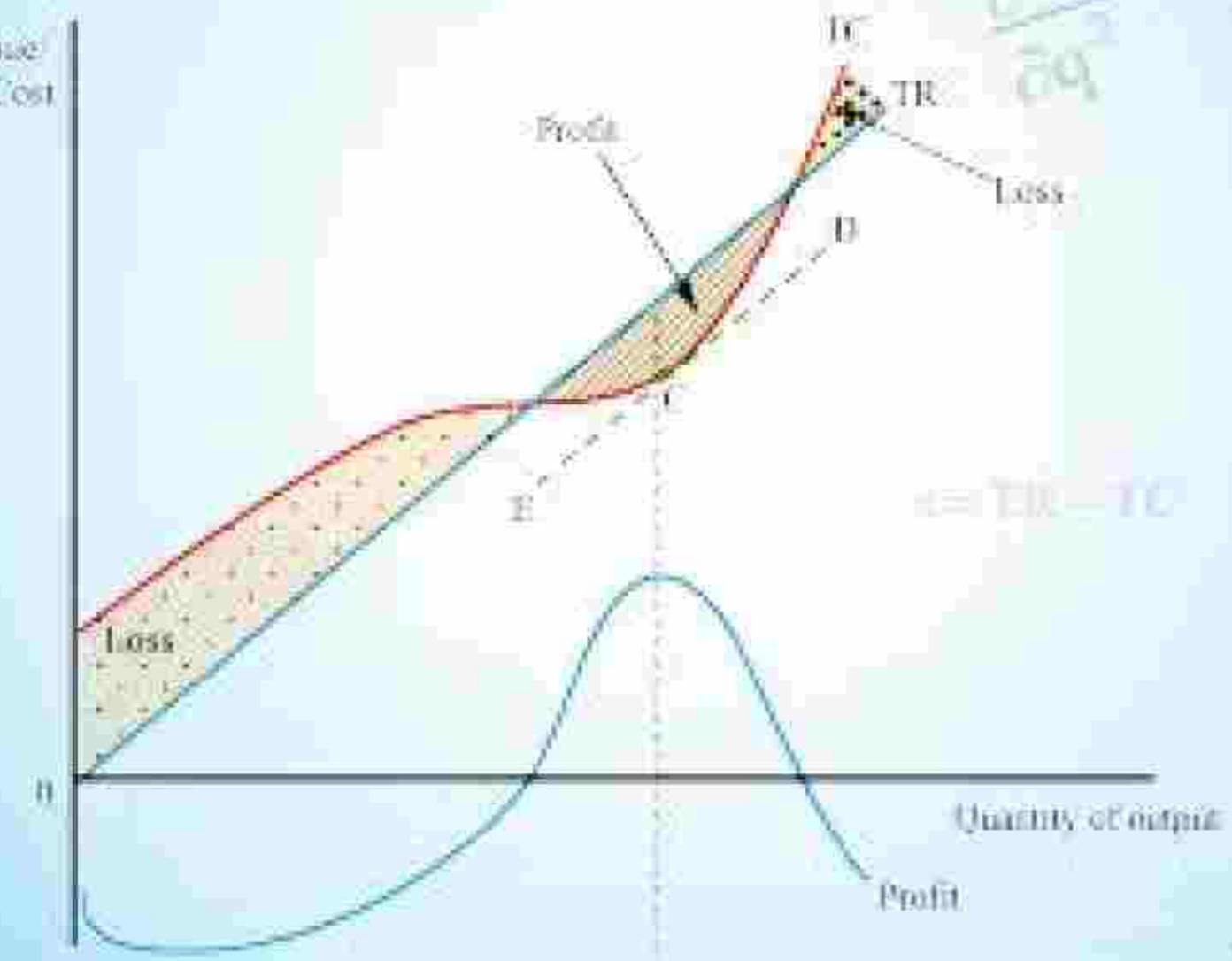
Microeconomics for Advanced Secondary Schools

Student's Book

Form Five



Tanzania Institute of Education

Property of the Government of the United Republic of Tanzania, Not for Sale

Microeconomics for Advanced Secondary Schools

Student's Book Form Five

THE UNITED REPUBLIC OF TANZANIA
MINISTRY OF EDUCATION,
SCIENCE AND TECHNOLOGY

Certificate of Approval

No. 742

Title of Publication: Microeconomics for Advanced Secondary Schools Student's Book Form Five

Publisher: Tanzania Institute of Education

Author: Tanzania Institute of Education

ISBN: 978-9987-09-466-0

This book was approved by the Ministry of Education, Science and Technology on 2nd March, 2022 as a Textbook for Form Five students in Tanzania Secondary Schools as per 2010 Syllabus.

Dr Lyabwene M. Mtahabwa
Commissioner for Education

Tanzania Institute of Education

©Tanzania Institute of Education 2022

Published 2022

ISBN: 978-9987-09-466-0

Tanzania Institute of Education
P. O. Box 35094
Dar es Salaam, Tanzania
Mobile numbers: +255 735 041 168/735 041 170
Email: director.general@tie.go.tz
Website: www.tie.go.tz

All rights reserved. No part of this textbook may be reproduced, stored in any retrieval system or transmitted in any form or by any means whether electronic, mechanical, photocopying, recording, or otherwise, without prior written permission of the Tanzania Institute of Education.

Table of Contents

List of Figures v
List of Tables viii
Acknowledgements ix
Preface x

Chapter One

Subject matter of economics 1

Economic terminologies 2
Definitions of economics 12
Branches of economics 18
The nature of economics 24
Economic laws 26
The economic problem 28
Economic systems 36

Chapter Two

Theories of demand and supply 46

The concept of demand 46
The law of demand 47
The concept of elasticity of demand 65
The concept of supply 88
The law of supply 89
The concept of elasticity of supply 98
Market equilibrium 105

Chapter Three

Production theory 120

Production 120
Factors of production 121
Scales of production 152

Production function 154
Returns to scale 166
Economies and diseconomies of scale 168

Chapter Four
Theory of the firm 177
Concept of the firm 177
Costs of production 183
Cost function 186
Revenue analysis and profit maximisation of a firm 203

Chapter Five
The market 219
Concept of a market 219
Classification of markets 222

Chapter Six
Market failure 259
Causes of market failure 260
Ways of controlling market failure 271

Glossary 280
Bibliography 284
Index 285

List of Figures

1.1: Production possibilities frontier (PPF) 34
1.2: Linear production possibilities frontier.....34
1.3: Shift of the production possibilities frontier (PPF)-economic growth ..35
2.1: Demand curve 51
2.2: Market demand curve (aggregate demand curve for Fatuma + John) ...52
2.3: Exceptional demand curve.....55
2.4: Change in the quantity demanded.....58
2.5: Change in demand 58
2.6: Competitive demand of two substitute goods.....61
2.7: Joint or complementary demand.....62
2.8: Increase in demand for sunflower oil.....63
2.9: Increase in demand for labour 63
2.10: Demand curve showing arc and point price elasticity of demand 67
2.11: Perfectly elastic demand curve 69
2.12: Perfectly inelastic demand curve 70
2.13: Unitary elastic demand curve 70
2.14: Elastic demand curve 71
2.15: Inelastic demand curve 72
2.16: Negative income elasticity of demand curve.....78
2.17: Positive income elasticity of demand curve 79
2.18: Zero income elasticity of demand curve.....80
2.19: Positive cross-price elasticity of demand.....83
2.20: Negative cross-price elasticity of demand curve 84
2.21: Zero cross-price elasticity of demand curve 84
2.22: Supply curve 90
2.23: Market supply curve for individuals A and B 92
2.24: Change in quantity supplied of a good 95
2.25: Change in supply-increase in supply 96
2.26: Change in supply-decrease in supply.....96

2.27: Joint supply of beef and leather 97
2.28: Competitive supply of banana and maize 98
2.29: Perfectly elastic supply 100
2.30: Perfectly inelastic supply 100
2.31: Unitary elastic supply 101
2.32: Elastic supply 102
2.33: Inelastic supply 102
2.34: Market equilibrium 105
2.35: Price floor 112
2.36: Price ceiling 113
2.37: Market adjustment due to increase in demand 114
2.38: Market adjustment due to decrease in demand 115
2.39: Market adjustment due to decrease in supply 116
2.40: Market adjustment due to increase in supply 116

3.1: The production process 121
3.2: Wage determination by the intersection demand for and supply of labour 141
3.3: Relationship between wealth and capital 144
3.4: A total product curve 156
3.5: Stages of production 164

4.1: Short-run cost curves 188
4.2: Cost curves 191
4.3: Relationship between average product and average variable cost 192
4.4: Relationship between marginal product and marginal cost 196
4.5: Long-run average cost curve 201
4.6: Optimum firm 202
4.7: Profit maximisation curve of a perfectly competitive firm 212
4.8: Equilibrium of the firm 213

5.1: Digital marketing 220
5.2: Demand curve under perfect competitive market 224

5.3: Short-run supernormal profit under a perfect competition market 226
5.4: Short-run normal profit under a perfect competition market.....227
5.5: Short-run loss under a perfect competition market.....228
5.6: Break-even and shut down point for a perfectly competitive firm 229
5.7: Short-run supply curve of a perfect competition market 230
5.8: Long-run profit maximisation under perfect competition.....231
5.9: Monopoly demand curve 236
5.10: Short-run profit maximisation under monopoly market 237
5.11: Long-run profit maximisation under monopoly.....238
5.12: Loss under monopoly market 239
5.13: Price discrimination in a monopoly market.....241
5.14: Monopolistic competition market demand curve 245
5.15: Supernormal profit under monopolistic competition 246
5.16: Normal profit under monopolistic competition market 247
5.17: Loss under monopolistic competition market.....248
5.18: Long-run normal profit under monopolistic competition 249
5.19: Kinked demand curve under oligopoly market.....251
5.20: Profit maximisation under oligopoly market 253

6.1: Negative externalities 264
6.2: Positive externality 267
6.3: Emission charges 272
6.4: Pigouvian taxation 273
6.5: Subsidy.....275

List of Tables

1.1: Differences between microeconomics and macroeconomics 20
2.1: Hypothetical individual annual demand schedule for boxes of pen 49
2.2: Hypothetical market demand schedule for boxes of pen 50
2.3: Hypothetical individual supply schedule of beef.....89
2.4: Hypothetical market supply schedule for beef 90

3.1: Rent charged for different grades of land 124
3.2: Difference between large-scale and small-scale of production 154
3.3: Production function.....155
3.4: Total Product (TP), Average Product (AP) and Marginal Product (MP).....157
3.5: Operations of law of diminishing returns 164
3.6: Phases of law of returns to scale.....167

4.1: The relationship between costs and production 193

6.1: Types of goods 269

Acknowledgements

The Tanzania Institute of Education (TIE) would like to acknowledge the contributions of all organisations and individuals who participated in designing and developing this textbook. In particular, TIE wishes to thank The University of Dar es Salaam (UDSM), Sokoine University of Agriculture (SUA), Mzumbe University (MU), University of Dodoma (UDOM) and Dar es Salaam University College of Education (DUCE), School Quality Assurance (SQA) Department, teachers' colleges, and secondary schools. Besides, the following individuals are acknowledged:

Writers: Ms Doreen C. Samuel, Mr Nehemia Kaaya & Mr Elineema G. Moshi (TIE), Dr Eliaza Mkuna (MU), Dr Gabriel Hinju (DUCE), Dr Gerald Kibira (UDSM), Mr Evance Komba (Tusiime S.S) & Mr Allan Kitunga (Minaki S.S)

Editors: Prof. Joseph Kuzilwa (MU), Prof. Michael O. A. Ndashau (UDOM), Dr Elibariki E. Msuya (SUA), Dr Robert Lihawa & Dr Coretha Komba (MU) & Dr Innocent J. Karamagi (UDSM)

Designer: Ms Rehema H. Maganga

Illustrators: Mr Godlove Kyando & Mr Fikiri Msimbe

Coordinator: Ms Doreen C. Samuel

TIE also appreciates the participation of the secondary school teachers and students in the trial phase of the manuscript. Likewise, the Institute would like to thank the Ministry of Education, Science and Technology for facilitating the writing and printing of this textbook.

Dr. Aneth A. Komba
Director General
Tanzania Institute of Education

Preface

This textbook, *Microeconomics for Advanced Secondary Schools*, is written specifically for Form Five students in the United Republic of Tanzania. It is written in accordance with the 2010 Economics Syllabus for Advanced Secondary Education, Form V-VI, issued by the then Ministry of Education and Vocational Training.

The book consists of six chapters, namely, Subject matter of economics, Theories of demand and supply, Production theory, Theory of the firm, The market and Market failure. Each chapter contains text, illustrations, activities and exercises. You are encouraged to do all activities and exercises as well as other assignments that your teacher will provide. Doing so will enable you to develop the intended competencies.

Tanzania Institute of Education

Chapter One

Subject matter of economics

Introduction

There are different cultures in every society. This is the same in various fields of studies. In the field of medicine, for a person to become a doctor, he or she needs to learn the doctors' language and their way of thinking. The same applies in the field of economics. Economics as a social science is concerned with human behaviour in the allocation of scarce resources to satisfy unlimited wants. Thus, for a person to become a good economist, he or she needs to learn economics. Therefore, the subject matter of economics enables a person to learn the economic language and think as an economist. In this chapter, you will learn about economic terminologies, various definitions of economics, the branches and nature of economics, economic laws, economic problems, and economic systems. The competencies developed will enable you to make rational economic decision in solving economic problems in different economic contexts. Moreover, you will be able to apply various economic terms accordingly.

The scope of economics

Economics is a study of how people allocate resources at their disposal to various requirements of individuals and their society. Allocation of resources involves decisions of individuals, businesses, governments, and nations in choosing how to use different types of resources in undertaking their economic activities.

The scope of economics encompasses all what economics is about. It includes issues of wealth, human behaviour and scarce resources, among others. It also, addresses whether economics is a science or an art, and whether its approach is positive or normative. This chapter aims at sheding light on the nature of economics by explaining to you terminologies used in economics, various definitions of economics, economic laws and economic problems, among other issues that you will find in the chapter.

Economic terminologies

The subject matter of economics involves understanding the main economic activities, which include consumption, production, exchange and distribution. Other terminologies that are important to understand the subject matter of economics include goods, wants, wealth, economic resources and utility.

Production

Nature does not provide freely things people need in their lives. Even with wild fruits, which people do not cultivate, they have to pluck them. Production involves decisions and activities of creating goods and services that people need in their lives. Such goods and services have value and when obtained, they satisfy individual wants. Some of the goods that people produce are not for sale, rather, they are for personal consumption, for instance, a farmer may produce maize for family needs only. However, economics is concerned mainly with production of goods and services for exchange in various markets, for example, farmers may produce maize for sale at a local market. In selling their products, producers aim to make profit.

A producer of any good or service uses technology to combine and process inputs to produce output. The inputs used in the production process are referred to as factors of production and the resulting outcome is an output. For example, a farmer uses the following factors of production to produce crops: labour, capital, land and entrepreneurship.

Distribution

Distribution as used generally in marketing, is a process of moving goods and services from one location to another. Transportation is used to move goods from the place of production to the market. For example, the transportation of manufactured sugar from Mtibwa Sugar Company in Morogoro Region to wholesalers and retailers located in various regions of Tanzania or to other countries.

In economics, the term distribution has a different meaning. Distribution means an act of rewarding factors of production, that is, the distribution of rewards to different owners of factors of production used in producing a certain product. It

involves payments for the use of factors of production. The payments are wages for labour, rent for land, interest for capital and profit for entrepreneurship.

Exchange

Exchange involves buying and selling of goods and services. The process of exchange involves the use of cash, debit or credit cards, shares, bills of exchange and bonds, among others. Exchange enables individuals to consume items they do not produce and sellers to produce what they do not have to consume. For example, a carpenter produces and sells furniture to earn money (income). With the earned income, the carpenter can use it to buy goods that he or she does not produce, such as food and clothes, or services, such as health.

Consumption

Consumption is the act of buying and using goods and services to satisfy human wants. Economists assume that an individual who consumes a good or service derives utility (satisfaction). In everyday life, individuals consume different kinds of goods and services to satisfy their wants. For instance, people consume food to satisfy hunger, buy cars for transportation and visit banks for financial services. Moreover, individuals consume different services provided by doctors, teachers, the army, taxi drivers and tourist guides.

Note that the term consumption is not used for goods bought for production of other goods, for example, seeds. Rather, consumption of a good or service refers to only a good or service for final use.

Goods

Goods are items sold in the market with a general assumption that the more quantity of a good an individual consumes the more utility (satisfaction) they derive. Goods could be physical objects or services. Examples of physical goods are cars, furniture, clothes and houses. Individuals also consume services, for example, the work performed by doctors, teachers, bankers, actors, lawyers and car repairers. However, goods can be categorised in different groups, for example, economic and free goods, consumer and producer goods, durable and perishable goods, public and private goods, and merit and demerit goods.

Economic goods and free goods

Economic goods: These goods are scarce such that quantity demanded exceeds quantity supplied when they are to be availed for free (at a zero price). Economic goods form a large proportion of all goods demanded by an individual. They include goods such as cars, land, computers and clothes. Because economic goods are scarce, the theory postulates how a consumer reaches the right decision regarding the quantity of goods to buy (consume). Economic goods have the following features:

(a) Must give utility or satisfaction when consumed. People consume an economic good to satisfy a want (a particular human need) and shun from buying something that is a nuisance or irritant and does not give them utility (satisfaction).Thus, the latter does not qualify as an economic good.

(b) An economic good must be limited in supply in relation to its demand, otherwise, people would not be willing to pay a price for it. It means that something scarce commands a price. People would not be willing to pay a price for something that is plenty in supply such that it would still be inexhaustible when anyone who needs it can get whatever share they need. Such a thing would be freely available.

(c) An economic good is transferable in terms of ownership. For example, transfer of land ownership takes place from one owner to a buyer, who becomes a new owner, through payment of land price. This qualifies land as an economic good. A good that is transferable in terms of ownership makes it possible for one person to sell it to another person.

(d) Economic goods have price or exchange value. Because economic goods are scarce, people have to pay a price to get them. Thus, the market price of economic goods is greater than zero. A price greater than zero means that when people shop for their needs, they have to pay some money for any item they buy. In this regard, sellers exchange goods for the money they receive and the money that a buyer pays for the goods is the exchange value of the bought goods.

(e) Economic goods are products of economic resources. People produce economic goods using factors of production, which include land, labour, capital, and entrepreneurship. Note that land is called a gift of nature, yet it is included as one of factors of production. Land does not become a factor of production in its natural state. It becomes so after it has been prepared to qualify as an economic resource. For example, farmers work on land

preparation to make it a factor of production, through clearing, cultivating, irrigating, fertilising and planting crops on it.

A Free good: This good is readily available in unlimited quantity such that every individual may be able to consume any quantity to their satisfaction without limiting other consumers' demands for the good. It is said that opportunity cost of a free good is zero. The term opportunity cost in this case means an additional supply of a free good does not come at any cost to society in terms of resource requirement. Thus, consumption of this type of good by a given consumer does not affect availability of the good to other consumers. Examples of free goods are breathable air, rainfall and sunlight. Free goods have the following features:

(a) Free goods are available in large quantities and satisfy the needs of everyone. Because free goods are abundantly available. They exceed the quantity that all consumers demand and they can be available to every consumer for free (without paying any price). Hence, free goods are not scarce.

(b) Free goods do not have an exchange value because they are in abundance and readily available, such that the quantity that all consumers demand do not fall short of the available quantity for all. Thus, people can obtain free goods at zero price. For example, society cannot impose a price to pay for breathing air. Exchange value comes from paying the price to the seller. Because free goods are not exchanged (traded), no exchange value is created. If such goods were to be availed through the market, their price would be equal to zero.

(c) Free goods can satisfy human wants even though they have no exchange value. For instance, breathable air is important for survival, and it is fortunate for us that it is available, abundant, and free.

Consumer goods and producer goods

Consumer goods: This category of goods make individuals to derive utility (satisfaction) from consuming them. Consumption of a good means to use it as an end-use product, rather than as an intermediary good in production. Consumer goods are categorised into perishable goods (for example, fruits and vegetables) and durable consumer goods (for example, television sets, cars, buildings and furniture).

Producer goods: Producer goods, also known as capital goods, are goods used for producing consumer goods or other producer goods. They include goods such as machinery, seeds and raw materials. Depending on their use, some goods fall under the category of producer goods and consumer goods. A good example of such goods is electricity when placed to domestic use, such as lighting, electricity becomes a consumer good. However, when used in a factory to enable machines to operate they fall in the producer good category.

Perishable goods and durable goods

Perishable goods: These goods have a short shelf life. Examples of perishable goods include fruits, bread, vegetables, meat, and milk. When handled, stored or transported improperly, perishable goods tend to loose their value and usefulness.

Durable goods: These goods last long, for example, buildings, machines, vehicles, and equipment. Hence, their use spans longer periods than perishable goods.

Public goods and private goods

Public goods: Public goods are identified from their two characteristics. One is that, consumption of such a good by one consumer does not reduce the quantity available to other consumers. That is, the good exhibits non-rivalry in consumption. A good example of a public good is the national defence, which protects us against foreign enemies. To exemplify the non-rivalry characteristic of national defence, if 100,000 babies are born this year, the same national defence will protect them equally, without affecting protection of the rest of the population.

Secondly, there is the characteristic of non-excludability in consumption, which means that once the good is provided, people cannot be prevented from consuming it. For example, it is impossible to exclude or separate someone or group of individuals in the country from enjoying protection provided by the national defence forces. The national defence protects all people in the country indiscriminately. Other examples of public goods include law enforcement, road signs and street lights.

Because public goods are non-rival and non-excludable, it is difficult to charge price to individual consumers of a public good. Collective effort to finance supply of any public good is a challenge because some people do not pay their fair

share. In economics, this is called dodging of their responsibility, a "free-rider problem" because their behaviour is similar to enjoying a free ride, at the expense of others. Given this situation, the government, sometimes in collaboration with the community, provides public goods to everyone using publicly generated resources.

Private goods: These goods have the characteristic of being rival and excludable in consumption, a contrary case of public goods. To exemplify these characteristics, let us use the previous example of 100,000 babies. If 100,000 are born this year, they will consume some nappies. The rival characteristic imply that if the production of nappies remains the same, there will be less nappies, on average, for every baby because new babies would consume some of the nappies, assuming that all earlier babies still use nappies. Thus, the consumption of a nappy by one new baby takes a nappy that one older baby would have used. Excludability in consumption hinges on the fact that producers of nappies sell them at a certain price. Buyers who cannot afford the price will not get any nappy, that is to say, they are excludable from consuming nappies.

A consumer of a private good has to pay a price of the good. Thus, it follows that all private goods have exchange value. In addition, both rivalry and excludability characteristics imply that private goods are scarce. Hence, they are obtained from the market and they are supplied mainly by private firms. Examples of private goods are sugar, rice, bicycles and houses.

Merit and demerit goods

Merit goods: These are goods with high social benefits to consumers. Examples of merit goods include education, health services, sports facilities and fire protection. Consumers think these goods should be widely available because of their usefulness.

Demerit goods: These are goods with social costs to the consumers. Examples of demerit goods include tobacco, cigarettes and alcohol. Many consumers generally believe that such goods should not be consumed.

Wants

A want is anything that a human being desires to have. Many things are necessary for a human being. However, not all human desires are needs. There is a difference

between a want and a need. A want can be anything an individual desires to have, even when it is unnecessary. Needs however, are a special kind of wants that an individual must have to survive. For example, food, shelter, cloth, and water. Human wants can be categorised according to how intense they are in use:

Necessities

Necessities are essentials for survival of a human being. Necessities include things such as food, clothes and shelter. Necessities may also include things that raise performance of work, such as use of computer and projectors in teaching and mobile phones for communication.

Comforts

These are goods and services, which make an individual's life more comfortable and easier. Examples of comforts include a motor vehicle when used as means of transport instead of walking, mobile phone for communication (instead of travelling to deliver a message) and courier services for delivery of letters and parcels.

Luxuries

These are goods or services consumed for ostentatious purposes rather than for necessity. Examples include expensive jewelleries (to beautify oneself) or expensive boats and cars.

Characteristics of Human wants

Human wants have various characteristics, which include the following:

Human wants are unlimited in number: Wants are endless, yet, resources required to satisfy these wants are always limited. For instance, a student's wants from a family budget include school uniforms, bags, books and exercise books. The same student requires food and soft drinks or water. The same student also needs services of teachers and sometimes of doctors and sports coaches. This means that, wants have a tendency of multiplying or increasing continuously, such that when an individual satisfies one, another one tends to arise.

Human wants vary among individuals, time and place: Individuals have different preferences and choices. Time and place may influence preference of individuals on different things. At the same time an individual may prefer a car as means of transport while another may prefer a motorcycle. With time wants differ, for

instance, In the 1990s people were satisfied with desktop computers, whereas in 2000s they demand laptop computers. With place, people living in cold areas, such as Mbeya and Njombe would want to buy more sweaters, but people in warm areas, such as Morogoro and Dar es Salaam, would want shirts and t-shirts.

Human wants differ in intensity: Wants are not required in the same intensity, because they are not equally crucial. Thus, some wants are more important than others. Resources are scarce and hence, an individual tends to prefer to satisfy urgent wants by postponing satisfaction of those not urgent (to be satisfied later). For example, an individual would prioritise wants such as food, clothes, shelter and health services, and postpone less important wants like movies.

Human wants are repetitive: Human wants are recurring in nature and once satisfied they tend to occur repeatedly. For example, an individual requires food daily to survive.

Wants are competitive: This competitive characteristic centres on human wants being endless, whereas the resources to satisfy them are limited. Thus, an individual must choose between competing wants in the order of priority. As pointed out before, urgent wants should be given priority over wants that are not urgent. For example, an individual may decide to buy a bed and postpone buying a television set.

Wants are complementary: Some wants must be satisfied together. One want gives rise to another want. For example, a person who wants to write a letter requires a paper and a pen. Similarly, a person who buys a car requires fuel to run it.

Human wants are subjective: Human wants cannot be gauged in absolute terms because they are subjective and relative. They vary from person to person, place to place and time to time. This is because wants are sometimes influenced by advertisements, publicity, tradition and customs of an individual

Wealth

In economics, wealth constitutes all things that have value to society. For any product to be wealth, it must possess characteristics of scarcity, transferability and validity and has exchange value.

There are different forms of wealth:

(a) Individual wealth (private wealth)

This form of wealth include things such as cars and houses, among others.

(b) Social or collective wealth

This is wealth enjoyed by the whole society, such as schools, colleges, roads, canals, mines and forests.

(c) National or real wealth

This sums up all individual wealth and social wealth. It comprises all material assets possessed by society.

(d) International wealth

This form of wealth includes the United Nations Organisation and its various agencies like the World Bank (WB), International Monetary Fund (IMF) and World Health Organisation (WHO). These are forms of international wealth because all countries contribute to their operations.

(e) Financial wealth

This form of wealth involves holding of valuable financial assets or physical possessions, which can be converted into a form that can be used for transactions such as money, stocks and bonds by individuals.

Economic resources

Economic resources are agents of production. They comprise factors of production (land, labour, capital, and entrepreneurship), which are used in production of goods and services.

Utility

Utility is satisfaction derived from consuming various quantities of a good or service. For example, when sick people get medical services, they derive utility from consuming medical services. Similarly, a thirsty person derives utility from drinking a glass of water. If one glass is not enough to quench the person's thirst, then a second glass will add to that person's utility. There are different types of utilities:

(a) Form utility

This is the type of utility, which involves changing the shape (or form) of a good. For example, when a carpenter makes a chair from wood, transformation of a

wood into a chair creates form utility. Another example of creating form utility arises when a company produces sugar from sugarcane.

(b) Place utility

This refers to the type of utility created for the product by virtue of its location. Thus, place utility enables physical availability of goods or services at the market place. Transporters create place utility when transporting goods from a place of production to a place of consumption. For example, transportation of bananas produced in Kagera to Dar es Salaam forms a place utility. Likewise, the importation of goods, such as clothes and utensils from other countries like China and Turkey to the local markets, such as Kariakoo in Dar es Salaam, provides place utility for consumers.

(c) Time utility

This type of utility arises when goods or services are provided to consumers during the time of need. Storage of goods in a warehouse until the time they are demanded in the market tend to create time utility.

(d) Possession utility

This type of utility comes with ownership of goods. A product has less utility to a producer, since it is for sale. However, ownership and use of a product by a consumer creates possession utility. A good example is a sewing machine, which has little utility for a seller, though for a tailor, it generates possession utility.

(e) Natural utility

The consumption of free goods such as air and sunlight, creates natural utility, because these goods have the potential to satisfy wants. For example, the sunlight is potential for drying wet clothes. Thus, it provides natural utility to people.

(f) Knowledge utility

Knowledge utility results from consuming knowledge in different forms, for example, reading books. In this regard, knowledge utility will increase with reading more books. For example, a textbook of microeconomics is expected to provide knowledge utility to advanced-level students studying economics.

(g) Service utility

Service utility results from providing a service to an individual, for example, when teachers, doctors and lawyers deliver services that satisfy human wants.

Exercise 1.1

1. Write short notes on the following economic terms:
   (a) Production
   (b) Distribution
   (c) Exchange
   (d) Consumption
   (e) Economic resources
   (f) Utility
   (g) Wealth
2. Describe categories of goods. Define and explain by using vivid examples the economic usefulness of each category of goods.
3. Distinguish between a want and a need. Explain features of a want.



Activity 1.1

Use various learning resources such as websites, magazines, textbooks and journals to search and read on application of basic economic terms at family and community level.

(a) Write down the information you have obtained on application of basic economic terms at family and community level.
(b) Share and discuss with your fellow students on application of basic economic terms at family and community level.

Definitions of economics

Definitions of economics by various classical and neo-classical scholars serve to explain the subject matter of economics. Arguably, economics is "the study of wealth, welfare and scarcity." With this definition, the subject matter of economics traditionally covered the study of wealth and material welfare of individuals in an economy in relation to how their limited resources were used to satisfy their numerous wants. Generally, no single definition of economics is precise enough for all economists to accept. Instead, different schools of thoughts define economics differently, with every definition having its pros and cons.

Classical definition of economics

Pioneers in economics, popularly known as classical economists, considered economics as a science of 'wealth.' **Adam Smith** (1723 – 1790) explained economics as a study of nature and causes of the wealth of nations. He emphasised on production and accumulation of wealth as the subject matter of economics. Smith believed in a free market economy, where producers sell products and consumers buy these products based on forces of supply and demand.

In the middle of the 19th century, another classical economist, **John Stuart Mill** (1806 – 1873), defined economics as "a practical science of production and distribution of wealth." Mill's definition of economics relies on two main points. One is "Economics as a study of wealth only," that is, it deals only with consumption, production, exchange and distribution aspects of wealth. Second is that "only those goods which are scarce are included in wealth." Thus, non-material goods such as services were not included in the category of wealth. Classical definitions of wealth had the following contributions:

(a) The definitions are correct about determination of income, employment and economic growth in that they have a causal relationship with production and distribution of wealth (material goods).

(b) The definitions indicate that poverty, unemployment, and underemployment of resources mostly in developing countries call for expanding production and accumulation of wealth as well as reducing income inequalities.

(c) The definitions have widened the scope of economics from being a "science of house management," as proposed by Socrates and Aristotle, to "the practical science of production and distribution of material wealth" at national level.

Criticisms of classical definitions of economics

The classical definitions of economics by Smith and Mill were criticised.

They put much emphasis on wealth. The definitions gave more importance to wealth, at the expense of people, who are economic agents. In economics, the study of wealth cannot exclude people's involvement in economic activities. For example, the classical definitions did not consider human welfare as an important aspect of economics. Hence, these definitions seem to have ignored enhancement of social and human welfare as one of the goals of economics.

The classical economists defined wealth narrowly. They expressed wealth as comprising only material goods, such as land and bullion of gold, which leave out services. In modern economics, wealth includes material and non-material

goods. Material goods include human capital, non-material goods include services provided by doctors, engineers, teachers, lawyers and soldiers.

Moreover, they are not explicit on resource use (means). The classical definitions of economics focused on accumulation of wealth as an end in itself. Thus, they seem to ignore the means and ways used to create that wealth. In addition, they were unclear as to what the term 'wealth' meant.

Furthermore, they focus solely on the economic man. By the "economic man," these definitions meant that only self-interest guided the motivations for an individual's prosperity, which suggest that individuals do not consider social interest. They work only for self-interest.

Neo-classical definitions of economics based on welfare

Towards the end of the 19th century, a new school of economic thought, neo-classical economics, emerged to compete with classical economics. Briefly, neo-classical economics focused on supply and demand as the basis of production and consumption of goods and services through price determination.

Some neo-classical economists, in particular, **Alfred Marshall** (1842 - 1924) gave economics a new definition, which emphasised on human welfare. Marshall defined economics as "a study of mankind in the ordinary business of life." In general, the neo-classical school define economics as "a study of human behaviour as related to allocation of scarce resources which have alternative uses, in achieving satisfaction." The neo-classical definition of economics considers economics as a study of wealth secondarily and primarily as a study of human economic behaviour in the pursuit of happiness. The definition examines individual and social actions connected with use of material requisites in attainment of wellbeing. Thus, Marshall shifted emphasis of the definition of economics by giving priority to economic agents rather than wealth thereby reducing wealth to be of secondary importance.

Followers of Marshall, particularly **Arthur Cecil Pigou** (1877 – 1959) and **Edwin Cannan** (1861 – 1935) defined economics as "a study of causes of material welfare." Emphasis on welfare meant that wealth was not an end to human activities, rather, it was only a means to fulfilment of an end, which is human

welfare. In addition, the neo-classical interpretation of welfare suggests that economics is a study of an ordinary person in a free society, which implies that a person who is not part of such society would be of no interest to economics. Thus, we deduce from various aspects of the neo-classical definitions of economics that economics is concerned with understanding ways in which people apply their knowledge and skills to the gifts of nature so as to improve their material welfare.

Marshall's definition of economics was widely accepted for a long time because it enlarged the scope of economics by bringing the welfare aspect into the definition, rather than considering wealth alone. However, Lionel Robbins (1898 – 1984) came to criticise this definition.

Criticism of the welfare definition of economics

Lionel Robbins argued that the focus on "material" in the definition of economics narrows the scope of economics, as it leaves out non-material things that are beneficial for promoting human welfare, for example, the economic services offered by teachers, nurses, lawyers, engineers, musicians, and chefs, among others. These services create utility from consuming them and are limited in supply. Thus, to ignore such services and concentrate on material goods only tend to limit the field of economics. In addition, Robbin argued that some of the products people produce do not promote human welfare, yet they take them to be economic goods. Examples in this regard include weapons for waging wars and cigarettes, which the study of economics incorporates.

Furthermore, Robbins argued that welfare is subjective, as it varies among individuals, age groups as well as across geographical locations. In addition, Robbins argued that the concept of welfare is ambiguous, as it is difficult to be measured quantitatively.

Robbins' criticism makes it difficult even for two individuals to agree on what constitutes their welfare in common and what does not. Hence, the definition of economics based on welfare is perceived only in theory and very difficult to observe it in practice. For example, one challenge in practice would be on how to quantify an individual's material and non-material activities using the same measure.

Robbins' definitions of economics (scarcity based definition)

Following his criticism of earlier definitions by classical and neo-classical schools of thoughts, Robbins came with his own definition of economics in the year 1932 and claimed that it was precise, scientific and superior. He defined economics as "a science which studies human behaviour as a relationship between ends and scarce means which have alternative uses." Robbins argued that his definition of economics was better than the previous ones because it did not contain any reference to the material well-being or welfare.

Main pillars of Robbins' definition

The first pillar of Robbins' argument is that human wants are unlimited. He pointed out that human wants increase, change over time and vary among individuals and within groups of people. It is difficult for an individual's wants for goods and services to be fully satisfied, because every individual has a continuum of wants, when one want is satisfied, another one arises.

The second issue is about scarcity of resources. Robbins explained that resources, such as land, labour, capital, and entrepreneurship required for production are limited in supply relative to what producers require.

Thirdly, scarce resources for satisfying unlimited human wants have alternative uses. For instance, people use land for settlements, grazing cattle, building schools, hospitals and playgrounds. In addition, resources like labour are used for one activity at a time. For example, labour used in production of cassava cannot go into production of any other product (for example, for a school building) at the same time. Therefore, an economic agent must choose the best way of utilising the scarce resources among alternative uses.

Lastly, Robbins' definition elaborated that although human wants are unlimited, they are not equally important. Thus, economic rationality requires that economic agents should prioritise wants that are more important than those less important.

Strengths of Robbins' definition of economics

Many economists accepted Robbins' definition because it centred on scarcity, which is the economic problem. As already pointed out, Robbins argued that the foundation of economics rests on satisfaction of human wants using scarce resources with alternative uses. This argument makes the study of economics analytical. In studying economics, we analyse how scarcity influences human behaviour in seeking to satisfy human wants.

Let us consider what Robbins meant with the statement "scarcity is the central problem of economics." It is already known that an economic problem arises because human wants are innumerable and the means to satisfy them are scarce or limited in supply, which necessitate choosing among competing wants or prioritising them. If the economy had unlimited resources to satisfy all human wants, scarcity would not arise and thus, all resources to satisfy all our wants would be freely available. Hence, there would be no economic goods. Without economic goods, one would not have the economic problem, since the economic problem arises due to a need to economise on resource use.

Robbins' definition has universal appeal because all economies in the world face the problem of unlimited wants and limited resources. As a result, any society and its individuals need to make rational choices on how to allocate their scarce resources efficiently. Robbins' definition serves to underscore the nature, scope, and subject matter of economics.

Criticisms on Robbins' definition

Robbins' definition of economics still enjoys acceptance by majority of economists worldwide. However, some economists, such as William Henry Beveridge (1879 – 1963), have faulted the definition on several grounds.

Robbins' definition of economics ignores economic growth. The definition tends to suggest that economic growth is outside the scope of economics, whereas it is through economic growth that living standards improve. If economics is conceived as a science of administration of scarce resources, then its scope becomes wider as to include all aspects of economic life and not merely the aspect that relates to the market price.

Furthermore, Robbins' definition ignores social values and social welfare. The definition does not consider aspects of economics that involve value judgment. Yet, opinions about social well-being influences our choices as individuals and as society. For example, Robbins' definition suggests that a society can produce anything to satisfy human needs or wants. However, in many societies, social value is against production of heroin, marijuana and other products that would disrupt the social setup.

Despite the criticisms, Robbins' definition is still the best definition to date, since economics is fundamentally a study of ways in which people provide for their well-being. Thus, economists attempt to understand how human behaviour

features in a relationship between ends and scarce means with alternative uses to satisfy individuals and society's wants. In this regard, economics as a study of how individuals allocate scarce resources to satisfy unlimited wants deals with choices that individuals and businesses make to manage the scarce resources.



Activity 1.2

1. Library and educational websites contain different pieces of information and stories that may help you to make your own definition of economics.
   (a) Write down your own definition of economics from reading various sources.
   (b) Share the definition of economics you have obtained in part (a) with your fellow students.
   (c) Discuss different definitions of economics that you and your fellow students have shared and, with the help of your teacher, choose the best definition.
2. Think about various economic activities taking place in your society. List them so that you can remember them well. Discuss with some of your classmates (not less than five individuals) usefulness of each economic activity that you have listed. Summarise the information and share it with the rest of the class.

Branches of economics

The subject matter of economics comprises two major branches of economics, namely, microeconomics and macroeconomics. The two branches of economics originated from the Greek words, namely, Micro, which means 'small' and Macro, which means 'large.'

Microeconomics

Microeconomics is a branch of economics that focuses on individual economic units. The study of economics at the micro-level focuses on analysing small part or segment of an economy, rather than the whole economy. The aim is to understand the behaviour and decisions of micro-level economic units such as firms, individuals or households. Specifically, microeconomics analyse how firms seek to maximise profits from producing goods and services and when competing

with other firms in the market and how individuals seek to maximise utility from buying and consuming goods and services from the market. Additionally, microeconomics analyse how households make decisions with regard to how much goods to produce at household level, how much to consume and how much labour to supply to the labour market. Furthermore, microeconomics analyse market characteristics of demand for and supply of goods and services to show how they determine prices. In summary, microeconomics is concerned with analysing and understanding the following issues:

(a) How resources for producing goods and services are allocated among various firms;
(b) How micro-level economic units achieve efficiency in production and distribution of goods and services; and
(c) How prices of various products are determined;

Thus, conventional topics in microeconomics include theory of consumers' behaviour, demand and supply characteristics, theory of a firm, production theory and market structures, most of which form chapters of this book.

The study of microeconomics is important for developing economies because it explains how producers and consumers in a free market economy make proper decisions on allocation of productive resources efficiently. In addition, it explains how prices are determined and how they are linked with efficiency. It also enables formulation of policies for enhancing economic efficiency and welfare of individuals and society. Besides, microeconomics offers measures for eliminating inefficiency in the economy.

Macroeconomics

Macroeconomics is a branch of economics that focuses on analysing mainly economic trends affecting the whole economy, especially key macroeconomic variables. These key macroeconomic variables are economic growth, inflation, unemployment, interest rate and balance of payments. Macroeconomists analyse alternative policies for managing these variables in order to promote economic growth and enhance people's living standards.

Macroeconomics deals with problems faced by the entire economy, by addressing the functioning of the economy as whole. Such questions include the following,

for example:

(a) What is inflation and why is inflation a problem for many economies, large and small?

(b) What are the causes of unemployment and how should they be addressed?

(c) Why do economies experience fluctuations between high rates of economic growth and low rates of economic growth?

Briefly, macroeconomics analyses broader economic aggregates such as the national income and economic growth, employment and unemployment, aggregate consumption, saving and investment, the general price level, interest rates and balance of payments.

Table 1.1: *Differences between microeconomics and macroeconomics*

S/N	Microeconomics	Macroeconomics
1.	Studies specific units of the economy, such as households and firms	Studies the economy as a whole
2.	Involves partial equilibrium analysis, which examines the equilibrium in production of a certain good and equilibrium in consumption of the good, which determines the market equilibrium for that good	It involves equilibrium analysis of aggregate and how they affect the whole economy
3.	Uses deductive method (from the general to the particular)	Uses inductive method (from the particular to the general)
4.	Known as theory of pricing with two subdivisions, namely, the theory of consumer and theory of production	Known as theory of income and employment determination
5.	Resulting policies affect economic units (such as effects of tax on a specific good on consumers)	Resulting policies affect the aggregates (such as economic growth)

The two subfields of economics, microeconomics and macroeconomics, differ in several aspects. However, they are complementary and interrelated. For example, employment as a macroeconomic variable is the sum of all persons employed in all micro production units. Hence, it can be said that macroeconomics has a micro foundation.

Conclusively, the subject matter of economics is dynamic and continues to broaden, as several factors, such as time, circumstances as well as social and economic life of individuals, among others, could bring about some changes in the subject matter of economics.

Exercise 1.2

1. Explain subfields of economics.

2. All the following topics fall within the study of microeconomics EXCEPT
   (a) The impact of taxes on alcohol on the drinking behaviour of teenagers.
   (b) The effect of government budget deficit on economic growth.
   (c) Effectiveness of government antipoverty programs in reducing homelessness.

3. Identify from the following, which topics relate to microeconomics and which to macroeconomics.
   (a) A farmer's decision with regard to how much income to save from the harvests.
   (b) Factors that led Tanzania to become a middle-income country.
   (c) A firm's decision on how many people to terminate to reduce rise in cost of production.
   (d) Effects of changes in the quantity of money supplied.

4. Explain the difference between microeconomics and macroeconomics.

Why do we study Economics?

In general, people study economics to get knowledge. Part of this knowledge is specific from what people learn, and another part comes in form of applying principles that people learn.

In today's world, many of the topical issues are partly microeconomic issues in character, for example, production and productivity of companies, environmental, education and health issues and consumption issues such as high fuel prices, among other topics. As people apply principles obtained from studying microeconomics, they facilitate a good understanding of such issues.

As explained earlier, the study of economics enables one to understand the implicit logic behind the behaviour of economic agents with regard to choices they make

in consumption and production. It explains how consumers make rational choices of goods to buy among many alternative wants, as they are limited by the amount of money they have to spend on goods and services for maximum happiness. Money in this regard serves as a scarce resource and so, it is necessary to manage it well. In addition, it explains about how producers reach decisions on how to allocate scarce resources efficiently to get maximum profit from their production activities.

Knowledge of demand and supply helps individuals to know how prices of goods and services are determined in their respective markets. This knowledge enables us to answer the basic economic questions of what, how and for whom to produce goods and services.

Application of economics in other subjects

Economic theories are part of the subject matter of economics. The other part which puts economic theories into practice is applied economics. Many subjects use economics as a tool to facilitate the understanding of some topics in those subjects. Subjects that use economics include agriculture, business, geography, civics and history, among others.

In agriculture, economics is applied in agricultural production and it guides farmers in making rational decisions. Farmers, like any other producers, must allocate efficiently factors of production in farming. In this regard, a farmer either maximises profit or minimises cost of production. Furthermore, knowledge of the market price and costs of production helps a farmer to determine the quantity needed in the market and to know how the quantity supplied will respond to price changes. In addition, economics is applied in the field of research and projects planning. It is used to compute optimum input usage, optimum production and distribution of output. Also, it is used in estimating prices of agricultural inputs and outputs. In applying knowledge of economics, farmers deepen their understanding on how prices affect quantities of purchased inputs, quantities of products produced for sale and profit earned.

In business, economic principles are also applied. Firms apply these principals to make rational choices on quantities of inputs required to maximise profit or minimise costs of production. For profit to be realised, total costs of producing a good or a service must be less than the generated revenue from the produced output.

In geography, economic principles are used to address environmental effects arising from various human economic activities. For example, economic principles are used to give solution to pollution and deforestation by applying economic measures such as property rights, taxes, emission fees and tradable permits.

In civics and general studies, various topics on human economics are covered. They include poverty, democracy and rule of law. Some of these topics belong to microeconomics branch. These topics raise issues that call for economic interventions. The implementation of these interventions contributes to enhance economic performance, which in turn, promotes the well-being of society.

Economic principles are also applied in history subject. For example, it is applied in topics of modes of production and colonialism. Among issues addressed in these topics is accumulation of wealth. For example, some writers contend that the main motive behind Africa's colonisation was the search for cheap and reliable supply of resources. Such inputs were used in colonial countries to increase production in their territories.

The relationship between economics and other subjects

Economics analyses the behaviour of an individual as a community or society. Thus, it is central, relative to other fields, such as, mathematics, statistics, agriculture, geography as well as information and communication technology (ICT).

Mathematics and statistics have produced a set of tools for economic analysis. Application of mathematical methods at lower levels of studying economics is necessary for explaining the behaviour of some variables. For example, the rate of inflation, unemployment, the price index and the national income of a country, among many others. At higher levels, economics utilises advanced mathematical tools in various topics, that is why there is a specific branch of economics called mathematical economics. Furthermore, at this basic level of studying economics, statistics helps to generate several measures of variables, which are summarised mainly in tables and graphs. At higher levels of studying economics, advanced statistical analyses are used especially in specific areas of economics such as econometrics.

Information and communication technology (ICT) comprises of software, hardware, networks and presentation of information in form of data, images and other relevant forms. The ICT enables the application of economic theory that broadens the fast acquisition and processing of economic data. It provides mobile phones and computers with the internet and search engines, which enable interaction of buyers and sellers online thereby making transactions easier.

With regard to geography, knowledge of human geography benefits economics in assessing some economic phenomena, such as population growth and the unemployment rate in the economy. Furthermore, geography knowledge enables economists to design and estimate location as well as distribution models. Such knowledge allows firms to determine their geographical location by considering availability of facilitating resources, such as infrastructure, particularly roads and electricity.

The nature of economics

Economists have different views regarding the nature of economics whereby pertinent questions is whether economics is a science or an art and if it is a science, is it a positive science or normative science? This sub-section elaborates on these concepts.

Question: Is economics a science or an art?
Answer: Economics is both a science and an art.

Economics as a science

Economics is a social science that employs similar methods as those used in other fields of science, such as chemistry, biology, and physics. Like other sciences, economics uses theories and models to enhance understanding and prediction of various economic situations in the real world. Paul Samuelson (1915 – 2009), a famous American economist claimed that, "economics is the queen of all social sciences."

Similarities of economics with pure science

Economics applies principles, theories, and laws in research, as is the case with pure science fields. For example, the principle of comparative advantage, the laws of demand and supply and the law of diminishing marginal utility. Economics uses scale of measurement and measurement units (SI Unit) as done in the field of

pure science. For example, economists measure output in units, income per person, Gross Domestic Product (GDP) per annum, the inflation rate and price per unit of output or input.

However, economists do not conduct experiments in a typical pure science laboratory. The reason is that economics deals with human behaviour and hence, economists conduct experiments in real-life situations. For example, economists observe what happens to demand of related commodities in a certain market when the price of a certain good changes. The principles and knowledge of economics can be improved and changed with time as in pure science. New research, with extensive data and advanced methods and tools, may produce results that may refute existing theories. Moreover, the study of economics forecasts the future as other science fields do. For example, the theory of demand predicts a decrease in demand of a given good based on the observed trend of the price of the good.

Economics as an art

Art is the practical application of principles, laws and theories in order to achieve particular goals. Whereas the principles, laws and theories come from science, one depends on art to turn them into reality. For example, application of different policies by the government to control the price level follows theoretical knowledge on factors that influence on price level. In this definition, it can be surmised that economics is an art, as it is geared at finding solutions to economic problems that societies face. Such solutions may include sustainable economic growth, stable prices, and low rate of unemployment, among others.

Question: Is economics a positive or normative science?

Answer: Economics is both positive and normative as it has both theoretical as well as practical aspects.

Positive economics

This is an analysis that is limited in making either purely descriptive statements or scientific predictions. It deals with what is actually happening in real situation, using empirical method, like 'What is.' It does not involve personal judgments or opinions. Positive statements are descriptive in nature. They make a claim about the present situation of the real world.

For example, "low income tax rates encourage workers to work and save more." This is a positive economic statement that claims how the world works. It is not

a statement of anyone's value judgment, opinion or subjective feelings. Such statement can be confirmed or refuted by analysing data on changes of tax and changes of savings over a certain period of time.

Normative economics

This is an analysis involving value judgment or opinion about economic policies or situations. These statements are prescriptive involving what should be done under certain circumstances. It is a statement of what ought to be (what should be done). For instance, "The government should ensure that all healthcare services are available to every citizen." This statement is based on personal perspective and satisfies the need for 'should be' or 'ought to be.'

Exercise 1.3

1. With relevant examples, (different from those given in the chapter), distinguish between a positive statement and a normative statement.
2. With relevant examples, analyse why is economics both a science and an art? (Use examples that are different from the one mentioned in the chapter).
3. Discuss how economics and other fields are related.
4. What are utilities of studying economics?

Economic laws

Economic laws are generalisations (conclusions and inferences) about economic facts. An economic law predicts results of certain initial conditions. It is a universally accepted fact used to make predictions. Economists formulate laws from observing particular human economic behaviours and use reason to analyse laws with regard to their consistency in everyday life.

Features of economic laws

Laws of economics are less exact than laws of natural sciences. Economists cannot predict precisely what will happen in future in the economic domain. They only say what is likely to happen in the short-term. The following reasons elaborate why economic laws are not exact as those of natural sciences:

Natural sciences deal with matter, which is lifeless while economics deals with human beings whose behaviour is erratic: It is difficult or even impossible to

predict with certainty future actions of a human being because human beings change their behaviours from time to time.

Economics is based on data with unproven reliability: Most of the data do not tell a very factual situation with respect to the examined economic laws. In addition, the already collected data cannot capture the dynamism in the post-data collection period, for example, changes in the tastes and fashion, attitudes and economic environment.

Economics deals with many variables: Some factors that affect a certain economic situation remain unknown because economic models incorporate the main variables. Hence, economists cannot explain results from their experiments with certainty as in the natural sciences, but they do so to a high degree or high probability that they are correct. Economists accept them with a probability (though small) that they could be wrong.

Laws of economics are to be compared with the laws of tides rather than with the simple and the exact law of gravitation (Alfred Marshall): The rise of tides cannot be accurately predicted. Hence, the laws of tides are also not exact as there are some uncertain factors that could affect predictions. For instance, wind may change its direction to the opposite side. So is the case with the laws of economics.

Economic laws are essentially hypothetical: Economic laws are hypothetical in that they are true under certain given conditions or assumptions, such as '*ceteris paribus*' (All other things remaining unchanged). The drawn conclusions are true, only if the conditions are fulfilled. Otherwise, in a situation where the conditions are not fulfilled, the economic laws do not apply.

Economic laws are qualitative rather than quantitative: Laws of economics are qualitative in nature, which means that they are not stated in quantitative terms. They tell the direction of change and nature of relationship between economic variables rather than the quantity of change.

Types of economic laws

The following are types of economic laws.

(a) Macroeconomic laws

These are laws used in the macro-management of the economy. For example, the fiscal and monetary policies.

(b) Individual economic laws

Individual economic laws, also known as specific economic laws, are laws that operate within the same mode of production (same economic system). When the mode of production changes, these laws are replaced by laws of the new system. For example, laws of distribution differ between the capitalist mode and the socialist mode of production.

(c) Overall economic laws

Overall economic laws, also known as general economic laws, are laws that have operated throughout the history of humankind in all economic systems. For example, the law of diminishing marginal utility, the law of diminishing returns, and laws of demand and supply.

Method of economic analysis

Laws become generalisable through two methods, the deductive method and inductive method. Whereas the inductive method aims at developing a theory from the data, the deductive method aims at testing the existing theory.

Inductive method or approach

With the inductive method, the analysis starts from specific to general, that is to say, specific observations lead to broad generalisations. A researcher makes many observations, which help her or him to discern a pattern and generalise, from which he or she infers an explanation or theory.

Deductive method or approach

The deductive method (deduction) is a method of analysis that involves reasoning or deriving conclusions from general truth to a specific or particular fact. It starts with a general statement or hypothesis, and works to reach a specific conclusion. The method predicts what the observation should be if the theories were correct. That is to say, it uses deduction to test hypotheses and theories. Hence, the coinage from the general to the theory to particular observations.

The economic problem

All societies face a central economic problem, which is how to make the best use of scarce resources. The essence of the economic problem is that the needs and wants of people are unlimited, whereas the resources needed to satisfy them are limited. The problem of scarcity has led producers and consumers to make choices

that give rise to opportunity cost. The opportunity cost means that making a choice creates a sacrifice because a person has to forgo the second best alternative. The forgone alternative is the true cost of the choice because it involves loss of benefits that an alternative would have provided. Paul Samuelson provided an explanation to the economic problem. He argued that for a society to solve the economic problem, it must endeavour to answer three basic questions. What to produce? How to produce? For whom to produce?

What to produce?

A society must decide on the best combination of goods and services it can afford to produce so as to satisfy the needs and wants for people because resources cannot suffice to produce all the goods and services to satisfy the needs and wants. However, what to produce changes over time, due to changes in several factors such as technology, consumers' taste and preferences, market price and available resources . The producers would determine what to produce based on the market price, available resources, costs of production and many other factors that influence the demand for a product.

How to produce?

The producer has to decide on the best combination of factors at its disposal to produce the desired output of goods and services. For example, how much of factors of production should go into production of consumer goods and capital goods.

For whom to produce?

This is a problem of distribution of output and benefits. Specifically, society has to address the question of who will benefit from the output generated from its economic activity and how much they will get. For example, producing luxury goods would imply targeting people with high incomes.

Scarcity

Scarcity is at the core of the economic problem. It refers to fact that resources are insufficient to satisfy all needs and human wants for all societies. In general, human wants are endless, whereas resources to satisfy them are limited. For example, an individual cannot have enough income, resources, or wealth to satisfy every desire. Because individuals or communities cannot obtain all they

want, they must make choices. Thus, choices occur because most of resources are scarce.

Choice

Economic choice involves selecting a particular option among alternative uses of scarce resources. Choosing is generally an unpleasant fact that one cannot avoid. For example, an individual can choose what career to pursue or what to have for lunch. Some decisions are more important than others. Society has unlimited needs and wants. However, the resources, such as time, money and land, among others, are limited. Moreover, resources like labour are scarce in a given time because they cannot be used in many activities simultaneously. For example, taking the next hour to write an economics assignment will prevent a student from using that hour to study another subject.

Scale of preference

This is an arrangement of wants from the most preferred to the least preferred one (descending order of preference). People tend to fulfil the most preferred wants first. For example, a boarding student with pocket money from the parents faces different choices, like buying an economics book, buying a newly fashioned shirt or a mobile phone. The likely scale of preference would be:

1. Buying economics book.
2. Buying a newly fashioned shirt.
3. Buying a mobile phone.

Opportunity cost

Scarcity of resources in relation to unlimited ends necessitates a choice, as already discussed. The choice implies forgoing some other wants. The real cost of choice one makes is the next highest ranked alternative that one forgoes. Thus, opportunity cost is second best alternative forgone because of making choice. Opportunity cost is the cost of something in terms of opportunity sacrificed. For example, a student may decide to study for one more hour. However, there may be many alternatives available for use of that hour. Suppose the student chose not to study, and decided to watch a football match on the television. The student's decision is the next-highest-ranked alternative. Hence, watching the football match on television is the opportunity cost of studying.

Another way to understand about opportunity cost is that any choice implies sacrificing something else. What is sacrificed is loss of benefit that one would have received had he or she made that choice. For instance, a person who chooses to save money incurs the opportunity cost of not consuming goods and services worth the amount of money saved. Likewise, the opportunity cost of sleeping instead of working is the money one could have earned if he or she had spent the time working. When making a decision, it is important to consider the opportunity cost.

Can opportunity cost be zero?

Opportunity cost is only positive with economic goods (scarce goods). However, there are exceptions where the opportunity cost is zero. For example, free goods, things provided by nature have no opportunity cost involved in their use. For example, there is no opportunity cost (real cost) in breathing (free oxygen) for there is enough air for everyone to breath.

Economic way of thinking

Economic way of thinking assumes that an individual must develop rational behaviour to provide a response to an economic problem of scarcity. This concept is explained by five core ideas, and these ideas form the basis of all microeconomics models. The following are core ideas:

(a) People make rational choices by comparing costs and benefits. The rational choice is a choice whereby a person uses resources available to him or her in the best manner possible to satisfy his or her wants.

(b) Cost is what an individual foregoes to get something. In economics, real cost is in terms of opportunity cost.

(c) Benefit is what you gain when you get something, which is measured by what you are willing to give up for it. The benefit of something is gain or pleasure that it brings. Economists measure benefits by what a person is willing to give up for it.

(d) Rational choice is at the margin. A choice on the margin results from comparing all relevant alternatives systematically and incrementally. Therefore, instead of just costs and benefits, an individual has to take into consideration marginal costs and marginal benefits:

   (i) Marginal cost: The cost that arises from one unit increase in an activity. The marginal cost of something is what you must give up to get one more unit of it.

(ii) Marginal benefit: The benefit that arises from one unit increase in an activity. It is the additional satisfaction or utility that a consumer receives when the additional goods and services are purchased.

People make rational choices and use scarce resources in the way that makes them as well off as possible when they take those actions for which marginal benefits exceed or equals marginal costs.

(e) In making choices, people respond to incentives. An incentive is an inducement to take a particular action. The inducement can be a reward in the form of an increase in benefit or a decrease in cost. On the other hand, an incentive can be a punishment with the opposite result. In general, a change in marginal benefit or a change in marginal cost brings a change in the incentives that people face and eventually, leads them to change their actions.

Therefore, to make a rational choice, an individual must determine the costs and benefits of the alternatives.

Production possibilities curve (PPC)

A production possibilities curve (PPC) or a production possibilities frontier (PPF) or a transformation curve is a curve that shows various combinations of outputs that the economy can produce with a given technology and available resources. The PPC represents the maximum possible level of output the economy can produce under efficient utilisation of its resources with a given state of technology. It is a boundary or frontier beyond, which the economy cannot exceed, unless there is discovery of new resources or improvement in state of technology or both.

Assumptions underlying production possibilities frontier

In real life, an economy can produce many goods and services. However, to simplify the analysis, it is assumed that an economy produces only two goods. Hence, when drawing the production possibilities curve (PPC), the following assumptions are made:

(a) The economy produces only two goods, for example, a producer good and a consumer good;

(b) Resources are fully used. There is full employment or efficient use of the limited resources;

(c) Production takes place over a specific time period, for example, one year;

(d) The resources or inputs in both quantity and quality used to the output are fixed over this period;

(e) Resources are substitutable in production, that is, they can be used to produce either of the goods; and

(f) Technology does not change over this period.

Figure 1.1 presents a production possibilities frontier (PPF) whereby an economy can produce any combination inside or on the frontier. The slope of the PPF measures the opportunity cost of a producer good in terms of a consumer good. The opportunity cost is likely to change depending on the amount of goods the economy is producing. The two ends on the PPF represent possibilities if the economy utilises all of its available resources to produce either consumer goods or producer goods. If the economy produces at point G, it will produce only consumer goods whereas at point H, it will produce only producer goods. At each end, all resources are used in the production of either good.

Moreover, the economy may decide to split its resources to produce both consumer and producer goods. For instance, at point A it can produce 500 units of consumer goods and 70 units of producer goods. All points of production along and under the PPF are attainable (G, F, A, B, E, H and D). All points above the PPF are unattainable such as point C. The economy cannot produce goods at point C due to limit of resources it has. Along the PPF, all points are efficient because there is full utilisation or employment of resources. However, point D is an inefficient point because the resources are underutilised or underemployed.

The movement along the PPF explains a trade-off (the opportunity cost) of the resources in the production of one good in terms of another. The concave shape of the PPF reflects increasing opportunity cost. That is to say, one needs to forgo an increasing quantity of consumer good to produce an extra unit of the producer good. If a country decides to move from G to H to produce more units of producer goods, then the economy has to forgo production of more and more units of consumer goods. For example, moving from point A to B, the economy gives up 100 units of consumer goods to add 20 units of producer goods. In addition, moving from E to B, an economy has to give up 20 units of producer goods to obtain 200 units of consumer goods.

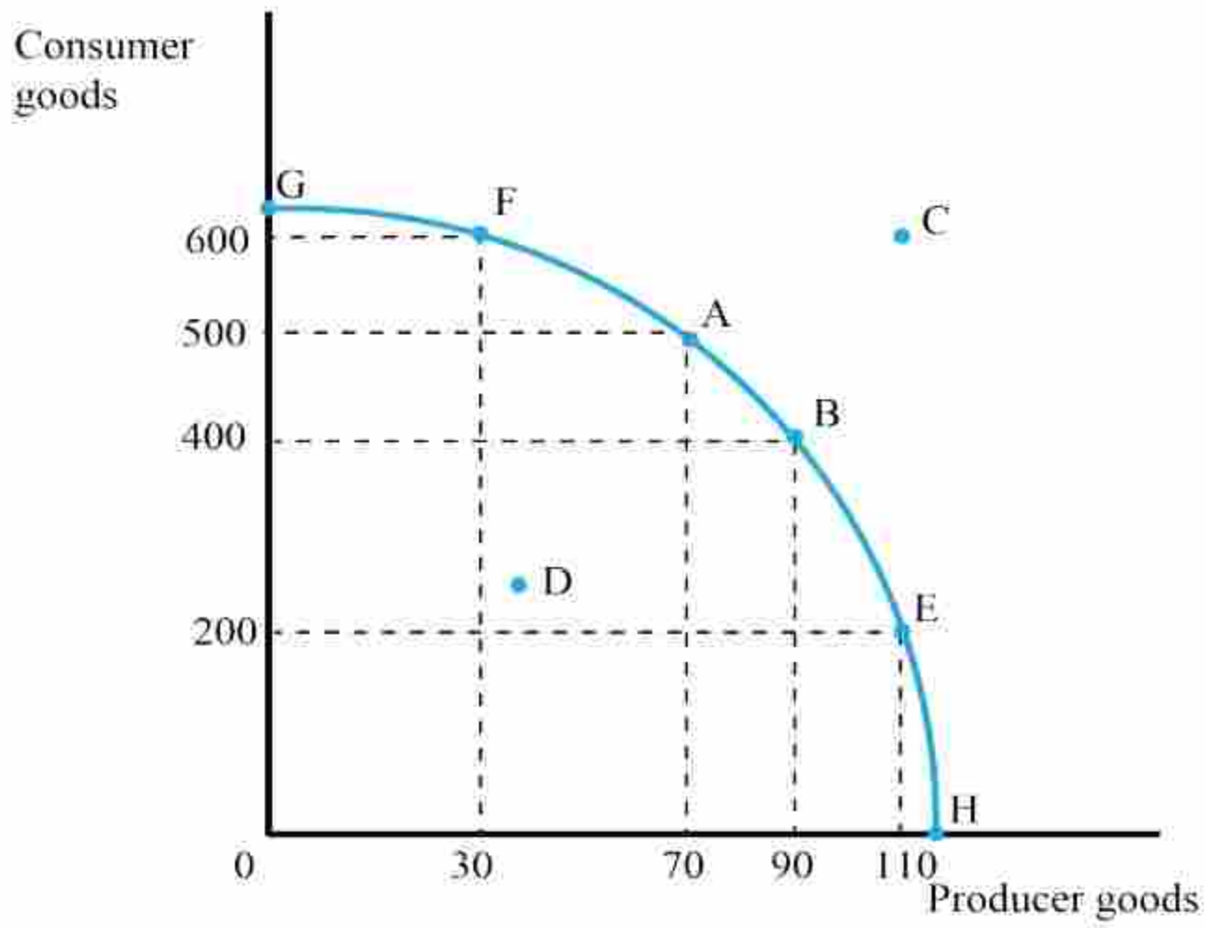


Figure 1.1: *Production possibilities frontier (PPF)*

The PPF would also be linear (Figure 1.2) when the opportunity cost is constant. This situation depicts a condition in which resources are not specialised in production of either good. Hence, resources can be perfectly substituted to produce either good. In Figure 1.2, if the economy decides to produce good Y the PPC will be at point A and if it decides to produce less of it the PPC will be at point B.

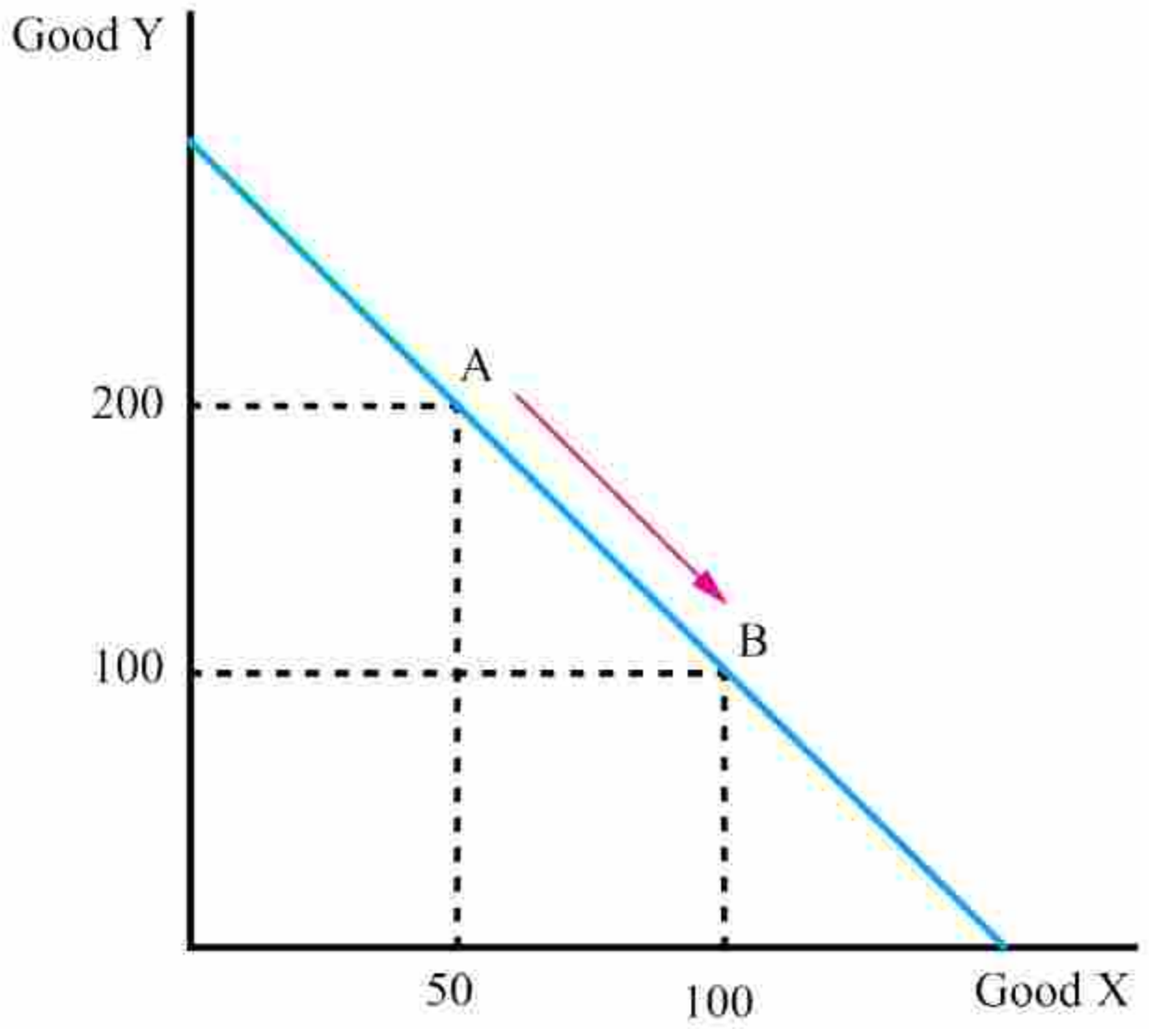


Figure 1.2: *Linear production possibilities frontier*

Significance of the concept of opportunity costs

The concept of opportunity cost is very important in any economic decision involving scarce resources which have alternative uses. It is used to depict levels of production and consumption in international trade to determine the country's comparative advantage. For instance, a country may decide to import products that have high opportunity cost and specialise in the production and export of products that have less opportunity cost. A rational producer or consumer will make a choice on how to allocate resources on a good that has low opportunity cost.

Moreover, opportunity cost aids in deciding on different choices when the scarce means have alternative uses. For instance, it shows how consumers could budget their expenditure by analysing what is most important for them and giving up other goods depending on the available resources.

Shift of the PPF curve

Progress in technology can enhance achievement of economic growth through a shift of the PPF. This is possible with relaxation of assumptions with respect to the PPF curve, for example, assumption of fixed resources. If an economy increases resources and/or advances technologically, production will increase and in turn there will be an upward shift of the PPF, which will result in economic growth. In Figure 1.3, when the supply of resources and/or improvement in technology occurs, the PPF curve shifts outwards from point AB to A'B.' On the outer PPF curve that is A'B,' the economy can produce more goods and services than on the lower PPF curve.

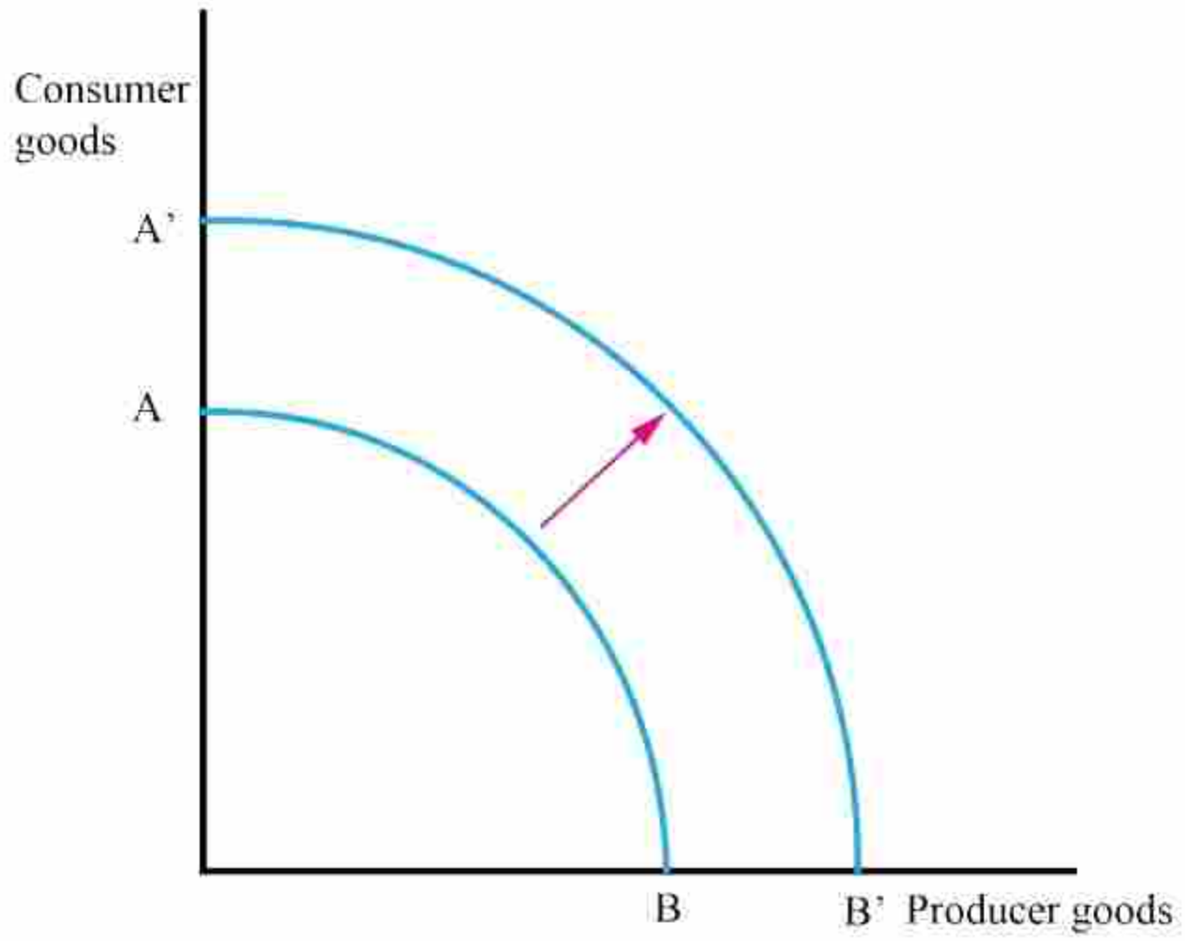


Figure 1.3: *Shift of the production possibilities frontier (PPF)-economic growth*

Economic growth (shift of the PPF) can be influenced by an increase in the quantity of natural resources such as discovery of minerals, oil, and gas or by increase in quantity and quality of human resources, technological development, innovations, and capital formation.

Exercise 1.4

1. Describe the economic problem.
2. Define scarcity and choice. Explain their economic relationship.
3. Elaborate on the concept of opportunity cost and explain its usefulness in making economic decisions.
4. Can opportunity cost be zero? Provide reasons for your answer.
5. Examine factors that would cause an outward shift of the PPF.

Activity 1.3

Construct a production possibilities frontier of an economy showing:

(a) Two goods which are mostly produced in your region and they use the same resources in production, with increasing opportunity cost.
(b) Assume one among the goods you have chosen undergoes resource shortage and show what happens to the frontier.
(c) Show points that depict scarcity, efficiency and inefficiency of the economy on the PPF and provide explanations for each point.
(d) What are the economic implications of the region represented by the PPF?
(e) Explain why the PPF is concave to the origin, rather than being convex?

Economic systems

An economic system is a mechanism by which an economy makes decisions on how to coordinate the production, distribution and consumption of goods and services. It is through an economic system that society tries to solve its central problems resulting from scarcity of resources and unlimited wants. Thus, every society faces three basic central problems concerning allocation of resources:

1. **What to Produce?**
   Due to scarcity of resources and unlimited wants, countries have to choose which among the goods and services demanded should be produced given limited resources.

2. **How to Produce?**
   Given the scarce resources, countries have to decide the best production methods (those that minimise the cost of production) with which they can attain the desired production level.

3. **For Whom to Produce?**
   Since countries cannot meet all individuals' wants, production decisions are made in order to meet specific needs of the market.

The way society solves these problems determines the type of economic system they have.

Types of economic systems

There are three major types of economic systems. These are the socialist (centralised or command) economic system; capitalist (price or free enterprise) system; and the mixed economy system.

Socialist economic system

A socialist economy is also called command economy or centrally planned economy. Socialism was advocated by Karl Marx who led a revolt against capitalism by arguing that allowing the economy to be controlled by individuals' self-interests would lead to crisis and disorder. As a result, he emphasised on the role of the government in controlling and regulating the economy. Thus, under socialism, the government owns and controls all productive resources in their society's overall interests.

Characteristics of a socialist economy

In a socialist economy all means of production are owned by the government in the interests of society. That is, individuals have no right to own any property from which they could earn profit. The government makes decisions at macro level aimed at maximising social welfare rather than maximising individual profit, meaning that forces of demand and supply have no influence in the market. Economic planning is done centrally whereby the government makes all decisions

concerning production and consumption, which ensure efficient utilisation of resources. For instance, the government makes decisions on what to produce and how much to produce, and these decisions are imposed on the society by law or by force, for example, by using production quotas. Socialist economies are able to reduce income inequalities through various means such as setting wages. This is possible because there is no right to owning private property and economic decisions are centralised. That is to say, no other economic agent other than the government is allowed to make economic decisions.

Most modern economies are practising both socialism and capitalism economic system. However, the most common examples of socialist economies are the Former Soviet Union, which was dismantled in 1991, China from the early 1950s to the late 1970s, Cuba from 1959 and North Korea.

Advantages of socialist economy

It provides economic security in the sense that the main objective of a socialist economy is to improve the welfare of all citizens and ensuring that all resources are fully employed. As a result even the disadvantaged are able to get their basic needs. Unnecessary competition and income equalities are reduced. That is, the government has the power to set the price levels that eliminate unnecessary competition and the levels of wages that reduce income inequality. It enables rapid transformation of societies to conform to the government vision. This may involve shifting production resources away from one sector to the sector of interest or increasing wages in sectors of interest in order to motivate workers. There are no industrial unrests in socialist economies because wages are determined by the government. In addition, socialist economies put great emphasis on improving the quality of life rather than increasing quantity of goods and services.

Disadvantages of socialist economy

Socialism leads to lack of economic freedom for producers and consumers. This is due to the fact that the government is the one that determines what to produce, how to produce and for whom to produce. Moreover socialism can cause economic inefficiency. This is because most government systems are rigid to change and therefore, they may fail to meet society's needs, leading to shortages. There is lack of innovation because workers and managers are not motivated to take risks where they can earn more. In addition, there may arise conflicts of interests between the government and its people because what is produced does not always match what is demanded.

Capitalist economic system

A capitalist economy is also called a free enterprise economy, or a market economy, or a laissez-faire economy. Earlier classical economists led by Adam Smith advocated laissez-faire policy, which means freedom in economic decision-making, with little or no government intervention. Adam Smith argued that allowing individuals to pursue their self-interests would lead to an outcome beneficial to all. Thus, Smith advocated capitalism such that individuals are free to make their own choices producers' choices are driven by the profit maximisation motive, whereas consumers' choices are driven by the utility maximisation motive. He referred to this coordination of economic activity the "Invisible Hand" because self-interest of individuals, namely, profit maximisation and utility maximisation, guided these interactions between producers and consumers; and regulated by market competition. Hence, no government interference.

In the free market economy, price mechanism works through forces of demand and supply. The forces of demand and supply determine what to produce, how to produce and for whom to produce as well as the rewards for the factor services.

Characteristics of a capitalist economic system

Within the capitalist economic system, there is freedom to ownership of private property. Individuals are free to own properties such as factories, machines and land from which they can earn profit. There is also freedom to make production decisions. Individuals are free to make their own decisions in terms of the occupation they want and the kind of goods and services they want to produce, with the main aim of maximising their incomes. Moreover, people have freedom to make consumption decisions. Consumers are free to make their own choices on goods and services on which their income will be spent in order to maximise utility. This freedom is called consumer's sovereignty.

However, in capitalist economic system, there is little or no role played by the government in economic activity. This helps markets function freely and efficiently. Entrepreneurs are guided by the profit motive, which motivates them to maximise revenue and minimise cost. This motive brings about efficiency and self-regulation in a capitalist economy. Moreover, there is presence of stiff competition in the market due to freedom of entry or exit. Market freedom allows prices to reflect what is actually happening in the market and what is desired

by economic agents. Thus, prices are determined by market forces of demand and supply. Furthermore, the basic central problems of what to produce, how to produce and for whom to produce are solved by the private sector because economic decisions are made by individuals.

Although there are no countries that practice pure capitalism in the world today, a few economies known to be capitalistic countries include United Kingdom, United States of America, Portugal, the Netherland, Australia, France and Canada with an active intervention of the government in economic development.

Advantages of the capitalist economic system

In the capitalist economic system, the decision-making process is greatly participatory because there is little or no government control. Producers are free to produce what consumers demand and consumers are free to spend their incomes according to their preferences. This allows producers and consumers to satisfy their self-interests. Moreover, there is market efficiency because resources are allocated to their most profitable use and prices are determined by market forces of demand and supply, what is produced is assumed to be entirely consumed. Therefore, the economy has neither shortages nor surpluses.

There is improvement in the quality of products. Stiff competition motivates firms to improve their products' quality in order to be highly competitive in the market Stiff competition prevents firms from selling their products at higher prices, and hence, consumers benefit from low prices. In addition, stiff competition stimulates firms to invent new techniques that will maximise their profits and make them highly competitive in the market.

Disadvantages of the capitalist economic system

Within the capitalist economic system, producers tend to produce only goods and services that yield the highest profit. This is because producers in capitalist economies are driven by a profit maximisation motive. Therefore, goods and services that do not yield the highest profit but are essential to the public, for example, traffic lights and national defence, may not be efficiently produced.

Consumers could be charged high prices even for essential goods. This could happen because there is little or no government interference in price determination. There are high income inequalities between the rich and the poor. Since these economies tend to be self-regulating, the rich are more likely to get richer and the poor are more likely to get poorer.

If the capitalist economy is not well regulated, there can be over-production and over-consumption of demerit goods such as drugs, cigarettes and alcohol. These are results of freedom of producers to decide what and how much to produce and consumers to decide how to spend their incomes. Since producers are free to decide on what and how much to produce, they are likely to focus only on profit maximisation motive and ignore the negative impacts of their production processes.

Mixed economic system

A mixed economy, also called hybrid economy, was introduced by John Maynard Keynes who argued that in order to expand aggregate demand and restore people's confidence in the economy, the government should take control over a capitalist economy. This is not done by controlling wages and prices like in socialism, but by running budget deficits and increasing government expenditure on public works. This type of economic system is referred to as a mixed economy because it combines some features of both free market and socialist economic systems. Individuals enjoy freedom of choice but the government also intervenes to protect the welfare of its citizens, especially the weaker ones.

Characteristics of mixed economic system

This economic system is characterised with co-existence of private sectors and public sectors. This implies that the economy is now driven by both profit maximisation motive and social welfare improvement motive. In these economies, the government regulates the private sectors through policies such as monetary and fiscal policies.

There is freedom of individuals to choose any occupation that maximises their income and any goods and services they want to consume and produce, keeping in mind people's welfare. Thus, the government intervenes by putting some restrictions on consumption and production of certain items that are harmful to people. Prices are determined by market forces of demand and supply.

The government can intervene in the market by setting either price ceiling or price floor in order to improve people's welfare. The government can also intervene in the market through redistribution of income to reduce income inequality. This can

be done through various means such as setting minimum wages and imposition of progressive taxes. Moreover, economic planning is done by the government by indicating roles to be played by both public sectors and private sectors. To motivate the private sector, the government provides incentives and subsidies to ensure that national priorities are met.

Most countries are now considered to be mixed economies because of the active role of the government and private sector in the economy. Some examples of countries that are now practising mixed economies include China, Vietnam, Tanzania, India as well as Scandinavian countries such as Sweden, Norway, Finland and Denmark are among the best examples of mixed economic system.

Advantages of mixed economic system

The government can protect its people from unfair practices by setting laws that restrict such practices and thus, improve their welfare. It also allows the government to play a big role in mobilising priority areas such as defence, which could otherwise be neglected.

The mixed economy encourages innovation because the most efficient producers are rewarded and incentivised so as to motivate them to further improve their performances. Moreover, there is improvement in service quality. Since public and private sector cooperate in providing certain services such as education and health care to people. Thus, the quality of these services is likely to improve and hence, improving the social welfare of people.

Disadvantages of mixed economic system

Too much government intervention can diminish the essence of free market. If the government always sets prices of commodities, markets will not clear, that is, there may arise excess demand or excess supply, which in the long-run may distort the whole economy. Mixed economies can lead to creation of monopolies, for example, the government can grant a monopoly power to a company to be a sole producer and supplier of electricity to improve provision of the services. In case the company fails to achieve efficiency in production, in the long-term the entire economy can be negatively affected by such inefficience.

Transition period

A transition period is a period which a country is changing from one economic system to another, mostly from centrally planned economy to a market or a mixed economy.

The economy in transition period is characterised by features of both, the new economic system, which is being introduced and the old system. There are different policy reforms in different economic aspects. On one hand, if an economy is moving from a command economy to a capitalist economy, policies like privatisation and liberalisation of the market are adopted. On the other hand, if the economy is moving from capitalism to socialism, nationalisation policy is adopted.

Most of economies are in a transition from a state controlled economy (socialism) to market controlled economies. This has necessitated these countries to embody on major programmes of economic reforms. The reform programmes involve the following key elements: freeing prices and withdrawal of subsides, adaptation of privatisation policy, budget deficit and development of budgetary and fiscal system.

Activity 1.4

Read newspapers and use various journals and websites that discuss about the economic matters of Tanzania and other countries (focus on at least one country from the America, Europe, and Asian continent).

(a) Assess features of the Tanzania economy and identify the economic system Tanzania is operating based on what you have learned.
(b) By using vivid examples, explain why Tanzania is operating under the economic system mentioned in section (a).
(c) Assess features of other countries economies' and identify types of economic system each country is operating based on what you have learned.
(d) Is Tanzania operating under the same economic system like the other countries you have identified? If yes or no, provide reasons with vivid examples to support your answer.

Chapter summary

1. According to classical economics school of thought such as Adam Smith (1723 -1790), economics is defined as "a study of the nature and causes of the wealth of a nation." Also, Neo-classical economists such as Alfred Marshall (1842 - 1924) defined economics as "a study of mankind in the ordinary business of life." Nevertheless, Lionel Robbins in 1932 introduced the scarcity concept and defined economics as "a science which studies human behaviour as a relationship between ends and scarce means which have alternative uses." Robbins' definition of economics is superior to that of other economists and philosophers because it does not contain any reference of the term material or welfare.
2. Economics is a social science that analyses how individuals allocate their scarce resources to satisfy their unlimited wants. It deals with choices that individuals and businesses make to manage the scarce resources.
3. Economics is a social science subject that employs similar methods used in other fields of science subjects, such as chemistry, biology, and physics. As an art, economics involves practical application of knowledge, principles, and theories in order to achieve desired ends, which may include sustainable economic growth, stable prices and low rate of unemployment.
4. To understand the pattern of society in decision-making under scarcity, economics is divided into two branches of Microeconomics and Macroeconomics. Microeconomics is a branch of economics, which focuses on decision making undertaken by individuals (or households) and by firms. On the other hand, Macroeconomics is a branch of economics that studies the behaviour of the economy as a whole.
5. The fundamental economic problem of any society is scarcity of resources. Scarcity means less than required or limited in supply. Due to scarcity of resources, producers must answer the three basic economic questions of what to produce, how to produce and for whom to produce. The concept of scarcity problem can be well illustrated and explained by using the production possibility frontier (PPF) also known as production possibilities curve (PPC), which shows combinations of output that the economy can produce given the available amount of factors of production and technology.
6. An economic system is an institutional mechanism for determining ways of utilising scarce resources to satisfy human needs and wants. The economic system is designed to answer basic economic problem questions what to produce? how to produce? and for whom to produce? The common economic systems known are Command, Capitalist, Mixed and Transition.

Revision exercise

1. Lionel Robbins' defined economics as "a science which studies human behaviour as a relationship between ends and scarce resource which have alternative uses." Explain why this definition is not the only definition of Economics.
2. Clearly differentiate between the wealth definition of economics and welfare definition of economics.
3. Why is it important to study Economics?
4. Sketch a production possibilities frontier for two goods of your own choice.
   (a) Using the concept of opportunity cost, explain why the PPF is most likely to have a bowed-out shape.
   (b) Explain the curve by indicating the points showing scarcity, efficiency and inefficiency of the economy and provide explanations for each outcome.
5. You were planning to spend Saturday studying but a friend asks you to go for a shopping. What is the cost of going shopping in this case? Explain.
6. Using a well-labelled diagram of the production possibility curve (PPC), discuss the contention that "Where there is scarcity choice and opportunity cost are inevitable.
7. What is the fundamental economic problem of any society and how does the society attempt to solve this problem?
8. Differentiate between want and need by using practical examples.
9. What are reasons for a country to adopt a mixed economy?
10. Present and discuss key weaknesses of a market economy as a system of resource allocation.
11. The concept of scarcity is the basis of any economic enquiry. How true is this statement?
12. Distinguish between the two branches of economics.
13. Critically explain types of utility.
14. Elaborate on the difference between free goods and public goods.
15. Justify why public goods should be provided by the government.

Chapter Two

Theories of demand and supply

Introduction

People meet in a market where they sell and purchase different goods and services. In the market, the sellers exchange goods and services for money and the buyers exchange money for the goods and services. In economics, buyers and sellers respectively, form the supply and demand sides of the market. In this chapter, you will learn about the concepts of demand and supply, laws of demand and supply, analysis of demand and supply by using schedules, curves and functions, and concepts of elasticity of demand and supply. The competencies developed will enable you to make rational decisions in allocating the scarce resources efficiently.

Demand and supply

Demand and supply are two sides of a given market. A market is any arrangement that brings buyers and sellers together to exchange goods or services. For any market, the demand side constitutes buyers of a good or service and the supply side constitutes the sellers of the same good or service. Thus, any market is comprised of consumers and producers. In brief, demand is associated with consumption, while supply is associated with production. Thus, the demand side focuses on the behaviour of consumers (buyers), while the supply side focuses on the behaviour of producers (sellers). This chapter discusses each side separately, first, starting with the demand side of the market and then moving to the supply side. Finally, it discusses both sides together to determine the market equilibrium.

The concept of demand

Demand is willingness and ability of a consumer to purchase a certain quantity of a good or service at a prevailing market price in a given period. This economic principle examines the behaviour of buyers in the market. In economics, demand

is not the same as desire, need or want. Desire is a wish to have a good based on its intrinsic qualities. Ability goes with the purchasing power of a consumer and hence, demand entails the willingness and ability to put one's desires into effect. The term demand is used as the short form of effective demand, which happens when a consumer desires to buy a good or service and has the ability to afford it.

Quantity Demanded

Quantity demanded is an amount of any good or service that consumers are willing and able to buy during a specific period at a given price. It is measured in units of a good or service within a given period, for example, daily, monthly, quarterly, semi-annually and annually. Price is the amount of money charged for one unit of a good or service. Many factors influence the quantity demanded for goods and services. They include price of that good or service, a consumer's income and prices of the related goods or services, among other factors. However, the price of a good or service remains the most important factor among them. Thus, in practice, demand is the relationship between the quantity demanded of a good and the price of that good. That is to say, holding constant all other factors that affect demand, the quantity demanded of a good varies negatively with its own price.

The law of demand

The law of demand states that, other things remaining constant, when price of a good or service rises, its quantity demanded will decrease, whereas, when price of a good or service falls, its quantity demanded will increase. Thus, *ceteris paribus* the higher the price, the lower the quantity demanded and vice versa.

The law of demand implies existence of an inverse (negative) relationship between the price of a good or service and its quantity that consumers are willing and able to purchase, other factors remaining constant.

Why does the quantity of a good or service demanded increases when its price falls? The answer is that with an income that cannot allow consumers to afford everything they desire, people always try to find the best deals they can get. If the

price of one good falls and the price of all other goods remain constant, the good whose price has fallen offers a better deal than before. Therefore, people would buy more of this item and may reduce the quantity demanded of the goods whose prices have remained unchanged.

Assumptions about the law of demand

The law of demand assumes that other factors affecting demand for a good or service do not change. However, the assumption will be implicit when demand is discussed. In this regard:

(a) The law of demand operates only when changes in consumers' incomes are not considered. That is, the consumer's purchasing power must remain constant.

(b) There should be no change of taste and preferences of the consumer.

(c) Prices of all other interrelated goods should not vary. Such interrelated goods include those could be a substitute to the good in consumption (for example, tea leaves and coffee) or those complement each other (for example, tea and sugar). Economists call the former substitute goods and the latter complement goods.

(d) The law assumes that consumers do not expect future changes in price.

Demand schedule

One way of presenting demand is to use a demand schedule. A demand schedule is a table that presents a list of quantities of a good or service demanded at a given price, when all other factors influencing demand for the good remains the same. There are two types of demand schedules, namely, the individual demand schedule and the market demand schedule.

Individual demand schedule

An individual demand schedule is a table that shows a list of quantities of a good or service demanded by an individual consumer at each price, when all other factors affecting its demand remain unchanged. Table 2.1 shows an individual demand schedule for pens in a given year.

Table 2.1: *Hypothetical individual annual demand schedule for boxes of pen*

Options	Price (Tshs) per box of pens	Quantity demanded of boxes of pen
A	1,000	10
B	2,000	8
C	3,000	6
D	4,000	4
E	5,000	2

Table 2.1 shows that when price of a box of pens is Tanzanian shillings (Tshs) 1,000 per box, a consumer purchases 10 boxes of pens. As the price per box of pens increases, the consumer's willingness and ability to purchase pens keep on falling and reach two boxes of pens when the price is Tshs 5,000.

Market demand schedule

Market demand schedule is a table showing a list of quantities demanded of a good or service by all consumers in the market at each price of a good or service, when all other factors that affect its demand remain constant. Therefore, market demand schedule sums up all individual demands at each price. For example, suppose there are only two buyers in the market, namely, Fatuma and John, whose quantities of boxes of pens demanded at various prices are as shown in Table 2.2. For example, when the price is Tshs 2,000, the quantity demanded is 8+4 boxes, which equals to 12 boxes of pen.

Note that to get the market demand schedule, one would sum up quantities demanded by all individuals at each given price, and not the prices. This is called horizontal summation, since the summation is across each row.

Table 2.2: *Hypothetical market demand schedule for boxes of pen*

Options	Price (Tshs per box)	Quantity demanded by Fatuma (boxes)	Quantity demanded by John (boxes)	Market demand (boxes)
A	1,000	10	5	15
B	2,000	8	4	12
C	3,000	6	3	9
D	4,000	4	2	6
E	5,000	2	1	3

Demand curve

Another way of presenting demand is to use a demand curve. The demand curve is a graphical representation of a demand schedule, which shows the relationship between price of a good or service on the y-axis and the quantity demanded on the x-axis, assuming other factors affecting demand remain unchanged. There are two types of demand curves, namely, individual demand curve and market demand curve.

Individual demand curve

This is a curve showing different quantities of a good or service demanded by an individual at each price, when all other factors affecting demand remain constant. A normal or typical demand curve slopes downward from left to right. A negative slope of the demand curve depicts an inverse (negative) relationship between price and quantity demanded. Table 2.1 shows options A to E of the quantities of boxes of pen demanded at different prices. Figure 2.1 shows an individual demand curve plotted for options A to E. The price per box of pens is plotted on the vertical axis and quantities of boxes of pens demanded on the horizontal axis.

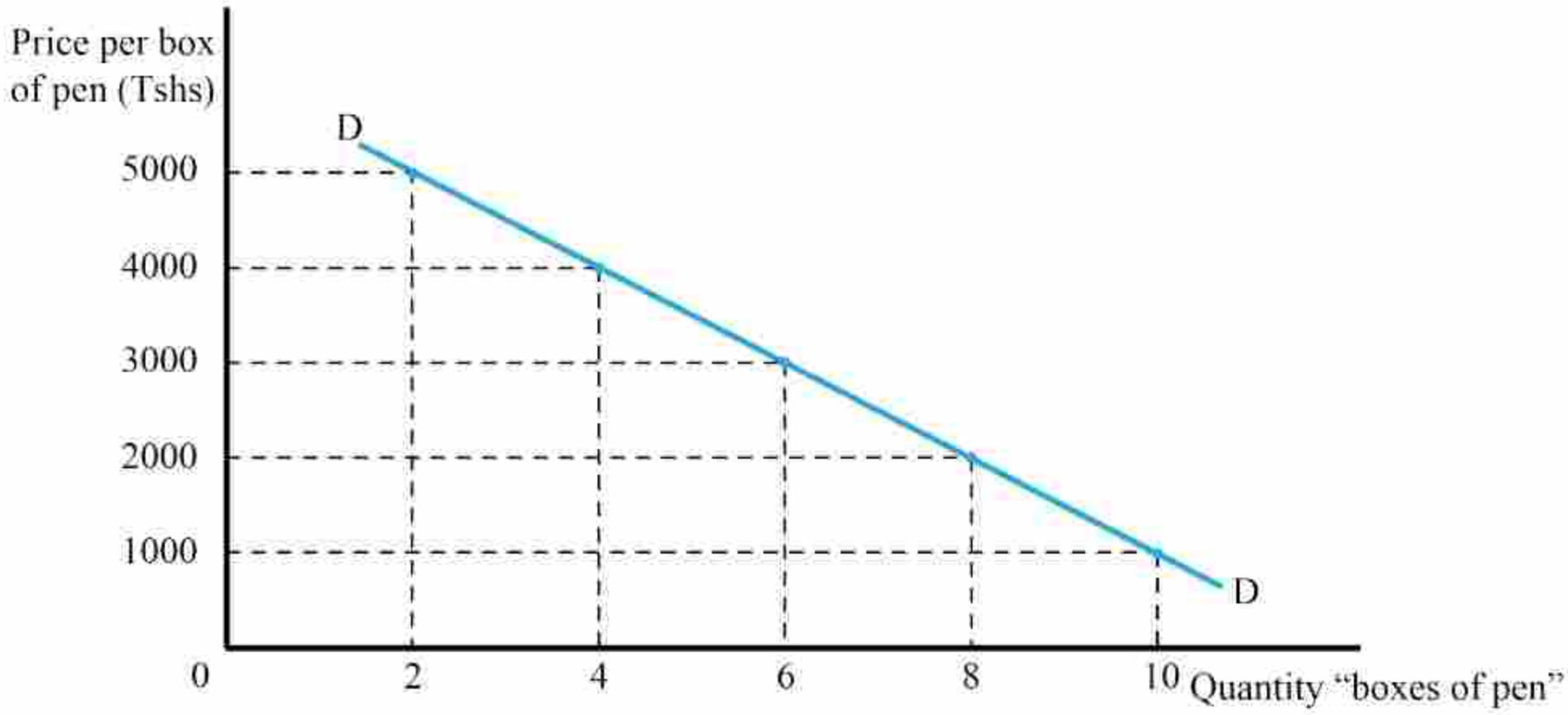


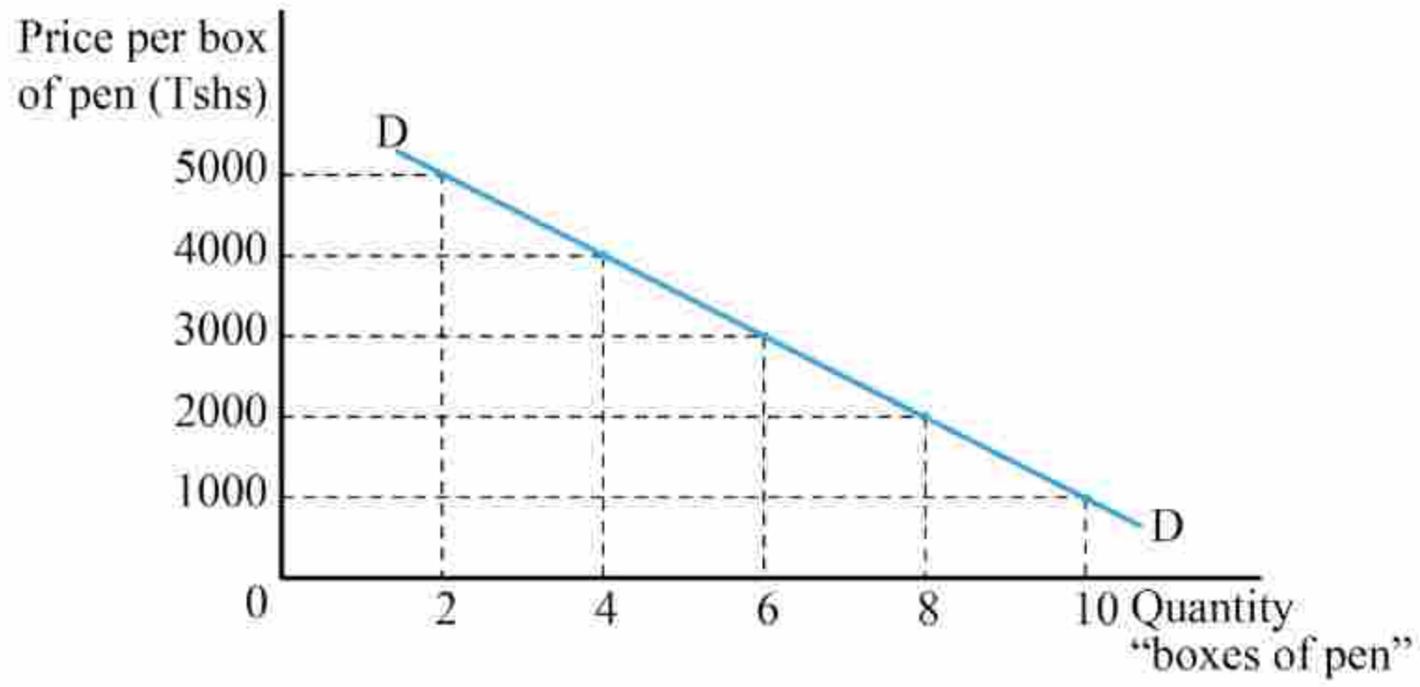
Figure 2.1: *Demand curve*

Individual and market demand curves

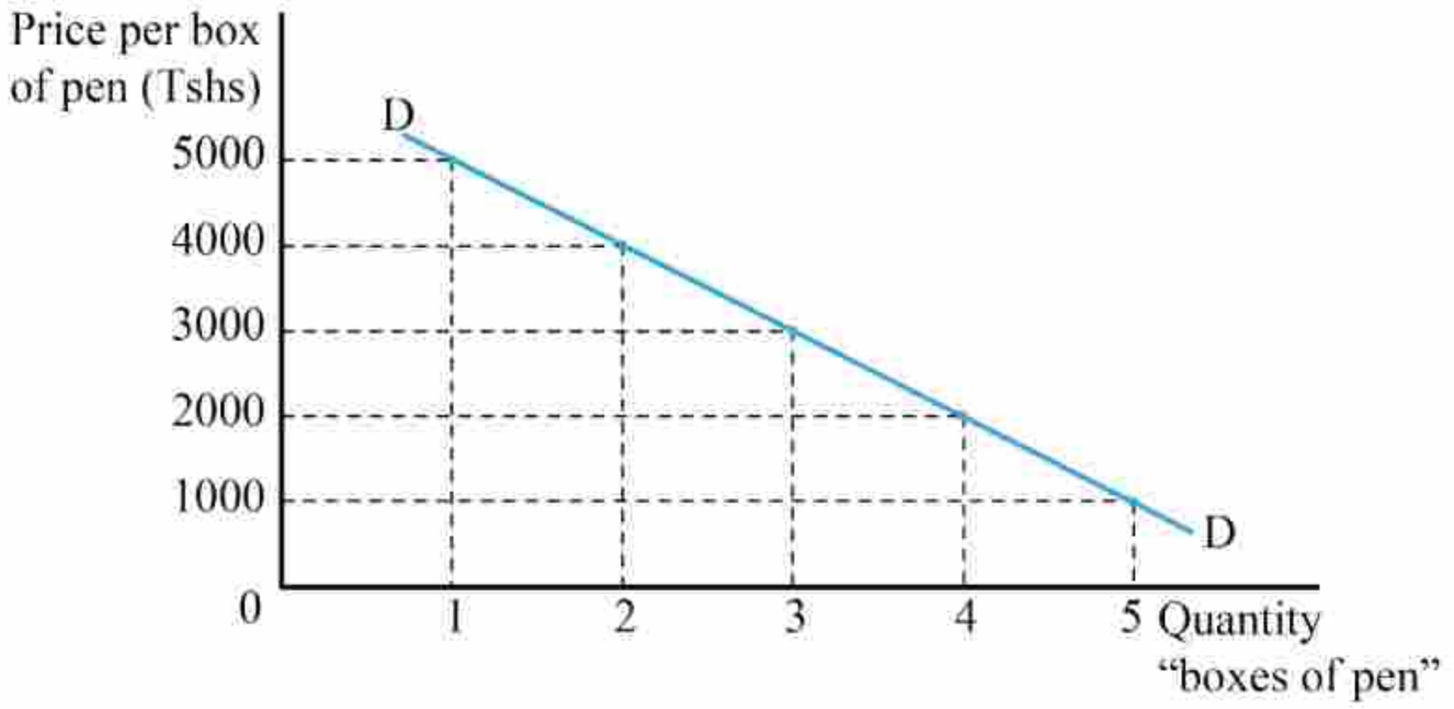
Market demand is used to analyse the way a market operates. A market demand curve is a graphical representation of the market demand schedule (Table 2.2). It is the horizontal sum of all individual demand curves for a particular good or service. It indicates how the total quantity demanded of a good will vary as the price of the good changes, while all the other factors that affect demand are held constant. Summing up horizontally, the individual demands for a good or a service across all consumers in the market means adding together all the quantities demanded by all individuals at each particular price.

For example, the market demand curve for boxes of pens is obtained by aggregating (adding together) quantities demanded by all individuals in the market. Figure 2.2 shows individual demand curves for boxes of pens for Fatuma and John, respectively, based on the data obtained from the market demand schedule in Table 2.2.

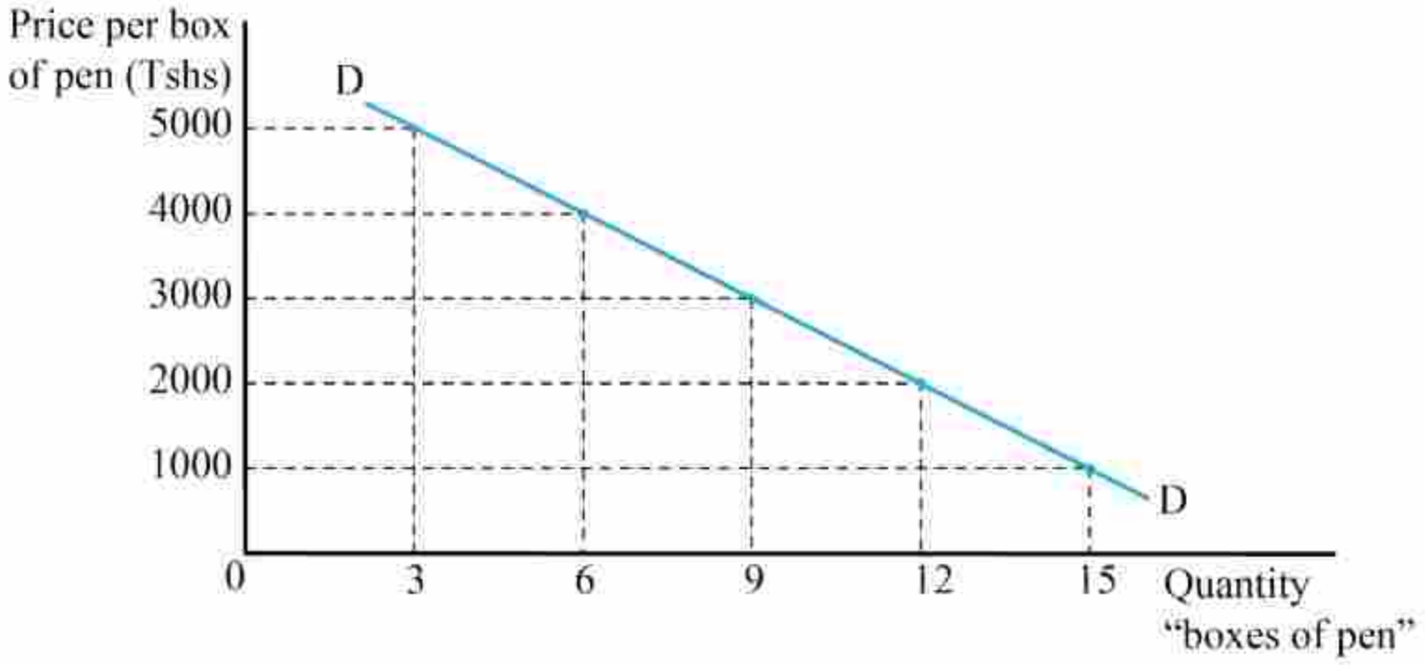
Fatuma's demand + John's demand = Market demand



Fatuma's demand curve



John's demand curve



Market demand curve

Figure 2.2: *Market demand curve (aggregate demand curve for Fatuma + John)*

Therefore, Figure 2.2 shows the market demand curve for Fatuma and John as derived from their individual curves.

Why is the demand curve downward sloping from left to right?

Consumers buy more of a good or service when its price falls, and less of it when its price rises. When the price of a product falls, a consumer buys a large quantity because of two effects, namely, the income effect and the substitution effect.

Substitution effect: This is a change in quantity demanded of a good or service that results from a change in the price of another good or service, without considering a change in quantity demanded that is caused by the change in the purchasing power. For example, when price of rice falls from Tshs 1,500 to Tshs 1,200 per kilogram and price of maize flour remains constant at Tshs 1,500 per kilogram, consumers will substitute rice for maize flour because rice has become relatively cheap.

Income effect: This is a change in the quantity demanded of a good or service due to a change in consumers' purchasing power. When the price of a good falls, the purchasing power of consumers increases and this may lead consumers to purchase more of the good because of the income effect. Conversely, when a price of a good rises, the decrease in the purchasing power of the consumers' incomes may also lead them to reduce the quantity of the good purchased.

A change in the quantity of a good or service that a consumer buys in the market when its price changes, is a combined effect of the substitution effect and the income effect, which is called the total effect of the price change. These effects are analysed in theory, but, in practice, the two effects cannot be observed separately.

Following up the example of rice and maize flour, a decrease in the price of rice will induce consumers to buy more rice instead of maize flour because rice will be cheaper relative to its substitute, maize flour. However, the decrease in the price of rice also increases the purchasing power of the consumer, which cause the consumer to buy more rice. When the purchasing power of income increase and the consumer uses part of it to increase the quantity of rice, in this case rice is referred to as a normal good, that is, a good whose demand increases as income increases. If the purchasing power of income increases and the consumer buys less quantity of rice, then rice is referred to as an inferior good, which is a good whose consumption decreases as income increases.

Exceptions of the law of demand

In certain circumstances, the law of demand does not hold. In such exceptions, the demand for a good or service varies directly with price such that the higher the price, the more the quantity that will be demanded, and vice versa. The following are some of the reasons for the existence of exceptions to the law of demand;

Giffen good: This is a specific type of inferior goods, whose consumption falls when a consumer's income increases. While other inferior goods obey the law of demand, Giffen goods do not. When the price of a Giffen good increases, its quantity demanded also increases, and vice versa.

Sir Robert Giffen (1837-1910) is credited for discovering this characteristic, and hence, the term Giffen goods. The demand curve of an inferior good slopes positively because the negative income effect on quantity of the good demanded is larger than the substitution effect. Thus, a decrease in price that cause the decrease in quantity demanded in the market is observed. Note that all Giffen goods are inferior goods but not all inferior goods are Giffen goods.

Veblen goods: The second exception to the law of demand are the Veblen goods (named after an American economist Thorstein Veblen (1857-1929) who discovered their characteristic). According to Veblen, there are certain luxury goods that become more valuable as their price increases. This is snob value, since it emanates from the ostentatious behaviour of people with very high income. The more expensive the good is, the more it becomes attractive to some consumers, causing its demand to increase as its price rises. This happens mostly in the case of luxurious goods such as cars, precious metals and gemstones such as gold, diamonds and Tanzanite. For such goods, the higher the price the higher it will be demanded.

Expectation of future change in price: If consumers expect the price of a commodity to rise in future, a current increase in price may not lead to a decrease in its demand. This is because people will wish to buy more today and avoid paying more in the future, and the vice versa. For example, during harvest season people tend to buy more cereals because they expect prices of cereals to rise in the future.

Figure 2.3 shows that the exceptional demand curve has a positive slope, as opposed to that of normal goods, whose slope is negative. This is because as the price of a good increases the quantity demanded increases, and vice versa.

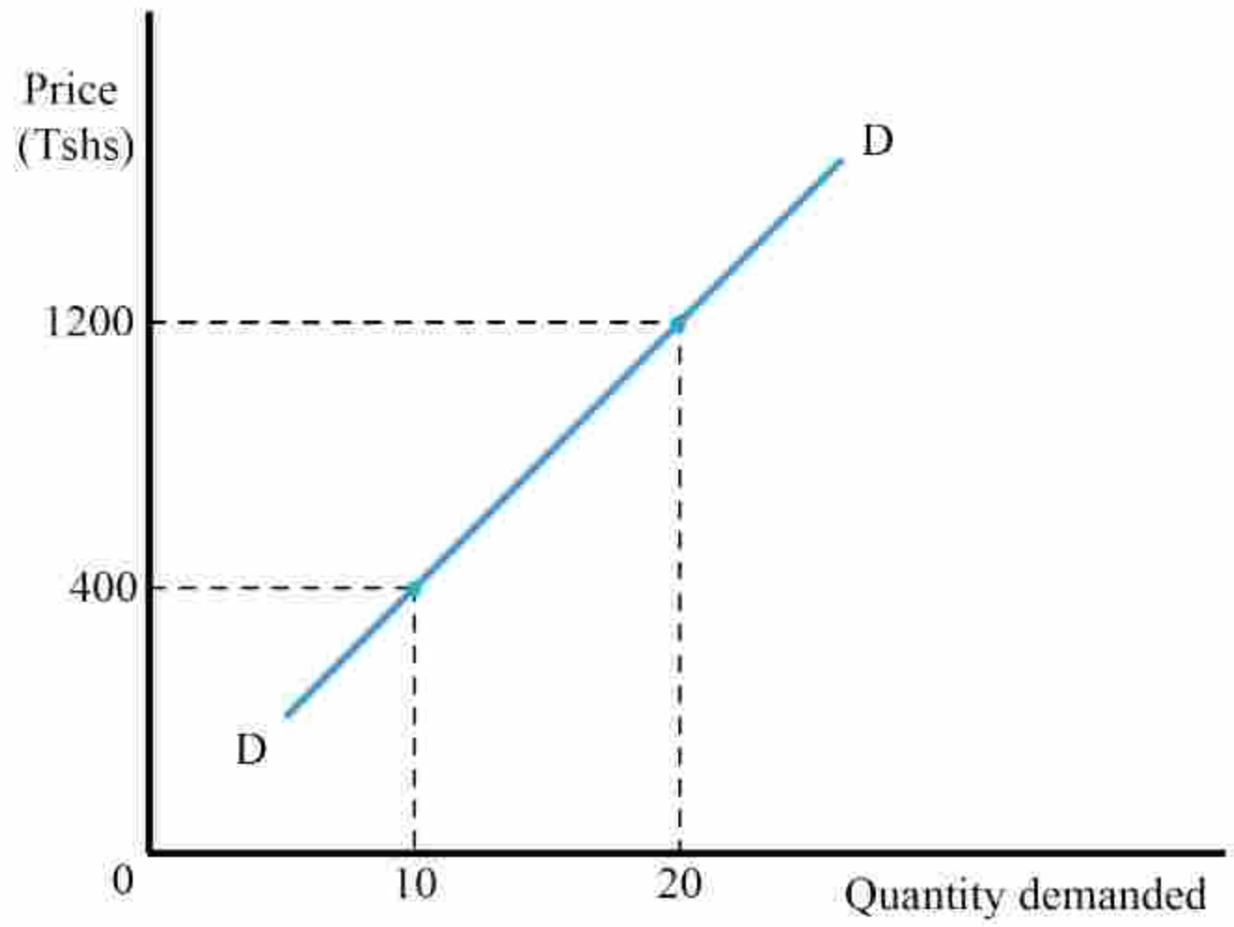


Figure 2.3: *Exceptional demand curve*

Necessities: Another exception to the law of demand is that of necessities. Even if price of the necessities rises, people will continue to buy them. Necessities include goods such as medicines, clothes and food. These goods have no substitutes and a consumer cannot reduce the required quantity, that is, the demand for such goods at any price remains the same. Thus, prices of necessities do not affect their demand.

Activity 2.1

There are various types of goods that you demand every day and you are willing to buy certain quantities at different prices.

(a) Choose one good of your preference, create its individual demand schedule, and then sketch its demand curve.
(b) With a fellow student, choose two goods of your preferences (choose two different goods). Generate a market demand schedule and plot the market demand curve.
(c) Distinguish between the two demand curves in parts (a) and (b).

Demand function

A demand function is a functional/mathematical relationship between the quantity of a good demanded and its determining factors. It can be expressed mathematically as

$$Q_x^d = f\left(P_x, P_o, Y, U, ...\right) \quad \text{.....(2.1)}$$

Where

Q_x^d is quantity demanded of good X

f is a function of independent variables contained within the parentheses

P_x is price of good X

P_o is price of other related goods

Y is income of a consumer

U is taste and fashion

For simplicity, it is assumed that factors other than the good's own price remain constant. Thus, it is only the price (P_x) of the good that affects the quantity demanded $\left(Q_x^d\right)$. Hence, demand function is expressed as

$$Q_x^d = f\left(P_x\right) \quad \text{.....(2.2)}$$

The slope of the demand function is negative, implying an inverse relationship between the price of the good and the quantity demanded. Thus, based on the general function in Equation (2.1), a demand function can be specified as

$$Q_x^d = a - bP_x \quad \text{.....(2.3)}$$

Where;

'a' is the intercept of the curve

'b' is the slope of the demand function

Note that in mathematics, the price should normally be plotted on the x-axis and quantity on the y-axis. However, in economics, price is plotted on the y-axis and

quantity on the x-axis. Therefore, mathematically, the demand curve is an inverse demand function, that is

$$P = f(Q) \quad \text{.....(2.4)}$$

Therefore, given the demand function in Equation (2.3) an inverse demand function is expressed as

$$P_x = \frac{Q_x^d - a}{b} \quad \text{.....(2.5)}$$

$$P_x = \frac{a}{b} - \frac{Q_x^d}{b} \quad \text{.....(2.6)}$$

$$P_x = \frac{a}{b} - \frac{1}{b}Q_x^d \quad \text{.....(2.7)}$$

Equation 2.7 suggests price (P_x) of commodity determines its quantity demanded (Q_x^d).

Change in quantity demanded and change in demand

Is there any difference between change in quantity demanded and change in demand? The answer is yes. Change in quantity demanded depends on the price of the good, whereas change in demand depends on factors other than price that determine demand.

Change in quantity demanded

Change in quantity demanded refers to movement along the demand curve. It refers to changes in quantity demanded as the price of the good changes. You may recall that change in quantity demanded results from change in the price of the good, while other factors (such as income of the consumer, price of other related commodities, taste and fashion) remain constant.

An increase in the quantity demanded is due to fall in price of a commodity, while other factors remain constant. Figure 2.4 shows that a decrease in price from Tshs 400 to Tshs 200 leads to an increase in the quantity demanded from 20 to 40 units.

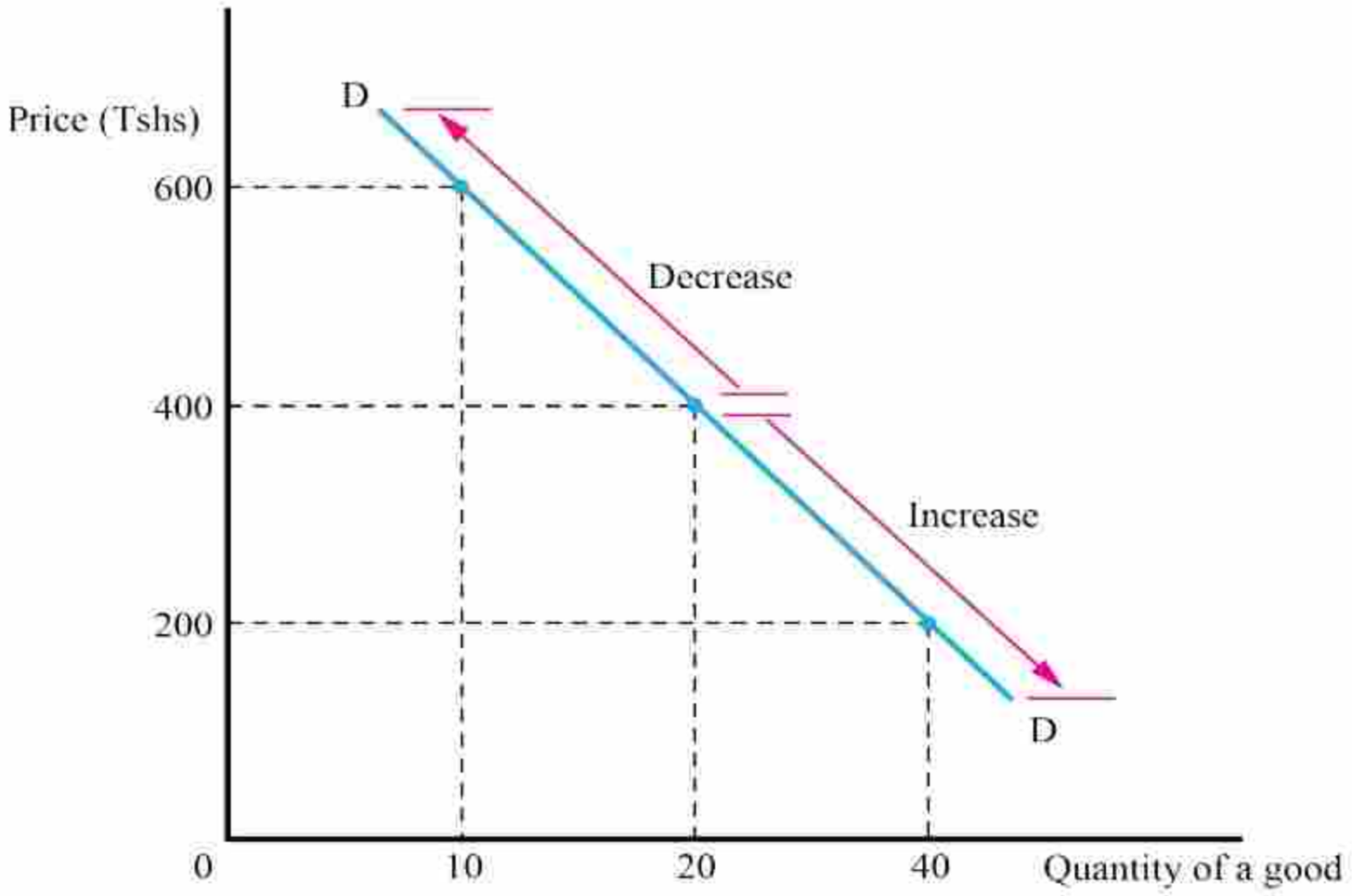


Figure 2.4: *Change in the quantity demanded*

On the other hand, *decrease in quantity demanded* results from a rise in price of the commodity while other factors remain constant. In Figure 2.4, when the price increases from Tshs 400 to Tshs 600, the quantity demanded decreases from 20 to 10 units.

Change in demand

Change in demand is presented by a parallel shift of the whole demand curve, either rightwards or leftwards. In Figure 2.5, the leftward shift of the curve is referred to as decrease in demand and the rightward shift is referred to as an increase in demand.

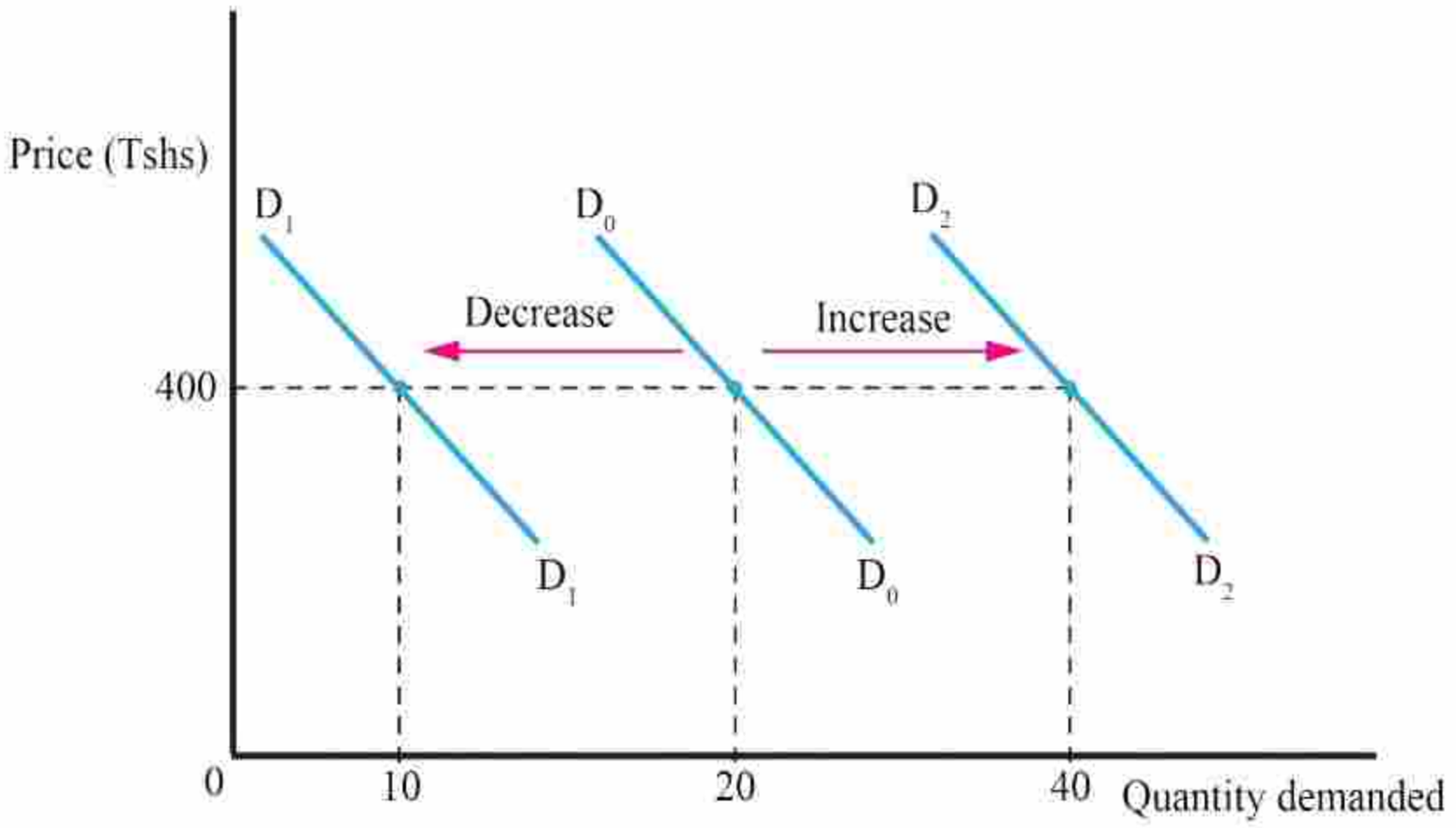


Figure 2.5: *Change in demand*

The increase or decrease in demand is brought about by change in other determinants of demand such as consumers' income, price of other related products, expected changes in future prices and tastes, while price of the good remains constant. It should be noted that a change in any factor other than the price that affects the demand of a good will shift the demand curve to right or left of the original demand curve if the commodity is a normal good.

In Figure 2.5, the price of a good remains unchanged at Tshs 400. But other factors that affect demand have changed, for example, consumer's income, taste, preference and price of other related goods. Suppose the initial demand curve is D_0D_0, price of the good is Tshs 400, and the quantity demanded is 20 units. Assume one of the other factors that influence demand has changed, for example, consumer's income. The demand for the good will increase if income increases (in this case, to 40 units) or decrease if income decreases (in this case, to 10 units). The increase in demand and decrease in demand mean that the demand curve will shift from D_0D_0 to D_2D_2 and from D_0D_0 to D_1D_1, respectively. Remember, an inward shift of demand curve represents a decrease in demand, whereas a total outward shift of the demand curve represents an increase in demand.

Determinants of demand

Holding price constant, there are many factors that influence demand for a good or service. Some of the most important factors include the following:

Income of the consumer: Increase in income normally means the consumer has more money to spend on goods and services at given prices (increase in purchasing power). Ordinarily, consumers with low incomes spend less on goods and services and those with higher incomes spend more. However, the effects of income on demand for goods and services depend on nature of the good. For a normal good (necessities and luxury goods), the higher the income the higher its demand, and the lower the income, the lower its demand. For inferior goods, however, an increase in consumer's income will lead to a decrease in the demand for a good, and vice versa.

Price of related goods: Goods are either complements or substitutes. Complementary goods (complements) are goods consumed together. Examples of complementary goods are tea and sugar; car and tyres; and automobiles and diesel or petrol. Perfect complements behave as a single good. If the price of sugar goes up then demand

for sugar goes down, which in turn, make demand for its complementary good (in this case tea) to go down as well. What causes the demand for tea to go down is the increase in the price of sugar.

The other category of related goods are substitutes. These are alternative goods that satisfy the same want of a consumer. Examples of substitutes are Pepsi and Coca Cola. The relationship between the price of one substitute good and the quantity demanded of the other substitute good is positive. If price of Coca Cola goes down, it means that the quantity demanded of Coca Cola will increase, which implies that the quantity demanded of Pepsi will decrease. Alternatively, if the price of Coca Cola increases, the quantity demanded of Pepsi increases. Therefore, two goods are substitutes if a fall in the price of one of them makes consumers to buy less of the other good, *ceteris paribus*.

Change in consumer's tastes and habits: When tastes and habits change, the demand for a good or services changes as well. Favourable change in tastes leads to an increase in demand, whereas unfavourable change in tastes leads to a decrease in demand. For instance, with time an individual may prefer to consume more milk than soft drink. The change in tastes is likely to cause an increase in demand for milk and reduce that of soft drinks. Likewise, demand can change because of change in the habits of a consumer. For example, an individual may change his or her eating habits after acquiring more knowledge on health eating, which lead to a change in the demand for certain food products and services.

Consumer expectations about future changes in prices and income: If a consumer believes that price of a good will increase in the future, he or she is more likely to demand more of the good at the present time. Thus, demand will increase now if consumers expect price of the good to increase in the future. The opposite will be the case if a consumer expects the price of a good to decrease in the future. In this regard, the consumer will postpone buying the good in the present time, which would cause a decrease in demand.

On the other hand, if the consumer expects income to increase in the future the consumer may borrow or dissave to purchase more of the good at present thereby causing its demand to increase. In other words, positive expectations about future income may encourage present consumption to increase, whereas when the consumer expects income to decrease in the future, demand in the present will decrease.

Interrelated demand

This is the demand for two or more related goods, such that the demand for one commodity affects the demand for the other. There are several types of interrelated demands, namely, competitive, complementary, derived and composite demands.

Competitive demand (for substitute goods)

Competitive demand occurs in a market with substitutes such that one good can be purchased instead of the other good. A substitute is a good consumed instead of another good to give the same utility, for example, Coca Cola and Pepsi or margarine and butter. If two goods are substitutes, an increase in the price of one good will lead to an increase in demand for the other good. Figure 2.6 shows the relationship between the price of one good and quantity demanded of the other competitive good, namely, Coca Cola and Pepsi. The figure shows that an increase in the price of Coca Cola causes an increase in the quantity demanded for Pepsi.

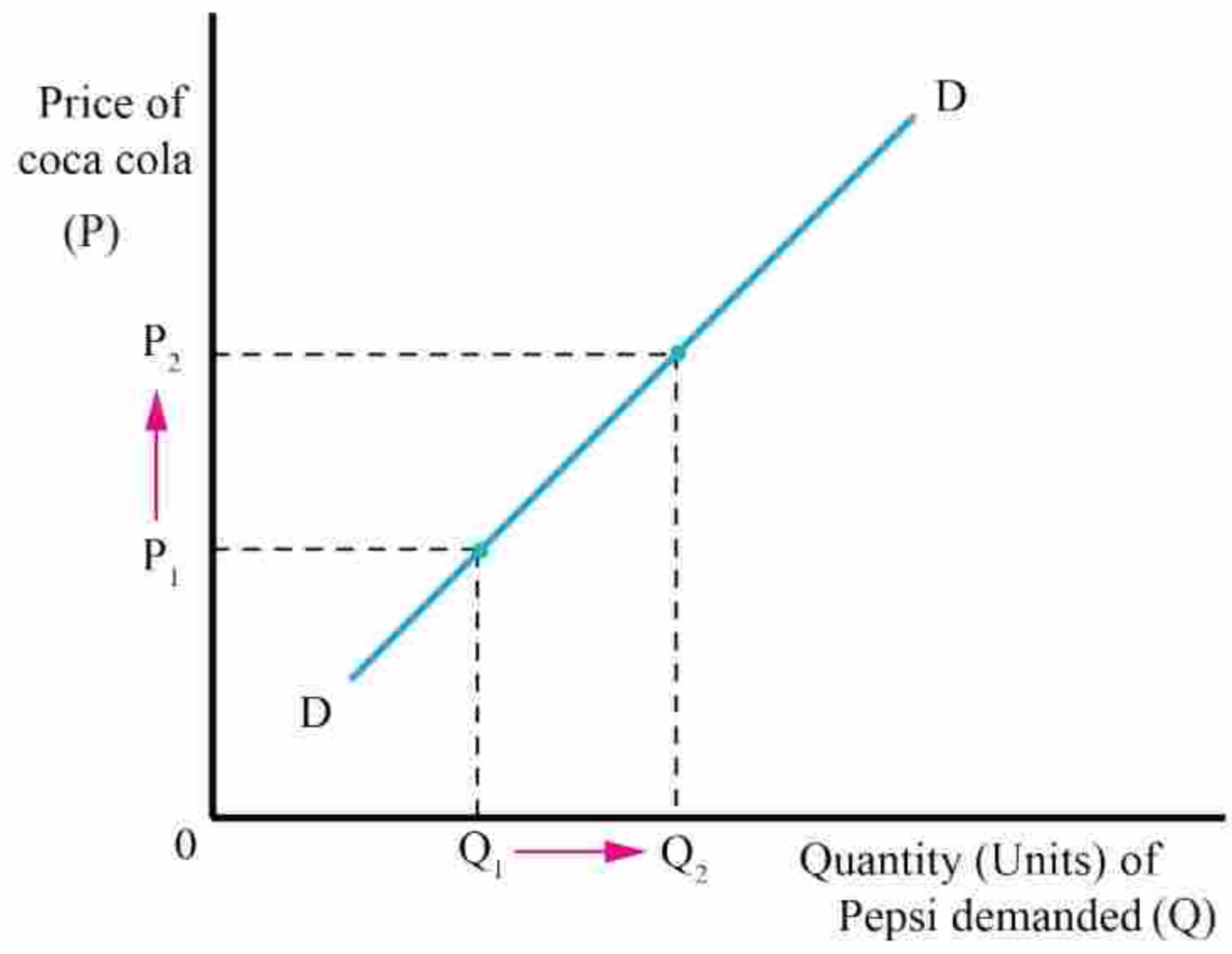


Figure 2.6: *Competitive demand of two substitute goods*

Joint or complementary demand

This is the demand for two goods that are used together (jointly) to satisfy the same want, for example, car and tyres. Two goods are complements if an increase in the

price of one of them results in a decrease in the quantity demanded of the other good, and vice versa. Figure 2.7 presents a curve showing joint or complementary demand. The figure shows that a rise in price of cars leads to a decrease in the quantity demanded for tyres. Since consumers will buy less cars at a higher price, less tyres will be needed.

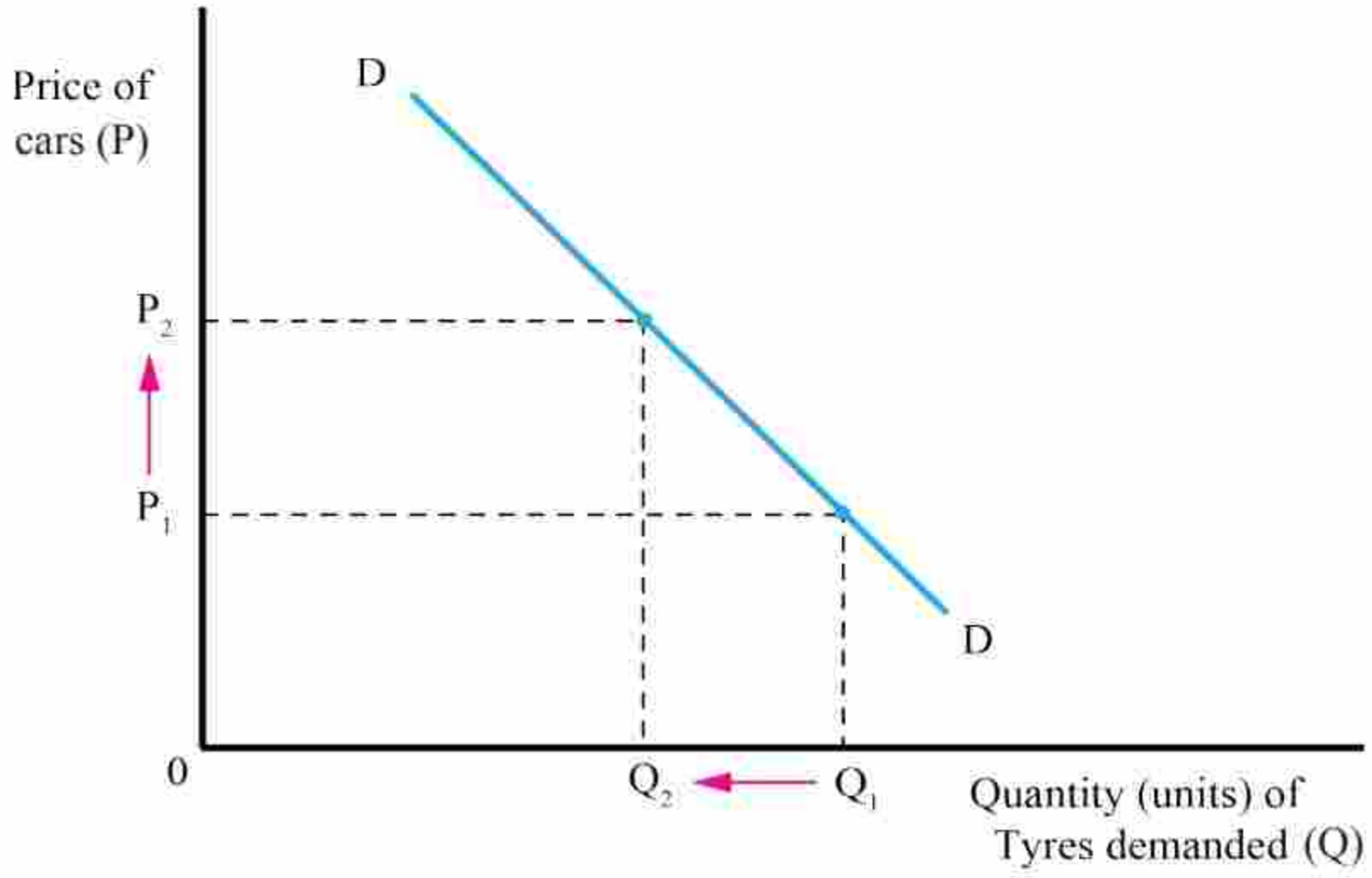


Figure 2.7: *Joint or complementary demand*

Composite demand

This is the demand for goods that have multiple uses. For example, demand for crude oil used to produce petrol, diesel and kerosene, and the demand for sheep for the supply of mutton and wool. Therefore, if the price of mutton increases the demand for sheep will also increase and similarly, if the price for wool increases demand for wool will also increase.

Derived demand

A derived demand is the demand for goods that are not demanded for their final use, but as inputs in the production of other goods. For example, an increase in the demand for sunflower oil will increase the demand for inputs used to produce it, for example, sunflower seeds. Thus, demand for sunflower seeds is a derived demand, derived from the increased demand for sunflower oil.

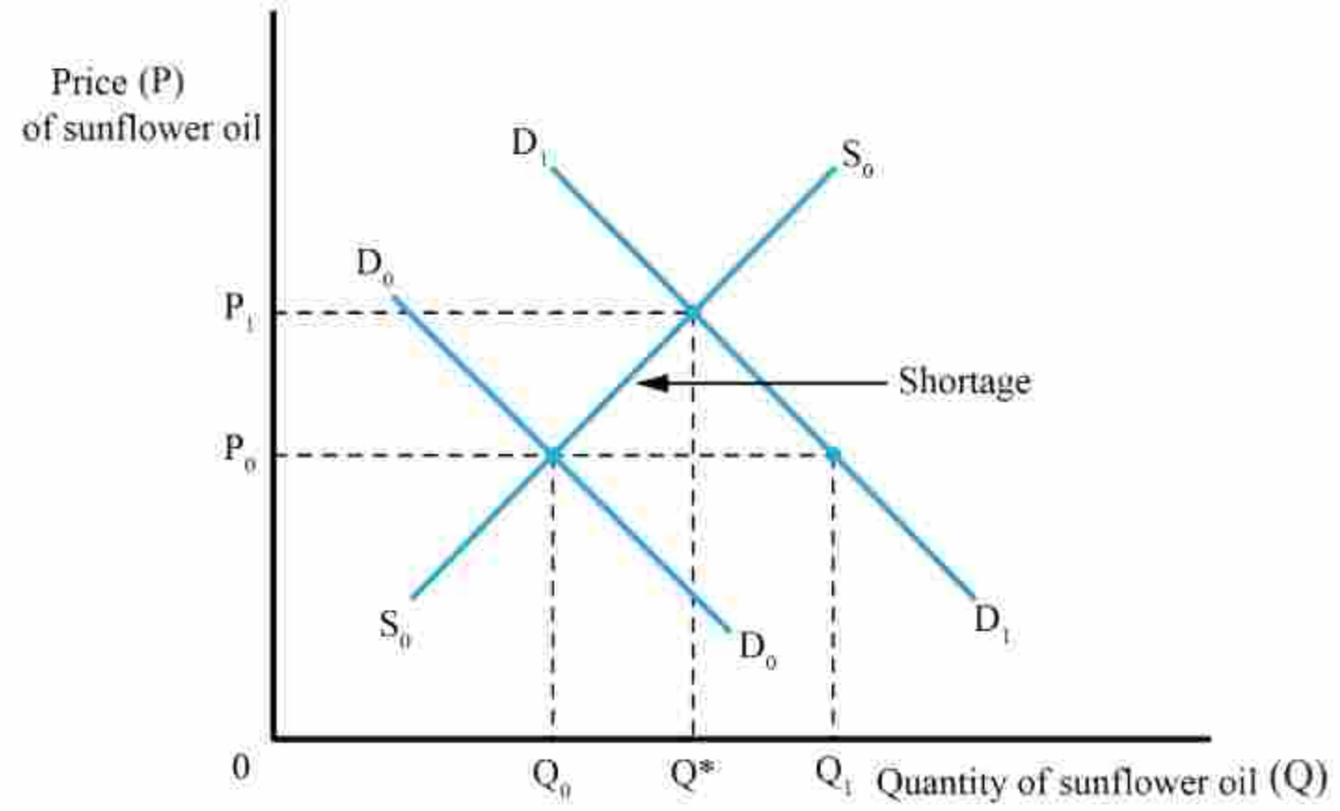


Figure 2.8: *Increase in demand for sunflower oil*

Figure 2.8 shows the increase in demand for sunflower oil from D_0D_0 to D_1D_1 and proceeds to show how the increase causes the demand for labour employed in the production of sunflower oil to increase as shown in Figure 2.9.

The increase in demand for sunflower oil, given its supply (presented by S_0S_0) will cause a shortage (Q_0Q_1) of sunflower oil in the market at P_0, which will cause the price to rise from P_0 to P_1.

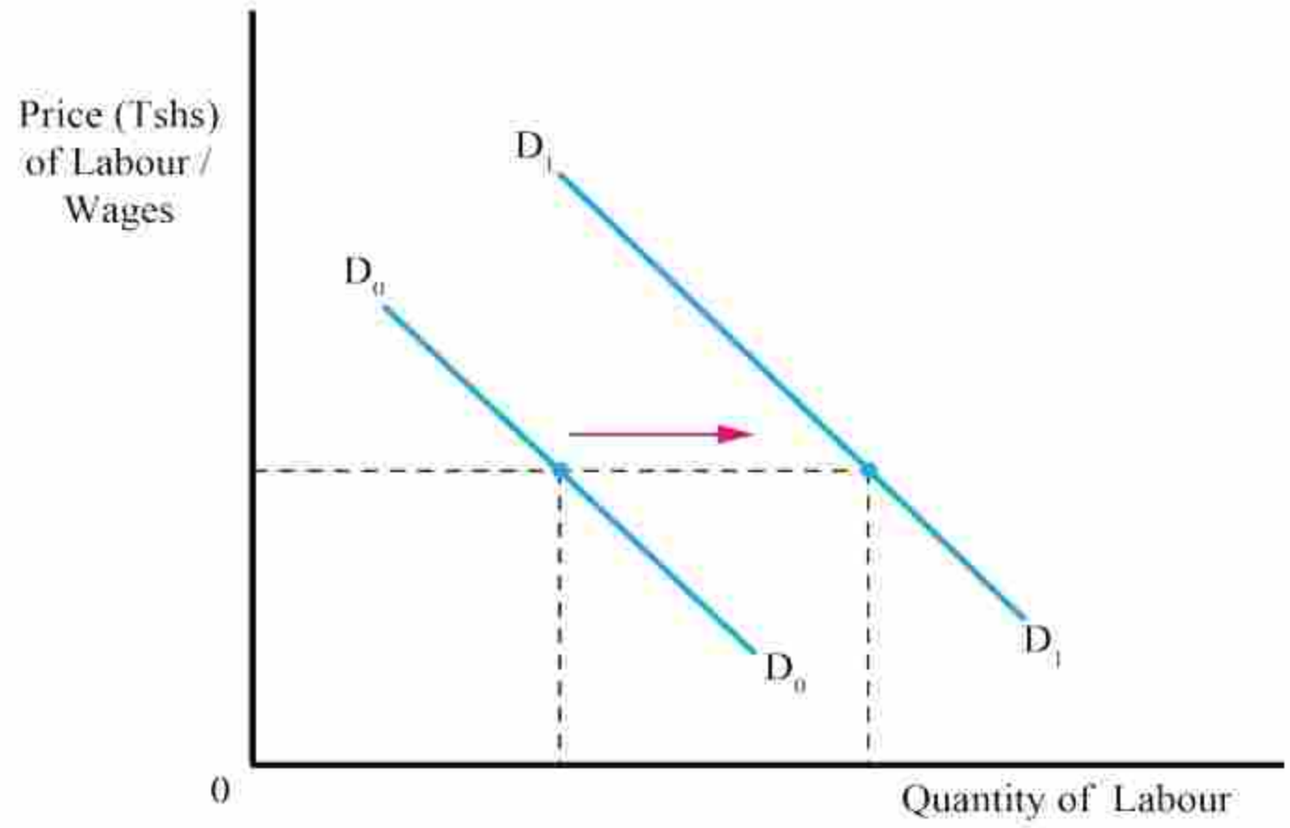


Figure 2.9: *Increase in demand for Labour*

Suppliers will increase supply of sunflower oil in order to take advantage of the opportunity caused by the rise in the price of sunflower oil due to its shortage. In order to increase the supply of sunflower oil, producers will demand more

labour for employment at the existing wage (Figure 2.9). As a result, the increase in demand for labour will shift the labour demand curve from D_0D_0 to D_1D_1, as shown in Figure 2.9.

The increase in demand for labour caused by the increase in demand for sunflower oil is a derived demand for labour. Derived demands often apply to markets for raw materials and labour, although they may occasionally apply to final or intermediate goods.



Activity 2.2

Think about various types of goods and services demanded by members of society you live in.

(a) Analyse types of goods and services demanded in the society.
(b) Examine if goods and services demanded (presented in (a)) appeal to the law of demand. If your answer is yes or no, provide reasons supported by vivid examples.
(c) Explain and illustrate the concept of interrelated demand using goods and services that you have analysed in (a).

Exercise 2.1

1. By using vivid examples, explain the meaning of demand and its determinants.
2. Discuss the relationship between a demand schedule and a demand curve and explain why the demand curve slopes downward from left to right.
3. State the law of demand. Explain its assumptions.
4. If there is a fall in the price of beef in an economy, then there will be an increase in the quantity of beef demanded and a fall in the quantity demanded for its substitutes, such as chicken. Draw a diagram depicting the demand curve for beef and show an increase in the demand for beef.
5. Analyse reasons for an abnormal demand curve.

The concept of elasticity of demand

Changes in factors that influence demand for a good or service can cause a large, small or no change in the quantity demanded for the good or service. The elasticity of demand addresses how much the quantity demanded for a good or service responds to a change in its price or to a change in any other factor that influences the demand for the good or service. It is a measure of the extent to which the quantity demanded of a good responds to a change in its price or that of any other determinant of demand. The most common measures of elasticity of demand include own-price elasticity of demand (or simply, price elasticity of demand), income elasticity of demand and cross-price elasticity of demand.

Types of elasticities of demand

The elasticity of demand for a good can be calculated with respect to price of a good, income of the consumer and price of other related goods.

Price elasticity of demand

This is a measure of degree of responsiveness of the quantity demanded of a good with respect to a change in its price, other factors being constant. The law of demand does not give the magnitude of changes in quantity demanded of a good in response to change in its price. Price elasticity of demand expresses how much the quantity demanded would increase or decrease as a result of a unit decrease or increase in the price of the good. The price elasticity of demand is calculated as follows:

$$\text{Price elasticity of demand } \left(e_p^d\right) = \frac{\text{Percentage change in quantity demanded}}{\text{Percentage change in price}}$$

$$e_p^d = \frac{\%\Delta QD}{\%\Delta P} \quad \ldots\ldots\ldots\

$$\%\Delta QD = \frac{Q_1^d - Q_0^d}{Q_0^d} \times 100$$

and the percentage changes in price (%ΔP) of the commodity are computed as follows:

$$\%\Delta P = \frac{P_1 - P_0}{P_0} \times 100$$

Therefore, the price elasticity of demand of a commodity is expressed as;
Rearranging the terms:

$$e_p^d = \left(\frac{\frac{Q_1^d - Q_0^d}{Q_0^d} \times 100}{\frac{P_1 - P_0}{P_0} \times 100} \right)$$

$$e_p^d = \left(\frac{Q_1^d - Q_0^d}{P_1 - P_0} \right) \times \left(\frac{P_0}{Q_0^d} \right) \quad \text{.....(2.9)}$$

Where; $Q_1^d - Q_0^d = \Delta QD$ and $P_1 - P_0 = \Delta P$, and therefore, Equation 2.9 can be expressed as:

$$e_p^d = -\left(\frac{\Delta QD}{\Delta P} \times \frac{P_0}{Q_0^d} \right) \quad \text{.....(2.10)}$$

Where

ΔQD is change in quantity demanded

ΔP is change in good's own price

P_o is original price and P_1 is the final price

Q_o is original quantity demanded and Q_1 is the final quantity demanded.

The price elasticity of demand normally is not a negative value, but an absolute value, which ranges from zero to infinity $0 < ed < \infty$. Because of this a negative sign or an absolute value symbol is plugged in the formula so as to obtain a

positive value. However, the negative sign of the price elasticity of demand is considered when interpreting an inverse relationship between price and the quantity demanded of a good.

$$e_p^d = -\left(\frac{\Delta QD}{\Delta P} \times \frac{P_0}{Q_0^d}\right)$$

Price elasticity of demand as computed in Equation 2.10 is at a particular point on the demand curve, such as point A or B in Figure 2.10. Hence, it is referred to as **point price elasticity of demand.**

Recall that price elasticity of demand expresses how much the quantity demanded would change (increase or decrease) as a result of a unit change in price. If there are two points A and B, the point elasticity of demand will give different answers depending on the starting point, whether the price increases or decreases.

Alternatively, one can use the mid-point (arc) elasticity of demand method. In this method, the price elasticity of demand is measured as an average elasticity of demand between two points on the demand curve, such as A and B in Figure 2.10 which is referred to as **arc price elasticity of demand.**

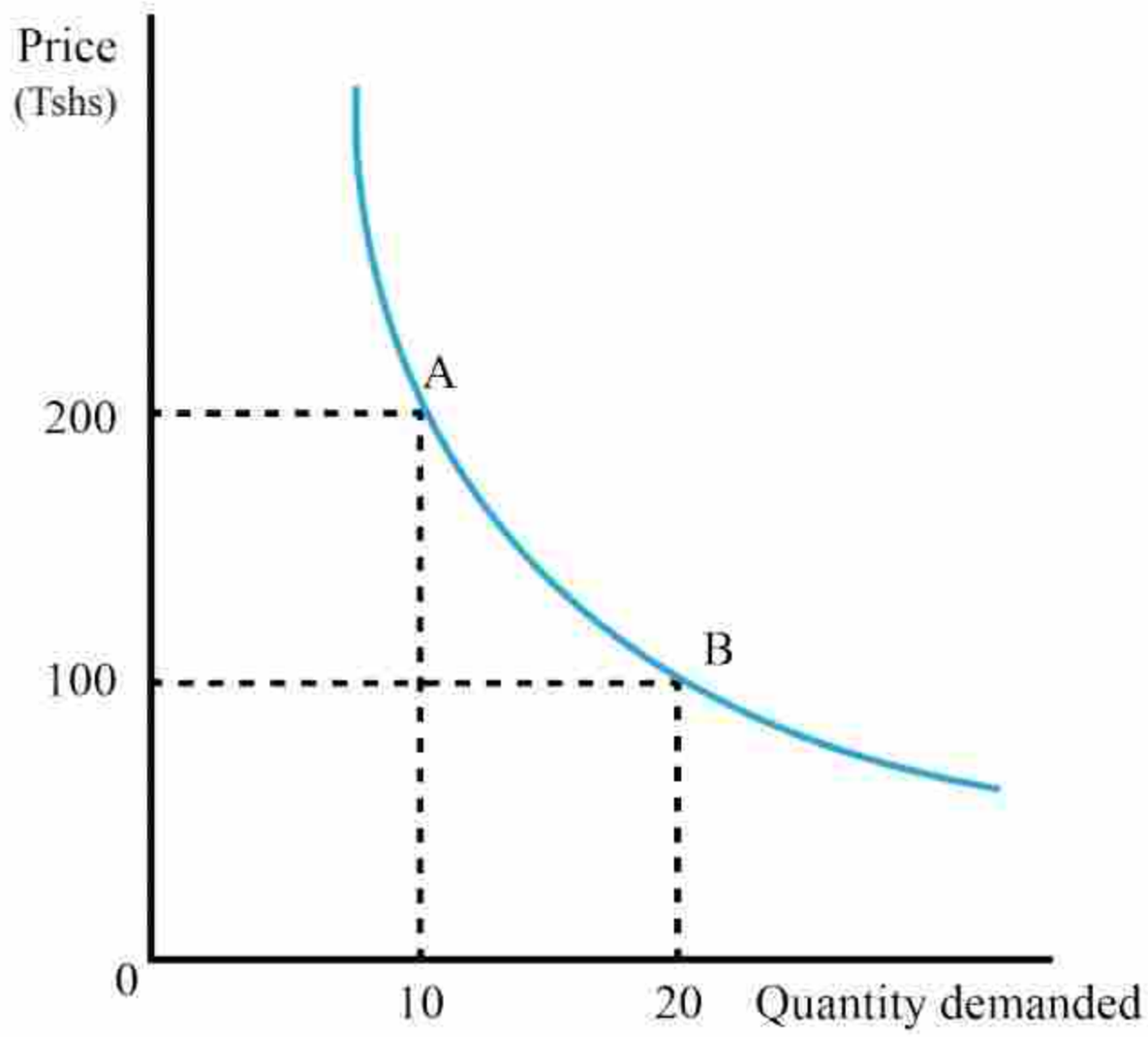


Figure 2.10: *Demand curve showing arc and point price elasticity of demand*

Thus, arc elasticity of demand is calculated as

$$e_p^d = -\left(\frac{Q_1^d - Q_0^d}{P_1 - P_0}\right) \times \left(\frac{P_1 + P_0}{2} \times \frac{2}{Q_1^d + Q_0^d}\right)$$

$$e_p^d = -\left(\frac{Q_1^d - Q_0^d}{P_1 - P_0}\right) \times \left(\frac{P_1 + P_0}{\cancel{2}} \times \frac{\cancel{2}}{Q_1^d + Q_0^d}\right)$$

$$e_p^d = -\left(\frac{Q_1^d - Q_0^d}{P_1 - P_0}\right) \times \left(\frac{P_1 + P_0}{Q_1^d + Q_0^d}\right)$$

$$e_p^d = -\left(\frac{\Delta QD}{\Delta P}\right) \times \left(\frac{P_1 + P_0}{Q_1^d + Q_0^d}\right) \quad \text{.....(2.11)}$$

Relationship between elasticity of demand and slope of the demand curve

Note that the price elasticity of demand is not the slope of the demand curve. However, the slope influences the elasticity of demand.

If price elasticity of demand is infinity $\left(e_p^d = \infty\right)$, it implies that a one percent change in the own-price of a good leads to an infinite change in the quantity demanded.

If price elasticity of demand is zero $\left(e_p^d = 0\right)$, it means that a one percent change in price leads to no change in the quantity demanded.

If price elasticity of demand is equal to one $\left(e_p^d = 1\right)$, it means that a one percent change in the own price of good leads to a one percent change in quantity demanded.

If price elasticity of demand is greater than one $\left(e_p^d > 1\right)$, it means that a one percent change in price leads to greater than one percent change in the quantity demanded.

If price elasticity of demand is less than one $\left(e_p^d < 1\right)$, it means that a one percent change in the own-price of a good leads to less than one percent change in the quantity demanded.

Interpretations of magnitudes of the own-price elasticity of demand

In calculating the price elasticity of demand, the answer can be 0, 1, ∞, greater than 1 or less than 1. The magnitudes of price elasticity of demand have different interpretations, which are consistent with following demands:

(a) Perfectly elastic demand $\left(e_p^d = \infty\right)$

Demand for a good is termed as perfectly elastic when an infinite quantity could be demanded even though the price of the good remains unchanged. Figure 2.11 represents a perfectly elastic demand curve. Its demand curve is a straight line parallel to the x-axis (horizontal line), implying the own price elasticity of demand is infinity (∞).

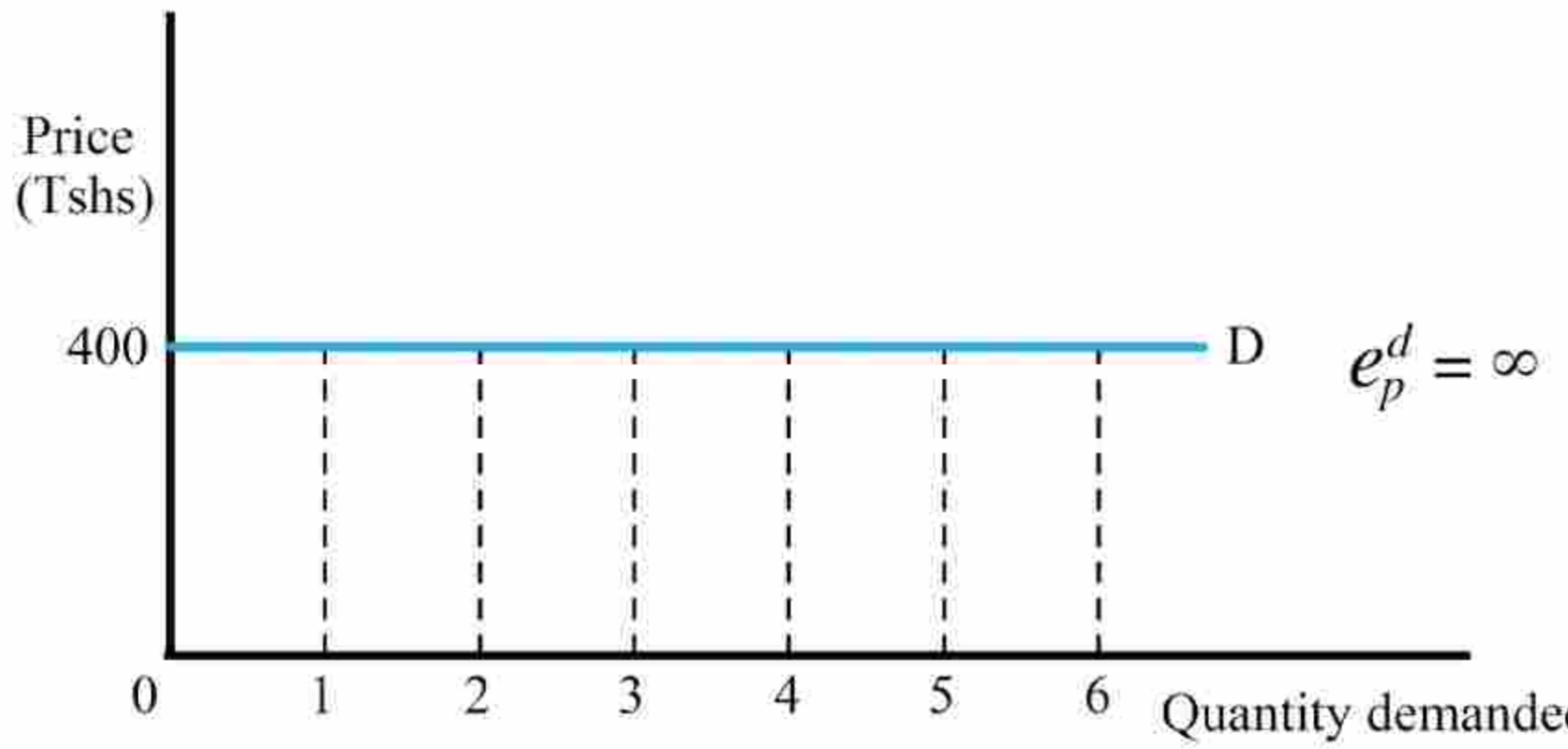


Figure 2.11: *Perfectly elastic demand curve*

Figure 2.11 shows that at the price of Tshs 400, the quantity demanded of a good will reach infinity because the percentage change in price (%ΔP) is zero and percentage change in quantity demanded (%ΔQD) is positive.

(b) Perfectly inelastic demand $\left(e_p^d = 0\right)$

The demand for a good is termed as perfectly inelastic if its quantity demanded does not change at all in response to whatever change in its price. That is, any percentage change in the own-price of a good leads to zero percentage change in the quantity demanded. Figure 2.12 represents a perfectly inelastic demand curve.

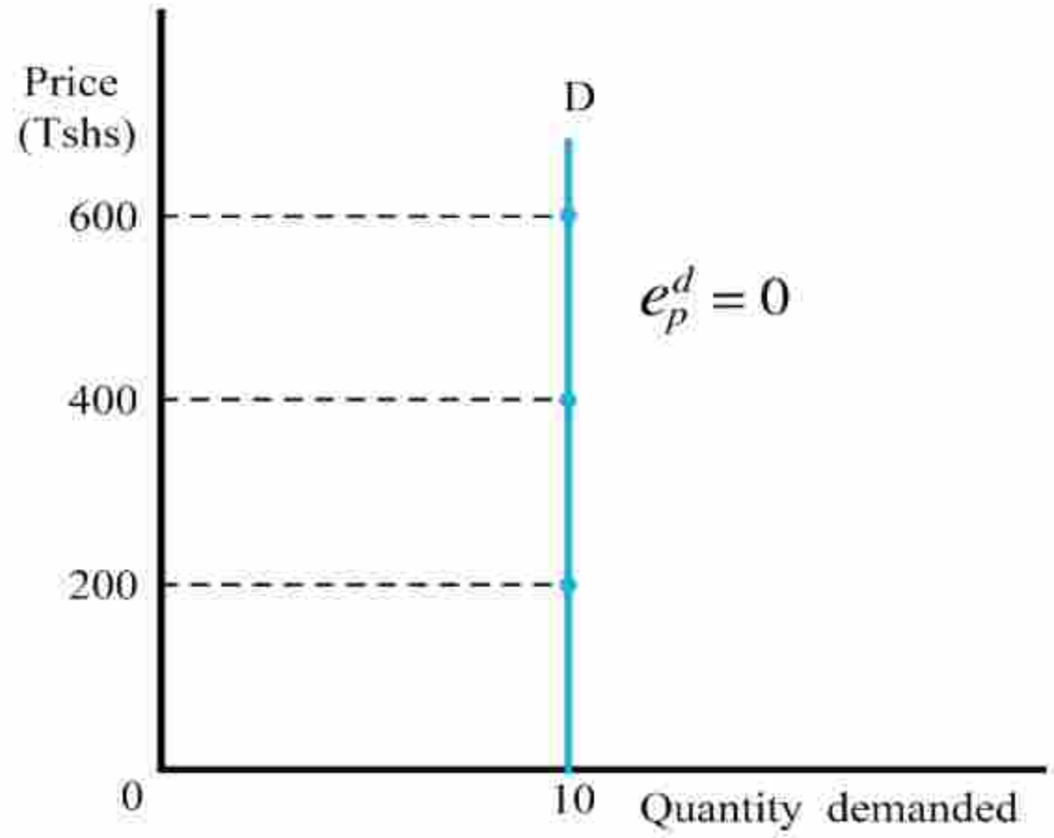


Figure 2.12: *Perfectly inelastic demand curve*

This demand curve is a vertical straight line parallel to the y-axis, implying that the own-price elasticity of demand is zero.

(c) Unitary elastic demand $\left(e_p^d = 1\right)$

The demand for a good is termed as unitary elastic if a given proportionate change in the own-price causes an equal proportionate change in the quantity demanded. As Figure 2.13 shows, when price decreases from Tshs 300 to Tshs 150, the quantity demanded increases from 100 to 150 units. Thus, proportionate changes in the price and quantity demanded of the good are the same.

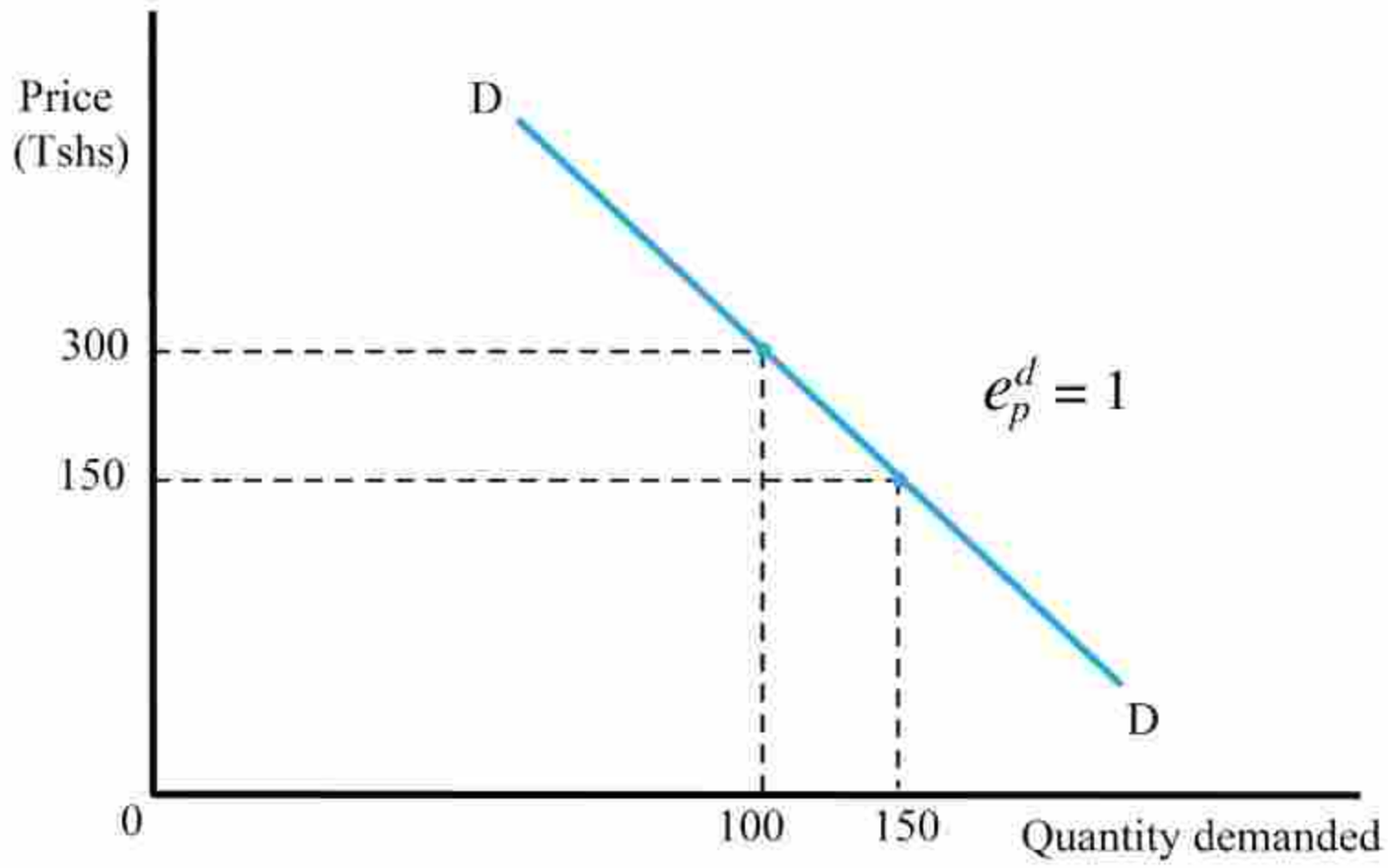


Figure 2.13: *Unitary elastic demand curve*

(d) Relatively elastic demand $\left(e_p^d > 1\right)$

The demand for a good is termed as relatively elastic if a given proportionate change in the own-price of a good leads to greater than proportionate change in the quantity demanded of a good, which is sometimes termed elastic demand. To exemplify elastic demand, Figure 2.14 shows that when the own-price of the good decreases from Tshs 600 to Tshs 400, the quantity demanded of the good increases from 10 kilograms to 18 kilograms. The percentage change in the quantity demanded is greater than the percentage change in the price of the good. The disproportionate changes in price and quantity demanded of a good lead to a relatively flat demand curve, as shown in Figure 2.14.

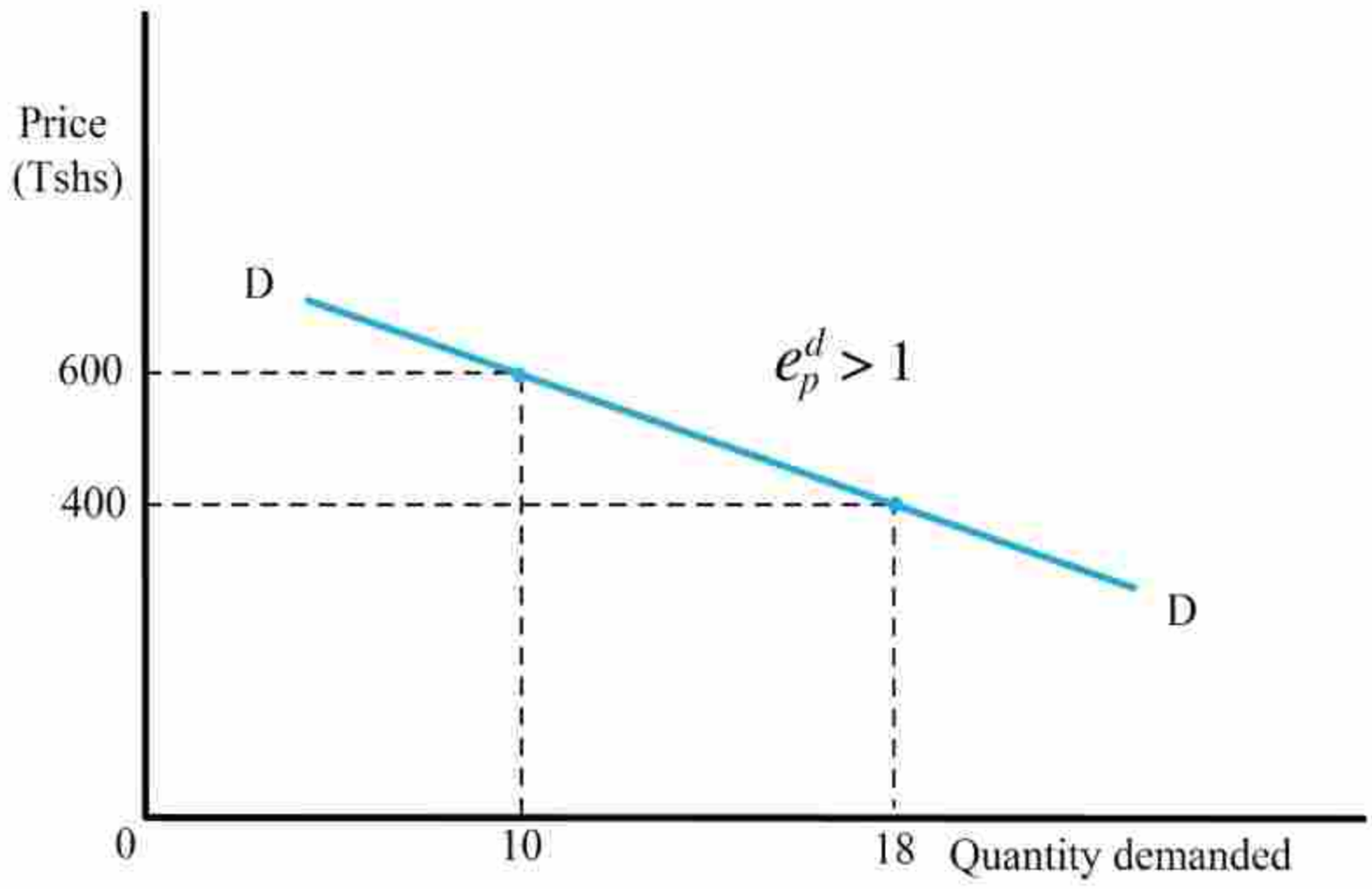


Figure 2.14: *Elastic demand curve*

(e) Inelastic demand $\left(e_p^d < 1\right)$

The demand for a good is termed inelastic if a given proportionate change in its price leads to a less than proportionate change in its quantity demanded. Figure 2.15 depicts an inelastic demand curve. When price of the good decreases from Tshs 900 to Tshs 200, the quantity demanded increases from 10 to 12 units. The percentage change in quantity demanded is less than the percentage change in the price of the good.

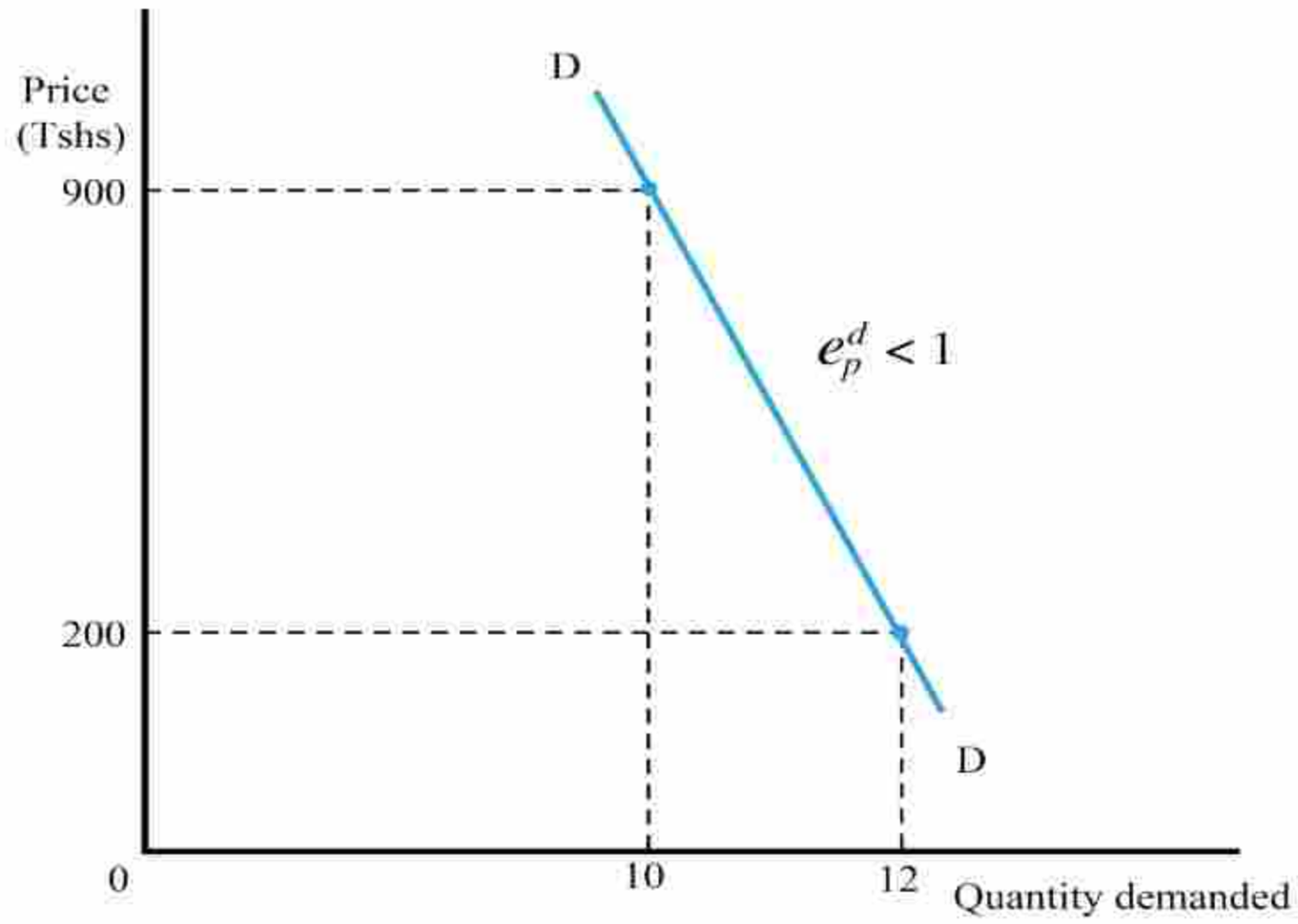


Figure 2.15: *Inelastic demand curve*

Example 2.1:

When price of 1 kilogram of beef was Tshs 6,000 in Arusha Municipality, the quantity demanded per day was 100,000 kilograms. Suddenly, the price of beef per kilogram increased to Tshs 8,000 due to increase in exportation of cattle to the United Arab Emirates (UAE). The increase in the price of beef caused the quantity demanded in Arusha Municipality to decrease from 100,000 to 80,000 kilograms per day. Use the information provided to calculate:

(a) The price elasticity of demand for beef and give its economic interpretation.

(b) Point and arc price elasticity of demand for beef and give economic interpretations of the answers.

Solution

Given:

Initial price = Tshs 6,000

Initial quantity= 100,000 kilograms

New price= Tshs 8,000

New quantity= 80,000 kilograms

(a) Recall, the formula for calculating the price elasticity of demand as follows:

$$e_p^d = -\left(\frac{\%\Delta QD}{\%\Delta P} \times \frac{P_0}{Q_0^d}\right)$$

$$e_p^d = -\left(\frac{80,000-100,000}{8,000-6,000} \times \frac{6,000}{100,000}\right)$$

$$e_p^d = -\left(\frac{-20,000}{2,000} \times \frac{6,000}{100,000}\right)$$

$$e_p^d = -\left(\frac{-120,000,000}{200,000,000}\right)$$

$$e_p^d = 0.6$$

The price elasticity of demand is 0.6.

The interpretation is that a one percent change in the price of beef leads to 0.6 percent change in quantity demanded of beef. The magnitude of the price elasticity of demand suggests that the demand for beef in Arusha Municipality is inelastic, that is, a proportionate increase in the price of beef leads to less than a proportionate decrease in the quantity demanded of beef. Also, it means that change in quantity demanded is less than the change in price.

(b) Recall, the formula for calculating the arc price elasticity of demand as follows:

$$e_p^d = -\left(\frac{\Delta QD}{\Delta P}\right) \times \left(\frac{P_1 + P_0}{Q_1^d + Q_0^d}\right)$$

$$\Delta Q = 80,000 - 100,000 = -20,000$$

$$\Delta P = 8,000 - 6,000 = 2,000$$

$$e_p^d = -\left(\frac{-20,000}{2,000} \times \frac{8,000+6,000}{80,000+100,000}\right)$$

$$e_p^d = 10 \times 0.077$$

$$e_p^d = 0.77$$

The arc price elasticity of demand is 0.77.

The interpretation is that a one percent change in the price of beef leads to 0.77 percent change in quantity demanded of beef. The demand for beef is inelastic, that is, a proportionate increase in the price of beef leads to less than a proportionate decrease in the quantity demand of beef. Also, it means that change in quantity demanded is less than the change in price.

Example 2.2:

Calculate price elasticity of demand given that the change in the price of beef by 50% led to a decrease in the quantity demanded of beef by 80%. Interpret your answer.

Solution

$$\text{Price elasticity of demand } \left(e_p^d\right) = \frac{\text{Percentage change in quantity demanded}}{\text{Percentage change in price}}$$

$$e_p^d = \frac{\%\Delta QD}{\%\Delta P}$$

$$e_p^d = \frac{80\%}{50\%}$$

$$= \frac{80}{100} \div \frac{50}{100}$$

$$= \frac{80}{100} \times \frac{100}{50} = \frac{80}{50}$$

$$e_p^d = 1.6$$

The price elasticity of demand for beef is 1.6, which means that, a one percent change in the price of beef has led to 1.6 percent change in its quantity demanded. This means that the demand for beef is elastic. That is, a percentage change in the price of beef has led to a larger percentage change in its quantity demanded.

Example 2.3:

Calculate and interpret the price elasticity of demand given the following demand equation for good X.

$Q_x^d = 20 - 4P_x$ and $P_x = 2$

Where, Q_x^d is the quantity of good X bought

P_x is the price of good X

Solution

$$e_p^d = -\left(\frac{\Delta QD}{\Delta P} \times \frac{P_0}{Q_0^d}\right)$$

From the demand equation $Q_x^d = 20 - 4P_x$,

$$\frac{\partial QD}{\partial P} = -4$$

Thus, $Q_x^d = 20 - 4(2) = 12$

$$e_p^d = -\left(-4 \times \frac{2}{12}\right) = 0.667$$

Therefore, a percentage change in price of the good leads to 0.667 percentage change in the quantity demanded of good X. This magnitude indicates that the demand for good X is inelastic.

Determinants of price elasticity of demand

The price elasticity of demand is determined by the following factors:

Nature of commodity: A necessary good has few substitutes (after all, it is a necessity), for example, food, whose demand is inelastic. A luxury good has many substitutes, for example, diamonds, gold and Tanzanite are substitutes. Hence, the demand for a luxury good is elastic.

Goods having several uses: Goods with alternative uses such as coal, electricity, water and milk have an elastic demand for a decrease in price, but inelastic demand for the rise in price. For example, from milk, one can produce cheese and yoghurt. Thus, the demand for milk will be elastic when its price decreases and inelastic when its price increases.

Level of income: The demand for goods by low-income earners is elastic to price changes, while that for high-income earners is inelastic. The reason for this is that more often people with high incomes can still purchase the same quantity even when the price of the good increases. On the other hand, people with low incomes are more likely to reduce the quantity demanded when the price rises.

Time horizon: With the passage of time, consumers can adjust to price changes and find new substitutes. Hence, as long as longer periods bring about change in the price of a good, the more elastic the demand for the good will be.

Deferred consumption: Goods and services whose consumption can be deferred are demand-elastic, for example, the demand for durable consumer goods, such as clothes, air conditioners, vehicles and buildings. If the price of such a good rises, people will postpone their consumption.

Proportion of income spent on the commodity: This is associated with the following statement: "The greater the proportion of income spent on a good, the larger the impact of a rise in its price on the quantities that people can afford to buy." It implies that a large proportion of income spent on a good invokes elastic demand.

Importance of price elasticity of demand

The importance of the price elasticity of demand is two-fold. One, it is useful for developing a price policy of a firm. If the demand of the product produced by the firm is price-elastic, the firm would find it advantageous to increase price. Alternatively, the firm can reduce price to increase revenue from sale of goods, whose demand is price-elastic. Two, elasticity of demand is important for designing tax and foreign trade policies. On one hand, the government can impose a tax on a product, whose demand is price-inelastic to raise revenue. On the other hand, can impose export tax on goods and services with inelastic demand abroad, that is, to the rest of the world. Or else, in order to raise tax revenue, the government can effectively impose tax on imports with inelastic demand in the domestic economy.

Income elasticity of demand

Income elasticity of demand measures the degree of responsiveness of the quantity demanded of a commodity to a change in the consumer's income. It is calculated as follows:

$$\text{Income elasticity of demand } \left(e_y^d \right) = \frac{\text{Percentage change in quantity demanded}}{\text{Percentage change in income}}$$

$$e_y^d = \frac{\%\Delta QD}{\%\Delta Y} \quad \ldots\ldots\ldots\

When $e_y^d = 1$, the income elasticity of demand is unitary elastic, implying a one percent increase in income leads to a one percent increase in quantity demanded.

When $e_y^d < 1$, demand for a good is income-inelastic. Hence, if income increases by one percent, demand for a good increases by less than one percent.

Interpretation of the income elasticity of demand

In calculating the income elasticity of demand, the coefficient of elasticity can be negative, positive or zero. The magnitude of the income elasticity of demand has different interpretations, which are consistent with following demands:

(a) When income elasticity of demand is negative

Goods with negative coefficient of income elasticity of demand are known as **Inferior goods**. The negative sign means that an increase in a consumer's income leads to a decrease in the quantity demanded of the good, which implies a negative relationship between income and quantity demanded of the good. Some goods are inferior but the coefficient of income elasticity is less than negative one and such goods are termed as **Giffen goods**. Many inferior goods depend on cultural habits, for example, some people categorise sardines (*dagaa*) as an inferior good. Figure 2.16 shows that when a consumer's income increases from Y_0 to Y_1, it leads to a decrease in the quantity demanded of the good from Q_0 to Q_1. This depicts the good as an inferior one, whose income elasticity of demand is negative.

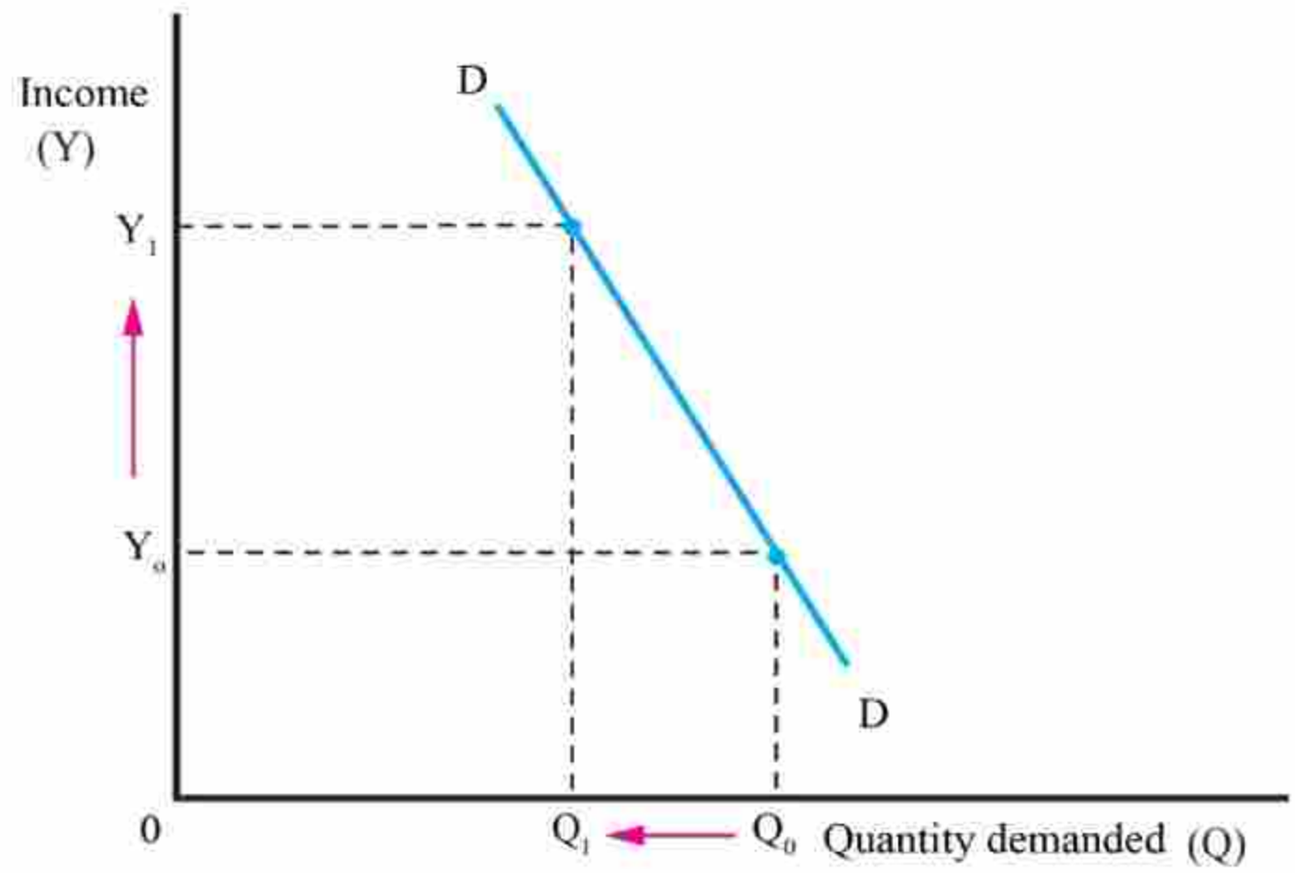


Figure 2.16: *Negative income elasticity of demand curve*

(b) When income elasticity of demand is positive

Goods with positive income elasticity of demand are termed as **normal goods,** which fall in two categories. Goods with income elasticity of demand less than one are called **necessities** and those goods with income elasticity of demand greater than one are called **luxury goods**. The income elasticity of demand may depend on income level. The income elasticity of demand for basic goods is likely to be higher for people with low incomes because an increase in their incomes gives them a chance to increase consumption of those goods. The income elasticity of demand for basic goods is likely to be higher for people with high incomes because as their incomes increase, they will spend more on non-basic goods, having satisfied the needs for basic goods. Figure 2.17 shows the effects of an increase in a consumer's income from Y_0 to Y_1,which leads to an increase in the quantity demanded of a good from Q_0 to Q_1, indicating that the income elasticity of demand for good X is positive.

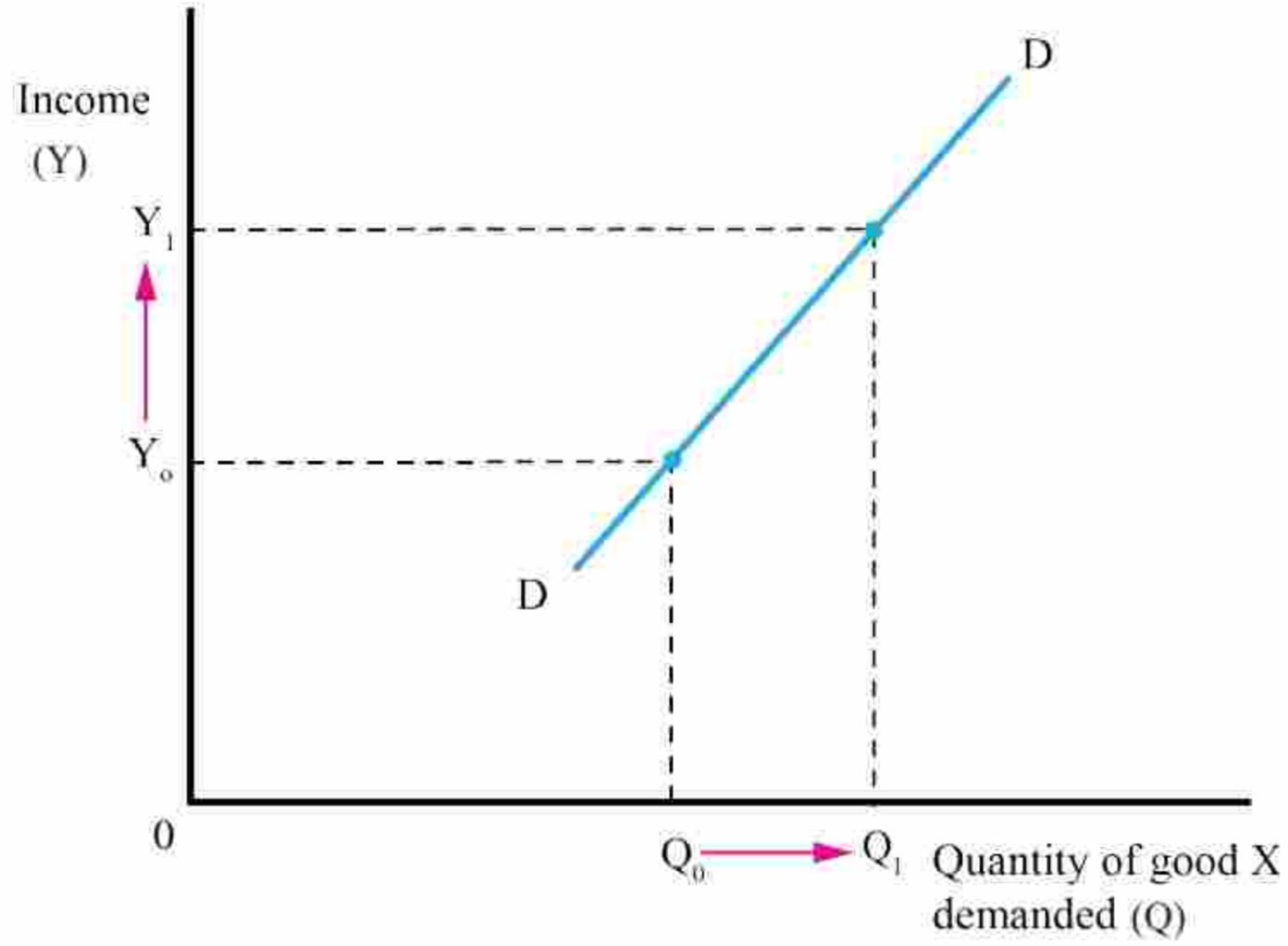


Figure 2.17: *Positive income elasticity of demand curve*

(c) When income elasticity of demand is zero

For some specific goods, the change in consumer's incomes has no effect on the quantity demanded. The goods whose demand is income-inelastic, include essential goods, for example, salt. The increase or decrease in income does not affect the quantity demanded of salt that a consumer buys. Figure 2.18 shows that there is no effect on the quantity demanded of the good when

income increases from Y_0 to Y_1 and then to Y_2, that is, the quantity demanded of the good is the same at different levels of the consumer's income. Hence, the demand curve D depicts zero income elasticity of demand.

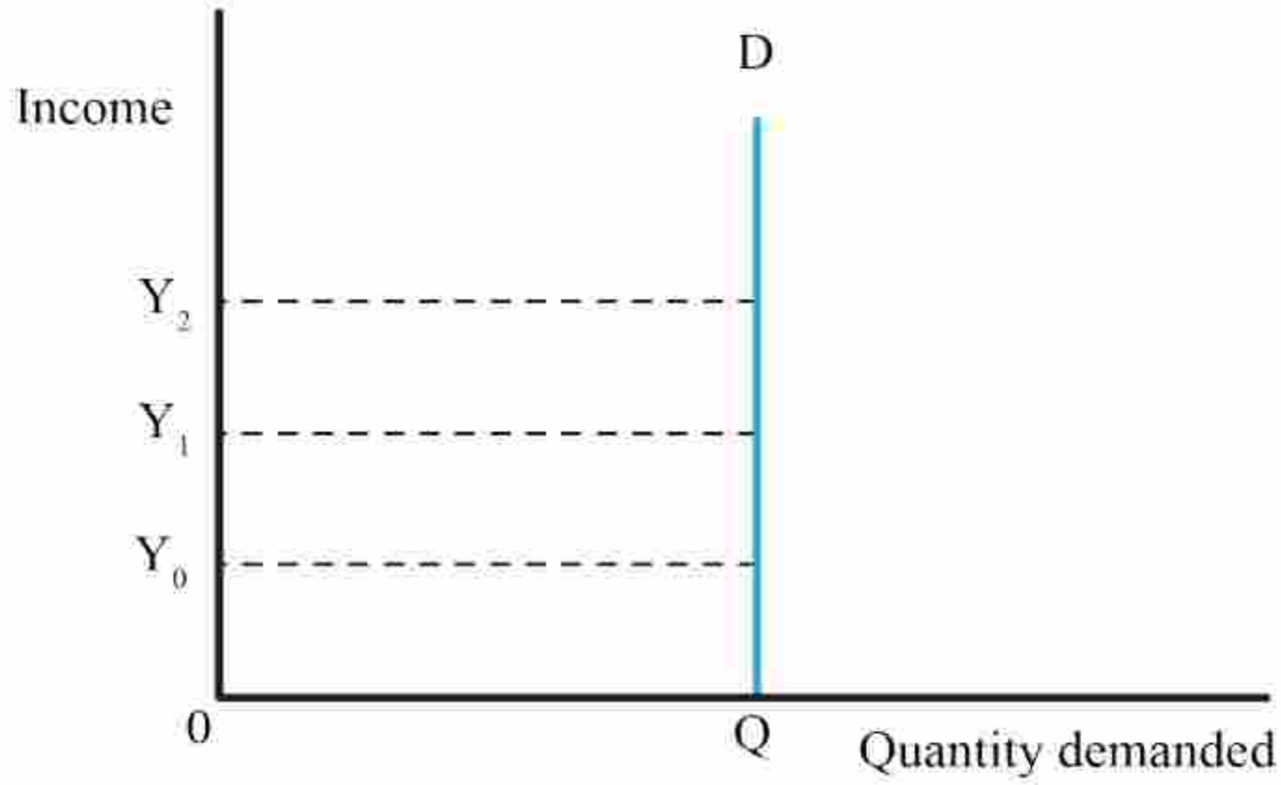


Figure 2.18: *Zero income elasticity of demand curve*

Example 2.4:

The consumer's income increased from Tshs 100,000 to Tshs 200,000 per month. At the initial income level, the consumer was consuming 40 units of commodity X, whereas after the increase in income, the consumer's consumption decreased to 20 units per month. Calculate the income elasticity of demand, give its economic interpretation and comment on the type of good.

Solution

Given;

$Y_0 = \text{Tshs}\,100,000$

$Y_1 = \text{Tshs}\,200,000$

$Q_0 = 40\,\text{units}$

$Q_1 = 20\,\text{units}$

Recall, the formula for calculating income elasticity of demand

$$e_y^d = \left(\frac{Q_1^d - Q_0^d}{Y_1 - Y_0}\right) \times \left(\frac{Y_0}{Q_0}\right) = \frac{\Delta QD}{\Delta Y} \times \frac{Y_0}{Q_0}$$

$$e_y^d = \frac{20-40}{200,000-100,000} \times \frac{100,000}{40}$$

$$e_y^d = \frac{-20}{100,000} \times \frac{100,000}{40}$$

$$e_y^d = \frac{-2,000,000}{4,000,000}$$

$$e_y^d = -0.5$$

The income elasticity of demand is – 0.5.

The interpretation is that a one percent increase in income leads to 0.5 percent decrease in the quantity demanded of a good. Since e_y^d is less than 0, this is the case of an inferior good because when the consumer's income increased, the quantity demanded of a good decreased.

Importance of income elasticity of demand

In practice, the income elasticity of demand is useful in an economy in the following ways:

(a) The income elasticity of demand is useful for decision-making of both business firms and industries. Firms producing products whose income elasticity of demand is high will benefit from the level of overall economic activity or business cycle. For example, during the boom period, the demand for luxuries increases tremendously and decreases sharply during recession.

(b) The income elasticity of demand plays a role in formulating market strategies of firms. For instance, firms producing products whose income elasticity of demand is high will be located in areas where people's incomes are high and probably are increasing rapidly.

(c) The income elasticity of demand helps tax authorities to distinguish between goods that are essential or necessities (with low-income elasticity of demand or are income inelastic) and the non-essentials (whose income elasticity of demand is high are income elastic goods). This will further help in taxing the commodities depending on the objective. For example, if the government's objective was to improve welfare of consumers, essential goods would be taxed less than luxurious goods. However, if the objective was to raise government revenue, increasing tax rates would be on income-inelastic goods.

Cross-price elasticity of demand

The cross-price elasticity of demand shows the degree of responsiveness of the quantity demanded of one good to change in the price of a related good. In this regard, the cross-price elasticity of demand measures the percentage change in the quantity demanded for one good, X, over the percentage change in the price of a related good, Y. Cross elasticity of demand is calculated as:

$$\text{Cross-price elasticity of demand}\left(e_c^d\right)=\frac{\text{Percentage change in quantity demanded of good X}}{\text{Percentage change in price of good Y}}$$

$$e_c^d=\frac{\%\Delta Q_X}{\%\Delta P_Y} \quad \text{.....(2.14)}$$

$$e_c^d=\frac{\Delta Q_X}{\Delta P_Y}\times\frac{P_Y}{Q_X}$$

$$e_c^d=\frac{Q_1^X-Q_0^X}{P_1^Y-P_0^Y}\times\frac{P_0^Y}{Q_0^X} \quad \text{.....(2.15)}$$

Where;

e_c^d is cross elasticity of demand

P_0^Y is initial price of good Y

P_1^Y is new price of good Y

Q_0^X is initial quantity of good X

Q_1^X is new quantity of good X

Interpretation of the cross-price elasticity of demand

In calculating the cross-price elasticity of demand, the coefficient of elasticity can be positive, negative or zero. These magnitudes of cross-price elasticity of demand have different interpretations, which are consistent with the following demands:

(a) When cross-price elasticity of demand is positive

The cross-price elasticity of demand is positive if the increase in price of good X leads to an increase in the quantity demanded for good Y. In this case, the goods are termed as substitutes or considered competitively demanded. Figure 2.19 depicts a positive cross-price elasticity of demand. The figure shows that the price of good X has increased, which has led to the quantity

demanded of good Y to increase and shows a positive relationship between the price of good X and the quantity demanded of good Y. Examples of substitutes include beef and chicken, Coca Cola and Pepsi, as well as Colgate and Whitedent toothpastes.

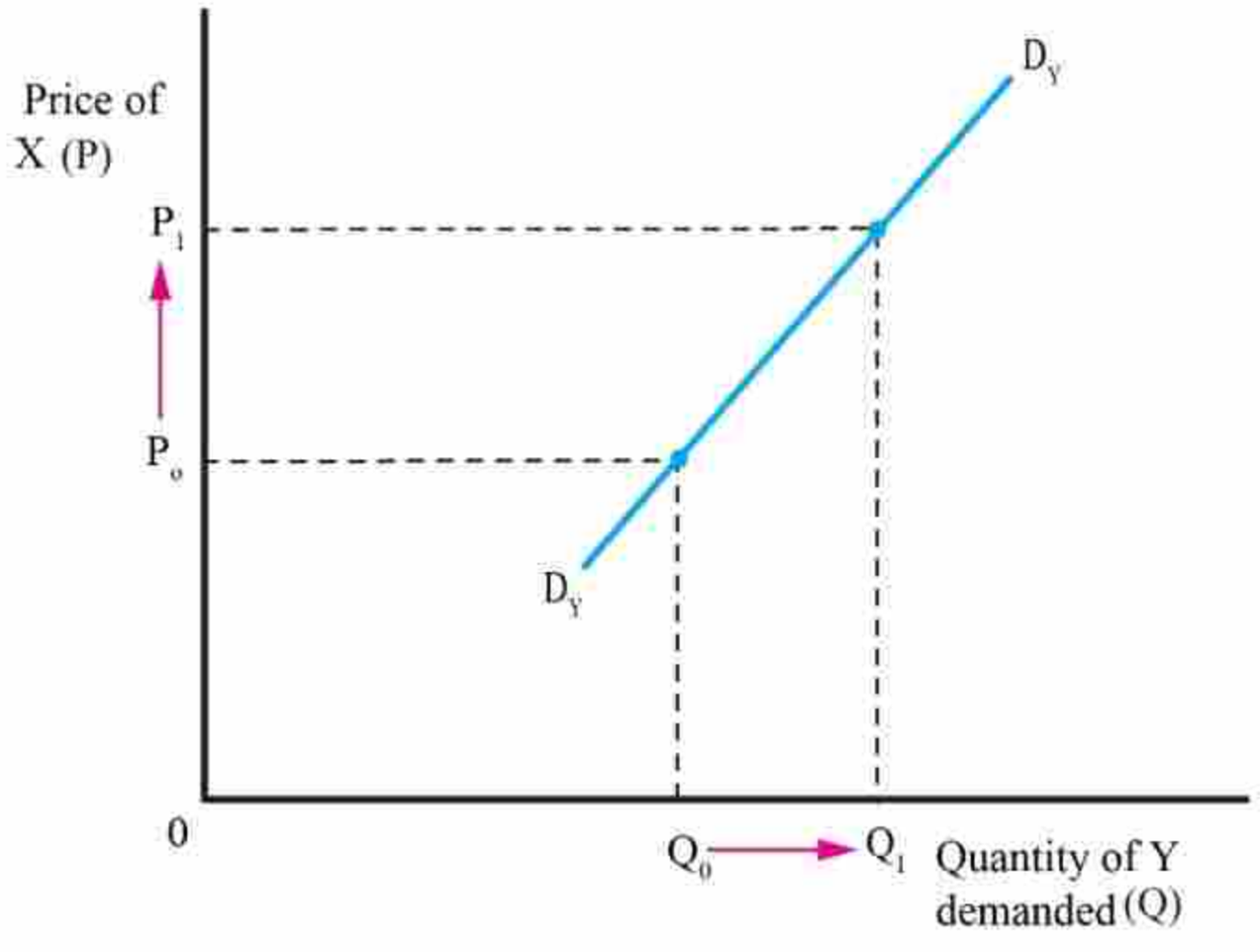


Figure 2.19: *Positive cross-price elasticity of demand*

(b) When cross-price elasticity of demand is negative

Cross-price elasticity of demand is negative if an increase in the price of good X leads to a decrease in the quantity demanded of good Y. Hence, the cross-price elasticity of demand for good Y with respect to the price of X is negative. Such goods are termed complements or jointly demanded goods. Examples of complements include cars and petrol, golf clubs and golf balls, and tea and sugar. Figure 2.20 depicts a negative cross-price elasticity of demand of good Y with respect to the price of good X. The increase in the price of good X leads to the quantity demanded of good X to decrease. Now, because good X and good Y are complementary in consumption, the quantity demanded of good Y will also decrease. Hence, the effect of an increase in the price of good X is a decrease in quantity demanded of good Y, which indicates a negative relationship between the price of good X and the quantity demanded of good Y.

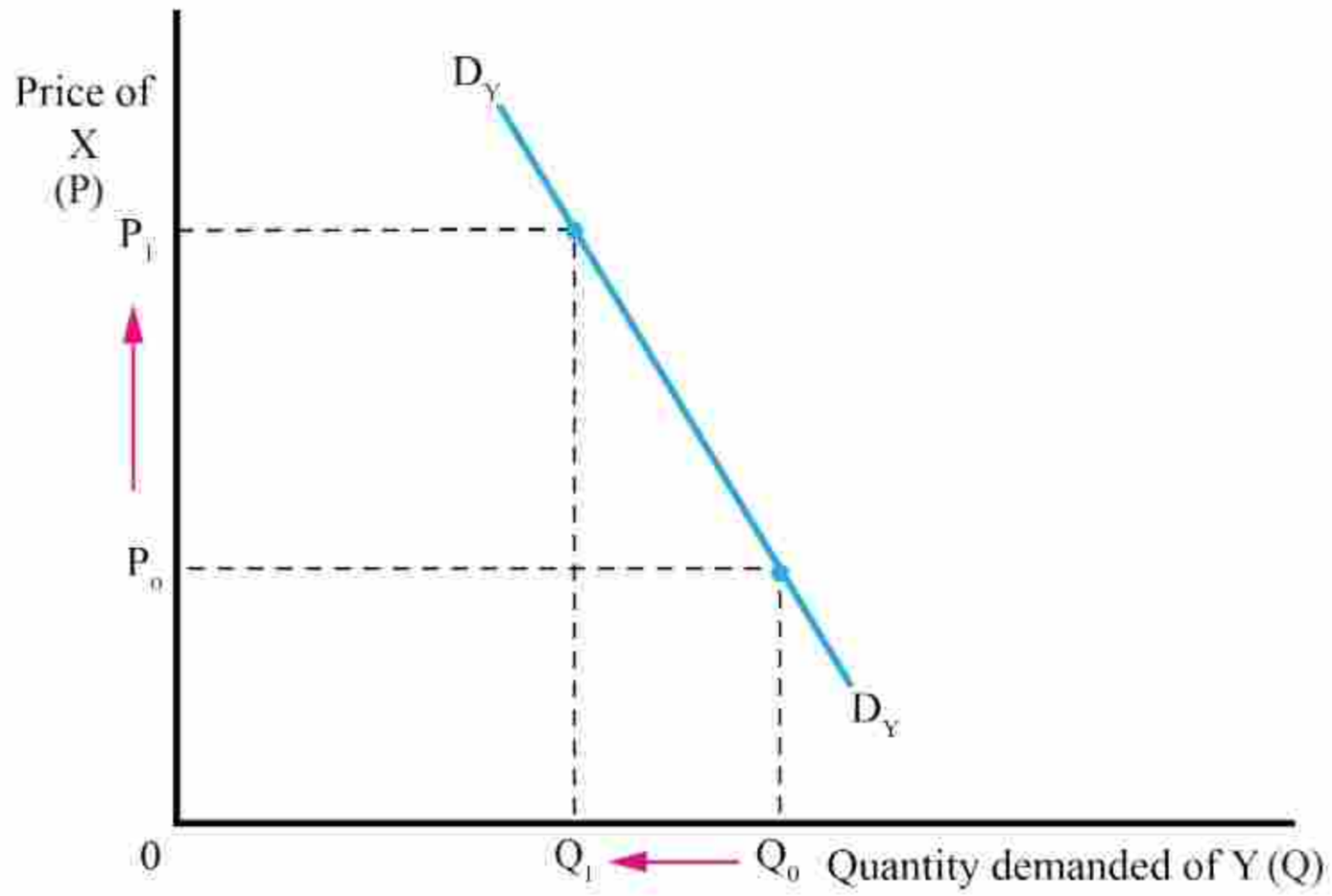


Figure 2.20: *Negative cross-price elasticity of demand curve*

(c) When cross-price elasticity of demand is zero

Cross-price elasticity of demand is zero if the change in price of a good does not lead to any change in demand of another good. In such a situation, the two goods are unrelated in consumption. Thus, a one percentage increase in the price of good X will lead to a zero change in the quantity demanded of good Y.

Figure 2.21 depicts zero cross-price elasticity of demand. A change in the price of good X does not affect the quantity demanded of good Y since consumption of two goods is unrelated. For example, cars and clothes, and computers and cement.

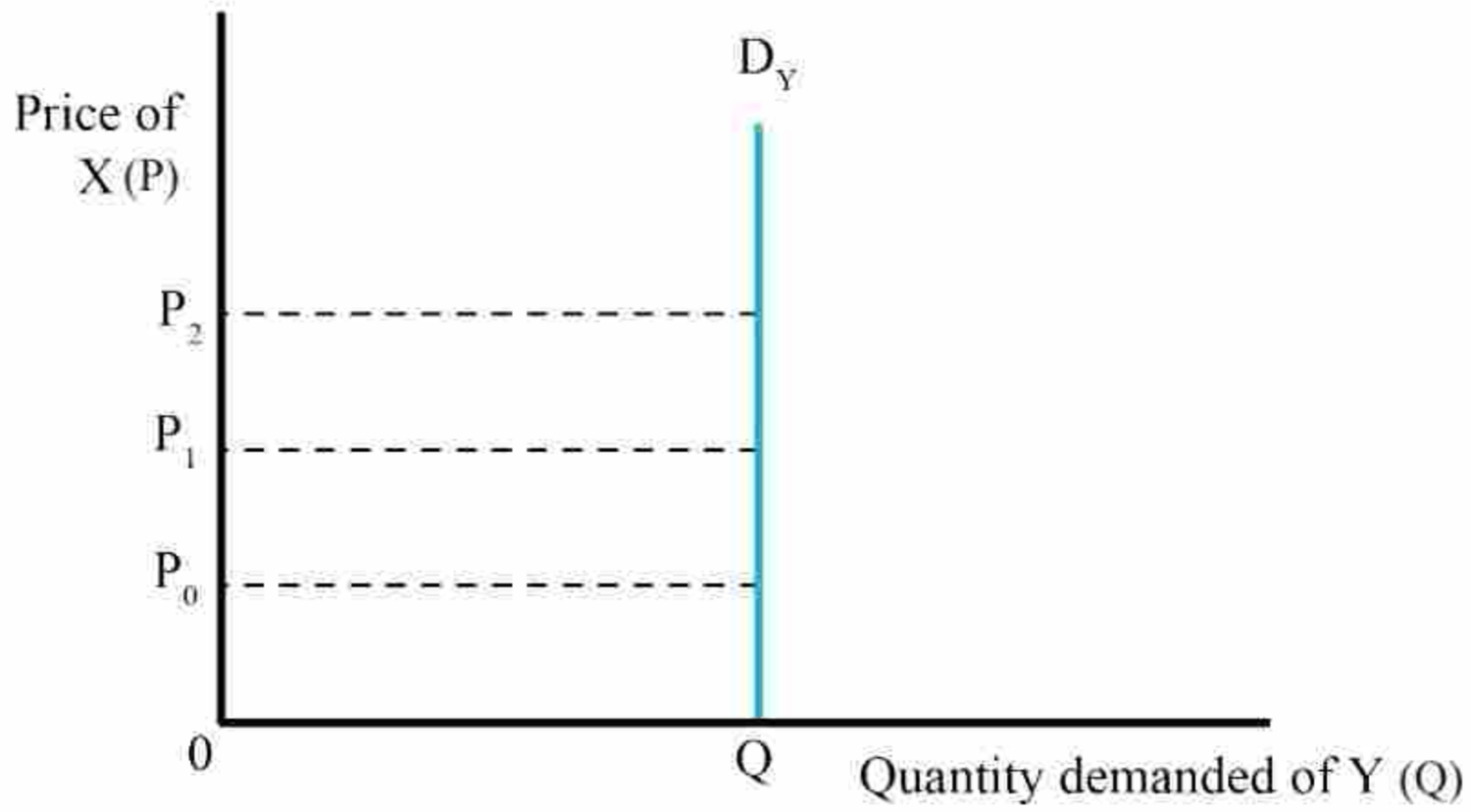


Figure 2.21: *Zero cross-price elasticity of demand curve*

Example 2.5:

Price of good X in Mererani town in July 2021 changed from Tshs 1,000 per unit to Tshs 800 per unit and the quantity demanded of the related good Y changed from 200 units to 180 units.

(a) What type of elasticity of demand is this? Why?
(b) Calculate the elasticity of demand.
(c) Give economic interpretation of your answer in (b) above.
(d) Give one example of commodities X and Y.

Solution:

(a) This is cross elasticity of demand because there are two interrelated goods (X and Y) for which a change in the price of one good affects the quantity demanded of the other good.

(b) Given:

$$P_0^X = \text{Tshs } 1,000$$

$$P_1^X = \text{Tshs } 800$$

$$Q_0^Y = 200 \text{ units}$$

$$Q_1^Y = 180 \text{ units}$$

$$e_c^d = \frac{Q_1^Y - Q_0^Y}{P_1^X - P_0^X} \times \frac{P_0^X}{Q_0^Y}$$

$$e_c^d = \frac{180 - 200}{800 - 1,000} \times \frac{1000}{200}$$

$$e_c^d = \frac{-20}{-200} \times \frac{1,000}{200}$$

$$= \frac{20,000}{40,000}$$

$$e_c^d = 0.5$$

(c) Goods X and Y are substitutes because the cross-price elasticity of demand is positive. A decrease in price of good X leads to the quantity demanded of good Y to decrease. This situation is illustrated in Figure 2.19.

(d) Examples of goods X and Y could be Pepsi and Coca Cola, rice and maize flour.

Importance of cross-price elasticity of demand

In practice, the cross-price elasticity of demand is useful in an economy in the following ways:

(a) In managerial decisions, the cross-price elasticity of demand is used for formulating a proper pricing strategy for appropriate prices of various firm's products. For example, a company produces both smartphones and sim-cards, which are complementary goods with high cross-price elasticity of demand such that if it decides to reduce the price of smartphones, it will increase the demand for sim-cards, as more people will purchase smartphones.

(b) The cross-price elasticity of demand is useful in defining boundaries of industry and measuring interrelationship between industries. For instance, a firm cannot increase the price of its product without losing sales to another firm for products within the same industry with high positive cross-price elasticity of demand. For example, cross-price elasticity of demand between Samsung, Apple and TECNO smartphones is positive because these firms belong to the same industry (telecommunications industry). As a result, if one firm increases the prices of its products, consumers will shift to its substitutes produced by other firms.

Applications of elasticity of demand

Elasticity of demand has the following application:

(a) In order to decide whether price of a good should be increased or reduced one needs to be aware of the nature of elasticity of demand for the good. To maximise revenue, the price should be reduced if the good is own-price elastic and price of the good should be increased if demand is price-inelastic.

(b) Monopolists apply the concept of elasticity of demand in adopting price discrimination. Price discrimination is an act of charging different prices for the same commodity to different buyers. Price discrimination is possible when price elasticity of demand for a commodity in the two markets is different.

(c) The government or other stakeholders can use the concept of elasticity of demand to assess effects of currency depreciation. Depreciation of the currency makes domestic goods to be cheaper in the world market thereby promoting exports. If demand for the country's exports is elastic, depreciation will have a positive effect on the balance of payments through increasing exports, assuming there is no effect on the value of imports.

(d) It helps the government to generate tax revenue, when it targets to raise taxes on the goods whose demand is inelastic, such as cigarettes, alcohol and fuel.

(e) The elasticity of demand helps the government to determine the tax incidence. Tax incidence involves distribution of the tax burden between buyers and sellers. The tax burden depends on the degree of elasticity of demand:

(i) If demand is elastic, an increase in price will lead to a large percentage decrease in the quantity demanded. Hence, the producer would pay a large proportion of the tax burden because they hardly shift a large part of tax to consumers because of informed increase in prices. The tax will be more effective in reducing demand than in raising revenue for the government.

(ii) If demand is inelastic, a higher tax rate will lead to a smaller decrease in demand. In this regard, imposing tax on a good with inelastic demand implies that a large proportion of the tax burden will be borne by the consumers in the form of increase in prices, which will lead to a significant increase in tax revenues.

(f) The concept of elasticity of demand is useful in determination of wages for a particular type of labour. For example, if the demand for labour is inelastic, employers are likely to increase the wages.

Exercise 2.2

1. When is the demand for a good price-elastic? Explain with an illustration.
2. Suppose a 40% increase of a consumer's income leads to an increase in the quantity demanded of good M by 20%.
   (a) Calculate the income elasticity of demand.
   (b) Name three examples of commodity M and give reasons for your choice of the goods.
3. Consider demand for a good. When its price is Tshs 400, the quantity demanded for the good is 2,500 units. Suppose price of the good increases to Tshs 500

and as a result, the quantity demanded for the good falls to 2,000 units. Calculate the price elasticity of demand and give the economic interpretation.

The concept of supply

Supply is the ability and willingness of sellers to produce and sell different quantities of goods in the market at a given price, in a given period. Quantity supplied is the amount of goods that sellers are able to sell in a market at an existing market price in a given period.

Determinants of supply

There are various factors that affect the supply of a good. The following are some of the important factors:

Prices of factors of production: If prices of factors of production increase, the cost of production of a firm will also increase. The increase in the cost of production will prompt a firm to produce and sell less at a given price. On the other hand, if the cost of production decreases, the firm may produce and sell more at the same price.

Level of technology: Use of technology that reduces the cost of production or raises productivity will lead to increase in supply. The reason for increase in supply of a good is that lower costs implies higher profit, which create an incentive for a firm to produce more of the product. Development of a new method of production or invention of a new machine may also reduce costs thereby leading to an increase in supply.

Expectations of future change in price: In case producers expect an increase in price of a good in the near future, they will withdraw goods from the market or delay production. Thus, the firm will reduce supply of the good, the intention is to sell larger output at a higher price in the future.

Change in price of other related goods: Holding other factors constant, change in price of a good can lead to change in quantity supplied of another interrelated good. For example, an increase in price of beef may lead to an increase in supply of leather and vice versa.

The law of supply

The law of supply states that the quantity supplied varies directly with the price of a good, other things remaining constant. It means that at a higher price, more of a good or service will be supplied and vice versa.

As is the case for the law of demand, the law of supply assumes that, all other factors affecting supply, other than the price of the good, remain unchanged. The assumptions of the law of supply include no change in the cost of production, constant technology, constant number of firms and constant price of other related goods.

Supply schedule

A supply schedule is a table that shows the relationship between the price of a good and its respective quantity supplied in a given period. For example, Table 2.3 shows the hypothetical individual supply schedule of beef. There are two types of supply schedules, namely, the individual supply schedule and the market supply schedule.

Individual supply schedule

Individual supply schedule is a table that shows the quantity supplied of the good to the market by one firm at each price level in a given period. Table 2.3 presents an example of a firm's supply schedule for beef. For example, when the price of beef is Tshs 1,000 per kilogram, the producer will supply 10 kilograms of beef. As the price rises, the quantity supplied increases.

Table 2.3: *Hypothetical individual supply schedule of beef*

Price of beef (Tshs per kilogram)	Quantity of beef supplied (Kg)
1,000	10
2,000	20
3,000	30
4,000	40
5,000	50

Market supply schedule

The market supply schedule is a table that shows quantities of a good supplied by all suppliers or producers in the market at different price levels in a given

period. Table 2.4 presents an example of a market supply schedule with two beef suppliers (A and B). The market supply schedule is a sum of individual producers' quantities supplied at a given price. Note that one does not sum up the prices, rather, he or she sums up the quantities of goods horizontally, that is, at the prevailing price.

Table 2.4: *Hypothetical market supply schedule for beef*

Price of beef (Tshs) per kg	Quantity of beef supplied by A (kg)	Quantity of beef supplied by B (kg)	Market supply of beef (A + B) (kg)
1,000	10	5	15
2,000	20	10	30
3,000	30	15	45
4,000	40	20	60
5,000	50	25	75

Supply curve

A supply curve is a graphical presentation of a supply schedule, which is an upward-sloping curve relating price of a good to the quantity supplied of that good. For example, Figure 2.22 presents a supply curve. When the price is Tshs 400, the quantity supplied is 10 kilograms. When price of the good increases to Tshs 1,200, the quantity supplied increases to 20 kilograms.

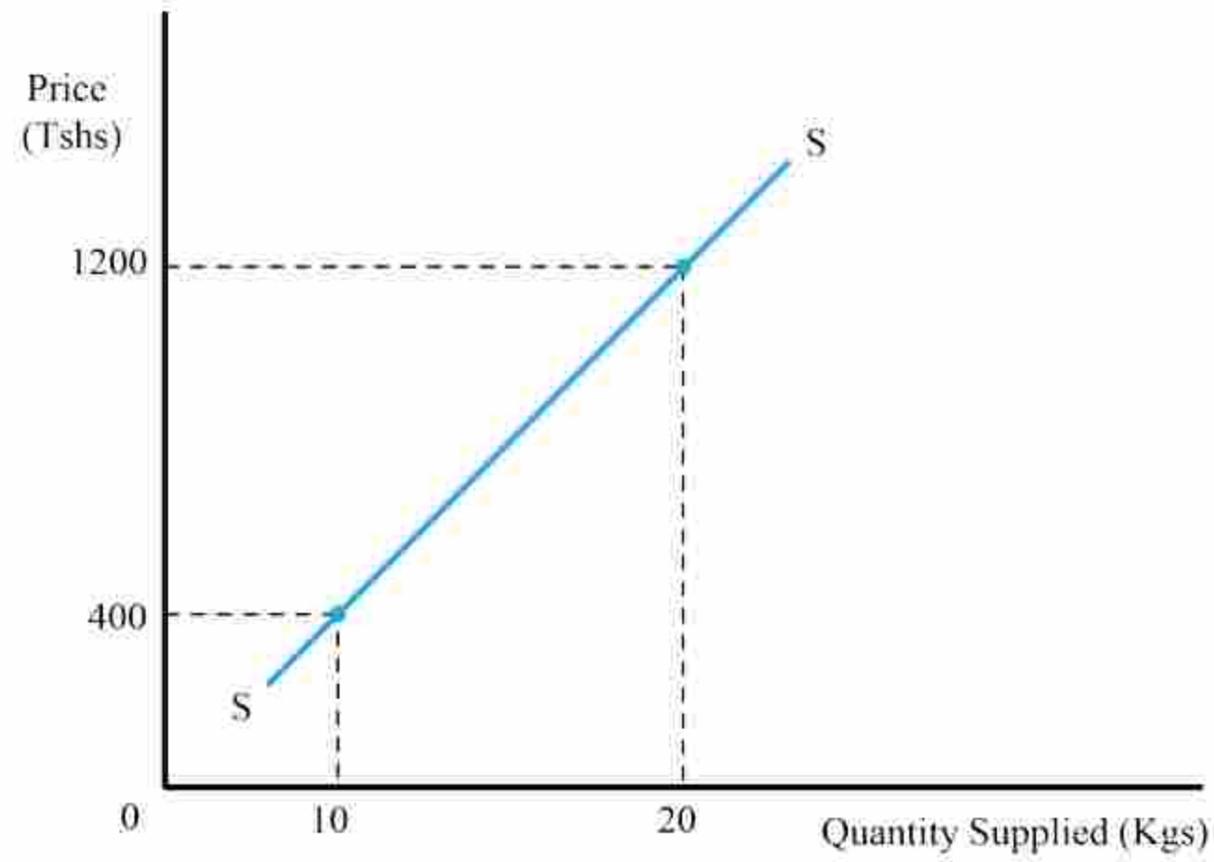


Figure 2.22: *Supply curve*

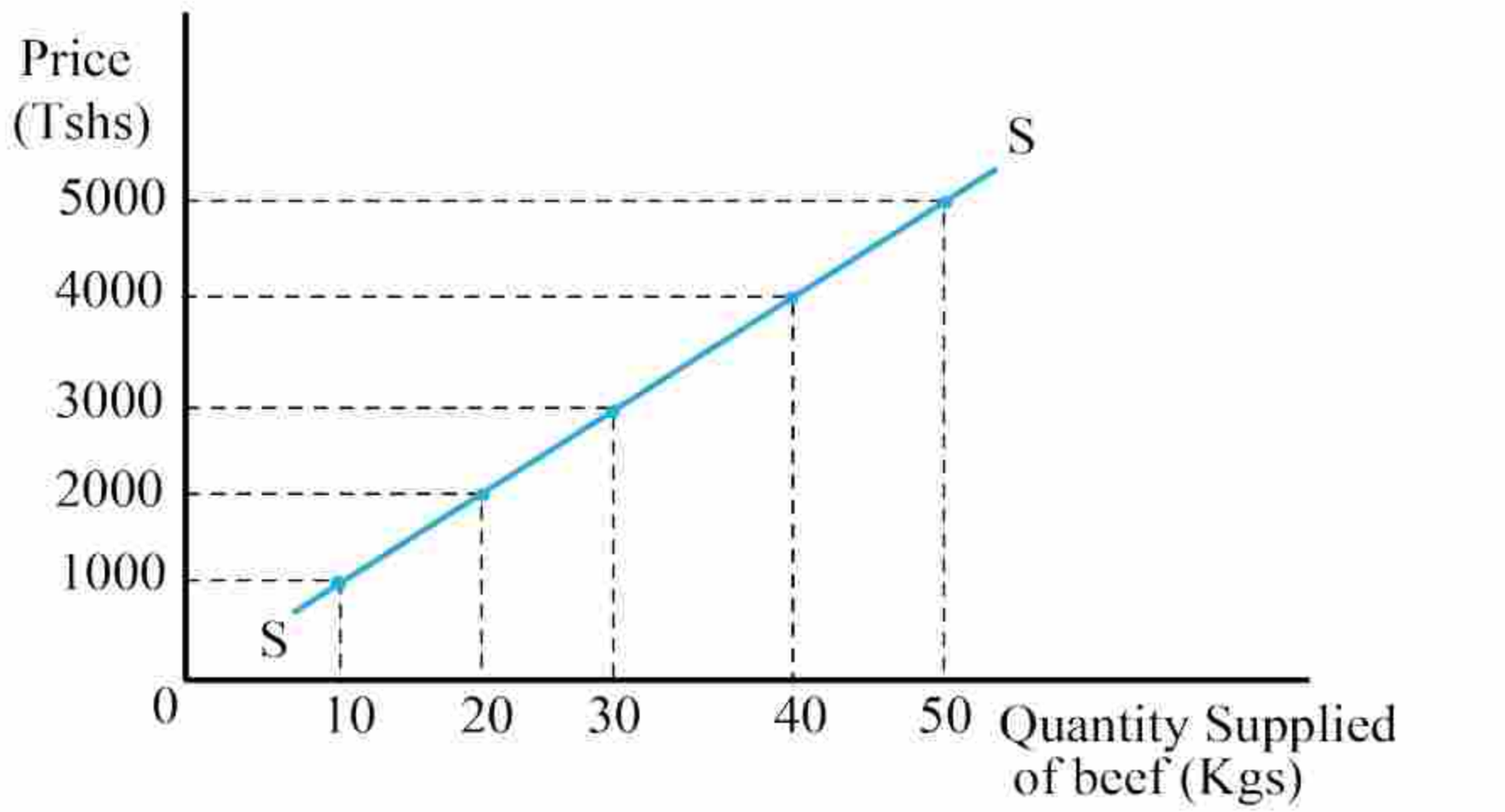
Individual and market supply curves

Figure 2.22 presents an individual supply curve for a good. However, to analyse how the market operates, one needs to determine the market supply curve. A market supply curve is a sum of all individual supply curves of a particular good or service at a given price, that is, the horizontal summation of individual supply curves. It indicates how the total quantity of a good supplied on the market will vary as the price of the good changes, while all the other factors that affect supply are held constant.

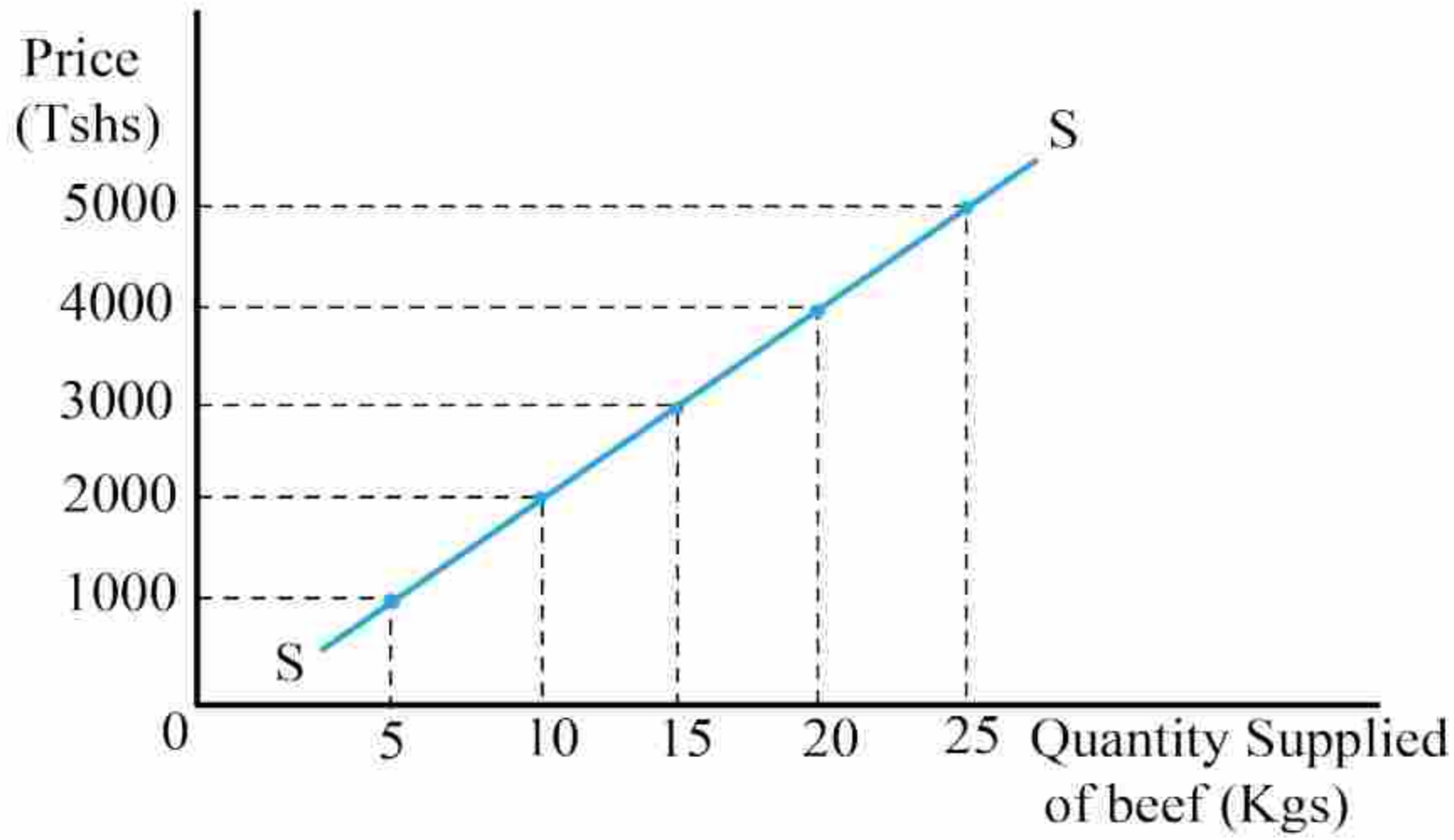
For example, to obtain the market supply curve for beef you need to have information on all individual supply curves. Figures 2.23 shows individual supply curves of beef (Kilograms) for producers A and B, using information from the market supply schedule in Table 2.4.

To obtain a market supply curve, one has to sum up horizontally the supply curves of producers A and B. Figure 2.23 represents the market supply curves of A and B, which is a sum of the two individual supply curves.

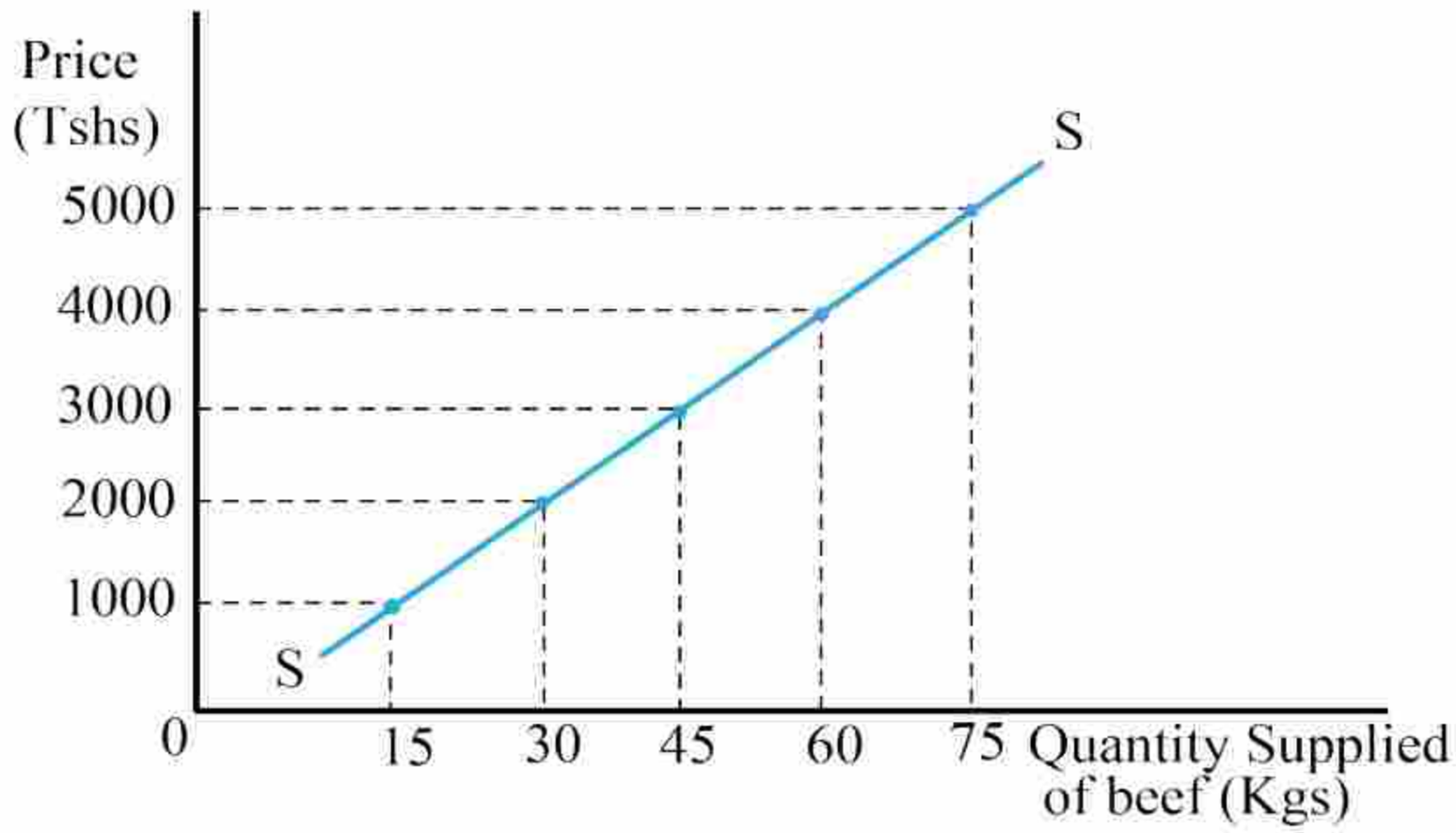
Individual A's supply + Individual B's supply = Market supply



Individual A's supply curve



Individual B's supply curve



Market supply curve

Figure 2.23: *Market supply curve for individuals A and B*

Why is supply curve upward sloping from left to right?

Supply curve slopes upward from left to right because there is a direct or positive relationship between the quantity supplied of good and its price. When price of a good rises, its quantity supplied increases and vice-versa. The main reasons for this producer's (or seller's) behaviour are as follows:

Profit: If the price of a good rises, producers generally increase production due to motivation of higher profit possibilities from the increase in the price of the good. Thus, because the increase in price of a good increases the quantity supplied, an upward sloping supply curve is obtained.

Change in stock: If price of a good increases, sellers are ready to supply more from their stock of goods. On the other hand, when the price of the good decreases, sellers would increase their stock to avoid losses thereby selling less of the good.

Entry or exit of firms: When price of a good increases, new firms enter into the industry with a view to earn profits, which in turn, increase the quantity of the good supplied. On the other hand, when the price of a good starts to decrease, marginal firms (or relatively less efficient firms) exit the market to avoid expected losses, which thereby decreases quantity supplied.

Supply function

A supply function is a mathematical expression of the relationship between quantities of the good supplied and factors affecting the supply, expressed as follows:

$$Q_x^s = f\left(P_x, P_0, P_e, C, T, U\right) \quad \text{.....(2.16)}$$

Where;

Q_x^s is quantity supplied of good X

f is a function

P_x is price of good X

P_0 is price of other goods

P_e is expected price

C is cost of production

T is level of technology

U is other factors

Assuming factors other than price of good (P_x) that affect the supply of good (X) remain constant, the supply function can be expressed as:

$$Q_x^s = f\left(P_x\right) \quad \text{.....(2.17)}$$

A supply function has a positive slope, which implies a positive relationship between price of a good and its quantity supplied. Normally, a supply function is expressed generally as:

$$Q_x^s = a + bP_x \quad \text{.....(2.18)}$$

Where;

'a' is the intercept of the curve

'b' is the slope of the supply function

Supply function is related to one good. An individual supply function expresses the relationship between the price of a good and its quantity that a firm sells, for example,

$$Q_x^s = 15 + 3P$$

The market supply function expresses the relationship between the price and the quantity supplied of a commodity by all suppliers in the market. For example, if a market consists of 10 identical suppliers each with $Q_x^s = 15 + 3P$ as a firm's supply function, then the market supply function will be given as $Q_x^s = (15 + 3P)10$ equal to $Q_x^s = 150 + 30P$.

Change in the quantity supplied and change in supply

The following explanations gives the difference between change in supply and change in the quantity supplied.

Change in the quantity supplied

A change in the quantity supplied is an increase or decrease in the quantity supplied in response to a change in the price of the good or service when other determinants of supply remain constant. When the quantity supplied changes due to a change in the price of a good, it is a case of expansion or contraction of supply whereby movements occur along the same supply curve.

Expansion of supply: This is an increase in quantity of the good supplied due to increase in price of the good. Figure 2.24 exemplifies expansion of supply with the increase in quantity supplied of a good from 15 units to 20 units due to an increase in the price of a good from Tshs 800 to Tshs 1,200 (movement from A to B).

Contraction of supply: It occurs when the quantity supplied of a good decreases due to a decrease in the price of the good. In Figure 2.24, it is depicted by a movement along the supply curve from point A to C whereby the quantity supplied decreases from 15 units to 10 units due to a decrease in price from Tshs 800 to Tshs 400.

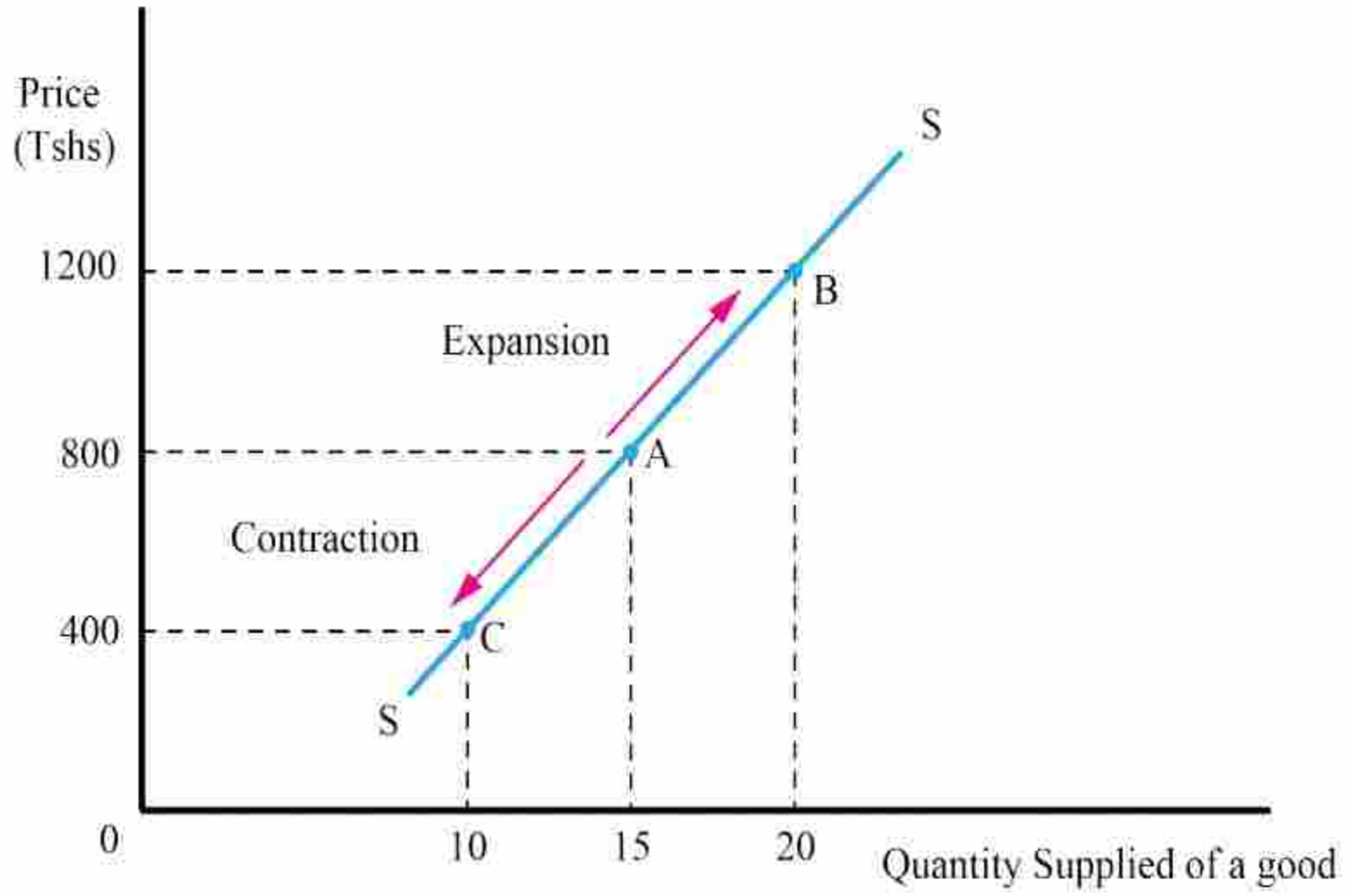


Figure 2.24: *Change in quantity supplied of a good*

Change in supply

A change in supply is the increase or decrease in supply due to a change in factors that affect supply of the good, other than price of the good, which is called a shift of the supply curve to the left or to the right. The shift of the supply curve to the right shows the increase in supply of a good, whereas a shift of the supply curve to the left shows a decrease in the supply of a good.

Increase in supply: An increase in the supply of a good is due to changes in factors affecting supply other than the price of the good, such as a rise in the productivity of inputs or a decrease in the price of inputs. Figure 2.25 illustrates the increase in supply such that firms supply more quantities of a good while price of the good remains the same (shift from S_0S_0 to S_1S_1).

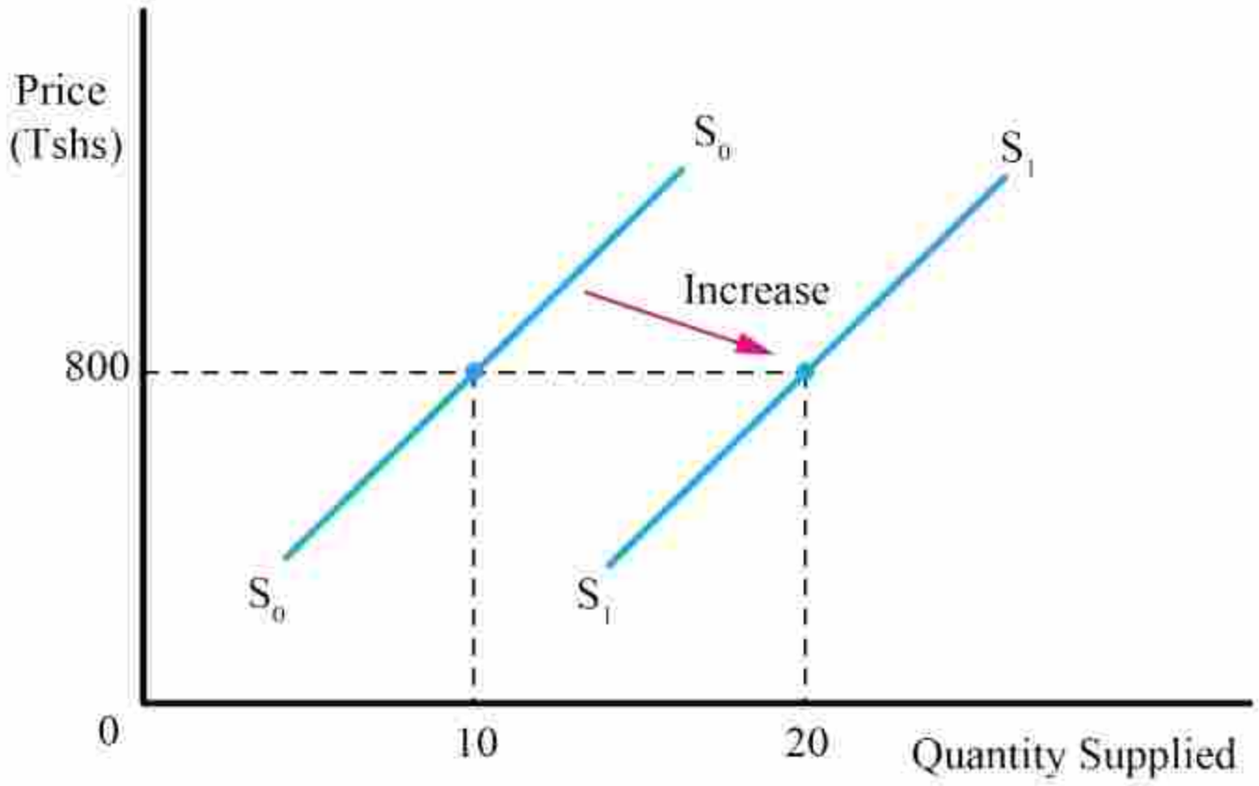


Figure 2.25: *Change in supply-increase in supply*

Decrease in supply: A decrease in supply is due to a change in factors that affect supply, other than the price of the good. This decrease is termed a shift in the supply curve to the left. For example, due to an increase in the price of labour, producers supply less quantity of beef at the same price (shift from S_0S_0 to S_2S_2), as shown in Figure 2.26.

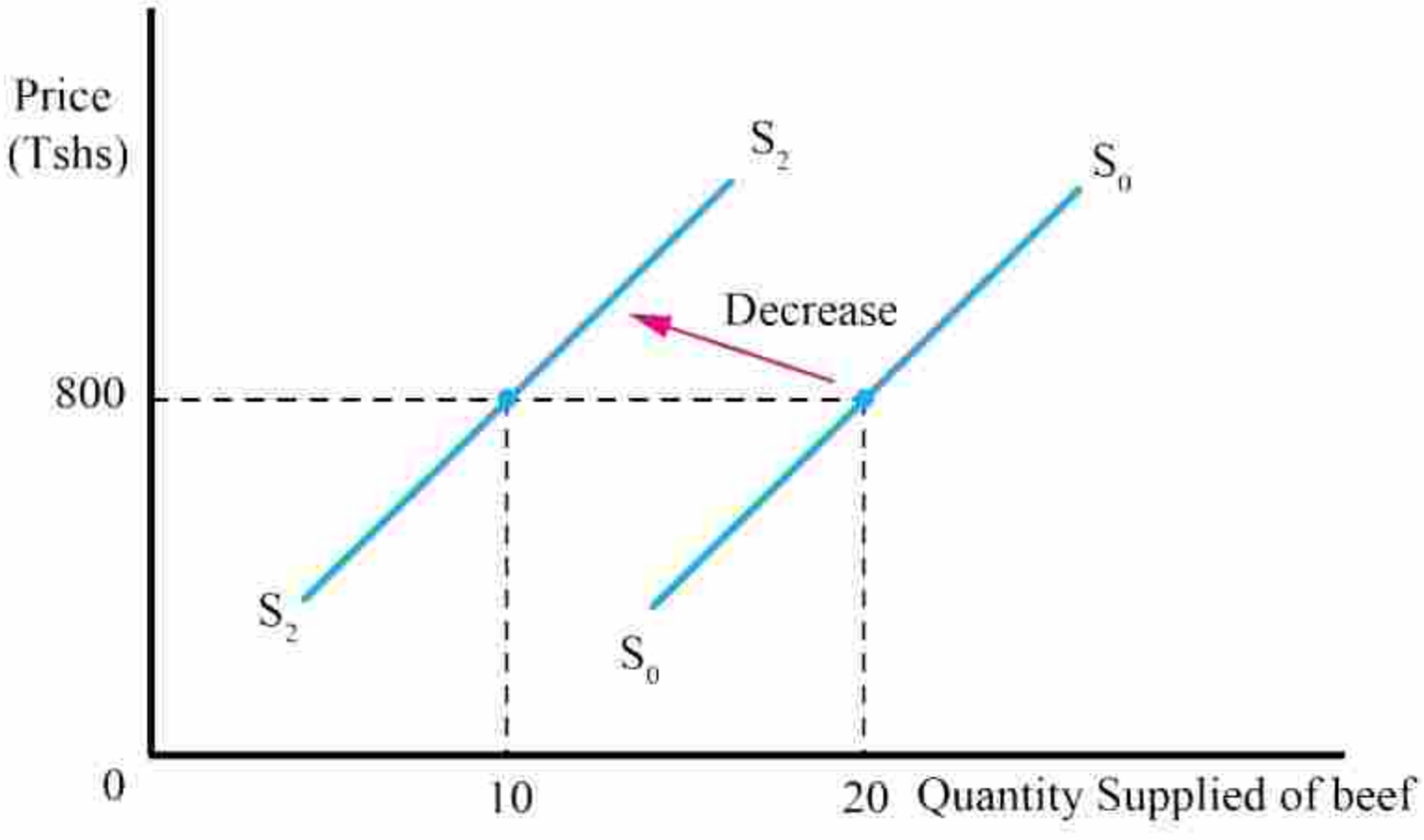


Figure 2.26: *Change in supply-decrease in supply*

Exercise 2.3

1. Discuss and present the three major things that are fundamental in the definition of supply.
2. Differentiate between a supply curve and a supply schedule.

3. With the aid of graph, explain how producers of a good would adjust their current supply if they expect the price of the good to rise in future.
4. By using relevant examples, distinguish between change in supply and change in the quantity supplied.
5. Explain at least three reasons for the outward shift of the supply curve.

Interrelated supply

A supply of one good may affect the supply of another good in the market. Such effect is possible if the two goods are interrelated. This relationship is either positive or negative. The interrelated supply may be joint, composite or competitive supply.

Joint supply

A joint supply involves goods produced together. Hence, the supply of such goods increases or decreases simultaneously. An example of joint supply goods is beef and leather. An increase in the price for beef will lead to an increase in the supply of beef (Figure 2.27). In turn, because the supply of beef and leather move together, the supply of leather (a product of hides) will also increase as shown in Figure 2.27.

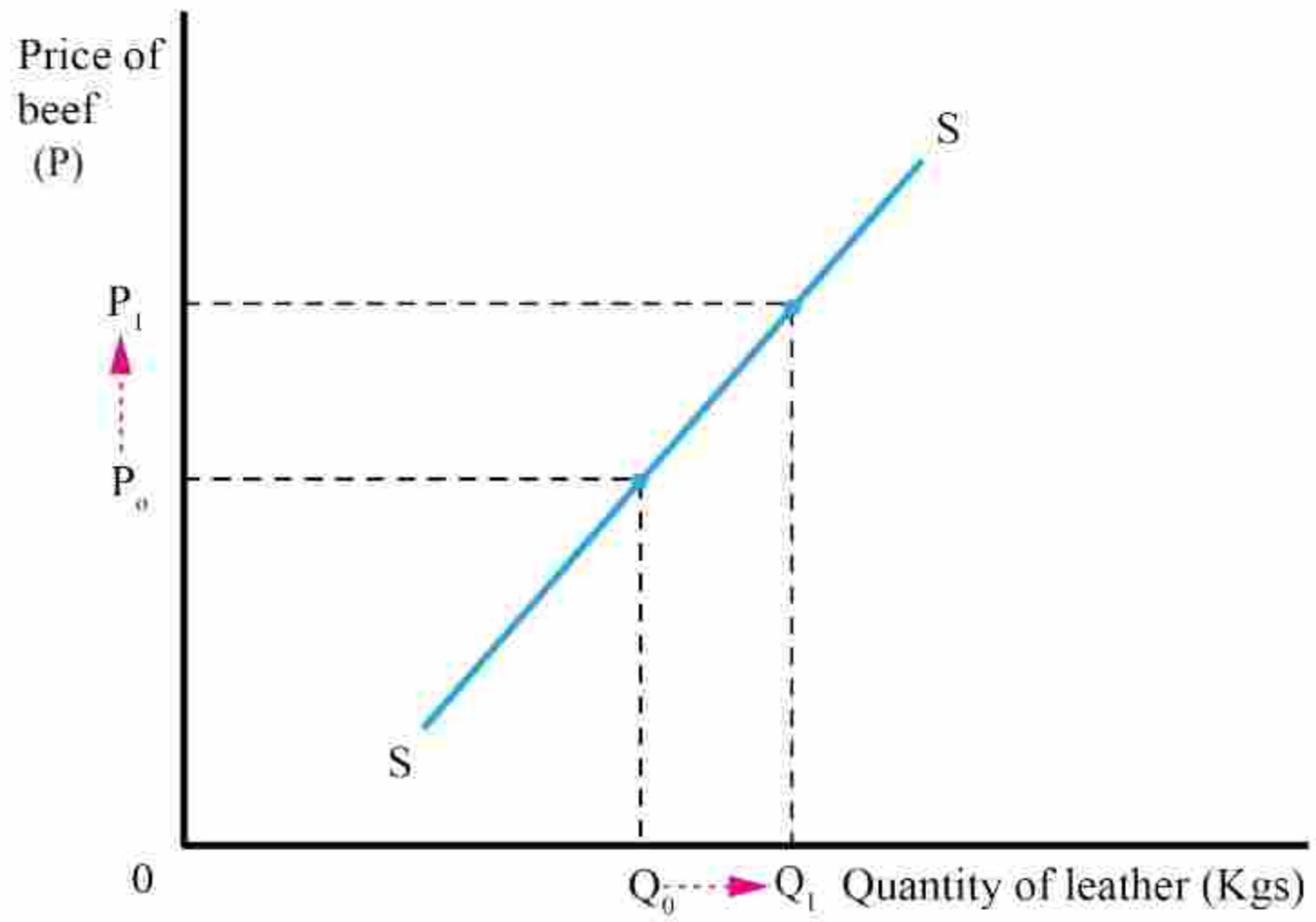


Figure 2.27: *Joint supply of beef and leather*

Composite supply

Composite supply is the total supply of the goods that are substitutes such that an increase in the price of one good leads to a decrease in the quantity supplied of the other good. An example of the composite supply is the supply of beef and chicken or the supply of soft drinks, such as Coca Cola and Pepsi.

Competitive supply

Recall that in Chapter One, the PPF analysis showed that factors of production have alternative uses. If the economy uses factors of production, such as land, labour and capital to produce a particular good, they will affect the supply of other goods that they could have produced. For example, when more land is used to produce bananas, it will lead to the decrease in supply of maize, since less land will be apportioned to a maize production. Similarly, if workers involved in the production of banana with skills to produce both bananas and maize shift to producing maize, then there will be less production of bananas, relative to maize as shown in Figure 2.28. This type of interrelated supply is termed competitive supply.

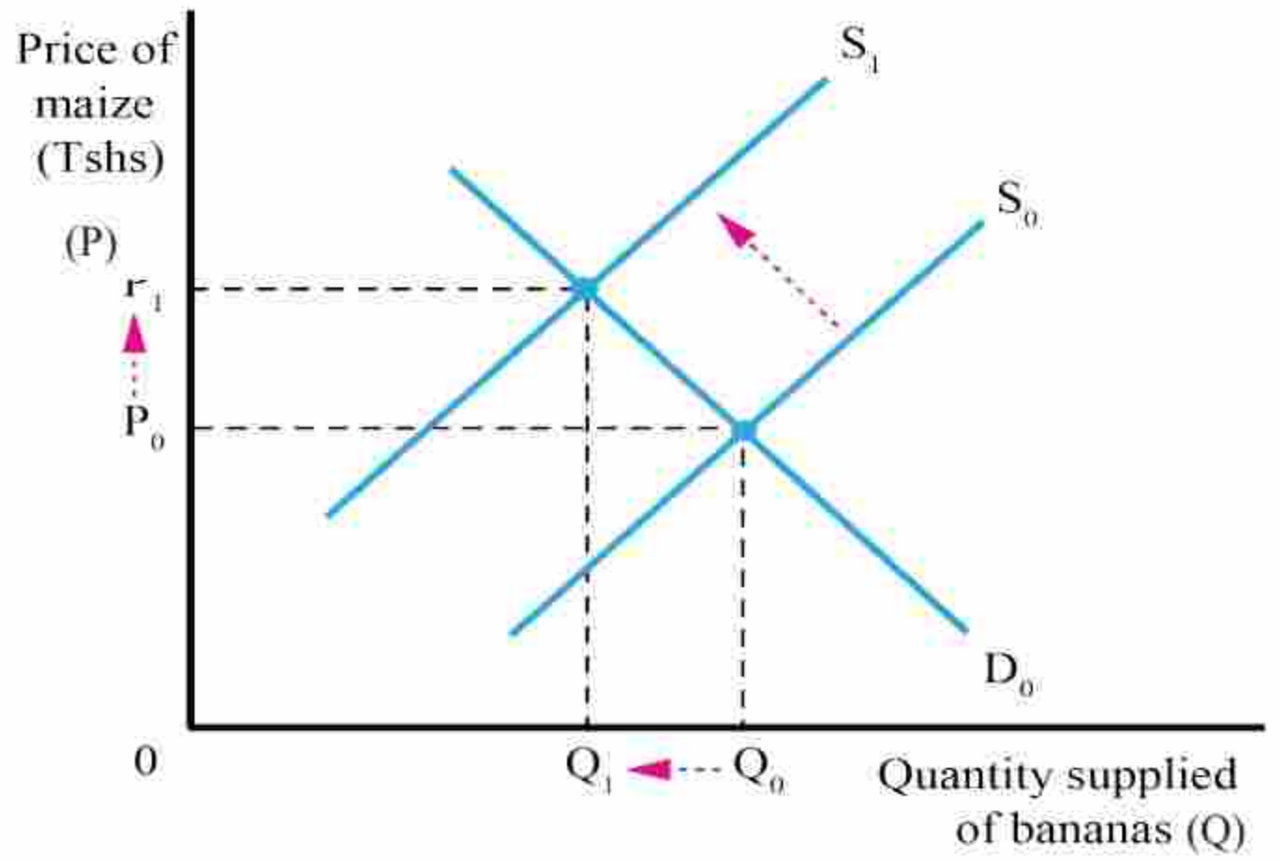


Figure 2.28: *Competitive supply of banana and maize*

The concept of elasticity of supply

Elasticity of supply is a measure of the extent to which the quantity supplied of a good responds to proportionate changes in the own price of a good.

Price elasticity of supply

Price elasticity of supply is the degree of responsiveness of quantity supplied of a commodity to change in the price of the good. It is calculated as follows:

$$\text{Price elasticity of supply } \left(e_p^s\right) = \frac{\text{Percentage change in quantity supplied}}{\text{Percentage change in price}}$$

$$e_p^s = \frac{\text{Change in quantity supplied } (\Delta QS)}{\text{Initial quantity supplied (S)}} \div \frac{\text{Change in price } (\Delta P)}{\text{Initial price (P)}}$$

$$e_p^s = \frac{\%\Delta QS}{\%\Delta P} \quad \ldots\ldots\ldots\

Interpretations of price elasticity of supply

On the basis of the degree of responsiveness of supply to change in the price of the good, elasticity of supply is interpreted as:

(a) Perfectly elastic supply $\left(e_p^s = \infty\right)$

Supply of a commodity is perfectly elastic when any amount can be supplied at a given market price. The Supply curve will be a straight line parallel to the x-axis (horizontal line), as shown in Figure 2.29. In this case, the price elasticity of supply is infinity (∞).

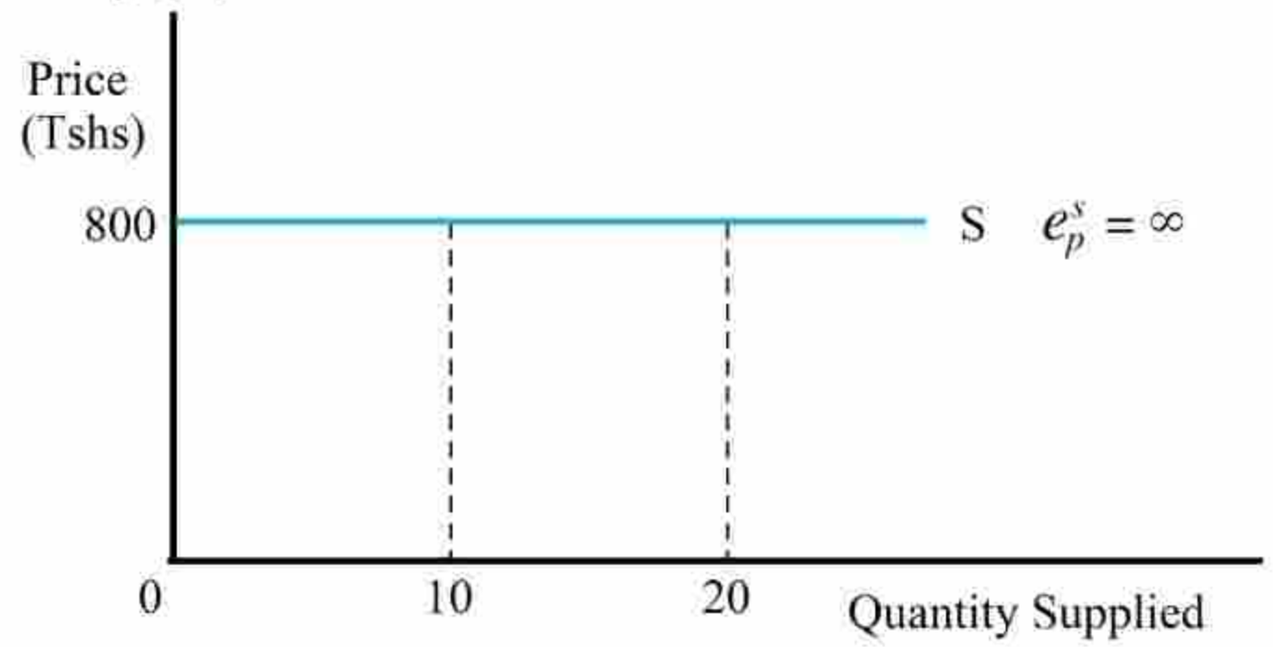


Figure 2.29: *Perfectly elastic supply*

(b) Perfectly inelastic supply $\left(e_p^s = 0\right)$

Supply of a commodity is perfectly inelastic if the quantity supplied of a good does not change due to a change in its price of the good. The supply curve is thus a vertical straight line parallel to y-axis, as shown in Figure 2.30. Here the price elasticity of supply is zero.

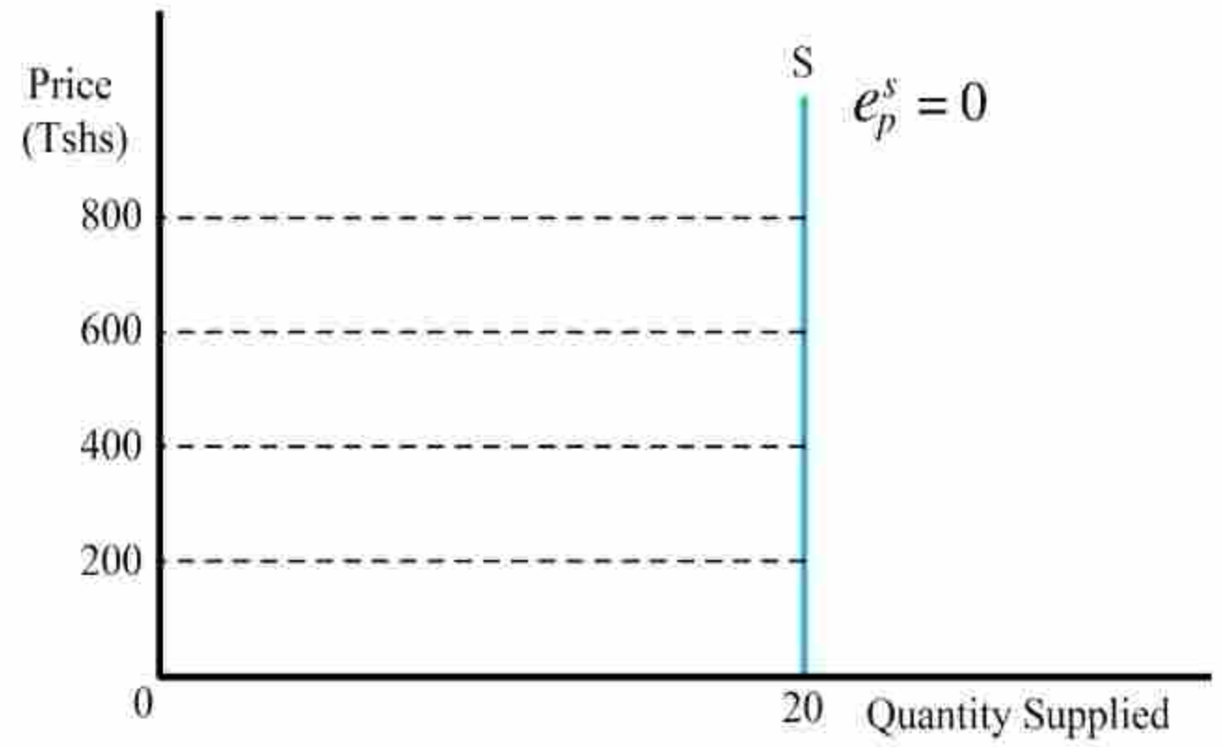


Figure 2.30: *Perfectly inelastic supply*

(c) Unitary elastic supply $\left(e_p^s = 1\right)$

Supply of a commodity is unitary elastic if a proportionate change in the price of a good causes an equal proportionate change in the quantity supplied of that good. Figure 2.31 illustrates a unitary elastic supply curve whereby when the price of a good decreases from Tshs 50 to Tshs 40, the quantity supplied of a good decreases from 40 units to 32 units, and hence the same proportional change for both the price and the quantity of a good.

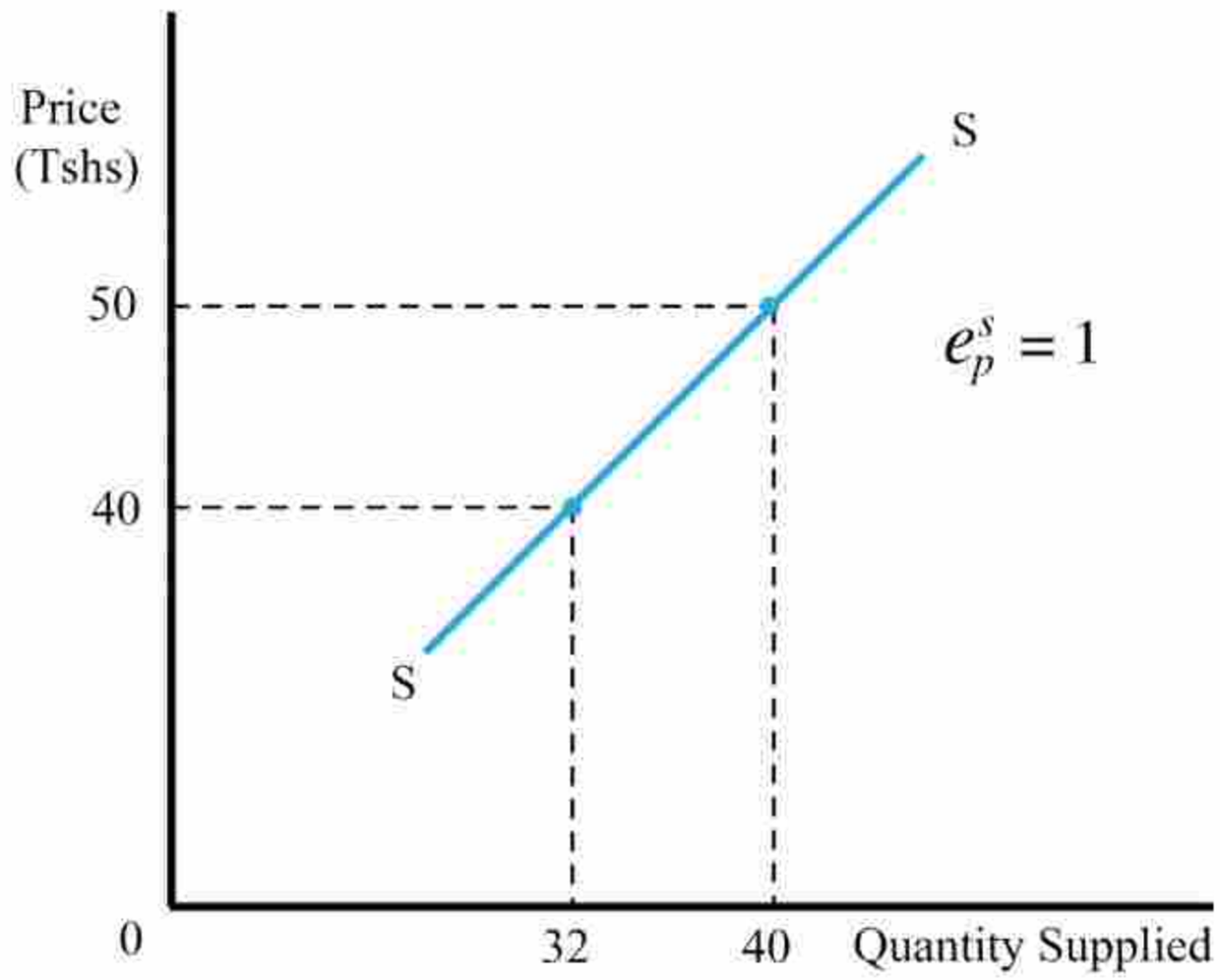


Figure 2.31: *Unitary elastic supply*

(d) Elastic supply $\left(e_p^s > 1\right)$

Supply of a good is more elastic if a given proportionate change in the price of a good leads to a more than proportionate change in the quantity supplied of the good, which is termed elastic supply. In Figure 2.32, when price of a good rises from Tshs 600 to Tshs 800, the quantity supplied of a good increases from 10 units to 20 units. The proportionate change in the quantity supplied of a good is more than the proportionate change in price of the good.

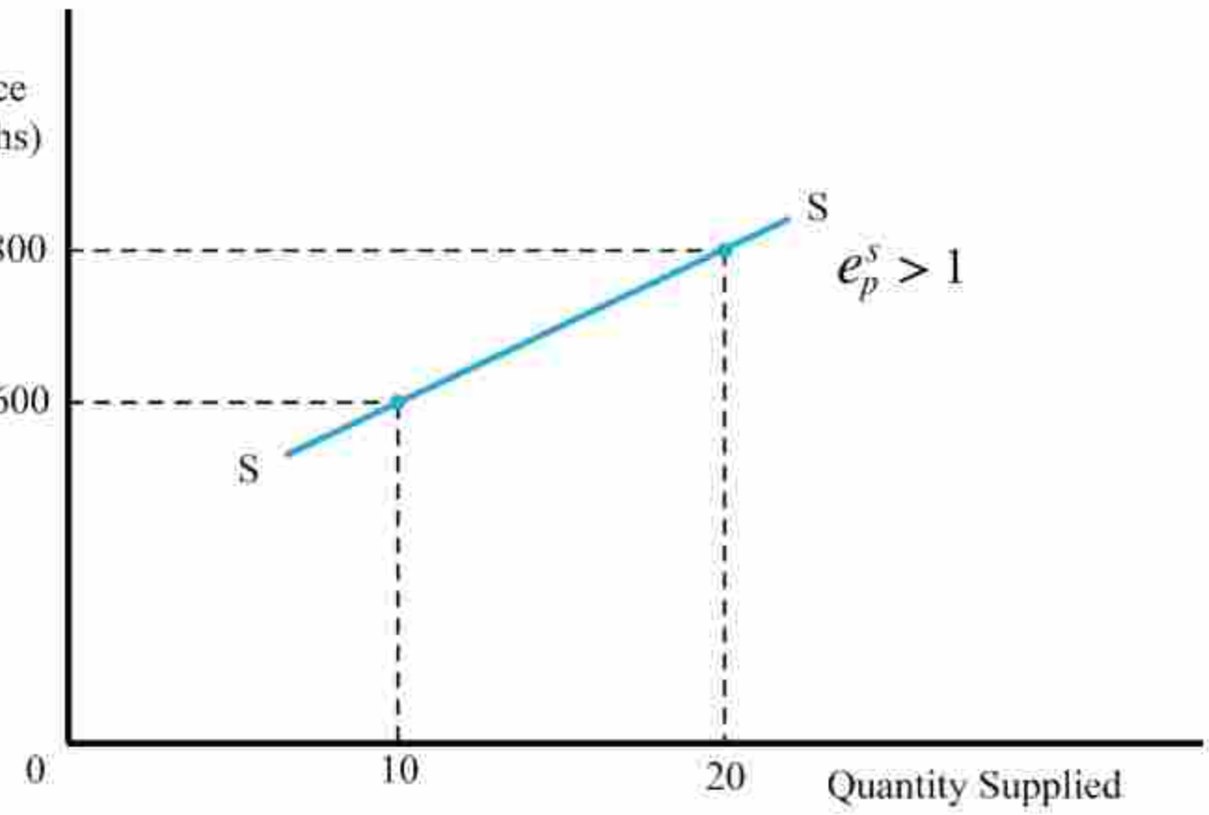


Figure 2.32: *Elastic supply*

(e) Inelastic supply $\left(e_p^s < 1\right)$

Supply for a commodity is less elastic or inelastic when a given proportionate change in the price of a good leads to a lesser than the proportionate change in quantity supplied of a good. In Figure 2.33, when the price of a good rises from Tshs 200 to Tshs 900, the quantity supplied of a good increases from 18 units to 20 units. The proportionate change in supply is less than the proportionate change in the price of the good.

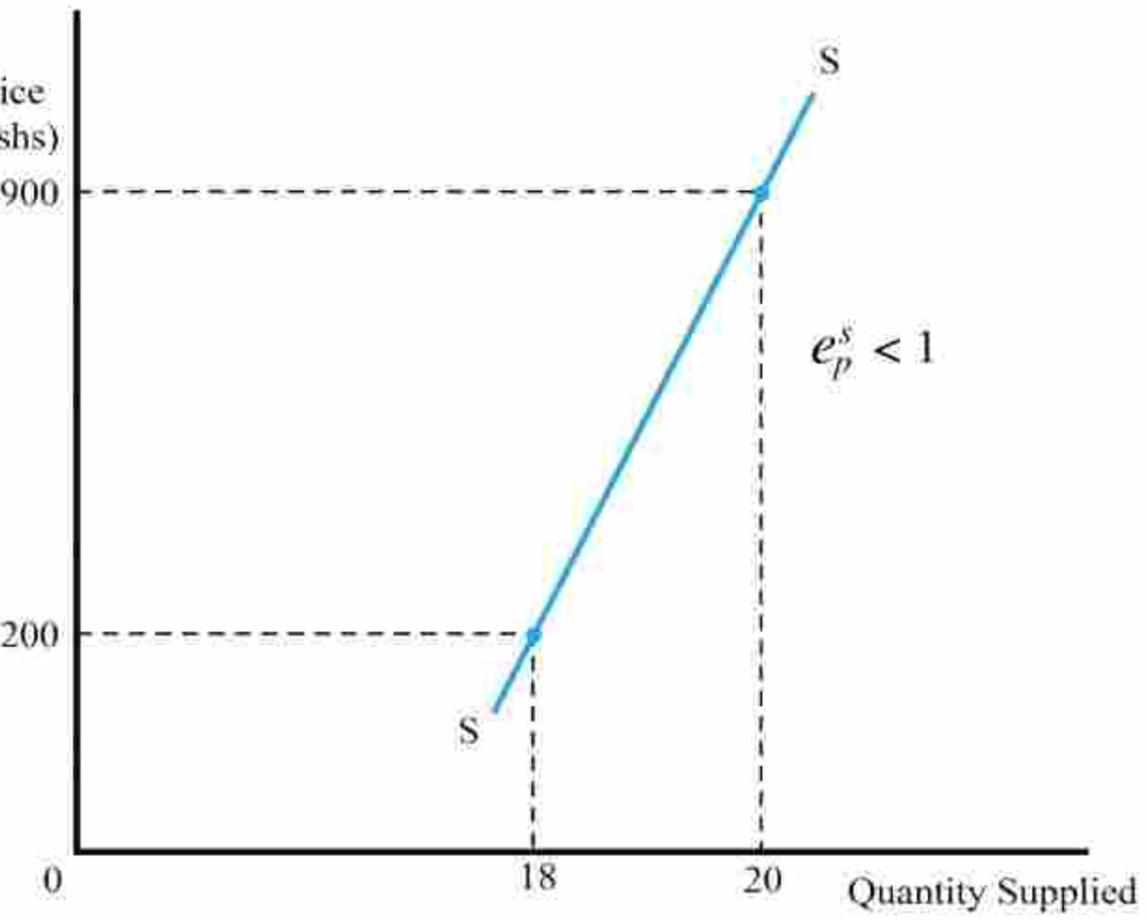


Figure 2.33: *Inelastic supply*

Example 2.6:

Suppose that when the price of one kilogram of beef in Mbeya was Tshs 4,000, the quantity supplied per day was 50,000 kilograms. All of a sudden, the price per kilogram of beef increases to Tshs 7,000 and causes the quantity supplied per day to increase from 50,000 to 150,000 kilograms. Using the information given, calculate the price elasticity of supply for beef and give its economic interpretation.

Solution

Given;

$P_0 = \text{Tshs } 4,000$

$Q_0 = 50,000 \text{ kgs}$

$P_1 = \text{Tshs } 7,000$

$Q_1 = 150,000 \text{ kgs}$

$$e_p^s = \frac{Q_1 - Q_0}{P_1 - P_0} \times \frac{P_0}{Q_0}$$

$$e_p^s = \frac{150,000 - 50,000}{7,000 - 4,000} \times \frac{4,000}{50,000}$$

$$e_p^s = \frac{100,000}{3,000} \times \frac{4,000}{50,000}$$

$$e_p^s = 2.67$$

The elasticity of supply is 2.67, implying that a one percentage increase in the price of beef leads to a 2.67 percentage increase in the quantity of beef supplied. The price elasticity of supply is greater than a unit change in the price of beef, implying the supply of beef is elastic.

Factors determining elasticity of supply

The following factors cause the elasticity of supply to be either zero, greater than one, less than one, one or infinity.

Cost of production: If cost of production per unit rises as output rises, the supply will tend to be inelastic. If, on the other hand, when unit costs rise slowly as production increases, supply will tend to be elastic.

Techniques of production: Goods produced using simple techniques of production will cause supply to be elastic. On the other hand, goods produced using complex techniques of production will cause less elastic supply.

Risk bearing capacity: If producers have a large risk bearing capacity, their goods will have elastic supply but if producers are risk averse (that is, they are not inclined to bear risk), the supply will be less elastic.

Availability of inputs for expanding output: The extent to which the producers increase supply of their products depends also on availability of inputs. If the inputs used in the production process are easily available at the going market prices, then output can expand with a small increase in price, which will lead to elastic supply.

Response of producers to a change in price: The elasticity of supply depends on responsiveness of producers to a change in the price of a good. If producers do not respond positively to the increase in the price of a good, the price rise would not cause the quantity of output supplied for a product to increase. This would lead to inelastic supply.

Excess capacity and unsold stock: In the short-run, it may be possible to increase supplies considerably if there is a pool of unemployed labour and unused machinery in the industry. If the industry accumulates a large stock of unsold goods, supply can increase quickly. Thus, supply will be more elastic when there is excess capacity in the industry and high level of unsold stock.

Practical usefulness of price elasticity of supply

The price elasticity of supply is useful in the economy in the following ways:

(a) It can be applied to labour to show how the quantity of labour supplied responds to changes in wages and salaries.

(b) It helps the government to determine policies to adopt. For instance, if the government wants to promote exports, it must increase prices for goods whose elasticity of supply is elastic so as to increase the volume of supply for exports.

(c) It is used by producers in deciding what to produce. Producers tend to produce more goods whose elasticity of supply is elastic so as to enable them to increase output at a larger proportionate as the price increases.

Exercise 2.4

1. Suppose the price of good A changes from Tshs 600 to Tshs 500 per unit while the quantity supplied of the good changes from 40 to 20 units per week. Compute the elasticity of supply and interpret your answer.
2. Given that the supply of a good X has risen from 200 to 300 units per week due to the rise in price from Tshs 600 to Tshs 800 per unit, calculate the price elasticity of supply and interpret your answer.
3. A firm sells 200 units of good X when its price is Tshs 50. At what price will it sell 300 units of the commodity if its price elasticity of supply is 0.4?
4. How does the cost of production of additional units of output influence the elasticity of supply of a good?

Market equilibrium

A market is a concept encompassing (incorporating) the forces of supply and demand, and the interaction of buyers and sellers aiming at exchanging goods or services. A market brings together those who are willing and able to sell goods and services and those who are willing and able to purchase the goods and services.

Market equilibrium is a state of balance between supply and demand in the market. When a market is in equilibrium, the decisions of consumers and producers are consistent with the forces of demand and supply. The supply curve and demand curve intersect to form market equilibrium price and quantity in the market. At the equilibrium price, the quantity demanded of a good is equal to quantity supplied in the market.

Market equilibrium is illustrated in Figure 2.34 where point E is the equilibrium point where ($Q_d = Q_s$), which is the point of intersection between the demand and supply curves. At that point, the equilibrium price is Tshs 1,000 and the equilibrium quantity is 18 units.

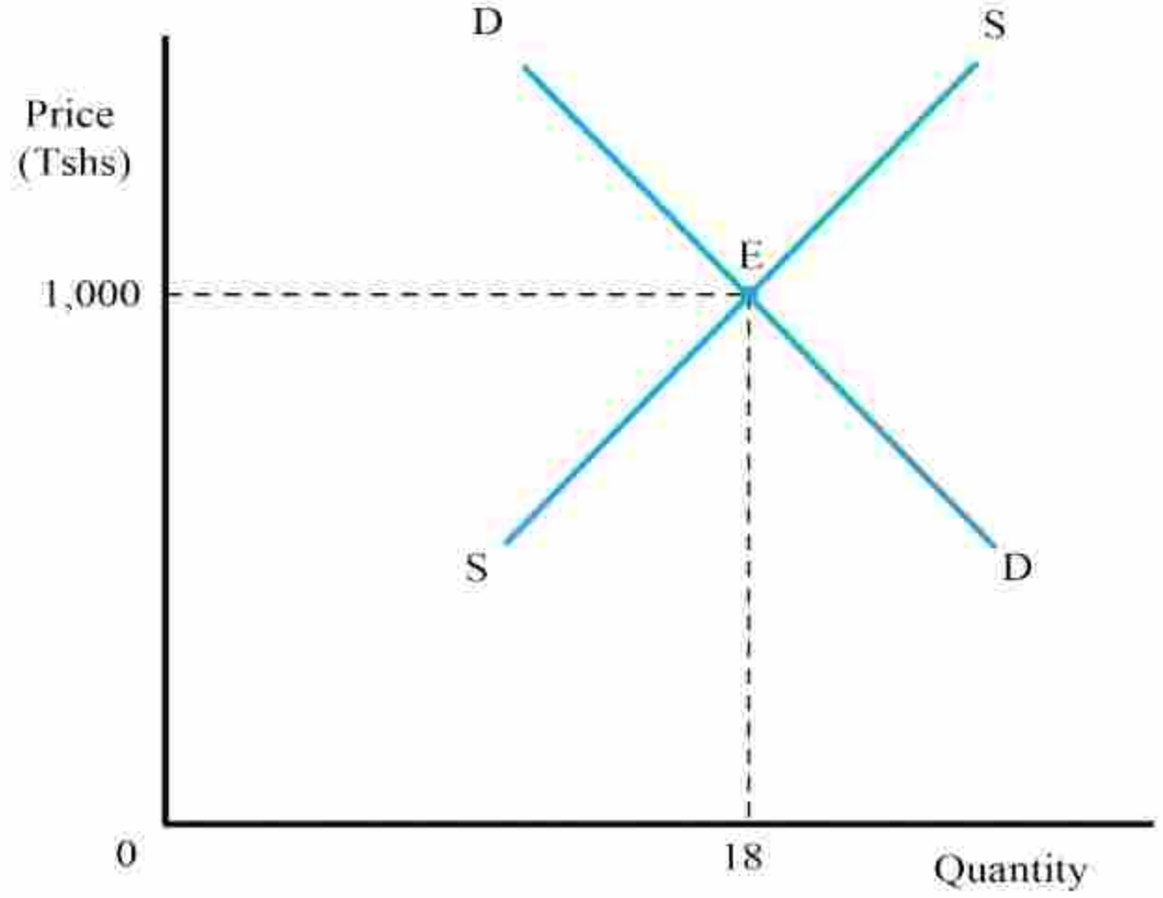


Figure 2.34: *Market equilibrium*

There are several ways of determining the price of a good in the market. Some of the known methods include the following:

Interaction of forces of demand and supply: Through this method, the price of a good is determined at a point where demand and supply curves intersect.

Haggling: This is a method of price determination whereby the seller and the buyer negotiate or bargain for the price of the good. This process culminates in a mutually agreeable price.

Price control: With this method, the government authorities fix either the maximum price (price ceiling) or the minimum price (price floor).

Auction: With auction, the prospective buyers bid or compete for the price of the commodity and the buyer who claims or bids the highest price wins the product.

Example 2.7:

Suppose a demand function is $Q_d = 20 - 0.5P$ and the supply function is $Q_s = -10 + 1.5P$

(a) Find the equilibrium price and quantity.

(b) Using the graphical method, sketch a graph showing equilibrium price and quantity.

Solution

Remember that at the equilibrium price, the quantity demanded of a good is equal to the quantity supplied.

Thus, $Q_d = Q_s$

Therefore, by substitution:

$$20 - 0.5P = -10 + 1.5P$$

To find the value of price:

$$20 + 10 = 1.5P + 0.5P$$
$$30 = 2P$$
$$P = 15$$

Given the equilibrium price (P) is Tshs 15, by substituting it in either demand or supply function

$$Q_d = 20 - 0.5P$$
$$Q_d = 20 - 0.5(15)$$
$$Q_d = 12.5$$

The equilibrium quantity (Q) is 12.5 units.

Therefore, the market is in equilibrium where 12.5 units of good are demanded and supplied at a price of Tshs 15.

The graphical method of calculating equilibrium price and quantity.

Given $Q_d = 20 - 0.5P$

$Q_s = -10 + 1.5P$

By applying intercepts

From $Q_d = 20 - 0.5P$

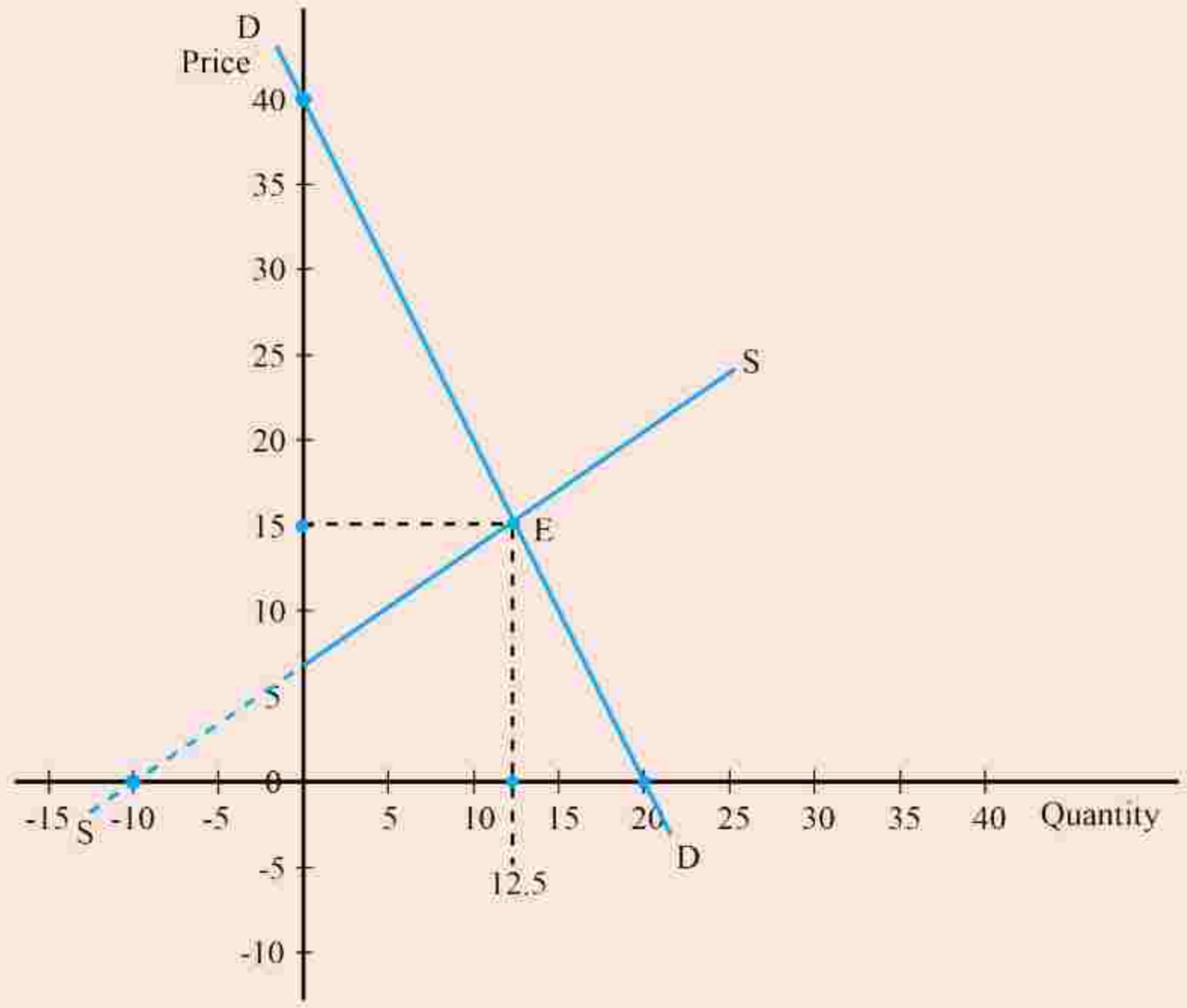
when $P = 0$, $Q_d = 20$

$Q_d = 0$, $P = 40$

From $Q_s = -10 + 1.5P$

when $P = 0$, $Q_s = -10$

$Q_s = 0$, $P = 6.67$



Equilibrium price = Tshs 15

Equilibrium quantity = 12.5 units

Note that the negative quantity implies that at a given level of price the suppliers are not willing to supply goods or services in the market and hence, it is likely for shortage to occur.

Price mechanism

This is a system which addresses the economic problem of what to produce, how to produce and for whom to produce. To facilitate the efficient allocation of scarce resources, the price mechanism acts as "an invisible hand" or signalling mechanism.

The "father of economics" Adam Smith introduced the **invisible-hand principle,** which is a tendency of market forces to channel the actions of self-interested individuals into activities that promote the general betterment of society. That is to say, price mechanism is the system that engineer economic decisions through the market forces of demand and supply.

For example, assume that a fashion changes in favour of a particular good (for example, X). If the demand for the good increases, it will lead to an increase in the price of X. The rise in price of X will induce many firms or producers to produce or supply more of good X. This leads to efficient alloction of resources, as there is no any government plan to increase the supply of X, rather, it is the invisible hand of the price mechanism that is working to ensure scarce resources are allocated to their efficiency use.

Price mechanism enables efficient allocation of productive resources, such as land, labour and capital to produce goods and services profitably. Consumers get the most and the best for their money and suppliers get a good price for their products. The mechanism encourages competition, which leads to efficiency in production and consumption. With efficient utilisation of resources, competition leads to improvement in production, the quantity of goods and services and the quality of delivery.

Advantages of price mechanism

Price mechanism determines automatically what to produce, how to produce and for whom to produce. Prices provide an incentive for growth. High prices encourage production and discourage consumption. Therefore, it is important to assign resources to produce what is beneficial to both producers and consumers.

The price mechanism encourages competition, which leads to increased efficiency. Only the efficient firms survive the competition allowed by the price

mechanism. Price mechanism also facilitates distribution of scarce resources and spatial distribution of goods and services.



Activity 2.3

Visit the library and review library materials and use various websites to study about advantages and disadvantages of the price mechanism.

a) Write down information gathered from the sources you read.
b) With vivid examples, analyse advantages and disadvantages of the price mechanism in Tanzania.
c) Suggest ways to address the negative effects of the price mechanism in the economy.
d) Share your work with other students to broaden your understanding.

Exercise 2.5

1. By using the demand and supply schedule, determine the equilibrium price and equilibrium quantity and then sketch the graph for demand and supply and indicate the equilibrium point.

Price (Tshs)	Quantity demanded (tons)	Quantity supplied (tons)
1	13	8
2	10	10
3	7	12
4	4	14
5	1	16

2. Use the demand and supply functions to find the equilibrium price and quantity then sketch the graphs for demand and supply that indicate the equilibrium price and quantity.

   Equations: $P = 50 - 2Q_d$,

   $P = 10 + 2Q_s$

3. Given the following pairs of equations, answer the questions that follow:

   $Q = 13 - 3P$, and $Q = 8 + 2P$

   (a) Which of the equations represents a demand function? Which represent a supply function? Explain why.

   (b) Solve for the equilibrium price and quantity.

   (c) Present the results graphically.

4. The demand and supply functions for monthly library membership are given as:

   $Q_d = 6,000 - 100P$ and

   $Q_S = 100P - 3,000$, where P is the monthly price, in Tanzanian shilling. Answer the following questions:

   (a) If the current price for memberships is Tshs 50 per month, what is going to happen in the market?

   (b) Would you expect the membership price to rise or fall in order to attain the equilibrium? If so, by how much?

   (c) From the two equations above, calculate price elasticity of demand and price elasticity of supply.

Price controls

Following economic reforms in Tanzania in the mid-1980s, markets were liberalised and prices were allowed to move freely. However, not all markets have been free. There are cases of government intervention to regulate prices, which is termed as price control. Price control is the government attempt to set the price of a good or service either above the equilibrium price to protect the interest of producers or below the equilibrium price to protect the interest of consumers of a particular good or service. Buyers often complain that prices are too high, whereas sellers complain that prices are too low. When unhappy with the prices established by the market forces, various groups might try to persuade the government to intervene and impose price controls. Therefore, price control is a government-mandated act to impose prices known as maximum or minimum prices.

Price floor (Minimum price)

Price floor is a legally established minimum price, which buyers must pay for a commodity. The minimum price is set by the government above the equilibrium price to protect suppliers or producers when the market price is too low. This is done to generate enough profit to producers and/or motivate production of a particular commodity. For example, government authorities tend to fix price floor for some agricultural products such as coffee, cashew nuts and cotton every season in order to protect farmers against falling prices. Figure 2.35 shows the concept of price floor. When the government sets a price above the equilibrium price P_e, at P_2 (Price floor or minimum price) suppliers supply more quantity (Q_s) whereby the quantity supplied (Q_s) exceeds the quantity demanded (Q_d). Since suppliers are unable to sell all they want at the going price, a surplus (excess supply) arises as shown in Figure 2.35.

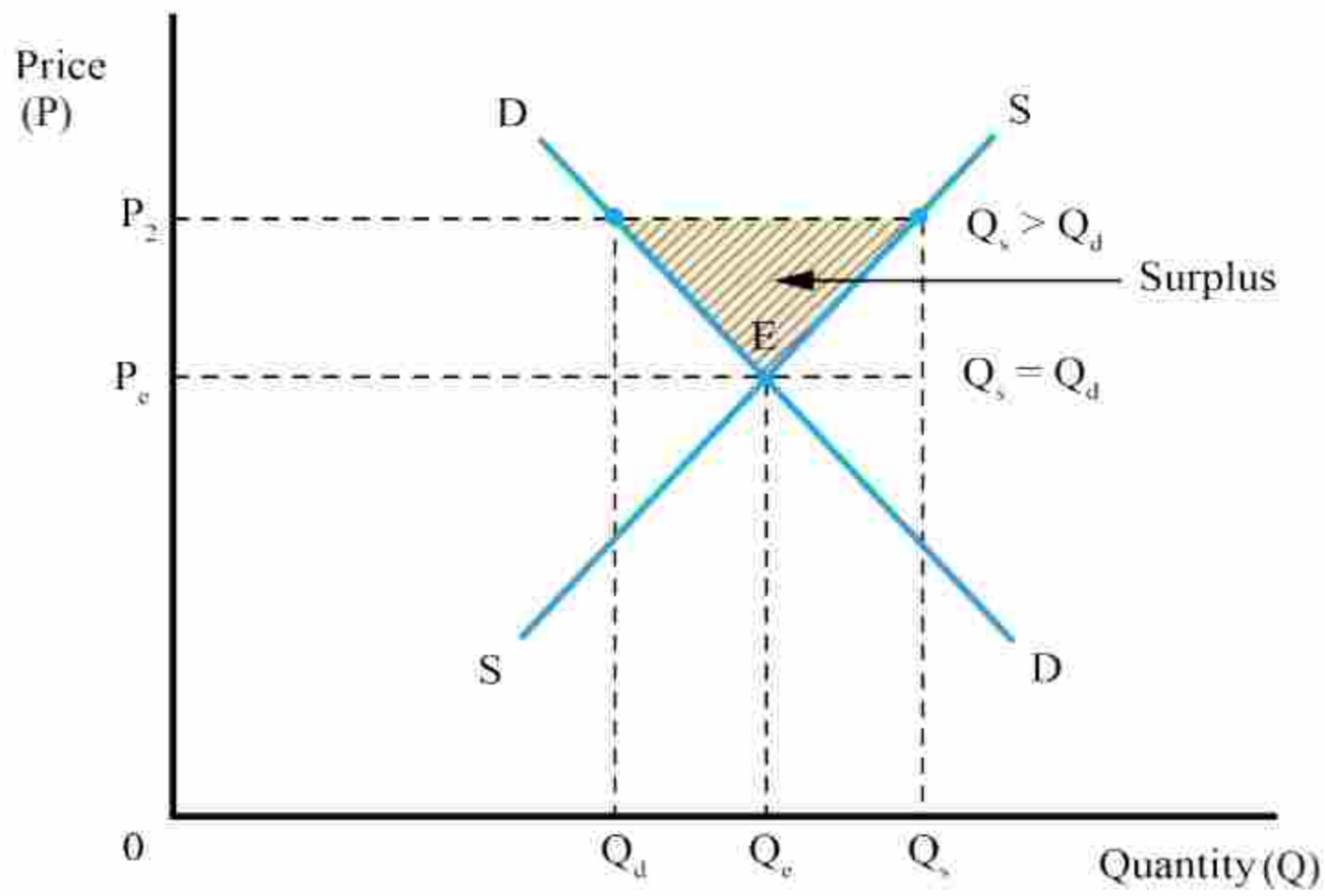


Figure 2.35: *Price floor*

Surplus refers to a condition in which the amount of a good offered for sale by a producer is greater than the amount that buyers purchase at the existing price. A decline in price would eliminate the surplus. This is because when price is above the equilibrium, suppliers or producers will be motivated to supply more and consumers will be discouraged to buy more because of the high price (refer to law of demand and law of supply). The resulting surplus in the market will influence producers to reduce production as well as the price in order to clear the market. As the price falls, the quantity demanded increases since consumers are willing to buy more of the product at a low price.

Price ceiling (Maximum price)

Price ceiling is a legally established maximum price that sellers can charge for a commodity. This is the maximum price set by the government below the equilibrium price in order to protect consumers, when the government thinks that the market price is too high to be afforded by consumers. For example, government authorities like Energy and Water Utilities Regulatory Authority (EWURA) tend to fix maximum prices of fuels such as petrol and diesel regularly. Figure 2.36 elaborates the concept of price ceiling. When the government decides to set price below the equilibrium price P_e, for example, at P_1 (maximum price) suppliers' supply less quantity of goods at Q_s, causing shortage in the market whereby the quantity demanded (Q_d) exceeds the quantity supplied (Q_s). Suppliers being unable to supply enough goods for the consumers results into shortage (excess demand). This is another effect of market regulation by the government.

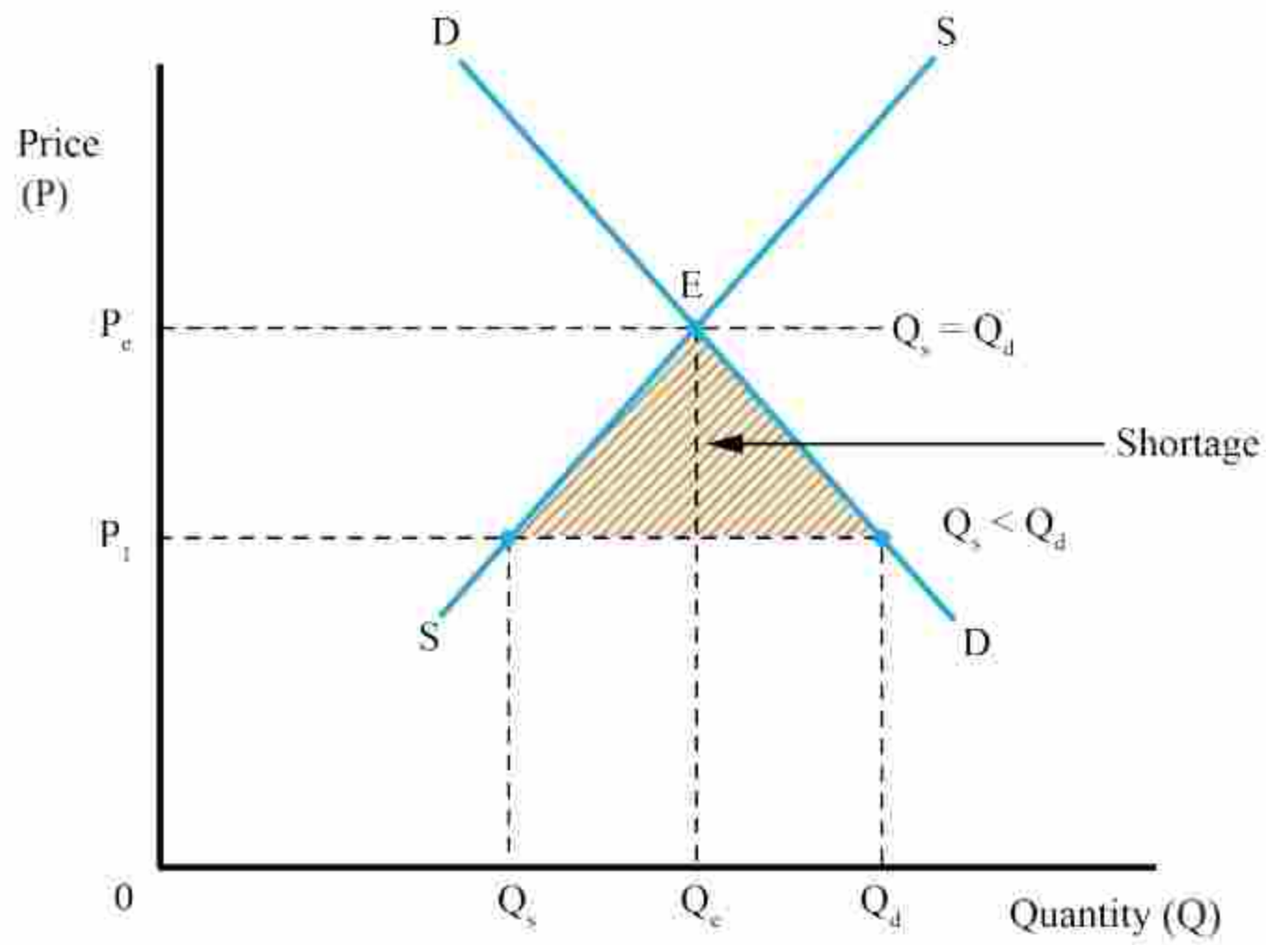


Figure 2.36: *Price ceiling*

Shortage exists when amount of a good offered for sale by producers is less than the amount demanded by buyers at the existing price. An increase in price would eliminate the shortage. If the market price is too low, suppliers or producers will be discouraged to supply the commodity, whereas consumers are encouraged to purchase the commodity, '*ceteris paribus.*' Because of shortage, consumers will offer a high price for the product. As the price increases, producers will be willing to supply more of the good, but the quantity demanded by consumers will decrease. Forces in the market will continue to drive price up until the quantity supplied equals quantity demanded.

In case the government sets price ceiling, there will be shortage of commodity because of an increase in demand and a decrease in supply. The government will be forced to ration the commodity, occurrence of social unrest and smuggling is also likely to happen.

How markets respond to changes in demand and supply

The following are ways the market responds to changes in demand and supply.

Market adjustment due to increase in demand

When demand for a good increases in a market economy and its supply remain unchanged it will lead to a rise in its price. The increase in demand may be due to factors like increase in consumers' income, change in weather, rise in price of the good's substitute, fall in price of its compliment and change in fashion towards the good. Figure 2.37 shows market adjustments due to increase in demand. The increase in demand shifts the demand curve to the right or upwards. Initially, the demand and supply curves intersect at point A where equilibrium price is P_1 and quantity is Q_1. The increase in demand from D_1 to D_2 leads to establishment of a new equilibrium point at B where the equilibrium quantity is Q_2. As a result of the increases in demand, the quantity supplied increased from Q_1 to Q_2.

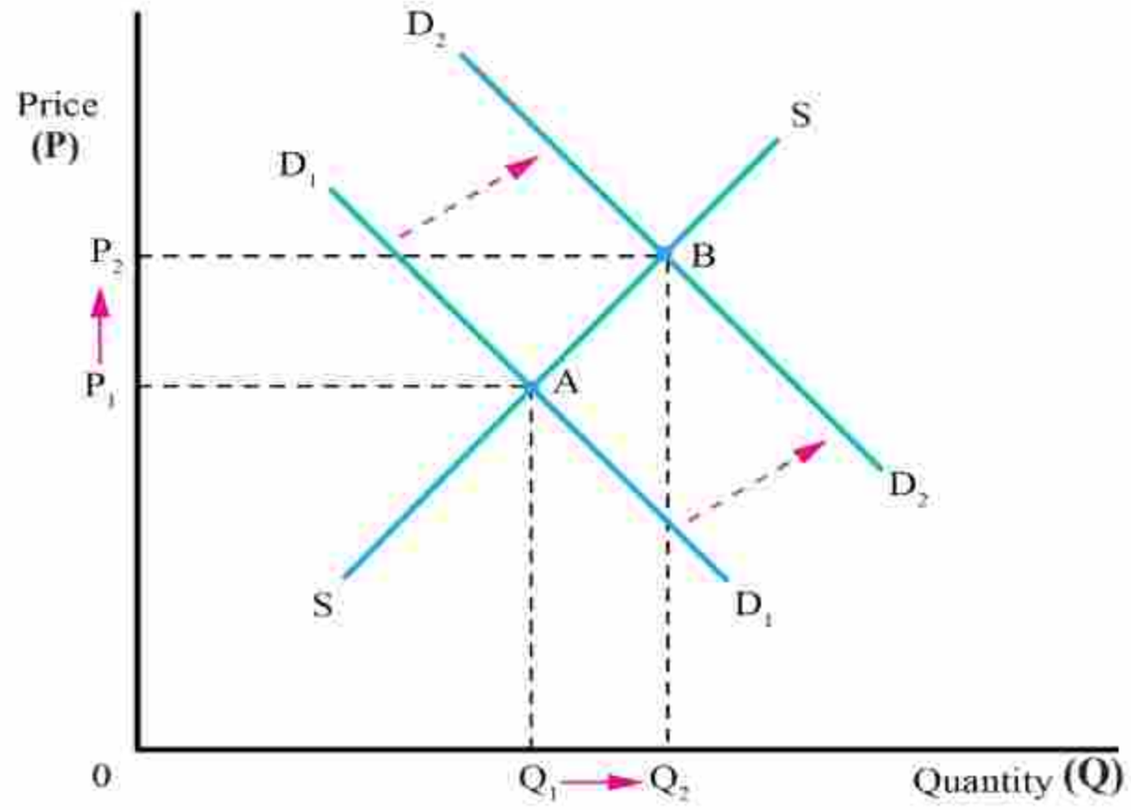


Figure 2.37: *Market adjustment due to increase in demand*

Market adjustment due to decrease in demand

When the demand of a particular good decreases (as shown in Figure 2.38, a shift from D_1 to D_2), the price of that good falls and the quantity bought and sold decreases, respectively. The decrease in demand may be due to factors like change in preference and habits, fall in consumers' income, decline in price of its substitute and increase in price of its complement.

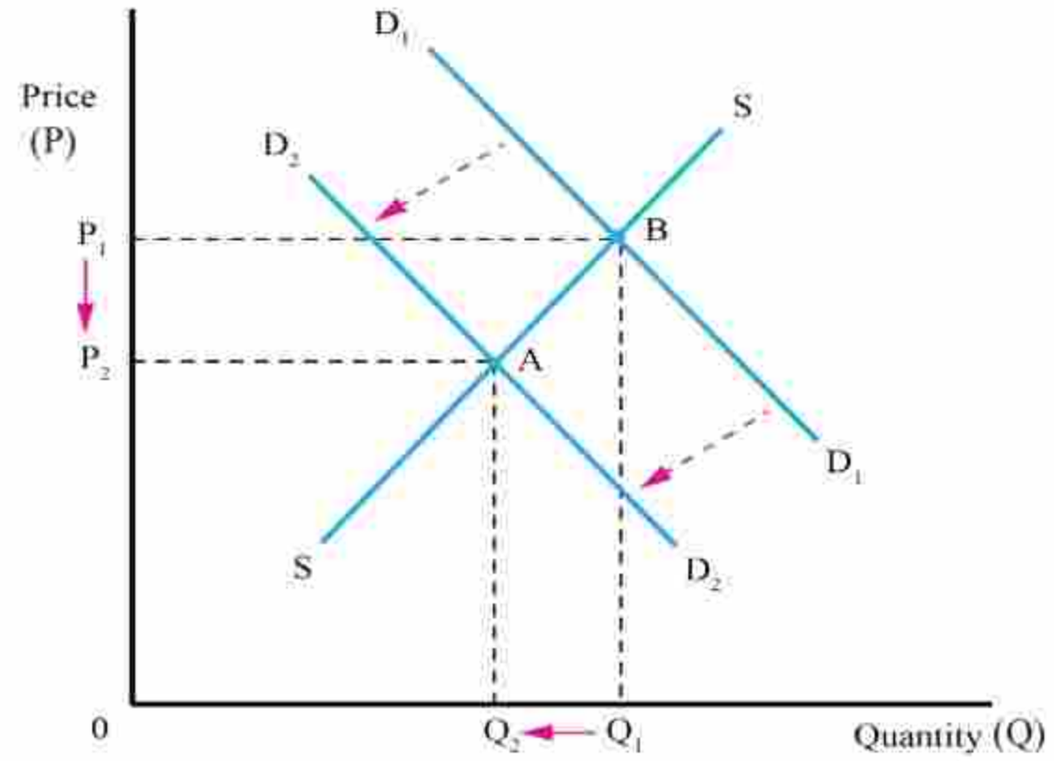


Figure 2.38: *Market adjustment due to decrease in demand*

A decrease in demand (a shift of demand curve to the left) will lead to a lower price and lower quantity traded in the market. Figure 2.38 shows that lower price reduces the incentive of producers to supply goods. Lower demand leads to fall in market prices which force producers to decrease production from Q_1 to Q_2. The reduced output allows these resources to be freed up to go into production of other goods consumers want more. Initially, the demand and supply curves intersect at point B where the equilibrium price is P_1 and equilibrium quantity is Q_1. The decrease in demand from D_1 to D_2 leads to establishment of a new equilibrium point at A. As a result of this, the equilibrium price falls from P_1 to P_2 and the equilibrium quantity supplied and bought decreases also from Q_1 to Q_2.

Market adjustment to decrease in supply

How will markets respond to changes in supply? Decrease in supply will lead to higher prices and a lower equilibrium quantity. The decrease in supply may be brought about by an increase in price of inputs, bad weather like droughts (or flooding and severe rains), low level of technology and, small number of suppliers.

For instance, the harvest or production of maize drops by a large portion due to drought. Figure 2.39 shows a reduction in supply (shift from S_1 to S_2), which causes the price of maize to increase sharply (from P_1, to P_2). Because of the higher price, consumers will reduce consumption of maize (the movement along the same demand curve from A to B) as some will switch to substitutes (let say sorghum). The higher price also will encourage the other maize suppliers to take additional steps, like use of irrigation techniques that allow them to produce more maize than otherwise would have been the case. The higher prices will rebalance the quantity demanded and quantity supplied at the new equilibrium.

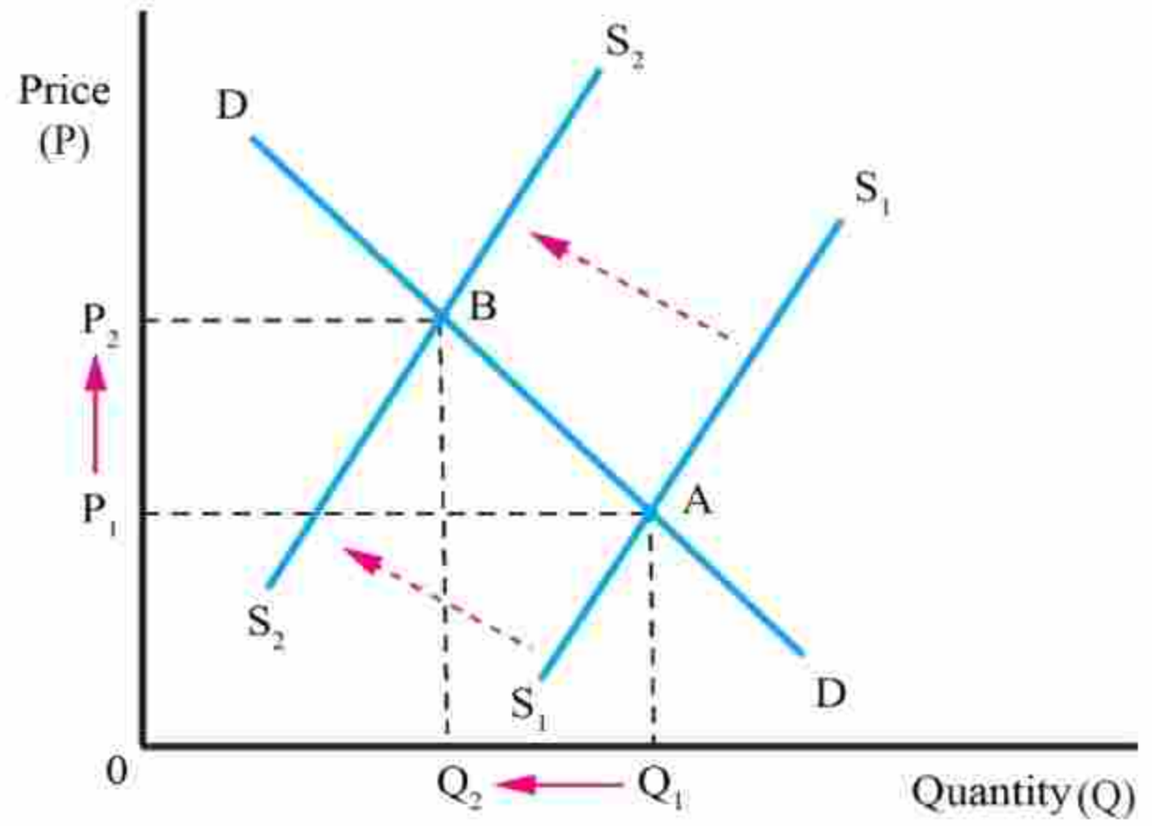


Figure 2.39: *Market adjustment due to decrease in supply*

Market adjustment to increase in supply

When supply of a good increases, it leads to a decrease in its price in the market but results to an increase in its quantity bought and sold. The increase in supply can be brought about by advancement in technology, favourable weather condition and an increase in its number of suppliers. In Figure 2.40, an increase in supply from S_1S_1 to S_2S_2 leads to decrease in price and increase in quantity of the commodity bought in the market. Besides, low price leads to an increase in quantity demanded of the commodity from Q_1 to Q_2, while fall in price discourages producers from increasing production.

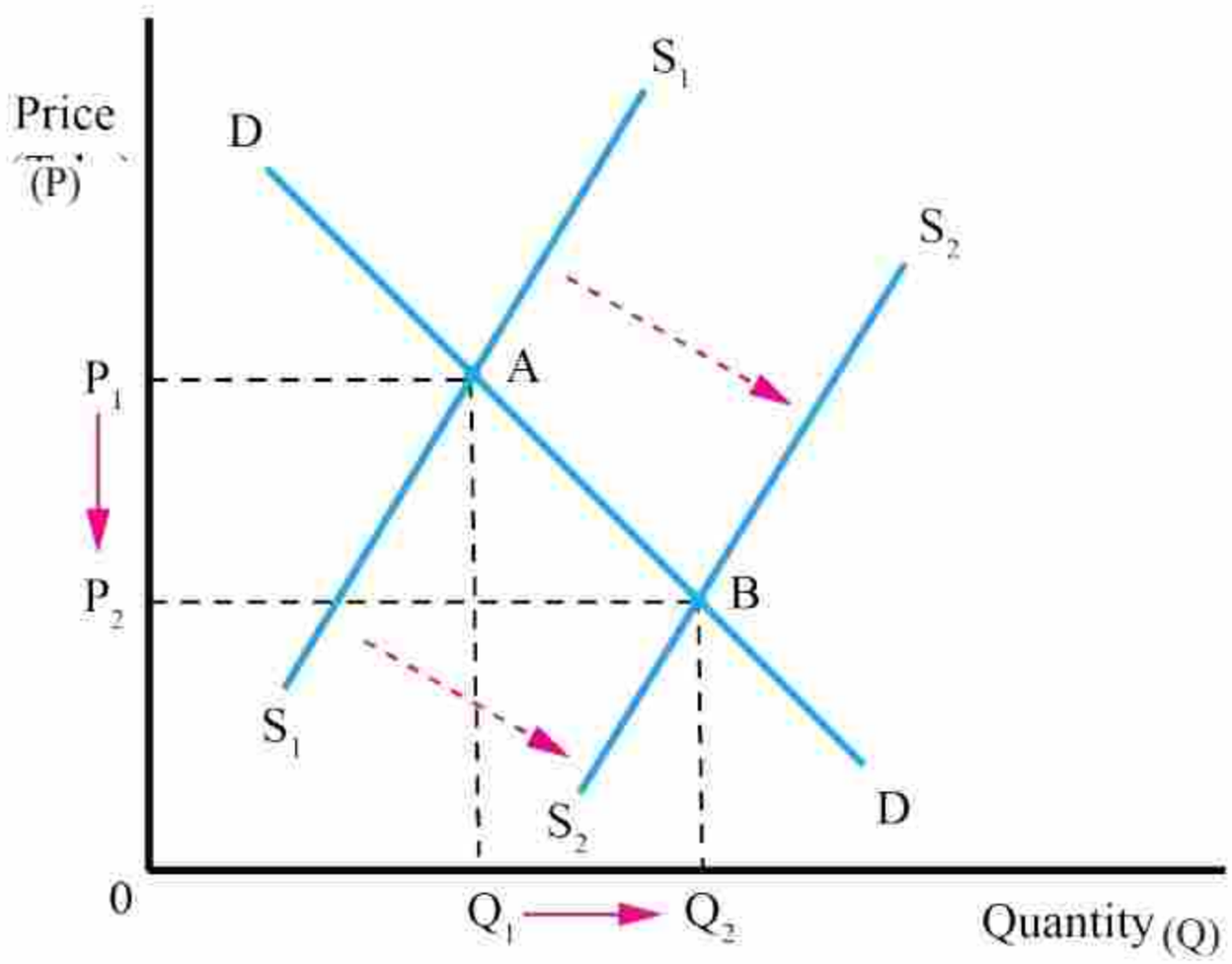


Figure 2.40: *Market adjustment due to increase in supply*

Activity 2.4

Suppose there is an increase in the level of technology used to produce maize or fall in the price of inputs. How do you think the market price and quantity of maize would adjust? Sketch the appropriate supply and demand curves to illustrate this case. Share your work with your classmates.

Chapter summary

1. Demand is willingness and ability of a consumer to purchase goods and services at a certain price in a particular period of time.
2. Supply is the willingness and ability of a supplier or producer to offer for sale different quantity of goods and services at different prices at a certain period of time.
3. Willingness to pay is the maximum price that a consumer is ready to pay for a unit of a product.
4. The law of demand states that "*At a lower price more of a commodity is demanded and at a higher price less of the commodity is demanded and vice versa is true "ceteris paribus."*
5. The law of demand assumes all other factors that affect demand remain unchanged other than price of the commodity.
6. The law of supply states that "*Other things being equal, at a higher price more of the commodity is supplied and, at a lower price less of the commodity is supplied.*"
7. The law of supply assumes that the quantity of a commodity supplied varies directly with a price of goods, *ceteris pa*ribus.
8. Both demand and supply can be expressed in terms of individual and market schedules, curves and functions.
9. Demand curve has a negative slope because of an inverse relationship between price of the commodity and the quantity demanded.

10. Supply curve has a positive relationship because of the direct relationship between price of the commodity and the quantity supplied.
11. Change in quantity demanded or supplied is the movement along the same demand or supply curve (contraction or expansion) due to change in price of the commodity.
12. Change in demand or supply is the shift of demand or supply curve (increase or decrease) due to changes in other factors that affect demand or supply other than price of the commodity.
13. Elasticity of demand or supply is the degree of responsiveness of a change in quantity demanded or supplied due to change in one of factors that affect demand or supply.
14. The government may use price ceiling or price floor to control fluctuations of prices, which may disturb both consumers and producers.
15. Price mechanism acts like an invisible hand in the economy in which the forces of demand and supply determine the price of commodities in the market.

Revision exercise

1. "An increase in the demand for computers raises the quantity of computers demanded but not the quantity supplied." Is this statement true or false? Explain and illustrate the answer.
2. Describe factors for change in demand. Provide explanation for each factor.
3. Explain the role of price in the economy.
4. Why the demand curve slopes downwards from left to right?
5. Elaborate the relationship between the supply schedule and the supply curve. Explain the reasons for a positive slope of the supply curve.
6. Given: $Q_d = 40 - 2.5P$ and $Q_s = 20 + 4P$
   (a) Find the equilibrium quantity and price
   (b) Calculate the price elasticity of demand and interpret the answer.
7. Given products A and B in Kariakoo market. Suppose price of product A changes from Tshs 600 to Tshs 500 per unit and the quantity demanded of product B changes from 40 to 20 units per week.

(a) Compute elasticity of demand

(b) Name the concept above

(c) Give two examples of goods A and B

(d) Draw a graph to show the two goods

8. With the aid of graphs, explain how supply and demand may be used to determine price.
9. Assume that the market consists of three consumers A, B and C, whose individual demand functions are given below

   (a) $P = 35 - 0.5Q$

   (b) $P = 50 - 0.25Q$

   (c) $P = 40 - 2Q$

   Required:

   (i) Find the market demand function for the commodity

   (ii) If the market supply function is given by $Q = 40 + 3.5P$. Determine the equilibrium price and quantity.

   (iii) Calculate the price elasticity of supply at the equilibrium.
10. "Elasticity of demand is not useful at all. We only calculate it for leisure" Discuss.

Chapter Three

Production theory

Introduction

In a construction industry, various materials such as sand, steel, wood, stones and bricks are used to build houses and other infrastructure like roads, bridges and railways. This implies that production process requires inputs to produce an output. Thus, the production theory shows how labour, capital and other inputs are combined to produce output. In this chapter, you will learn various concepts related to production, including factors of production, scale of production, the production function, the law of diminishing returns and economies of scale. The competencies developed will enable you to choose the optimal level of factors of production that will minimise cost of production.

Production

Producers create goods and services consumed by the society and goods required for production of other goods. Production of goods and services requires the use of various resources at their disposal. The required resources are scarce, which is at the core of the subject matter of economics. In this regard, producers have to make rational choices in allocating efficiently the scarce resources.

The production process involves transformation of inputs to outputs in form of goods and services to satisfy human wants or for producing other goods and services. Such transformation involves changing materials from one form to another. For example, carpenters use wooden materials as inputs to produce outputs in form of furniture such as tables, chairs, beds and cupboards. The production process creates value for firms, as depicted in Figure 3.1.



Figure 3.1: *The production process*

A firm is an entity or organisation (company, enterprise) that produces goods or provides services with the aim of making profit. It is assumed that the fundamental objective of the firm is to maximise profit. However, the firm may have other objectives, for example, acquiring the largest market share, enhancing public relations and creating employment. Profit maximisation is the overriding objective because most firms that stay in business over a long-run period focus on maximising profits, regardless of their other goals.

Inputs are productive resources used by a firm to produce output of goods and services. Inputs are commonly known as factors of production that include land, labour, capital and entrepreneurship.

Outputs are goods or services produced by a firm.

Factors of production

Factors of production are inputs used in the production process. They are also known as agents of production or productive resources. Generally, they are classified into four major groups, namely, land (including all natural resources), labour (including all human resources), capital (including all man-made resources), and entrepreneurship (which brings and organise all resources together for production).

Land

In economics, land comprises all natural resources used in production. It includes all that nature provides, such as minerals, forest products, water and other land based resources. Land is a factor of production where almost all kinds of productive economic activities are conducted. Activities are in many sectors, such as agriculture, manufacturing, mining and construction.

Features of land as a factor of production

The following are features of land as a factor of production:

Land is a primary factor of production: Although all other factors are required for production, land provides the foundation for the production process to take place. For example, to produce maize flour, a producer requires land for construction

of a factory, whereas a farmer requires land for maize cultivation, without which one cannot have maize flour.

Supply of land is fixed: Since the size of the earth surface cannot change, land can neither be increased nor decreased. However, to an individual or a productive entity, supply of land is relatively elastic because economic agents can buy or sell it in the market.

Land differs in quality: The quality of land differs from one area to another. Thus, uses of land vary depending much on its quality. For example, the land in Kilimanjaro region is suitable for production of food and cash crops, such as bananas and coffee. However, the land in Singida Region is suitable for production of drought-resistant crops, such as millet and sunflower.

Land is a passive factor of production: Land itself cannot produce anything without application of other factors of production. To make it a factor of production, producers work to develop it using capital and labour. For example, beans cannot be grown on land without existence of a labourer (farmer) and capital (seeds and equipment).

Land utilisation is a derived demand: Demand for land is a derived demand. Its demand depends on the demand for other goods or services. For instance, the demand for land depends on demand for agricultural products, industrial products and settlements.

The price of land is rent: Rent is a reward for use of land. Classical economists such as David Ricardo connected rent with fertility of land, whereas Alfred Marshall emphasised that land earns rent because of its scarcity.

Factors affecting productivity of land

The following are factors affecting productivity of land:

Quality of land: Productivity of land depends on its natural characteristics, such as topography and altitude. For example, flat or levelled land can be more productive because it is conducive for economic activities (such as construction of settlements, agricultural activities and development of infrastructure).

Means of irrigation: Land with good means of irrigation system such as drip, sprinkler and tube wells is more productive than one depending on rainfall. The reason is that irrigation improves availability of water for crops to grow.

Proper use of land: Productivity of land depends directly on its proper utilisation. For example, black soil is suitable for cultivation of cotton and groundnuts. However, if such land is used for cultivation of rice, its productivity may be low. Also, land becomes more productive when small portion of it is used for building high-rise buildings for offices and settlements in urban areas than using it for a single storey building.

Improved methods of cultivation: Improved methods of cultivation enhance productivity of land. For example, use of fertilisers, quality seeds and mechanised ploughs in the crop cultivation increases productivity of land.

Usage with other inputs: Productivity of land depends directly on other inputs. For example, skilled labourers are more efficient than unskilled ones. Thus, if labourers are efficient in sowing seeds, watering plants and spraying pesticides at the right time, the productivity of land will increase. In addition to skilled labour, the productivity of land depends on extent to which it is combined with capital.

Pricing of land

Price of land is rent. In economics, rent is explained by two theories, namely, the Ricardian theory of rent and the Modern theory of rent.

Ricardian theory of rent

This is the earliest theory of rent proposed by David Ricardo. It explains rent as a portion of produce of earth that is paid to the landlord in exchange for the use of the original and indestructible powers of soil.

The Ricardian theory of rent assumes that:

(a) Supply of land is fixed because it is a gift of nature;
(b) Rent arises because of inequalities in soil fertility;
(c) Land contains indestructible powers;
(d) Land is characterised with diminishing returns.(Although the overall cost of production in each plot of land is the same, the yield from the plots with low fertility of land is low);
(e) Land is the only factor of production that earns rent;

(f) The most fertile land will be used first, followed by the less fertile land; and
(g) Land has only one use of cultivation.

David Ricardo believed that land could be divided into different categories referred to as grades on the basis of its fertility. Also, David Ricardo believed that human beings begin cultivation by using land of the best quality and subsequently, go to the less fertile/quality land. According to Ricardo, the process continues until an individual reaches that piece of land where the amount of money invested in it equals the revenue earned. At that point, the piece of land would have no residual amount and rent would be zero. Such land would be referred to as marginal land.

Table 3.1 shows the rent charged for different grades of land. Plots A, B and C have similar size and cost incurred on each piece of land is the same. The table shows grade A of land is of the highest quality, as its rent is higher than the other two grades of land. Moreover, from land grade A to B and C, the rent for the same pieces of land keep on decreasing. The cost incurred on land grade C is equal to the value of output produced. This is referred to as marginal land. Marginal land produces no surplus over the cost of production used on it.

Table 3.1: *Rent charged for different grades of land*

Land grades	A	B	C
Incurred cost (Tshs)	500	500	500
Output	1000	800	500
Rent charged	200	110	0

The Ricardian theory of rent has been criticised by some modern economists in that rent is not derived solely when land is used for farming. Rent is not paid because of soil fertility, but because of its scarcity. Moreover, Ricardo's statement that the most productive land is cultivated first and followed by the less fertile one is not supported with evidence. In practice, the nearest land is cultivated first. Nevertheless, Ricardo's concept of marginal land is idealistic because paying rent is not only for cultivated land, but also for various productive activities.

Modern theory of rent

In the modern theory of rent, rent is a surplus that is not limited to land. It can arise with respect to any factor of production. Thus, the demand and supply of land is the main determinant of rent, which is a reward for use of land. Thus, in the market, rent is determined at the equilibrium point where demand and supply intersect. The demand for land is a derived demand, which depends on demand for goods it produces. When a farmer or a firm demands for land, it is beacuse of its productivity. Thus, the farmer or firm will pay rent that equals to the marginal value product of land. The law of demand applies to land as well, and thus, the demand curve for land slopes downwards, which implies that as rent decreases, more land is demanded, other factors remaining constant.

Transfer earnings

The modern theory of rent also defines rent as a surplus, which is the difference between actual and transfer earnings.

Rent = Actual earnings – Transfer earnings

Actual earning of a unit is the price received by a factor for selling or supplying its services. Its transfer earnings are minimum prices essential to keep it in use. Thus, transfer earnings is the minimum amount of money that a factor of production can earn to keep it in its present use rather than in alternative uses.

Quasi-rent

According to Alfred Marshall, rent is referred to as income derived from use of durable factors whose supply is fixed in the short-run period of production. Because durable factors are fixed in supply, when their demand increases such factors earn a surplus referred to as quasi rent. This is a temporary phenomenon, which does not apply in the long-run where supply of goods can be increased. For example, the normal wage of a pilot could be Tshs 1,200,000 but might be paid Tshs 2,000,000 per month due to shortage of qualified pilots in the short-run. Therefore, the pilot earns a quasi-rent of Tshs 800,000 per month.

Labour

Labour is a factor of production, which includes the human mental ability and physical effort employed in production of goods and services.

Types of labour

There are three types of labour, namely, skilled, semi-skilled and unskilled labour.

Skilled labour: Constitutes workers who use more mental effort in the production process. Examples of skilled labour include doctors, teachers, lawyers and accountants. Skilled labourers are highly educated, trained and well experienced in performing their tasks.

Semi-skilled labour: They are workers who combine both mental and physical efforts in production. Shopkeeper, tax driver and waiter or waitress are examples of semi-skilled labour. Semi-skilled labour require certain ability and training to perform specific tasks. They do not need advanced education or specialised skills.

Unskilled labour: This type of labour constitutes workers who use more of physical effort than mental effort in production. Cleaners and janitors are example of unskilled labour.

Features of labour

Labour has the following features:

Labour is inseparable from its labourer: Labour cannot be separated from its labourer. Workers sell their services, which cannot be separated from themselves.

Labour is a perishable factor: Labour cannot be stored. Once labour is lost, it cannot be recovered. For instance, unemployed workers cannot store their labour for future employment.

Cost of producing labour cannot be determined: It is easy to calculate cost of producing a commodity in an industry. But the cost of producing labour is a slippery concept because it includes even expenses incurred by parents to raise and educate their children. Thus, such costs cannot be established accurately.

Labour is an active factor of production: Labour is an active factor because a labourer is a human being. For instance, a firm can produce high and quality level of output when land and capital are combined with labour. Therefore, labour is regarded as an essential input in the production process.

Labour is a heterogeneous factor: Two persons cannot possess the same quality of labour. Skills and efficiency differ from one person to another. Some workers may be more efficient than others in the same job, inspite of acquiring the same education and training.

Labour supply is inelastic in general: Supply of labour depends on many factors such as size of population, age, gender, desire to work, quality of education and attitude on work. Thus, supply of labour cannot be changed easily in response to changes in its demand.

Labour demand is a derived demand: Like other factors of production, labour has an indirect demand. Demand for labour depends on demand for the goods labours are employed to produce. For example, the demand for furniture by consumers requires carpenters to devote skills and physical effort to produce the furniture. Thus existence of demand for furniture causes an increase in demand for carpenters which is a demand for labour.

Labour is a means as well as an end: Labour is a means of production for a firm. However, outside the firm, workers are consumers of outputs produced by firms. Therefore, the same people who supply labour in the production process are end users of the goods and services produced.

The reward or payment for labour is wage: Wage is per unit price of labour paid in monetary terms (for example, Tshs per hour, or Tshs per day) for supplying labour in the production of a good or service. The total wage bill for producing a good or a service is the cost of labour of the firm.

Productivity of labour

Productivity of labour is the production capacity of labour in a given period measured as output per unit of labour. The higher the output per unit of labour, the higher the productivity. Typically, productivity is influenced by various factors such as, hereditary, culture and individual qualities of a labourer. In addition, the level of education and training of the labourer influences the productivity of labour. Workers with higher level of education and training have higher skills for performing their tasks and therefore, they have higher labour productivity than workers with lower levels of education and training.

The productivity of labour may also be influenced by working hours and wages that workers earn. A worker who is better or highly paid will be more motivated to increase productivity than a lowly paid worker. Furthermore, a conducive working environment may lead to raise productivity of labour, whereas poor working conditions may reduce the productivity of labour.

Mobility of labour

Mobility of labour is the extent to which the workers are willing and able to move between jobs and across geographical areas. Mobility of labour helps to increase supply of labour. For example, when soldiers go to neighbouring countries to ensure peace and security or when doctors go to other countries to work.

Mobility of labour helps to solve unemployment problem and increases the income of workers when they move to places where they are required. It enables labour to increase efficiency and productivity when used in better suited occupations. Mobility of labour also enhances skills and experience of workers through on-the-job training.

Types of labour mobility

The following are types of labour mobility:

Geographical labour mobility: This refers to the case at which workers can to relocate in order to find gainful employment that reflects their training and occupational interests. It is the movement of labour from one geographical area to another.

Occupational labour mobility: This refers to an ease with which workers can switch between different careers to find gainful employment or to meet their goals. Higher levels of occupational labour mobility help to maintain strong employment and high levels of productivity. This has caused many governments to provide occupational re-training to help workers acquire necessary skills and become more mobile vertically, which means going up the ladder in terms of employment. Occupational mobility of labour has two types, namely, horizontal and vertical mobility.

(a) Horizontal mobility

This is the movement of labour from one employment to another employment without changing the worker's grading or position. For example, when a nurse leaves his or her present job in hospital A to join hospital B in the same grade or position.

(b) Vertical mobility

This is the movement of labour leading to a change in the worker's grading or status or occupation from a junior position to a senior position. For example, changing position from the normal teacher to academic teacher or to the headmaster or headmistress.

Factors affecting mobility of labour

Mobility of labour depends on the following factors:

The level of education and training: Mobility of labour depends on extent to which a labour is educated and trained. The higher the worker's level of education and training, the greater the chances of moving from one occupation to another.

Improved infrastructure: Well-developed transport and communication network encourages mobility of labour. Workers can easily communicate and travel back home in case of emergency.

Industrialisation: Mobility of labour is influenced by industrial development. Workers move from different occupations and places to seek jobs in established factories. Industrialisation also leads to development of towns, industrial centres and big cities which attract workers to move from rural to urban areas in search for employment in industries.

Political stability: Mobility of labour depends on existence of peace and security in a country or region. If people are sure that their lives and properties are safe, they will not be inclined to move to other areas. However, political instabilities also tend to promote labour mobility as labourers will move to find jobs in areas with peace and security.

Mind set: The mind set of workers determines their mobility. Ambitious, optimistic and broad-minded workers tend to move to other jobs and places. Differences in language, habits and religion, among others, will rarely limit their mobility.

Specialisation and division of labour

Specialisation exists when factors of production are employed to produce specific parts of a good or specific task of a service. In economics, specialisation of labour enables workers to concentrate on performing specific tasks during the production process. This entails an increase in productivity of a worker since specialisation enhances development of the worker's skills for specific tasks.

In order to utilise available resources efficiently, Adam Smith emphasised on the importance of division of labour, which entails categorisation of workers to

perform specific tasks that they do relatively better/efficiently during production of a good. Therefore, specialisation involves splitting up the production process into its component parts. Thus, the task of producing a good is sub-divided among workers as per their relative performances of such task.

In a television station, there are different workers who perform separate roles towards production of news. For example, there will be an anchor who reports the news, a camera operator, the director, news director, web master, audio engineer and a broadcast technician. Each worker is engaged partly in the production of the same product, namely, the news.

Forms of division of labour

Division of labour can be of the following forms:

Simple division of labour: This occurs when production is split into different parts and many workers team up to complete the production process. For example, in cloth manufacturing industries, workers are categorised to perform various functions, such as pattern making, cutting, sewing, printing, packaging and folding. Different groups of workers perform each of the aforementioned functions to ensure production of cloth is completed.

Complex division of labour: This occurs when the production process is split into different parts and a qualified person is needed to perform a specific part. For example, each secondary school teacher teaches specific subjects.

Occupational division of labour: This occurs when production of a commodity is specific to occupation of a worker, for example, the work of carpenters, medical doctors, farmers and fishermen.

Advantages of specialisation and division of labour

Specialisation and division of labour has the following advantages in the production process:

Time saving: Division of labour increases one's ability and efficiency in performing work. Hence, it saves time and reduces costs of production. Overtime, the worker performs the same task faster.

Increase in worker's skills: Division of labour enhances development of workers' skills due to repetition in performing the same task, which adds to perfecting the tasks. Specialisation enables workers to perform their assigned work in the best possible way, which facilitates production of quality goods.

Increase in employment opportunities: Division of labour leads to an increase in employment opportunities due to diversity of occupations, which leads to further creation of jobs. Specialisation increases employment because a number of people work to produce one good.

Increase in use or employment of machines: Division of labour paves way for use of machinery and mass-production methods.

Increase in work professionalism: Division of labour increases professionalism when a person regulary performs a specific task in production of a commodity. The worker becomes more professional, which might result to occupational mobility.

Disadvantages of specialisation and division of labour

The following are disadvantages of division of labour and specialisation:

Monotomy: Division of labour makes a worker to perform the same task repeatedly. The process may be boring to the worker.

Risk of unemployment: By knowing only a specific part of a job, a worker is exposed to the risk of being unemployable. If the worker loses the present job, it may not be easy to secure another job.

Increased dependence: Absence of one worker would affect the whole production process especially when production is split into a number of processes performed by different workers. This is because the workers depend on each other to accomplish the task. For example, in a hospital, if the laboratory technician fails to show up, the work of diagnosis will not be performed, and hence, patients will not get medical treatment.

Less creativity and mental development: When workers are trained to perform a specific task, they become less creative because they lack knowledge of the complete work.

Rising costs of inputs: Division of labour applies well in large-scale production, which may lead to an increase in demand for factor inputs. A firm has to hire a

labour for each specific task to be performed. The increase in demand for factor inputs may lead to an increase in their price, which subsequently, increase the cost of production.

Pricing of labour

The reward for use of labour is wage. The wage is payment for both mental and physical services offered by the labour in the production process. Wage includes fees, salaries and commissions. There are various forms of wages, however, the main ones are real and nominal wages.

Nominal wage: Nominal wage, also known as money wage, is the amount of money income a labour receives for services it provides during the production process.

Real wage: This is referred to as the purchasing power of the wage income earned by labour. It entails the basket of goods and services offered by labour out of its nominal wage. In the aggregate form, real wage is the ratio of nominal wage to the general price level.

Factors affecting real wages

Real wage can be low or high due to the following factors:

Level of prices of goods and services: When prices of goods and services increase, the purchasing power of money decreases, which causes a decline in the real wage of the worker and vice versa.

Amount of nominal wage for workers: The higher the nominal wage of a worker, the higher is the real wage and vice versa.

Extra benefits to workers: Existence of extra benefits, such as free medical services, free accommodation and transport increase real wages and vice versa.

Nature of the work: If work is enjoyable and pleasant, it could lead to an increase in real wages of workers because they would work for more hours.

Future prospects: An occupation that promises better prospects of promotion in future affects the real wage of the worker more than the one with no future prospects.

Social prestige: Employees engaged in prestigious occupations have higher real wage than employees working in ordinary occupations. Because of existence of extra benefits of workers in the prestigious occupations.

Working condition: Working condition also affects real wages. Lack of conducive working conditions, such as too much heat and noise can affect the health of workers, which may lead to low productivity and low real wages.

Systems of wage payments

Rewards or payment to labour can be done by using different systems. Some of the wage systems used are time rate system, piece rate system and profit sharing system.

Time rate system

Time rate system also called the flat rate system is used for wage payment for a specific period. For instance, a worker can be paid daily, weekly or monthly. The wages paid depend on the time that a worker spends in producing goods or services.

Advantages of time rate system

The following are advantages of time rate system:

(a) It increases efficiency of workers;
(b) It reduces wage differentials;
(c) It enhances the government in taxing the worker's income;
(d) If facilitates formation of trade unions for workers with the same occupation; and
(e) It facilitates easy estimation of labour costs.

Disadvantages of time rate system

The following are some disadvantages of the time rate system:

(a) It has high cost of supervision and record keeping;
(b) It may lead to low output;
(c) It may make an employer to pay a lot of money at once; and
(d) It is difficult to identify hard workers, since all workers with the same number of working hours are paid equally.

Piece rate system

This refers to the type of wage payment system in which a worker is paid

according to work performed. This means that the more the work done, the more the payment and vice versa. Sometimes, piece rate system of wage payment is known as 'payment by a result.'

Advantages of piece rate system

The followings are some advantages of the piece rate system:

(a) It reduces supervision cost;
(b) It may result into increase in output;
(c) It enables an employer to determine hard workers in the firm;
(d) The system facilitates creativity and innovation during production; and
(e) It makes it easier to determine the wage rate.

Disadvantages of piece rate system

The following are some disadvantages of the piece rate system:

(a) It creates difficulties for the government in imposing tax from income of workers;
(b) It may lead to low quality products due to rushing in completing the work;
(c) It may accelerate wage differentials among workers in the firm;
(d) It is difficult to be applied in the case of tasks that are difficult to measure; and
(e) Workers may be injured.

Profit sharing system

This refers to the type of wage payment system in which profit is shared among workers according to agreed rates. Employees are paid a share of all net profits of the company that employs them in accordance with a written formula defined in advance. The level of wages varies among employees according to their salary scale.

Advantages of profit sharing system

The following are some advantages of the profit sharing system:

(a) An increase in productivity;
(b) Reduction in labour turnover;
(c) Building team spirit among workers;
(d) Strengthens cordial relations between management and employees; and
(e) Provides workers with additional income.

Disadvantages of profit sharing system

The following are some disadvantages of the profit sharing system:

(a) Individual employees cannot see how their own work and actions impact on profitability of the company;

(b) Reduces individual performance;

(c) Uncertainty of profits;

(d) Profits may not be accurately shown by management; and

(e) Inadequate incentives for those who work hard.

Theories of wages

There are various theories of wage determination that have been put forward by various economists from time to time. These include the following: subsistence theory of wages, marginal productivity theory of wages, bargaining theory of wages, market (Modern) theory of wages, wage fund theory and the residual claimant theory.

The subsistence theory of wages

This theory is sometimes referred to as the iron law of wages. It examines labour power as a commodity, whose price is determined by its cost of production. Thus, wages should be determined at the level sufficient enough to meet basic requirements of workers at mere minimum subsistence. If at any time, wage exceeds the subsistence income, workers will find themselves in a good economic position. The situation might lead to an increase in birth rate, which will increase population, and consequently, its supply of labour. This situation will continue until the wage falls to the minimum subsistence level of income. However, if it happens that the wage falls below the subsistence income level, there will be a reduction in the population size, for example, due to starvation. As a result, the supply of labour will decrease.

Assumptions of subsistence theory

The subsistence theory assumes that:

(a) Wage is used to meet only the basic needs of labourers;

(b) All workers are paid the same wage rate at the subsistence level;

(c) Population is determined by the change in the wage rate. Thus, an increase in wage paid to a worker increases the population rate and decrease in the wage decreases population rate; and

(d) Workers demand the same requirements, such as food, clothes and shelter.

Criticisms of subsistence theory

The following are criticisms of subsistence theory:

(a) The theory was unable to explain the heterogeneity nature of wage rates, that is, changes in level of wages among workers that occur over a certain period of time.

(b) The theory explains wage determination by focusing on the supply side of labour and completely ignores the demand side.

(c) The theory ignores efficiency and productivity of workers, since it assumes that all workers earn the same wage at the minimum subsistence level. However, in practice, efficiency and productivity of workers differ.

(d) The subsistence theory of wage assumes wage is a determinant of population changes. In practice, it is difficult to explain the direct correlation between change in the population and wage rate.

Marginal productivity theory of wages

The marginal productivity theory of wage posits that reward for labour is determined at a level equal to value of its marginal product. The marginal product of labour is an additional output into the total product that results from the use of one additional unit of labour in production of a good or service. Thus, in a perfectly competitive market, the wage rate tends to be equal to the marginal product of labour. That is, the higher the productivity of labour, the higher the wage that will be paid, while the lower the productivity of the labour, the lower the wage that will be paid.

According to the marginal productivity theory, an entrepreneur will continue to employ workers up to the point where the product value of the last worker (the marginal value product) is equal to the additional cost of employing that additional worker. Such additional cost is equal to the wage.

Assumptions of the marginal productivity theory of wage

The marginal productivity theory of wage assumes:

(a) Perfect mobility of factors of production, such as labour and capital;

(b) Employers are able to predict the marginal productivity of a factor;

(c) There is an equal bargaining power between labour and the management;

(d) There is perfect competition in the market and thus, workers compete freely for jobs; and

(e) During the production process, labour can be increased or reduced and an employer can substitute one factor of production for another.

Criticism of the marginal productivity theory of wage

The following are criticisms of the marginal productivity theory of wage:

(a) The assumption of perfect competition in both product and factor markets is rigid because under imperfect competition rewards for factors of production are unequal to the value of marginal product.

(b) The theory ignores the role of labour unions in influencing on wage rate of workers.

(c) The theory ignores the interrelation between wages and productivity or efficiency of workers, since the rise in workers' wage has a positive effect on productivity of labourers.

(d) The theory does not explain the reward of factors of production used in fixed proportions. When factors of production are used in fixed proportions, an addition of one factor while other factors remain unchanged, will not lead to an increase in the total product. Thus, the marginal product of the factor will be zero.

(e) Productivity of labour does not only depend on the effort of labour but also on productivity of other utilised factors, such as capital.

Wage fund theory

This theory, which was proposed by John Stuart Mill (1806-1873), explains that the wage rate of workers depends on the proportion between capital available and the population. Thus, the workers' wages are paid from the wage fund available for wage payment. According to this theory, wage is determined by two factors, wage fund and population.

Wage fund: This is the amount of capital that is set by an entrepreneur for direct purchase of labour services.

Population: This determines the labour force available for employment. A large number of workers implies a low wage rate and a small number of workers implies a high wage rate. Therefore, wages depend on the ratio between wage fund and population.

Thus,

$$\text{Wage rate} = \frac{\text{Wage fund}}{\text{Population}} \quad \ldots\ldots\ldots\

labour market are able to raise wages of workers by reducing the supply of labour. This is because the forces of demand and supply determine wages.

Roles or objectives of a trade union

The following are some of the roles or objectives of a trade union:

(a) To improve the working conditions and wages of its members through bargaining;

(b) To advise the government with regard to economic planning, especially with reference to employment;

(c) To protect members against unfair treatments by employers and ensure job security of its members, for instance, a trade union may protect a member against unfair dismissal from an employer;

(d) To ensure constant and persistent supply of labour force to employers; and

(e) To improve members' skills, for example, through seminars organised by trade unions.

Approaches used by a trade union to achieve its objectives

The following are approaches used by a trade union to achieve its objectives:

(a) A trade union achieves its objective through threatening to strike, if their demands are ignored;

(b) Through collective bargaining that involves peaceful negotiations between representatives of a trade union and an employer about implementation of the trade union's demands, such as wage increments and improvement in workers' conditions;

(c) Through sabotage, that occurs when workers convince consumers not to buy the products produced by certain firms; and

(d) Through compulsory arbitration, when the government gives out directives to both employer and employees on what should be done without taking sides.

Determinants of strength of trade unions

A trade union is formed to protect the interest of its members. However, its strength (power) is affected by the following factors:

Productivity of labour: If the productivity of labour is high, the willingness of an employer to raise wages and to improve the working conditions will be high, which in turn, fosters the strength of a trade union. However, if productivity is low, then the employer will not have any motivation to raise wages, which in turn, may weaken the trade union.

Elasticity of demand for the final product: When the final product produced has inelastic demand, a trade union tends to becomes more powerful in defending the interest of its members. On the other hand, when the final product has elastic demand, the trade union becomes less powerful.

The size of membership: When a trade union has a large number of members, spread all over the country, it will have strong bargaining power so that its strength will be high.

Government policy: If a government policy supports labourers, the strength of trade union will be high and vice versa.

Substitutability of labour: If the labour offered by workers or members of trade union is substitutable, the strength of trade union tends to decrease and vice versa.

Scale of production: If the scale of production for a firm is large, the strength of the trade union will tend to increase and vice versa.

Economic or financial power: A strong trade union that is capable of hiring lawyers and financing advertisements in various media will be able to increase its power.

Circumstances under which workers demand for higher wages

The following are circumstances under which workers demand for more wages:

(a) When the nature of work entails high risk;
(b) When there is increase in the cost of living;
(c) When there is increase in labour productivity (or efficiency);
(d) When an employer is earning an abnormal profit; and
(e) When there are wage differentials among the workers in the industry.

Supply of labour

Supply of labour constitutes the number of workers that would offer themselves for employment at each possible wage rate. Usually, the higher the wage, the higher will be the quantity of labour supply. This implies an upward sloping labour supply curve.

The supply of labour depends on several factors, such as population size, number of working hours, the level of education and training, labour laws with regard to employment of children as well as women and attitude of society on employment of women. Other factors influencing on supply of labour are attitude of labour on work, leisure and mobility of labour.

From the market theory of wages, the wage rate in the industry will be determined at a point where the demand for labour equals the supply of labour. In Figure 3.2, OE is the number of workers employed at the wage W. Equilibrium wage occurs at point B where the demand for labour equals to the supply of labour. A wage above equilibrium level, for example W_1 will induce more workers to offer themselves for employment and the firm would be induced to cut down employment. Thus, wage will be brought down to W level. On the other hand, a decrease in the wage level to W_2 would lead to decrease in the number of workers and in a bid to stop this, the firm would offer higher wages and the equilibrium level of wages would be restored.

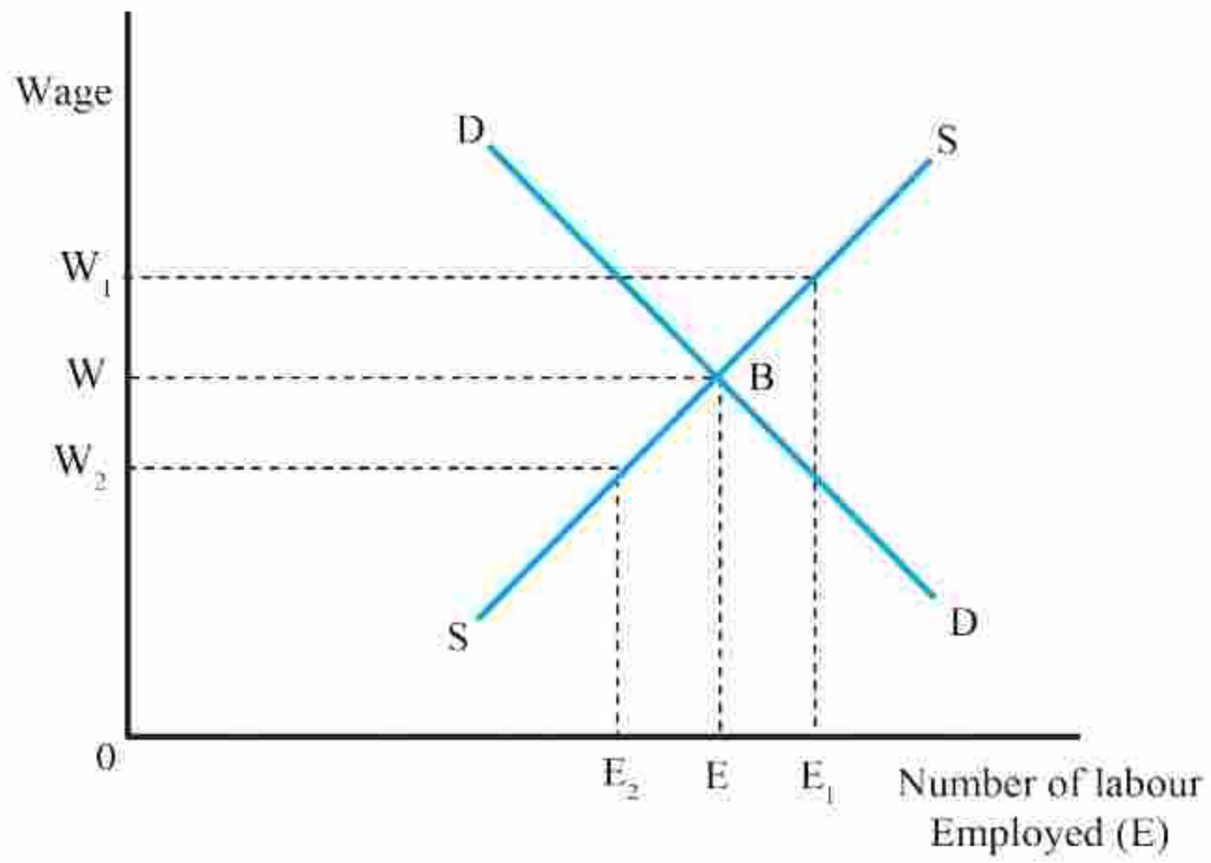


Figure 3.2: *Wage determination by the intersection of demand for and supply of labour*

Residual claimant theory of wages

Residual claimant theory of wages is accounted by an American economist Francis A. Walker. According to the theory, payment for labour is a residual after other factors of production, such as land, capital and entrepreneur are paid. Thus, if labour puts more effort and increases output without increasing the input of other factors, the residual amount that labour claims in form of wages will be enhanced.

Criticism of the residual claimant theory

The following are criticisms of the residual claimant theory:

(a) It lacks the positive method of determining what the level of rent, profit and interest would be. It means that the theory does not specify the way of determining distribution of rewards to land, capital and entrepreneurship from total output.

(b) It does not explain the role of trade unions in changing wages.

(c) A worker is not the residual claimant. Rather, it is the entrepreneur who is obliged to pay other factors of production before he or she gets anything.

(d) It ignores the influence of supply of labour on wages. This means that the theory leans on demand for labour as having full influence on wage.

Wage differentials

Wage differentials refer to differences in wages in various occupations or employments or grades of employments and also between different persons in the same employment or grade.

Reasons for wage differentials among workers

The following are some of the reasons for wage differentials among workers:

(a) Differences in levels of education and training;

(b) Level of experience of a worker;

(c) Productivity of a worker;

(d) Strength of a trade union;

(e) Factors influencing on demand for and supply of labour in the market;

(f) Regularity and irregularity of employment. Such that regular employment generally carries a low wage; and

(g) Risk involved. Generally, high-risk jobs entail high wages.

Modern theory of wages

This refers to demand and supply theory or market theory of wages. Modern theory asserts that forces of demand and supply of labour in the labour market determine wage. The demand for labour is a derived demand. For example, the demand for clothes stimulates demand for labour to produce them.

Several factors affect demand for labour, such as, price for labour, productivity of labour, substitutability of labour, government policy, technical progress and nature of demand for products produced by labour.

Assumptions of the modern theory of wages

The modern theory of wages assumes that:

(a) There is perfect mobility of labour;
(b) There is no government intervention in wage fixation;
(c) There are many workers and employers in the labour market;
(d) There is freedom of occupation, which means that an employee can choose to work with any employer and an employer can choose to work with any employee; and
(e) Labour is homogeneous.

Capital

Capital is a human-made resource such as machines, equipment and buildings used in production of goods and services. For example, capital for a coffee farm includes a pair of scissors used in pruning, warehouse for storing coffee harvests, trucks for transporting the coffee and funds used to purchase land on which coffee is grown. Since capital is owned by individuals, the owner earns income from use of capital in production.

Capital in relation to wealth

A person may use part of his or her wealth to purchase items used in production of goods or services. For example, a person may decide to use part of his or her wealth to purchase a motor vehicle for business, such as providing taxi services. In such business, the vehicle is used as capital that generates income to the owner. Hence, that part of wealth used in the production process constitutes capital. But if the vehicle is used for non-business activities, it is still part of an individual's wealth, though it is not capital (see Figure 3.3).

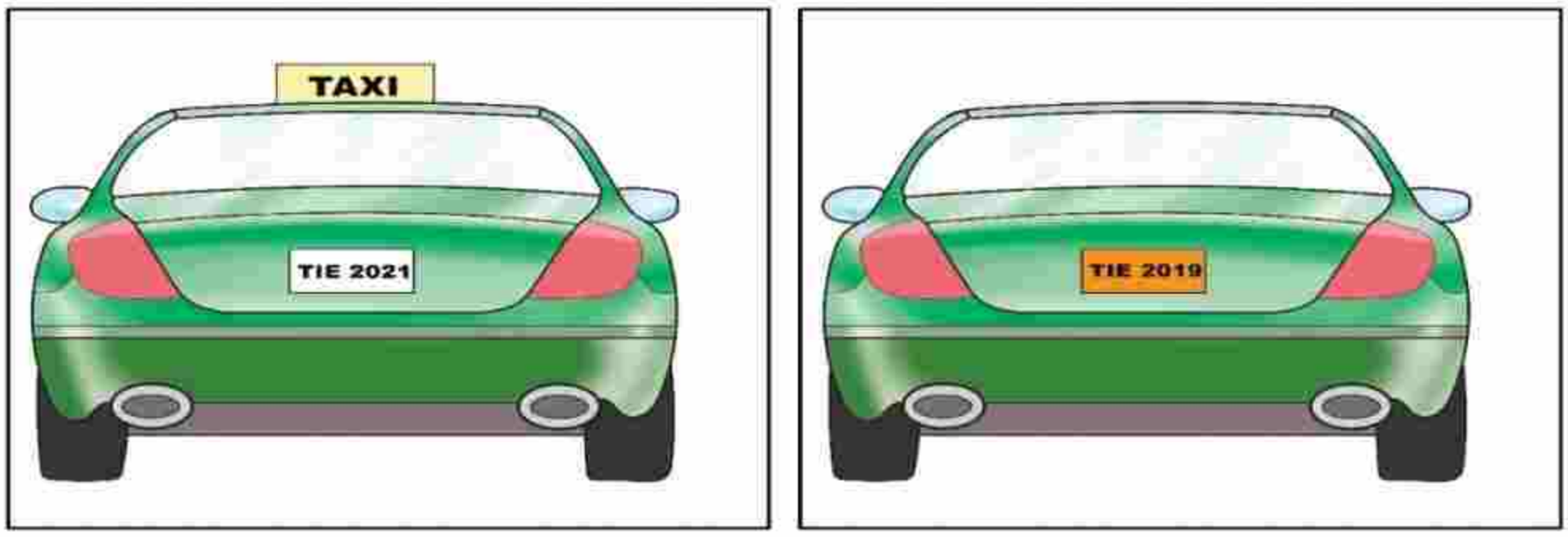


Figure 3.3: *Relationship between wealth and capital*

Capital in relation to money

Money is not a factor of production unless it is used to acquire a stock of capital goods used in production. If a firm uses money to purchase a piece of land to produce commodities for profit, then that amount of money used is referred to as capital. Similarly, money borrowed by a small scale enterprise to start business is also referred to as start-up capital.

Characteristics of capital

The following are characteristics of capital:

Capital is a human-made factor: Capital is not a gift of nature. It is secondary and a non-natural factor of production.

Capital has elastic supply: Compared to land, supply of capital depends on the capital formation process. Capital formation depends on savings and investment. Accelerating capital formation increases supply of capital over the long-run period.

Capital has a social cost: Resources have alternative uses. They can be used to produce capital goods or consumer goods. When resources are used to produce capital goods, it means the economy foregoes production of consumer goods.

Demand for capital is a derived demand: As a factor of production, demand for capital is a derived demand, as it is used to produce finished goods, which have direct demand. For example, the demand for raw cotton is derived from demand for cotton cloth. Therefore, people invest in capital goods first to produce consumer goods or other capital goods.

Capital has easy mobility: It is easy to transfer capital from one place to another. It is due to mobility of capital that there exists foreign direct investment (FDI) which facilitates generation of more income.

Capital is wealth: Capital has all features of wealth, which are utility, scarcity, transferability and externality. Even though capital is wealth, wealth is not necessarily capital.

The price of capital is interest: Interest is a reward for use of capital in an economic activity.

Capital is a productive factor: Capital helps to increase productivity and speed up production.

Types of capital

There are different types of capital, namely, fixed capital, working capital, money capital and real capital.

(a) Fixed capital

Fixed capital constitutes durable capital goods and assets used in production of goods and services. For example, machinery, equipment, means of transport and buildings.

(b) Working capital/ Circulating capital

Working capital includes inventory of finished goods, work in progress, raw materials and cash held by a firm and what it expects to receive from goods and services sold.

(c) Money capital

Money capital, also known as liquid capital, is the amount of money available to an enterprise for purchasing capital goods, such as raw materials. Money capital is required for the acquisition of fixed capital and for purchasing raw materials and payment of wages.

(d) Real capital

Real capital consists of assets used to produce goods. For example, farmers use land as capital to produce crops, which generate profit from their sale.

Advantages of capital

A producer cannot produce goods or raw materials without using capital. Thus, capital creates employment opportunities, as workers are needed to produce

capital goods. Capital enhances acquisition of modern technology in production. Technology facilitates an increase in productivity through an increase in output and it is the core of economic development in any economic system. The reason is that economic development is achieved mainly through using capital, such as machines, well-developed roads, improved agricultural tools and equipment, among other things.

Price of capital

There are various theories of capital that have been put forward by various economists. They include the following: classical theory of interest, loanable fund theory of interest and liquidity preference theory.

Classical theory of interest

According to classical theory of interest, the rate of interest is determined by demand for capital and supply of saving. Money used to buy capital goods is made available by those who save their current incomes. Interest paid to depositors is compensation for sacrificing current consumption for future consumption. If interest is high people will be motivated to save more. The supply curve of saving is upward sloping from left to right. The equilibrium level of interest rate is determined by the intersection of demand curve for capital and supply of saving.

Assumptions of classical theory of interest

Classical theory of interest assumes that:

(a) Households form the supply side of capital (depositors) and producers (borrowers) form the demand side of capital;

(b) The theory assumes full employment and capital is one of the factors of production assumed to be fully employed, which implies idle capital does not exist in the economy;

(c) No government intervention, for intervention will affect market forces, and hence, equilibrium interest rate will not exist;

(d) Perfect competitive markets, meaning both borrowers and lenders are interest rate takers, which implies that powers of individual borrowers or lenders have no effect in determination of market interest rate; and

(e) Capital is mobile across industries.

Criticisms of classical theory of interest

The following are criticisms of classical theory of interest:

(a) Assumption of full employment is unrealistic. In practice, there are idle resources in the economy.
(b) The theory ignores changes in the income level. By assuming full employment, the theory ignores changes in income, which affect investment and saving.
(c) The theory assumes that demand for loan depends on demand for goods produced (derived demand), which leaves out the government, as it can also borrow.
(d) The theory assumes that demand for capital is determined only by the rate of interest. However, in practice, the demand for capital may be influenced by other factors, such as profitability of businesses.
(e) It is not always true that savings is from current incomes only. It may be due to accumulation of previous incomes.

Loanable fund theory of interest

According to this theory, interest is determined by demand and supply of loanable funds. The theory explains that interest is the price of credit determined by supply of credit or funds. Demand for loans is formed by governments, businesspersons and consumers. Supply of loanable funds is formed by private individuals and bank credit.

Assumptions of loanable fund theory

Loanable fund theory assumes that:

(a) Perfectly competitive markets and hence, both borrowers and lender are price takers;
(b) Perfect mobility of funds in the markets;
(c) Existence of diminishing returns such that expected return on capital increases at a decreasing rate as units of capital goods continue to be employed;
(d) The theory uses partial equilibrium approach whereby demand and supply of funds determine interest rate, keeping other factors constant; and
(e) Money supply is exogenous.

Criticism of loanable fund theory

The following are criticisms of loanable fund theory:

(a) In practice, money supply can be exogenous or endogenous and thus, the theory is biased in favour of exogenous money supply.
(b) Perfect competition assumption is unrealistic because it is not always that borrowers and lenders are price takers, as sometimes a lender can be a price

maker.

(c) Sometimes people save not only to earn interest, but also for other reasons, for example, for precautionary motives, such as for unpredictable events, like sickness.

(d) The theory assumes that supply of loanable funds is interest elastic, whereas in practice, supply of loanable fund is fixed at a time and hence, it is highly inelastic.

(e) The assumption of perfect mobility of capital is unrealistic.

Entrepreneurship

An entrepreneur is a person with the capacity to organise other factors of production in order to produce goods and services. An entrepreneur bears the risk and performs organisational functions of combining required factors of production. Since reward for entrepreneurship is profit, an entrepreneur may face some factors such as risk and uncertainties that lead an entrepreneur to incur profit or loss.

An ideal entrepreneur should have certain qualities or skills. Such skills include the ability to organise various factors of production effectively. An entrepreneur has to understand all aspects of a business undertaking. The entrepreneur has to be ready to bear risk and uncertainties of a business. He or she should be innovative by adopting modern techniques of production. In addition, the entrepreneur has to make the right decisions at the right time by showing promptness, since a delay in decision-making can increase costs of business.

Roles of an entrepreneur

An entrepreneur has a role of coordinating and organising other factors of production since the entrepreneur takes the final responsibility for proper business operation. However, risk taking is a very important role played by the entrepreneur. Also, entrepreneurs should take risks and they should be able to gauge the risks.

In addition, the entrepreneur is an innovator, which implies that he or she has to introduce new methods of production or improve the old ones. Entrepreneurs are expected to introduce new products and discover new sources of materials.

Furthermore, the entrepreneur has to seek for opportunities to increase business sales. He or she has to identify which market to expand through proper research on customers' needs and advertisement.

Determinants of production

The following are factors that determine production:

Availability of factor inputs: If factors of production are available in abundance, firms will be able to produce large volumes of output, rather than when resources are scarce.

State of technology: Improved technology leads to higher productivity of inputs, which results in higher quantities of output.

Size of a firm: Large firms are capable of producing large quantities of output by using large quantities of resources and advanced technology. In contrast, small firms with small production capacities and limited resources cannot produce large quantities of output.

Price of factors of production: When factor prices are low, firms will be able to employ more factors of production (units of inputs) in order to increase output of goods and services. However, when factor prices are high, firms will only be able to employ few factors of production.

Political and social stability: Existence of stable political and social conditions facilitates the operations of firms and growth of economic activities in the economy.

Infrastructure: Good infrastructure smoothens and promotes growth of productive activities in an economy, whereas poor infrastructure limits the size of production. This is the case because infrastructure facilities are considered to be lubricants of production.

Market size: If there is a large market for a good or service (demand is high) firms will be motivated to produce more quantities to increase their output and volume of sales. On the other hand, a small market size limits production, sales and profit.

Theories of profit

Various economists have come up with theories such as rent theory of profit, risk bearing theory of profit, uncertainty theory of profit and dynamic theory of profit to explain why an entrepreneur is paid or rewarded by profit.

Rent theory of profit

American economist, Francis Walker (1840-1879), propounded the rent theory of profit in line with the Ricardian theory of rent. In this regard, Walker saw profit as

the rent of the least and most efficient entrepreneurs. The superior entrepreneur will earn more profit than the less efficient one. Hence, differences in profit arise due to different abilities of entrepreneurs. The rent theory of profit assumes:

(a) Existence of perfect competition, where all firms are assumed to have the same managerial ability;
(b) Profit is the wage for management ability; and
(c) Profit arises due to differential ability of entrepreneurs.

Criticism of rent theory of profit

The following are criticisms of rent theory of profit:

(a) The theory has been criticised in that there does not exist a marginal entrepreneur or non-profit entrepreneur. An entrepreneur may be inefficient, but other factors of production (labour, land and capital) may still be efficient.
(b) The theory fails to explain why firms make loss even with superior entrepreneurs.
(c) The theory ignores the possibility that profit may be due to the entrepreneur's monopoly in the market.

Wage theory of profit

According to wage theory of profit, profit is determined as wages. In this regard, profit is a type of wage for services rendered by the entrepreneur. According to the wage theory of profit, the entrepreneur is a superior kind of labour and the reward for the entrepreneur's service should be determined as wages are determined.

Criticism of wage theory of profit

The following are criticisms of wage theory of profit:

(a) Profits are rewards for those who take risks. Workers on the other hand, do not take risks.
(b) An entrepreneur takes responsibility for profit or loss, whereas reward for workers (wage) depends on contracts.
(c) Shareholders in different companies earn profit by virtue of their shares, whereas workers have to undergo physical exertion in production to earn their wages.

Risk bearing theory of profit

Risk bearing theory of profit advanced by Frederick Barnard Hawley (1843-1929) who argued that profit is a reward for those who are ready to take risk by investing their resources in a business. For those who are not ready to take risk of investing, their resources cannot get profit.

Criticisms of risk bearing theory of profit

The following are criticisms of the risk bearing theory of profit.

(a) An entrepreneur makes profit not only because of taking risk but also due to efficiency and productivity of other factors of production.
(b) The theory does not distinguish between insurable risk and non-insurable risk.
(c) Profit does not rise due to readiness of the entrepreneur in taking risk rather, a superior entrepreneur is able to reduce the risk.
(d) It is not necessarily that higher risk is associated with higher profit.

Uncertainty theory of profit

Prof. Frank Hyneman Knight (1885-1972) propounded uncertainty theory of profit by arguing that profit is a reward for uncertainty and not risk taking. According to the theory risk is of two kinds: foreseeable risk, such as accidents like fire, car accidents, which are mostly borne by insurance companies, and unforeseeable risk, such as natural calamities like floods, earthquake, drought and other uncertainties. These are borne by the entrepreneur. Hence, according to the theory, the entrepreneur gets profit for bearing uncertainties and not for risk taking because risk can be foreseen or predicted.

Criticisms of uncertainty theory of profit

The following are criticisms of the uncertainty theory of profit:

(a) Uncertainty is not the only factor that is faced by the entrepreneur, but there are factors that entrepreneurs have to consider and address.
(b) Uncertainty bearing is not like other factors of production. Rather, it is a psychological concept.

Dynamic theory of profit

The theory was advanced by Prof John. Bates Clark (1847-1938). He advocated his belief that profit arises in a dynamic economy not a static one. According to Clark, the world and economic circumstances are always changing, for example,

population, human wants (in both quantities and qualities). Hence, a clever entrepreneur who will foresee these changes and produce accordingly, such as new things or modern high quality goods, will earn profit.

Criticisms of dynamic theory of profit

The following are criticisms of the dynamic theory of profit:

(a) It fails to differentiate between a change that is foreseen and one that is unforeseen.

(b) Dynamic changes give rise to profit only when such changes and their consequences are unpredictable.

(c) The theory does not mention profit as a reward for risk taking.



Activity 3.1

You are now aware of factors of production. Select a product you prefer from the area where you live and then:

(a) Choose the most appropriate factor(s) of production suitable for producing the product you have chosen.

(b) Explain the reasons for choosing factor(s) of production for producing the product. In addition, explain why you left out other factor(s).

(c) Visit a place where a similar product is produced and ask the producer about factors of production he or she has used to produce the product selected. Make sure you note down the reasons why the producer decided to use the factor(s) of production.

(d) What is the difference between your answer in (b) and information you gathered from the producer?

(e) Share your work with the rest of the class.

Scales of production

Scale of production refers to quantities of output produced by a firm relative to input usage. Scale of production involves the size, number of plants and the technique of production used by a firm in production. The firm can be operating under a small-scale or large-scale production level.

Large-scale of production

A firm that invests in huge plants and machinery to produce goods and services using large quantities of factors of production is considered to be operating under a large-scale of production. For example, automobile firms. Note that in this case, large quantities of inputs are used to produce large quantities of output. Hence, an individual always considers size of the firm by looking at profit and output generated.

Large-scale production involves division of labour, mechanisation and production of large quantities of goods. Producers tend to expand their firms with the motive of increasing sales and maximising profit. Firms with large-scale production tend to benefit from both internal and external economies of scale. They are technical economies, which involve use of more modern machines and technology. The firm is also in the position of gaining more profit, since increased production will reduce costs.

Small-scale production

A firm that produces goods and services by using small quantities of factors of production is said to be operating under a small-scale production. Because of producing small quantities of output, the firm will not benefit from economies of scale. Such small-scale production activities include handcrafts, tailoring and photography, among others. Most of the small-scale production firms lack research and development capability that would contribute to their growth. However, in developing countries, small-scale firms are major sources of employment and income to a large chunk of the population, which cannot find employment in the formal sector of the economy.

In small-scale production, a producer requires a small amount of capital. This is advantageous for many aspiring producers with limited resources to invest in economic activities. Moreover, they supply raw materials to large-scale producers.

Most people especially those in developing countries are employed in the small-scale sector. Hence, the sector is the main source of employment and incomes to most people in the economy.

However, firms operating under small-scale production face high cost of production due to lack of economies of scale. There is no research and development and hence,

there is no innovation, which confines them to low technology in production. Table 3.2 presents differences between large-scale and small-scale production.

Table 3.2: *Difference between large-scale and small-scale of production*

S/N	LARGE-SCALE PRODUCTION	SMALL-SCALE PRODUCTION
1	Uses large quantities of factors of production	Uses small quantities of factors of production
2	Earns large profits	Earns small profit
3	Enjoys economies of scale	Does not enjoy economies of scale
4	Produces large quantity of goods and services	Produces small quantity of goods and services
5	Invests in large machinery and technologies	Uses labour intensive technique of production

Production function

A production function is a mathematical expression of a technical relationship between quantities of inputs and output used in production. This function specifies the firm's output for all combinations of factor inputs typically, capital and labour. It shows the maximum amount of output produced by a certain combination of factor inputs employed in production.

The production function assumes that the level of technology remains constant. The firm is assumed to use its inputs to produce maximum output from the used inputs, which means producing output at the maximum level of efficiency. A change in any of the variable factor produces a corresponding change in output. It is assumed that inputs are divisible into small units and it relates to a particular unit of time.

A production function can be presented in three forms, as a function (mathematical form), as a tabular (schedule) form and as a graph.

Functional form

A production function is generally expressed as

$$Q = f(X_1, X_2, X_3 ... X_n) \quad \ldots\ldots\ldots\

where:

$Q = \text{Output}$

$X_1, X_2, X_3 ... X_n = \text{Factor inputs (such as capital, labour, land or raw materials)}$

For simplicity, factors of production (inputs) are categorised into two categories, namely, labour (L) which sums up all human resources (except entrepreneurship) and capital that sums up all other non-human factors (K). Thus, a general production function with two factor inputs is given as

$$Q = f(K, L) \quad \ldots\ldots\ldots\

Graphical form

A production function can be presented graphically. By using information in Table 3.3, the production function is represented graphically as in Figure 3.4, which is called a total product (TP) curve.

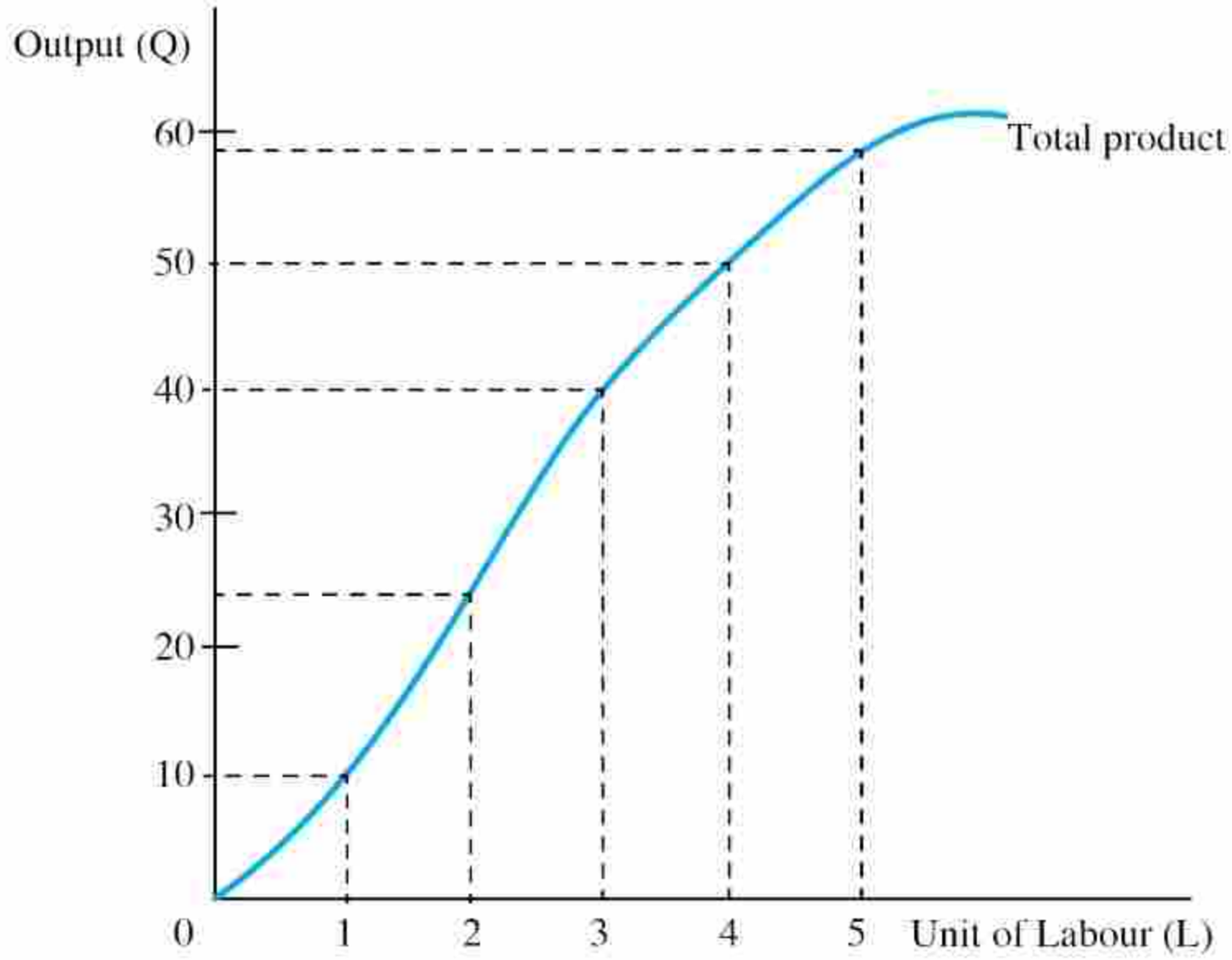


Figure 3.4: *A total product curve*

Production in the short-run

In the short-run, certain factors of production remain fixed. For example, a farmer may be unable to increase the size of land and similarly a factory owner may not be able to change the factory machinery.

Let us assume that labour is a variable factor and capital is a fixed factor. Since capital is fixed, it is constant in the short-run period. Thus, output is determined by the variable factor of production, that is, labour (L). The short-run production function is, therefore, given as:

$$Q = f\left(L, \overline{K}\right) \quad \ldots\ldots\ldots\

A specific (explicit) production function of a firm can be expressed as

$Q = 10L$ or

$Q = 2L + 5$ or

$Q = 48 + 12L - L^2$

Total Product (TP), Average Product (AP) and Marginal Product (MP)

There are three concepts that characterise output produced by a firm, namely, total product (TP), average product (AP) and marginal product (MP). To understand better the three concepts in the short-run period of production, let us assume a firm is producing computers by using four (4) machines as the fixed factor and labour as a variable factor. The hypothetical quantities of the output of computers produced by the firm are presented in Table 3.4.

Table 3.4: *Total Product (TP), Average Product (AP) and Marginal Product (MP)*

Capital (K)	Labour (L)	Total Product (TP)	Average Product (AP)	Marginal Product (MP)
4	0	0	-	-
4	1	5	5.0	5
4	2	13	6.5	8
4	3	23	7.7	10
4	4	34	8.5	11
4	5	45	9.0	11
4	6	51	8.5	6
4	7	55	7.9	4
4	8	56	7.0	1
4	9	54	6.0	-2

Total Product

This is the quantity of output produced by a firm by combining fixed and variable factor inputs. In the short-run period, factors such as capital (in this case, machines) remain fixed, and thus, total product (TP) is determined by the variable factor (labour). In Table 3.4, it is shown that as the amount of the variable factor increases, the TP increases although its rate of increase varies at different levels of employment of labour (L). When the first unit of labour is employed by the firm, the TP is 5 units and employment of a second unit of labour causes total

output to increase to 13 units. Table 3.4 shows that as the firm employs more labour, total product starts to increase at a diminishing rate.

Average Product

Average product (AP) of variable factor (labour) is the quantity of output per unit of the variable input used in production. That is, it is the total output (TP) produced by a firm divided by the amount of labour employed by the firm. Technically, the AP of labour can be expressed as:

$$AP_L = \frac{\text{Total product}}{\text{units of labour}}$$

That is,

$$AP_L = \frac{TP}{L}$$

or

$$AP_L = \frac{Q}{L} \quad \ldots\ldots\ldots\

Thus, $$MP_L = \frac{\Delta TP}{\Delta L} = \frac{\Delta Q}{\Delta L} \quad \text{.....(3.6)}$$

From Table 3.4, 23 units of computers are produced when three units of labour are employed. Suppose instead of the three labour units, the firm decides to employ four labour units, the result is that total output produced will increase to 34 units. This means the fourth unit of labour has led to an addition of 11 units of computers to the total output of computers. Thus, the 11 units added to the total output by the additional unit of labour are referred to as the marginal product of labour, presented in Table 3.4.

Mathematically, the marginal product of labour (MP) is obtained as a partial derivative of the production function presented in equation (3.7) with respect to labour, such that:

$$MP_L = \frac{\partial TP}{\partial L} \quad \text{.....(3.7)}$$

The MP in this equation is the slope of the TP curve.
In the case of capital, the marginal product of capital (K) is the increase in output that arises from an additional unit of capital employed by the firm. It shows by how much total product will increase when using one additional unit of capital, holding labour (and other resources) constant.

$$MP_K = \frac{\Delta Q}{\Delta k} \quad \text{.....(3.8)}$$

Example:

Given a production function $Q = 12L^2 - 0.8L^3$ whereby Q is output and L is quantity of labour used in production, determine
(a) Average product and marginal product of labour.
(b) The value of labour that maximises output.
(c) The value of labour at which Average product is maximum.
(d) Find the marginal product and average product using the value of labour that maximises output.

Solution

(a) The average product of labour (AP_L) is determined by dividing the total output by L. Thus

$$AP_L = \frac{Q}{L} = \frac{12L^2 - 0.8L^3}{L}$$

$$AP_L = 12L - 0.8L^2$$

Therefore,

$$AP_L = 12L - 0.8L^2$$

The marginal product of labour (MP_L) is determined by the first derivative of output with respect to labour. Thus:

$$MP_L = \frac{\partial Q}{\partial L} = 24L - 2.4L^2$$

$$MP_L = 24L - 2.4L^2$$

Therefore,

$$MP_L = 24L - 2.4L^2$$

(b) The value of labour that maximises output can be obtained by equating marginal product function to zero.

$$MP_L = 24L - 2.4L^2$$

$$0 = 24L - 2.4L^2$$

$$0 = (24 - 2.4L)L$$

$$\frac{0}{L} = \frac{(24 - 2.4L)L}{L}$$

$$0 = 24 - 2.4L$$

$$2.4L = 24$$

$$\frac{2.4L}{2.4} = \frac{24}{2.4}$$

$$L = 10$$

Therefore, the value of labour that maximises output is 10 units.

(c) The value of average product will be maximised where it is first derivative equals zero. Therefore, the value of labour at which AP is at maximum is determined by equating the first derivative of the AP function to zero.

$$AP_L = 12L - 0.8L^2$$

$$\frac{\partial AP_L}{\partial L} = 12 - 1.6L$$

$$12 - 1.6L = 0$$

$$12 = 1.6L$$

$$\frac{12}{1.6} = \frac{1.\not{6}L}{1.\not{6}}$$

$$L = 7.5 \text{ units}$$

$$\approx 8 \text{ units}$$

Therefore, the value of labour at which average product is at maximum is 8 units.

(d) The average product and marginal product by using the value of labour that maximises output can be obtained by substituting the value of labour in the marginal product and average product functions, such that:

$$AP_L = 12L - 0.8L^2$$

$$AP_L = 12(10) - 0.8(10)^2$$

$$AP_L = 120 - 80$$

$$AP_L = 40$$

and

$$MP_L = 24L - 2.4L^2$$

$$MP_L = 24(10) - 2.4(10)^2$$

$$MP_L = 0$$

Therefore, the average product determined using the value of labour that maximises output is 40 units and the marginal product is 0 units.

Exercise 3.1

1. ABC Company makes curtains and sells them door to door. The following table shows the relationship between the number of workers and output produced by ABC Company during a given day.

Capital (K)	Workers (L)	Total Product (TP)	Average Product (AP)	Marginal Product (MP)
5	0	0		
5	1	30		
5	2	60		
5	3	90		
5	4	110		
5	5	140		
5	6	170		

Calculate:

(a) Average product of labour

(b) Marginal product of labour

2. Given: $Q = 40\,L + 3L^2 - \frac{L^3}{3}$,

(a) Obtain average product (AP) and marginal product (MP) functions.

(b) Find the value of labour that maximises output.

(c) Find the marginal product and average product using the value of labour that maximises output.

3. Distinguish types of scale of production. Provide examples.

Law of diminishing returns or variable proportions

This is a short-run phenomenon that looks at how output relates to changes of only one factor input. The law of variable proportion is the modern approach to the 'Law of Diminishing Returns,' which analyses production in the short-run

period. The law of variable proportion summarises a production function with one variable factor input while keeping constant the other factor inputs.

The law of diminishing returns has been stated by most economists The law states that "as more of a variable input is employed on a fixed factor total output increases first at an increasing rate and later at a decreasing rate until a point is reached where additional quantities of input will yield diminishing returns assuming all other factors are kept constant."

The law explains the effect of variations in the variable factor on output. When one factor varies (labour) and the others remain fixed (land and capital), the proportion between the fixed factor and the variable factor will vary.

Assumptions of the law of diminishing returns

The law of variable proportion or diminishing returns as stated operates under the following conditions:

(a) The state of technology is assumed to remain constant. If it occurs that there are improvements in technology due to economic development and average and marginal product may increase instead of diminishing.

(b) There must be only one variable factor input (labour) and other factors remain fixed (such as capital and land). This law does not apply when all factors are proportionately varying. The outcome from variation of all inputs in production exists under what is known as returns to scale.

(c) All units of the variable factors are identical. They are of the same size in terms of efficiency and quality.

(d) The law operates in the short-run period of production only. In the long-run period, all input factors vary in proportionality.

The law of diminishing returns or variable proportions is explained in Table 3.5. Initially, the AP and MP increase and then starts to decline. The TP is increasing at an increasing rate until the employment of the 4th worker. Beyond the 4th worker, the MP starts to diminish, as the addition of the 5th worker reduces TP by 14 units of output. Beyond use of the 8th worker, the TP would diminish and, in turn, the MP would become negative. Thus, it would imply that when more and more units of labour are combined with other fixed factors, the MP increases first at an increasing rate, then starts to diminish and finally, it becomes negative. However,

production will not take place when MP is negative because producers are sensible and know that the same output could be produced with less quantities of inputs. In addition, in the negative range, they would incur loss that is unnecessary.

Table 3.5: *Operations of law of diminishing returns*

Number of workers (Variable Input factor)	Total Product (TP)	Average Product (AP)	Marginal Product (MP)	Remarks
1	10	10	-	Stage I
2	24	12	14	
3	39	13	15	
4	56	14	17	
5	70	14	14	Stage II
6	78	13	8	
7	84	12	6	
8	84	10.5	0	
9	81	9	-3	Stage III

The three stages of production

Economists identify three stages of production as a way of understanding the behaviour of output when one input is variable and others are fixed (see Figure 3.5).

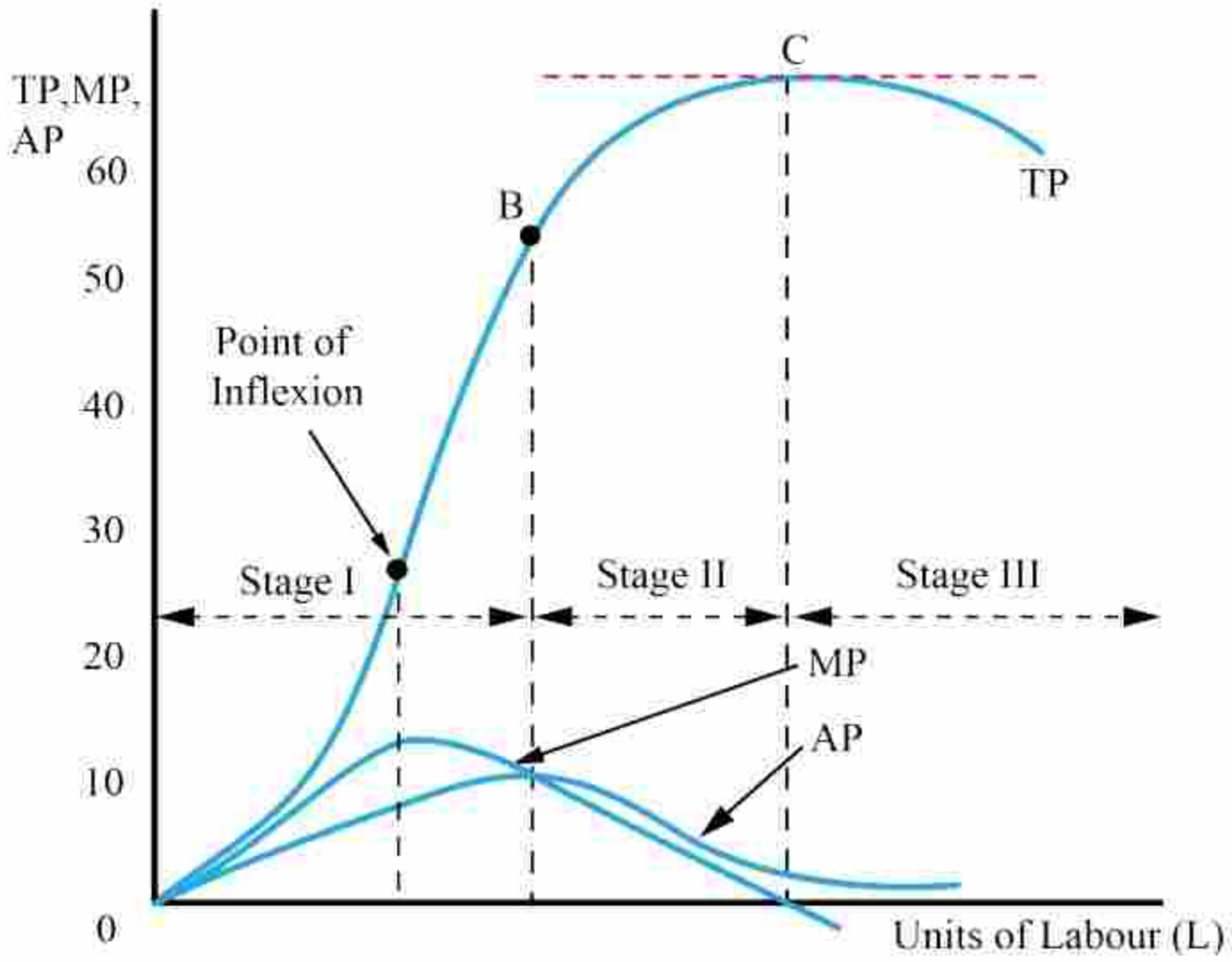


Figure 3.5: *Stages of production*

The three stages of production summarised in Figure 3.5 can be explained as follows:

Stage I

In this stage, TP increases at an increasing rate up to point B, which means the AP also increases and the MP of labour increases up to its maximum level but then starts to fall, as seen in Figure 3.5. At this stage, the MP of the variable input (labour) is higher than its AP, that is, MP>AP. Thus, TP and AP are all increasing. The AP curve is at maximum level at the point where the MP curve intersects the AP curve from above. Stage I is known as a take-off stage.

Firms in Stage I will experience an increase in productivity (or output) from each addition of the variable factor (labour). This may seem as a good stage a firm to operate because it will not make a loss as it increases the use of labour. But Stage I represents underutilisation of production capacity, implying a firm has the potential for increasing output.

Stage II

In this stage, the TP increases at a diminishing rate until it reaches its maximum point at C, while both AP and MP are decreasing, but still positive. At the end of Stage II, which is referred to as maturity stage, the MP becomes zero. Thus, MP is zero when TP is at the maximum level, while AP shows a steady decline throughout this stage. Also, in Stage II, the marginal product of the variable input (labour) falls below its average product, that is, MP<AP. Since both AP and MP decline, Stage II is also known as stage of rational or feasible region of production elsewhere referred to as a stage of diminishing returns.

Economists say this is a stage in which rational firms should operate. Conversely, Stage II is one that all firms strive to reach. Somewhere at the peak of the stage (where MP = 0 and TP is at maximum) is the exact ideal level of output firms should stop producing. However, depending on the price of output determined by the market, the output that maximises profit may not necessarily be the maximum output.

Stage III

In this stage, the TP falls, the AP shows a steady fall, but never becomes zero though, the MP becomes negative. Hence, Stage III is known as the stage of negative return. Therefore, a rational producer will never choose to operate in Stage III because the MP of the variable factor is negative. Thus, it is clear, the rational producer will stop production at the second stage where the MP of variable factors is zero.

Returns to scale

The law of variable proportions (law of diminishing returns) analyses the behaviour of output when one factor input varies and the other factors are held fixed. Thus, it is a short-run phenomenon.

When all factors of production are changed in the same proportion, the behaviour of output is analysed by applying laws of returns to scale. Thus, the **law of returns to scale** is a long-run concept in production theory. In the long-run period, output can be increased by varying all factor inputs (Factors of production). The law states that "when all factor inputs are increased in the same proportion, output may increase either by the same proportion, greater proportion or less proportion than the factor input."

With returns to scale, all factors of production vary with the scale of production. The production function with two factor inputs is technically expressed as:

$$Q = f(K, L)$$

In the long-run period, a firm can increase scale of production, for example, by a factor λ, which can be any positive number, for example two (2), that means that all inputs are doubled. Therefore, the production function will be expressed as:

$$Q = f(\lambda K, \lambda L)$$

For particular production functions, it can also be presented as:

$$Q = \lambda^s f(K, L) \quad \ldots\ldots\ldots\

Increasing returns to scale

This exists when inputs are increased by a given proportion and output increases by a larger proportion. In other words, proportionate increase in all factors of production results in a more than proportionate increase in output. For example, if increase of inputs by 1 percent results to an increase in output by 2 percent, there is increasing returns to scale in production. Therefore, if an industry enjoys increasing returns, then its marginal product will be increasing and the output expanding.

Constant returns to scale

This occurs when a given proportionate increase in all factor input lead to the same proportionate increase in production. That is, when inputs are increased by a given proportion, the output increases in the same proportion. For example, if inputs are increased by 1 percent and result to an increase in output by 1 percent, then the firm is said to experience constant returns to scale. In the case of constant returns to scale, the total output increases at the same proportion as the increase in inputs.

Decreasing returns to scale

This occurs when a proportionate increase in all factor inputs leads to less than proportionate increase in output. For example, if inputs are increased by 1 percent, but output increases by less than 1 percent (say 0.5 percent), a firm is said to experience decreasing returns to scale. Decreasing returns to scale implies increasing costs to scale. Table 3.6 presents hypothetical examples of returns to scale. The input levels K and L are doubled at each successive round. Initially, output shows increasing returns, then it doubles, to suggest a case of constant returns to scale, which is followed by the case of decreasing returns to scale.

Table 3.6: *Phases of law of returns to scale*

Units of capital (K)	Units of labour (L)	Total product	Scale
2	4	70	Increasing returns $s > 1$
4	8	210	
8	16	460	
16	32	920	Constant return $s = 1$
32	64	1840	
64	128	3680	
128	256	4416	Decreasing return $s < 1$
256	512	4857	
512	1024	5342	

Economies and diseconomies of scale

A firm may enjoy economies of scale or diseconomies of scale if the long-run average cost increase or decrease, respectively as output increases. Economies of scale arise when the long-run average cost decreases and diseconomies of scale arise when there is an increasing long-run average cost.

Economies of scale

In the long-run, when scale of production is increased, the firm gets economies of scale. The term "**economies**" refers to cost advantages. Therefore, economies of scale refer to advantages of large-scale production. Economies of scale can be divided into internal economies and external economies.

Internal economies of scale

These are advantages realised by a firm from internal factors when it expands its scale of production. For instance, use of greater degree of division of labour and specialised machinery at higher levels of output is referred to as internal economies. They are internal because they accrue to the firm due to its own efforts. The following are internal economies of scale:

Technical economies: Large-scale production is associated with technical economies. These economies accrue to a firm from use of better machines and techniques of production. As a result, production increases and cost per unit of production decreases.

Managerial economies: This refers to reduction in managerial cost. When output increases, division of labour can be applied in management. A large firm can divide its big departments into various sub-departments and each department could be placed under control of an expert. A brilliant organiser can devote time wholly into the work of managing the firm and routine jobs be left to the cadre of other workers.

For instance, a production manager would only be involved in production, sales manager in sales and finance manager will only be involved in financial matters. If scale of production increases further, each department would be sub-divided further.

Financial economies: Large firms are better placed to raise finance for expansion. For instance, it can offer security to bankers and it would be better known to

investors to command confidence of finances. This factor would enable the firm to raise funds, for example, by selling shares in the stock market because usually, investors have more confidence in investing their money in well-established large firms. In addition, shares and debentures of a large firm can be easily sold in the shares market.

Risk bearing economies: A large firm or business with diverse and a multi-production capability is in a good position to withstand or tolerate economic downturns and upturns, and therefore, enjoy economies of risk bearing. The larger the size of a firm, the more likely are its losses to be spread among its various activities. A big firm produces a large number of items and of different varieties such that its loss in one product can be counter balanced by the gain from any other product.

Economies of research: A large sized firm can spend more money on its research activities in order to innovate varieties of products and improve the quality of existing products. Innovations or new methods of producing a product may help to reduce its average cost.

Economies of welfare: A large firm can provide welfare facilities to its employees, for example, subsidised housing, subsidised canteens and nursery facilities for children of workers and recreation facilities. All these measures or facilities have an indirect positive effect on output produced by a firm and can reduce the cost of production as they increase productivity of workers.

Labour economies: A large firm employs a large number of labourers. Therefore, each person can be employed where he or she is most effective and efficient. Moreover, a large firm is in a better position to attract specialised experts and well-qualified workers in the firm. Likewise, specialisation saves time and encourages new inventions. All these advantages result in lower costs of production.

Advertising economies: A large firm spreads its advertising and marketing costs over a large output. Due to advertising economies, large firms may be at better position than smaller firms to promote sales. For instance, well-trained and efficient sales persons can be appointed for promoting sales of big firms.

Purchasing economies: These are gained when larger firms buy in bulk and obtain large purchasing discounts. When a large firm purchases materials in bulk, it can bargain on its purchases on favourable terms. That is, it can purchase its inputs in

bulk at negotiated discounted prices if it has sufficient negotiation power in the market. It can also ensure continuous supply of raw materials. For example, large food retailers have buying power when purchasing supplies from farmers as they can negotiate for reasonable prices.

External economies of scale

These are advantages which accrue to firms as a result of expansion of output of the whole industry rather than the output of individual firms. These are external economies because they accrue to all firms in the industry. The external economies are available to all firms in the following forms:

Economies of concentration: When a number of firms are located in one place, all of them reap some common economies. First, skilled and trained labour becomes available to all firms. Thus, with growth of an industry in an area, a pool of trained labour is developed, which has a favourable effect on the productivity and cost of the firms in the industry. Secondly, banks and other financial institutions may set up their branches so that all firms in the area can easily obtain credit facilities. Third, transport and communication facilities may be improved considerably.

Economies of information: Economies of information may arise from collective efforts of firms in an industry. This may occur when firms pool resources to search for information and new knowledge for new inventions for enhancing productivity in the industry. Fruits of invention can be shared by all member firms.

Economies of disintegration: When an industry grows, it becomes possible to split up production into several processes and leave some of the processes to be carried out more efficiently by specialised firms, making specialisation possible and profitable. For example, in the cotton textile industry, some firms may specialise in manufacturing of thread, others may specialise in producing vests, and some others specialise in weaving t-shirts. The disintegration helps the industry to avoid duplication of efforts that increase costs.

Cheaper raw materials and capital equipment: Expansion of an industry may result in exploration of new and cheaper sources of raw material, machinery and other types of capital. In addition, expansion of an industry results in increased demand for various kinds of materials and equipment required by the industry. This makes it possible to purchase on a large scale from other industries. This reduces their cost of production and prices. Thus, firms using the materials and capital equipment will be able to buy them at low price.

Technological external economies: When an industry expands, it may result in discovery of new technical knowledge, use of improved and better machinery than before. This will enhance productivity and reduce the cost of production of firms in the industry.

Growth of ancillary industries: With growth of an industry, a number of ancillary industries may specialise in production of raw materials, tools and machinery. That may become available at low prices to other firms in the industry. Likewise, some firms may process or rather, recycle their waste products to produce useful products. This will tend to reduce the cost of production in general.

Diseconomies of scale

Diseconomies of scale are disadvantages which a firm faces when the scale of production is expanded beyond a certain level. The diseconomies of scale are observed from the rising long-run average total cost. Diseconomies of scale may be of two types, which are internal diseconomies and external diseconomies of scale.

Internal diseconomies of scale

Communication diseconomies: Large firms often suffer from poor communication because they find it difficult to maintain an effective flow of information between departments, sections or between the head office and their subsidiary offices. Time lags in flow of information can also create problems in terms of speed of response to changing market conditions. For example, a large supermarket chain could be less responsive to changing tastes and fashions than a much smaller 'local' retailer.

Large firms may *experience inefficiencies related to the principal-agent problem*: This problem is caused by size and complexity of very large firms. The reason is that owners often have to delegate decision-making to appointed managers, which can lead to inefficiencies. For example, owners of a large chain of retail shops of clothes will have a manager for each shop, and delegate some of the jobs to managers who may not necessarily make decisions in the owners' best interest.

Managerial diseconomies: This occurs when management finds it difficult to exercise control and bring coordination among various departments. The managerial structure becomes more complex and affected by bureaucratic procedures. All these factors affect efficiency and productivity of the management and the firm itself.

External diseconomies of scale

External diseconomies may also arise if there are certain disadvantages that may work against advantages from expansion of an industry.

Rise in some factor prices: When an industry expands, requirements of factors of production increase. This may increase the prices of such factors of production, especially when they are in short supply.

Higher transportation cost: Existence of too many firms in one place (area), may triger high transportation costs, since the demand for transport services in the locality will increase. Consequently, price for transport services will rise.

High marketing cost: This is more likely when many firms seek to advertise their products, for example, through radio stations, television stations, newspapers and other means of mass media communication. Because of high demand, the advertising firms would raise the firm's advertising prices, especially when many firms are advertising.

High pollution control cost: When many firms are located in one area, there are possibilities of pollution to increase in the area and that may lead to an increase in the cost of controlling pollution.



Activity 3.2

1. In a group of three students, present a figure showing the three stages of production. What are the major characteristics and economic intuition of each stage?
2. Assume a firm is operating under large-scale production. Explain the economies and diseconomies of scale gained by the firm.

Chapter Summary

1. Production involves transformation of inputs into outputs in a form of goods or services that satisfy human wants. The transformation involves changing the original form of the materials to produce goods. Carpenters use wooden materials as inputs to produce outputs in form of furniture, such as tables, chairs, beds and cupboards.

2. A production function is an expression of the technical relationship that exists between quantities of inputs and outputs used to in production. This function specifies the output of a firm for various combinations of inputs.

3. A firm is a single business entity that produces goods and services with the aim of making profit through acquiring a share and creating employment. It buys input in order to transform them into output.

4. Factors of production are resources used in the production process or are resources that facilitate production. These factors are generally classified into four major groups, which are land (including all natural resources), labour (including all human resources), capital (including all natural and human-made resources), and entrepreneurship (which brings all the previous resources together for production).

5. Scale of production refers to level of quantities of factors of production used by a firm to produce output of goods and services. It concerns the size of the firm, number of plants and the technique of production needed in the production of goods and services.

6. The law of diminishing returns states that an increase in quantity of any variable input, which is combined with a fixed quantity of another input leads initially to a rise of its marginal product that eventually decreases as additional units of the variable inputs are used in production.

7. The law of returns to scale seeks to analyse effects of scale of production on the level of output. If a firm increases the units of both labour and capital in the same proportion, its scale of production increases. The return to scale could be increasing, constant or diminishing

8. Economies of scale refer to advantages of large-scale production. Economies of scale can be divided into internal economies and external economies.

9. Diseconomies of scale are disadvantages which a firm faces when the scale of production is expanded beyond a certain level. This is notable from the increase in the long-run average total cost when the quantity of output produced is increased. Diseconomies of scale can be divided into internal diseconomies and external diseconomies.

Revision exercise

1. Define economies and diseconomies of scale and explain why they might arise.
2. Study the table carefully and then answer the questions given.

Land	Labour	Total product	Average product	Marginal product
10	0	0		
10	1	20		
10	2	50		
10	3	90		
10	4	120		
10	5	140		
10	6	150		
10	7	150		
10	8	140		

(a) Calculate average product and marginal product of labour.

(b) Show the three stages of production graphically.

(c) Within which ranges increasing returns, constant returns and decreasing returns are notable?

3. Given the average product (AP) function: $AP = \frac{200}{L} + 20 - L$

   (a) Derive the total product function and the marginal product functions.

   (b) Find the maximum possible level of output and the units of labour required to maximise output.

4. Use the following table and find the values of **Q, R, S, T, U, V, W, X, Y** and **Z**

Unit of Labour	Total Product	Average Product	Marginal Product
1	200	Q	R
2	S	T	120
3	U	140	V
4	420	W	X
5	Y	80	Z

   By using a graph, explain the relationship between marginal product (MP) and average product (AP). Identify stages of production.

5. Clearly differentiate between laws of return to scale and the law of diminishing returns.

6. By using examples, discuss why Stage II of production is considered to be an ideal stage that every firm should strive to operate in.

7. Explain advantages of division of labour in any firm.

8. Why is entrepreneurship regarded as an independent factor of production and not part of labour?

9. Given that $MP_L = 40 + 6L - L^2$, where L=number of labourers and MPL=marginal product of labour. Find the number of labourers at

   (a) Stage I

   (b) Stage II

   (c) Stage III

10. Study the table and attempt the questions.

Number of workers	Total product	Marginal product	Average product
1		8	
2		12	
3		16	
4		12	
5		7	
6		5	
7		0	
8		-4	

(a) Fill the table

(b) Using the data from the table, draw a hypothetical total product (TP), average product (AP) and marginal product (MP) curves and indicate the three stages of production.

Chapter Four

Theory of the firm

Introduction

When you walk into a supermarket you find shelves with a variety of goods classified according to their uses. Such goods are produced by different firms, large and small ones. For a firm to produce goods and services it must employ various inputs and thereby incur costs from their purchases. In this chapter, you will learn about the concept of the firm, costs of production, the cost function, and revenue analysis and profit maximisation by a firm. The competencies developed will enable you to determine the level of output, which minimises costs and maximises profits.

Concept of the firm

The Tanzanian economy has a large number of firms, which produce various goods, for example, clothes such as khanga and kitenge; dairy products, such as fresh milk, yoghurt and butter; and flour such as wheat and maize flour, among others. The economy also has firms that produce various services, such as transportation, telecommunications, education and health services.

In general, a firm is any legal business undertaking, which produces or supplies goods and services in order to generate profit and meet other objectives, such as creating employment or maximising welfare. Firms play a key role in deciding what to produce, how to produce and for whom to produce. There are several categories of ownership of firms, which include sole proprietorship, partnerships, corporation and public ownership.

A sole proprietorship: This constitutes a business for which a firm is registered as a single entity for tax and liability purposes. A sole proprietor is a firm that is not registered with the state as a limited liability company or corporation. The owner

of the firm does not pay income tax separately for the company, but only reports the income or losses on his or her individual income tax return. Thus, for this category, the owner is inseparable from the firm. Hence, the owner is responsible for any business debts.

Partnership: This is a business relationship of two or more individuals or firms.

Corporation: This is a business undertaking owned by shareholders and managed by a board of directors. Shareholders are protected from personal responsibility for its debts or liabilities.

State owned firm: This is a business entity that is managed by the government on behalf of citizens.

Thus, economics explain a firm as an economic institution that transforms factors of production (inputs) into goods and services (outputs). The production process of a firm is constrained by the following:

(a) Scarcity of all factor inputs;

(b) Market prices of factor inputs; and

(c) Market prices of outputs.

The defined concepts apply in the analysis of formal businesses. However, in developing countries many small firms operate in the informal sector and as such, these firms are not registered.

Given the constraints that a firm faces, it has to make optimal decisions with regard to minimising costs of producing a specific quantity of output or to maximise profit from the production activity. A firm that does not seek to maximise profit or fails to do so may collapse or become acquired by other firms that are successful in their lines of business. For example, several banks in Tanzania have been acquired by other successful banks.

Millions of firms could be operating in any country, which differ in size, products and services they supply. While traditionally, most of economists presuppose that the fundamental objective of a firm is to maximise profit, many of them acknowledge that some firms have objectives other than profit maximisation. Some firms seek to enhance public relation, create jobs or to protect the environment.

Moreover, other firms work to improve social welfare by supporting provision of education, health care, security, water and electricity or protecting marginalised and underprivileged people.

An Industry

In economics, firms producing similar or closely related products comprise an industry. As is the case with any other country, there are many industries in Tanzania. For example, the soap industry, which comprises firms that produce soap, include Mbuni and Rungu soaps. Another example is the textile industry, which include the 21st Century Textile Limited in Morogoro that produces shirts and yarn dyed fabrics like 'Maasai' cloth; the Kibotrade Textiles Limited in Dar es Salaam, which produces t-shirts and jersey, mostly for exportation; and the Tanzania-China Friendship Company Limited, which produce fabrics and yarn. Note that agriculture is considered to be an industry such that the theory of a firm applies also to agriculture.

Location of industries

Location of industries in an economy is influenced by various factors. Consequently, the location of industries across regions or states in a given country is not uniform. Tanzania has administrative regions with varying degrees of intensities of business. For example, regions such as Dar es Salaam, Arusha and Morogoro have a larger proportion of industries relative to regions such as Njombe and Tabora, which have a smaller proportion of industries. Even within a given region, location of industries may vary. In Dar es Salaam Region, for example, the districts of Ilala and Temeke have a relatively larger proportion of industries than other districts, namely, Kinondoni, Kigamboni and Ubungo.

Factors influencing location of industries

The following are factors that influence on location of industries.

Adequate supply of raw materials for production: Many firms are located in areas where raw materials they use are produced. For example, the cement factory in Mbeya region is close to the limestone producing area, whereas the Mufindi Paper Mills is close to the Sao hills timber forest, and the Minjingu Mines and Fertilizer Limited is located near mines of Minjingu in Manyara Region. These materials are bulky, and hence, in determining location, firms using these materials aim to minimise the cost of transporting them to the production site.

Transport cost of final products: The transport cost for bulk or perishable goods in nature is high when the industries are located far from the market. As a result, most industries, particularly those producing consumer goods, are established near big market centres.

State of infrastructure: Industries are located in areas with good and extensive infrastructure. Well connected transport and communication networks (for example, railways, airports, harbours or ports, internet and telecommunications) may reduce the cost of producing and distributing goods and services.

Availability of financial services: Financial services include credit facilities, money transfers and savings, among others. Firms may require a loan for acquiring capital (machinery, equipment, tools). Financial services facilitate trade between firms and their customers. Firms use financial services when they have a need to transfer or deposit money in the course of their business operations at low transaction costs.

Attractive geographical conditions: Fertile soil, conducive climate and weather can determine the location of agricultural firms. Also, availability of raw materials can influence on location of agricultural firms. The characteristic of the involved raw material, for example, bulkiness and perishability, plays part in determining the location of many agricultural processing industries. For example, tea or coffee processing factories, Tanganyika Instant Coffee Public Limited Company is located in Bukoba, Kagera a region where coffee is produced. Chai Bora Limited, a firm that blends, packs and sells tea is located in Mafinga, Iringa Region, the place where tea is grown in bulk.

Existence of economies of scale: Industries located in places where they can benefit from external economies of scale, that is, where other firms already exist, with the potential of acquiring inputs cheaply and share the existing facilities/infrastructure. Other firms follow footsteps of the pioneer industrialists who created the initial conducive environment.

Availability of social services: These services include schools, hospitals, hotels, police stations, and places of worship, which tend to influence on investors' decisions of where to locate firms.

Availability of labour: Business entities are more likely to be located in areas where labour is readily available.

Localisation of industries

This is concentration of many industries in a particular area, locality, region or town to specialise in production of a particular commodity. For example, most beverage industries are located in Dar es Salaam and fish industries in Mwanza.

Advantages of localisation of industries

Attraction of firms or industries to locate in some specific areas has some advantages and disadvantages to the firm and the economy at large. The decision of firms to locate in a specific area may enable them to reap the following external economies.

Carrying out joint research may lead to development of modern technology that may reduce costs of production for firms. In case of adaptation of new technology, localisation also enables firms to conduct jointly workshops, short courses and training for their workers.

Attracting skilled and unskilled labour to the location would benefit firms in terms of availability of labour and expertise needed in production processes. However, this may lead to growth of financial institutions which provide financial services, such as banking and insurance services.

Localisation can cause firms in the same area to share some costs, for example, advertising costs, which can be undertaken jointly by all firms located in a particular area. Hence, firms are enabled to minimise the cost of advertising. For example, firms may decide to develop a joint publication through a magazine that advertises all of the firms in the industry.

Development of towns and cities may expand the market for industrial products and local raw materials available within and outside the area. Consequently, varieties of products produced result in improving the living standards and enhancing consumers' choice.

Localisation may expand establishment of social services institutions such as schools and hospitals for the community, which would enhance health of workers and improve productivity of labour.

Disadvantages of localisation of industries

Localisation of industries is likely to cause congestion in an area due to population increase resulting from the benefits (employment, financial institutions and health centres) available in the area. This may lead to emergence of social problems such as diseases and crimes, environmental problems, such as air pollution, water pollution, soil pollution and noise pollution from vehicles and industries that operate in the area.

Localisation of industries limits mobility of labour because labour specialises in performing specific tasks only, in such a way that workers may find it difficult to work in other areas. Moreover, it limits employment of other types of labour because of the specificity of industries in that particular area. However, in case industries are closed they may lead to high unemployment.

At the national level, if location of the firms/industries is not well planned, it can cause regional economic imbalances. Thus, some regions that have got many industries such as Dar es Salaam become more developed than regions that have fewer industries, for example, Kigoma, Mtwara, and Rukwa. To address this situation, the government has established special economic zones and export processing zones across the whole country. There are ten central government designated Special Economic Zones (SEZ), four local government SEZ and eleven private SEZ in Tanzania.

Objective of a Firm

The main objective of a business firm is to maximise profit. Profit is calculated as the difference between total revenue from the sale of a firm's output and its total costs of production.

That is; $\pi = TR - TC$.

Total Revenue (TR): The price per unit of a good or service multiplied by the quantity of the good or service sold at that price.

Total costs (TC): Sum of all opportunity costs related to the production process of a firm.

Exercise 4.1

1. Distinguish between a firm and an industry. Provide vivid examples.
2. Explain the meaning and objective of a firm in economics.
3. Describe the difference between location and localisation of industries.
4. Discuss the advantages and disadvantages of localisation of industries. Provide vivid examples from your area.

Costs of production

Cost of production is a wide concept as it involves other fields such as costs of insurance, the cost of advertisement and the cost of utilities (water and electricity bills).

In economics, costs are payments for factors of production. They include labour, land, capital and entrepreneurship. Firms pay wages and salaries for the services of labour, interest as cost of capital, rent for land use and profit as a return to the entrepreneurship used in production. Land or natural resources are treated as owned capital, if the firm owns them. Otherwise, they are treated as a purchased or rented input. Economists categorise costs into various types, such as explicit and implicit costs; accounting and economic costs; and private and social costs.

Explicit cost versus implicit costs

Firms incur both implicit and explicit costs in the production process.

Explicit costs: These costs are quantifiable and are usually incurred as actual payments in form of money. Examples of these costs include wages and salaries paid to workers, rental prices of capital equipment and buildings, and cost of raw materials.

Implicit costs: These costs are not directly quantified because they do not arise from actual payment of money for use of factors of production during production. They refer to non-pecuniary costs measured by what a factor currently used in production could earn in the best alternative uses. Examples of implicit costs are depreciation costs, costs of owned capital goods, costs of owned land, costs of owned labour and normal profit.

Depreciation costs: Depreciation costs arise from change in the economic value of capital over a given production period. Depreciation costs are usually included

in the accounting costs to approximate wear and tear from use of capital, such as equipment, machinery, buildings and infrastructure.

Opportunity cost of owned capital: When the owner of a firm uses own capital assets, there arises a forgone interest income that could have been earned by investing the funds in the portfolios, for example, in bank deposits, government bonds, shares (stocks), treasury bills and other investment schemes.

Opportunity cost of owned labour: This cost sums the forgone wage or salary and other benefits that owned labour could have earned from the best alternative employment elsewhere, not in own firm. If the owner of the firm gets compensation from the firm for his or her labour input that would count as explicit cost.

Opportunity cost of owned land: This is the highest rent a firm would have to pay for the use of land if it were to rent it. In general, the opportunity cost of owned resources is payment from selling resources to the highest payer, if they were to be used elsewhere.

Normal profit: This is a return to an entrepreneur for engaging in a business venture. It may include a return of running another firm or conversely, the opportunity cost of not running another business. A firm will remain in an industry if it is earning at least a normal profit, which is a positive rate of return. The owner of a firm often supplies entrepreneurship, the factor of production that organises the business and bears the risk of running the business. In corporations, the top leadership of a company, such as the Chief Executive Officer (CEO) and directors of a company, act as entrepreneurs, rather than owners (the shareholders). Normal profit is computed as a minimum return just enough to keep a firm in business. According to economists, a firm earns normal profit when the difference between the firm's total revenue and the combined implicit and explicit costs is equal to zero.

Accounting costs versus economic costs

A distinction exists between accounting and economic costs. Accountants concentrate on operational costs of a firm, partly because it is a major component in the profit and loss equation, which they refer to as accounting costs (expenses). These costs include tangible money spent by the firm and any provision for losses or depreciation that the firm makes in the accounting period. A good accounting cost has the following elements:

(a) It ensures that a firm pays the correct amount of income tax;

(b) It shows a bank how the owners of a firm use a bank loan; and

(c) It gives proper calculations of profit or loss and dividends paid to a firm's shareholders.

Accounting profit is the excess of a firm's total income or revenue over total expenditure. Typically, any firm earns income after selling goods or services. If a firm earns from the sale of goods or services, it produces more income or revenue than the total monetary expenses on factor inputs plus depreciation costs, then that the firm has earned is known as accounting profit.

$$\text{Accounting profit} = \text{Total revenue} - (\text{Explicit cost} + \text{Depreciation cost})$$

$$\text{Economic profit} = \text{Total revenue} - (\text{Implicit cost} + \text{Explicit cost})$$

However, economists distinguish the accounting costs and accounting profit from economic costs and economic profit. Economists measure costs by the foregone benefits from the next best alternative use of a resource, which they refer to as opportunity costs. In general, firms seek to allocate resources at their disposal in the most efficient way, which is the optimal way. Any optimal allocation involves sacrifice of some benefits, which are referred to as opportunity costs weighed against economic profit or returns.

For example, a firm can employ factor inputs such as land, labour and capital to produce either computers or mobile phones. Thus, the decision by the firm to produce mobile phones implies that production of computers is the foregone opportunity in production. That is to say, the firm would have used the same resources to produce the highest ranked alternative product, which is computers. The highest valued forgone alternative is the opportunity cost of the firm's production. Therefore, the firm's opportunity cost of production is the cost of factors of production it employs.

Private cost versus social cost

Private costs are actual expenses paid by a firm. To the firm, such costs may include expenditure on raw materials, salaries and wages, rent and various overhead expenses.

Social cost includes private costs and costs borne by the society over and above the private costs. For example, pollution costs by a firm, whose productive activities pollute the environment may not be borne by the polluting firm rather,

such costs are borne by the entire society. Hence, social costs would include such costs in addition to the private costs of producing the good. Such costs are also known as external costs. Thus, social costs of a productive activity include all private costs incurred by a firm and all other unaccounted costs due to damages that production of a good or service by the firm causes to the society.

For instance, if a firm dumps its toxic wastes into a river flowing near a factory, such decision leads to water pollution and environmental damage. The cost of such damage or loss to people using water from the river and to the environment in general it is referred to as external cost, which is added to the private costs incurred by the firm to obtain the social cost of producing the product. Therefore:

$$\text{Social cost} = \text{Private cost} + \text{External cost}$$

Exercise 4.2

1. What is the difference between:
   (a) Social costs and private costs?
   (b) Accounting profit and economic profit?
2. Distinguish between explicit costs and implicit costs. Explain why for optimal business decision a firm should consider both of them.

Cost function

A cost function of a firm is a mathematical relationship between the cost of producing output and the produced output. It depicts how the cost of production changes with output in production. This function is expressed as:

$$C = f(q) \quad \ldots\ldots\ldots\

The short-run versus the long-run production period

A short-run period to a firm is a period under which some of the factor inputs, for example, capital and land are fixed. In the short-run, a firm can change only the quantity of the variable inputs, usually labour. In economic theory, collection of capital and land as factors of production is called a firm's plant. Thus, in the short-run, the firm's plant is fixed.

For a firm to increase output in the short-run, it must increase the quantity of the variable input. Labour is mainly categorised as the variable input, except for high skilled labour, which is very scarce. Hence, in the short-run a firm can increase or decrease output by increasing or decreasing the amount of labour it employs.

A long-run period is the period that is long enough to allow a firm to change (increase or decrease) all quantities of factor inputs to produce its desired quantity of output. Notably, in the long-run, the firm can change its plant size. Long-run decisions of a firm are not easily reversible. Once a firm buys a plant, its resale value is usually much less than what it paid for it. Thus, in practice, long-run is perceived as a planning horizon for a firm whereby it considers several scenarios of expansion of its output in future.

The relationship between output of a firm and its costs can be explained in relation to several concepts of costs, including fixed, variable, average and marginal costs.

Short-run costs

In the short-run period of production, a firm incurs total cost, average cost and marginal cost of production. These costs are explained as follows:

Total cost (TC)

The total costs of a firm are costs incurred by a firm for using factors of production. Thus, total cost comprises the variable cost of inputs, such as labour and raw materials and fixed costs of inputs, such as buildings and machinery. A firm computes its total cost as a sum of the total fixed cost (TFC) and total variable cost (TVC).

$$TC = TFC + TVC \quad \ldots\ldots\ldots\

incurred for using a firm's plant in production. Fixed costs do not vary with the quantity of output produced. Thus, in the short-run, the firm incurs fixed costs even when it does not produce any output. For example, a shoe-manufacturing firm will still incur costs on plant and machinery even if no shoes are produced. Hence, the fixed cost are constant at all levels of output in the short-run. Therefore, TC = TFC when output is zero (q = 0).

$$TFC = TC - TVC \quad \ldots\ldots(4.3)$$

The **total variable costs (TVC)** are costs of using the variable inputs in production, for example, wage for labour employed by a firm. The variable costs of a firm change with changes in quantity of output. Thus, when q = 0, TVC = 0. Variable costs are also called direct costs, prime costs or operating costs.

$$TVC = TC - TFC \quad \ldots\ldots(4.4)$$

Figure 4.1 illustrates curves for the three types of cost. The TFC curve is horizontal, since the fixed costs do not vary with changes in output. Hence, it remains constant at Tshs 20. The TVC curve is upward sloping because costs increase as output increases. The TC curve also slopes upwards because the TC increase as TVC increases. It can be noted that in Figure 4.1, when the firm is not producing any output, that is when q = 0, the TVC = 0 and TC = Tshs 20 because TFC = Tshs 20.

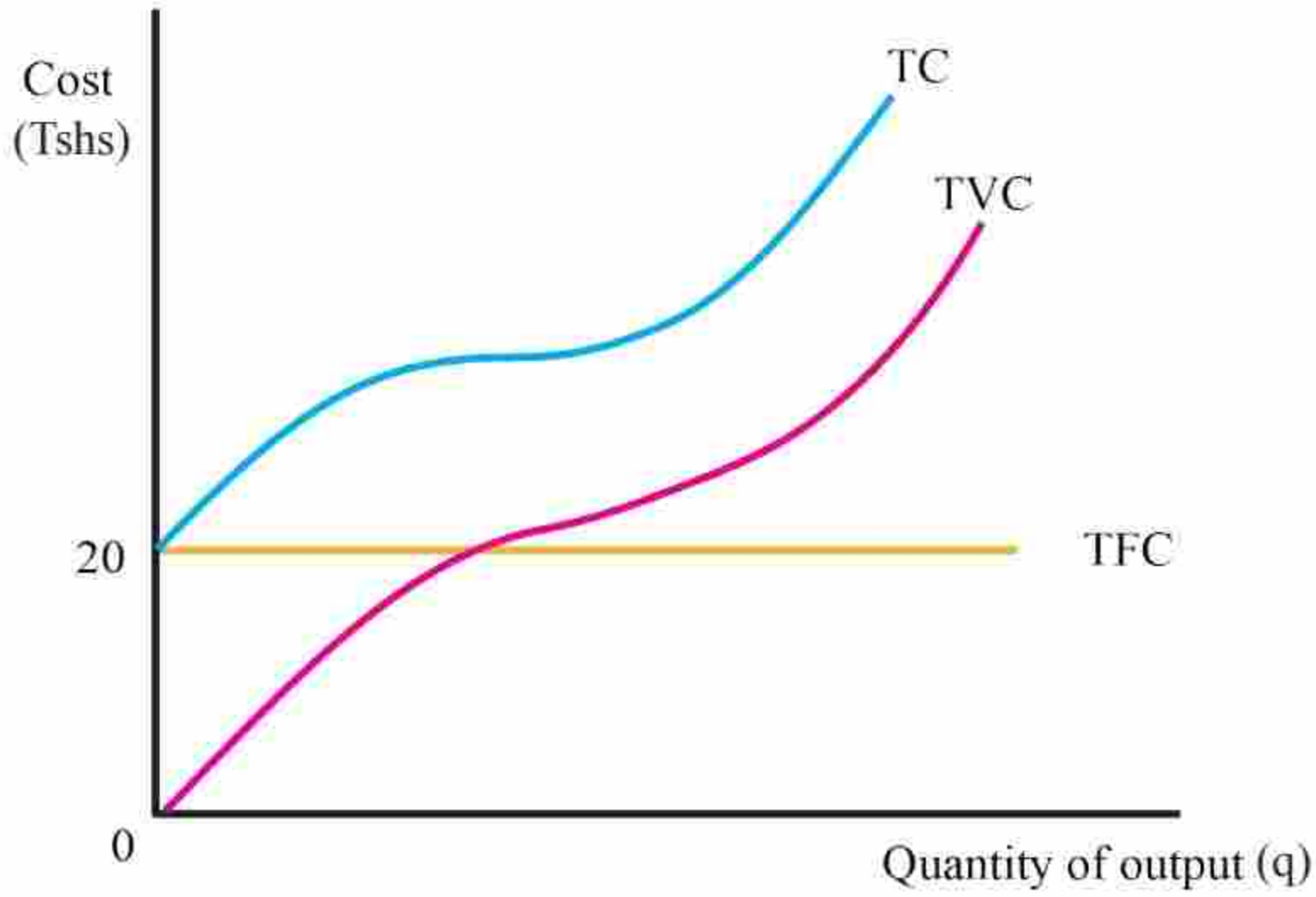


Figure 4.1: *Short-run cost curves*

When a firm makes a decision on how much to produce, it must consider how the level of production will affect its costs. The firm must determine how much it costs to produce one pair of shoes and its impact on total and average costs of the firm. In particular, when the firm makes a decision on the expected quantity of output to produce, it will seek to minimise the cost of producing that quantity of output.

Average cost

Average cost (AC) also known as average total cost (ATC) is the cost per unit of output. It is the ratio between total cost and the output produced.

$$ATC = \frac{TC}{q} \quad \ldots\ldots\ldots\

Marginal cost

A firm must consider how much it costs to produce an extra unit of output, for example, an extra pair of shoes. In this regard, a firm computes the marginal cost of production.

The marginal cost (MC) is an addition to the total cost that results from an additional unit of output produced by a firm. Given the total cost in Equation 4.2, the MC of the firm can be derived as change in total cost (TC) over change in total output (q).

$$MC = \frac{\Delta TFC}{\Delta q} + \frac{\Delta TVC}{\Delta q} \quad \ldots\ldots\ldots\

then AVC is decreasing and if MC > AVC then AVC is increasing. Likewise, if MC < ATC, then ATC is decreasing and if MC > ATC, then ATC is increasing. Also, the MC rises with the quantity of output.

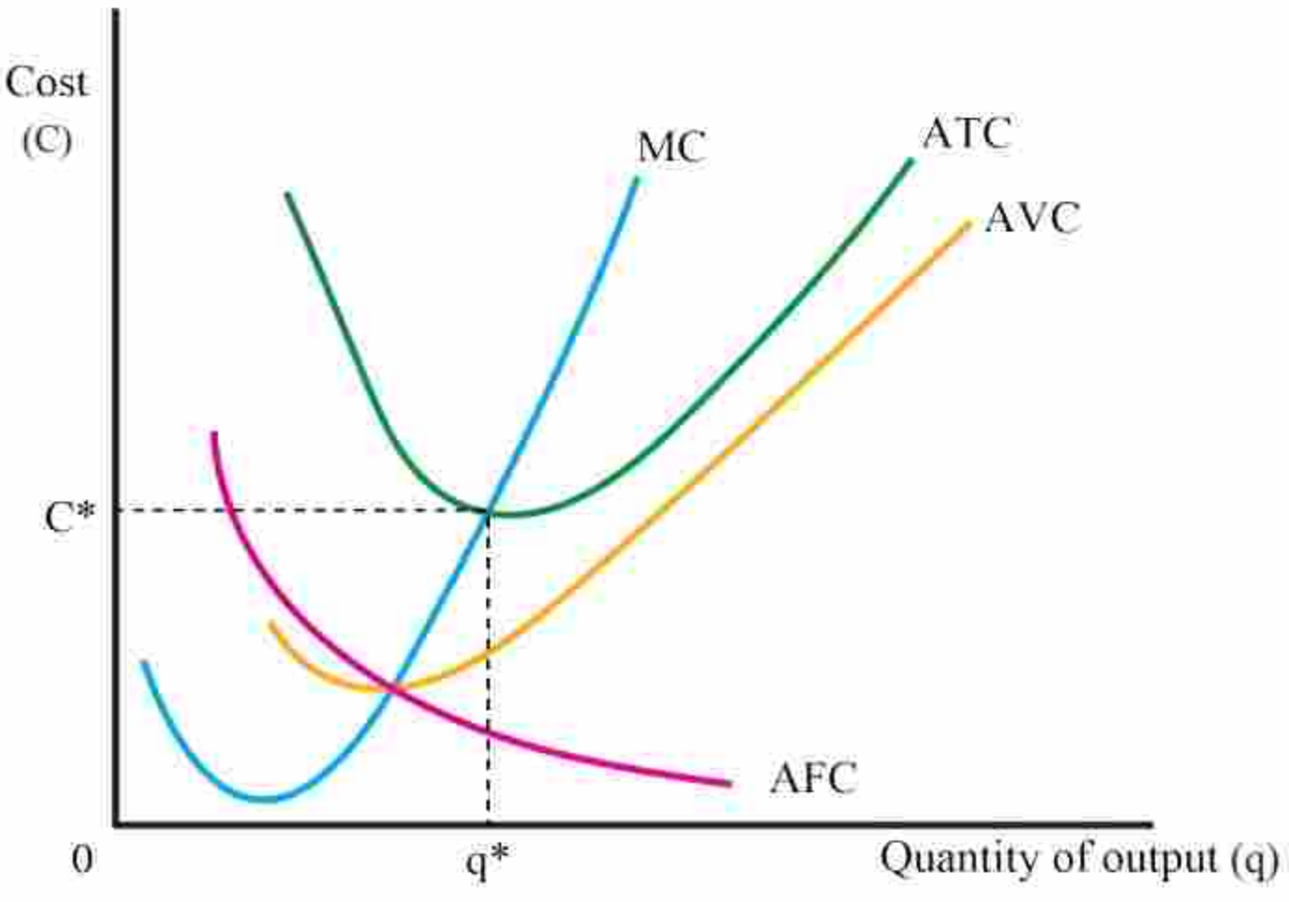


Figure 4.2: *Cost curves*

Shapes of ATC, AFC and AVC curves

The ATC, AFC and AVC curves are related as follows:

$$ATC = \frac{TC}{q}$$

Equation (4.7) shows that the shape of the ATC curve depends upon the shapes of the AVC and AFC curves. Figure 4.2 shows that both the AFC and AVC curves are declining in the early stage of production, and hence, the ATC curve also declines. The AFC curve continues to fall with the increase in output, whereas the AVC curve first declines, reaches a minimum and then rises. If the rise in the AVC curve is less than the decline in the AFC curve, the ATC curve will continue to decline. When the rise in the AVC curve is more than the fall in the AFC curve, the ATC curve will be rising. At first, the ATC curve is high due to a large AFC. As output increases, the ATC curve falls, which reduces the ATC until the AVC rises.

The gap between ATC and AVC curves at the beginning is large but it tends to decrease as the quantity of output increases. This is because as the quantity of output increases, the AFC curve decreases and approaches zero such that AVC becomes nearly equal to ATC.

The least-cost combination of inputs is at the lowest point of the ATC curve where the average total cost is the lowest. Nevertheless, the optimum q* in the short-run may not be the maximum level of output produced by the firm. It is just the point where per unit cost of production will be at the minimum.

Relationship between productivity of factors of production and cost

This part presents the relationships between average product and average variable cost, marginal cost and marginal product, and average total cost and total product of a firm.

Relationship between average product (AP) and average variable cost (AVC)

The AVC curve in Figure 4.3 is U-shaped because of its relationship with the average product (AP) of the variable input. The AVC curve is the mirror image of average product curve. As the AP increases, the AVC decreases due to the inverse relationship between AVC and AP.

The U-shape of the AVC results from increasing and diminishing marginal returns to variable factor as more output is produced.

The AP is at maximum level when additional employment of labour leads to no change in the AP, which corresponds to minimum AVC. However, an employment of more labour decreases the AP such that AVC will start to increase.

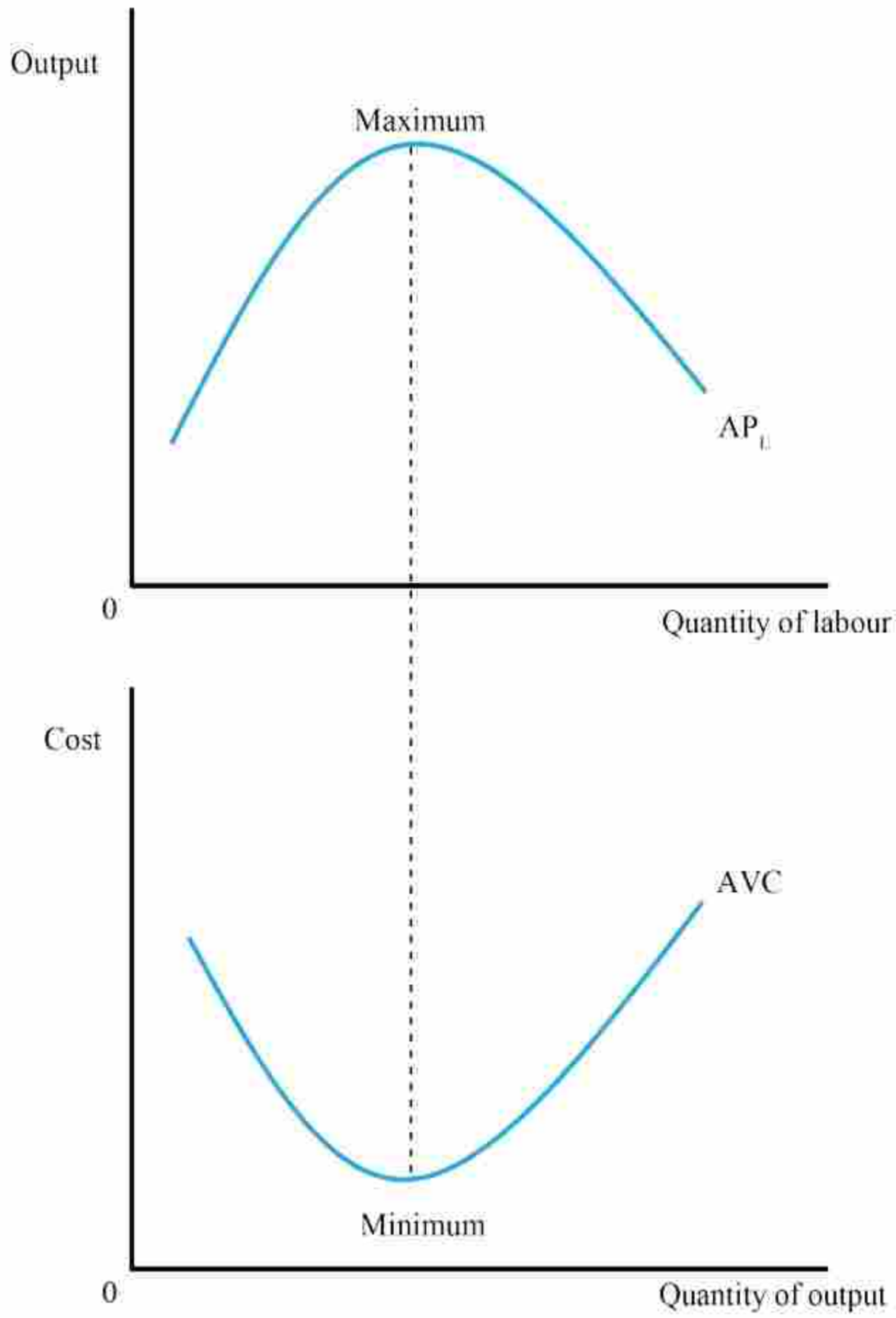


Figure 4.3: *Relationship between average product and average variable cost*

The relationship of the curves in Figure 4.3 is elaborated in Table 4.1 where it is assumed the wage of labour is Tshs 200,000. As shown in Table 4.1, the average product of labour (AP) is rising up to fourth unit of labour used in production, which suggests the existing of increasing returns to labour. The increasing return to labour is reflected in the AVC whereby it decreases up to the fourth unit of labour. The decrease in AVC is due to the fact that labour produces more than average returns even though it is paid the same wage. Therefore, when average product of labour (AP_L) reaches its maximum, the average variable cost (AVC) is at minimum as shown in Figure 4.3.

Table 4.1: *The relationship between costs and production*

Num-ber of workers	Total product (TP)	Average product (AP)	$\frac{1}{AP_L}$	Marginal product (MP)	$\frac{1}{MP_L}$	Total variable cost (TVC)	Average variable cost (AVC)	Marginal cost (MC)
0	0	0		-		0	-	-
1	50,000	50,000	$1/50,000$	50,000	$1/50,000$	200,000	4.0	4.0
2	140,000	70,000	$1/70,000$	90,000	$1/90,000$	400,000	2.9	2.2
3	250,000	83,333.3	$1/83,333.3$	110,000	$1/110,000$	600,000	2.4	1.8
4	350,000	87,500	$1/87,500$	100,000	$1/100,000$	800,000	2.3	2.0
5	420,000	84,000	$1/84,000$	70,000	$1/70,000$	1,000,000	2.4	2.9
6	450,000	75,000	$1/75,000$	30,000	$1/30,000$	1,200,000	2.7	6.7

Note: The digits have been rounded to one decimal place

From the fifth unit of labour, the Average product of labour (AP_L) starts to decrease. Specifically, after the fifth unit of labour employed by the firm, labour produces less output on average, but the wage is still the same, which leads to the variable cost per unit (AVC) to increase.

Assuming that labour is the only variable input in production, the relationship between AVC and AP is mathematically expressed as:

$$AVC = \frac{TVC}{q} \quad \text{.....(4.11)}$$

To a firm, the total variable cost (TVC) is:

$$TVC = L \times w \quad \ldots\ldots\ldots\

The additional output from employing the third unit of labour is 110,000 units, which is more than 90,000 added by employing the second unit of labour. The additional cost from employing the third unit of labour is calculated as:

$$\text{MC} = \frac{200,000}{110,000} = 1.8$$

Thus, the MC decreases from 2.2 to 1.8. Hence, with units of labour increasing in production, the quantity of output increases at an increasing rate, which leads to a decrease in the MC that depicts an inverse relationship between MP_L and MC. When the labour employed reaches 4 workers, the MP starts to decrease while the MC per unit starts to increase.

For example, Table 4.1 shows that the MP from employment of the fourth unit of labour is 100,000 units and the increase in TVC from employment of the additional unit of labour is Tshs 200,000.

$$\text{MC} = \frac{200,000}{100,000} = 2$$

Thus, as before, when a fifth unit of labour is employed, the MP is 70,000 units, such that:

$$\text{MC} = \frac{200,000}{70,000} = 2.9$$

Note that from the fourth unit of labour employed by the firm, output increases at decreasing rate such that the MP decreases and the MC increases.

The MC curve is a mirror image of the MP curve. When the MP curve is at its maximum, the MC curve is at its minimum, as illustrated in Figure 4.4. This relationship between the two curves depicts an inverse relationship between productivity and cost.

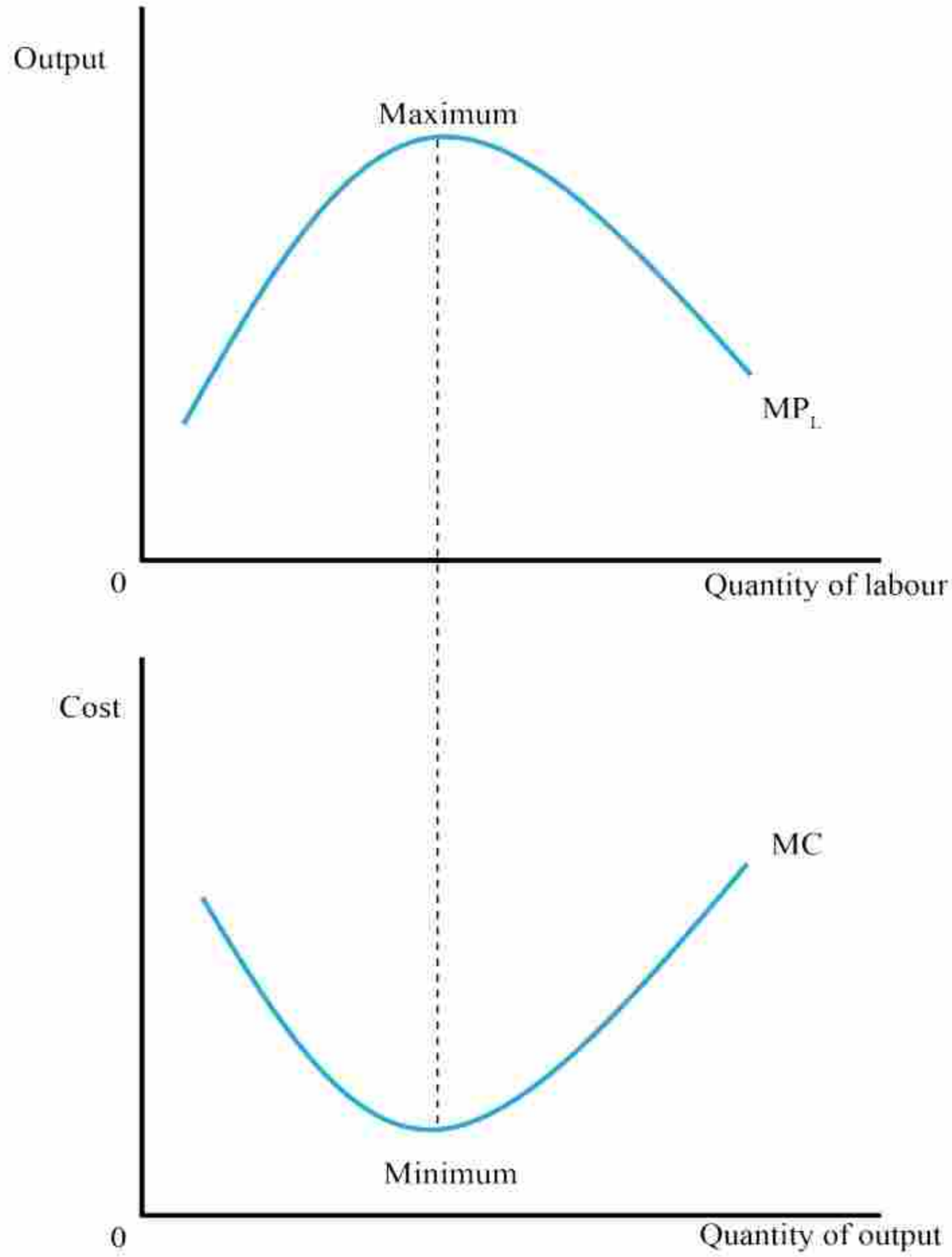


Figure 4.4: *Relationship between marginal product and marginal cost*

Mathematical relationship between marginal cost and marginal product

Technically, the relationship between marginal cost and marginal product is expressed as:

$$MC = \frac{\Delta TVC}{\Delta q} \quad \text{.....(4.18)}$$

Given Equation 4.12, it follows that:

$$MC = \frac{\Delta(L \times w)}{\Delta q} \quad \text{.....(4.19)}$$

But wage (w) per each unit of labour employed in a firm is constant such that:

$$MC = \frac{\Delta L}{\Delta q} \; w \quad \ldots\ldots\ldots\

Since 'a' is a constant value then,

$$MC = \frac{\partial(TC)}{\partial q} = b \quad \ldots\ldots\ldots\

To calculate the marginal cost, find the first derivative of total cost function.

$$MC = \frac{\partial(TC)}{\partial q} = \frac{\partial\left(4 + 2q + q^3\right)}{\partial q}$$

$$MC = 2 + 3q^2$$

Alternatively, marginal cost of the firm can be calculated by finding the first derivative of the total variable cost function. Thus:

$$MC = \frac{\partial(TVC)}{\partial q} = \frac{\partial\left(2q + q^3\right)}{\partial q} = 2 + 3q^2$$

Example 4.2:

A milk producing company is operating at a fixed cost of Tshs 200 and has a variable cost function of $TVC = 100q - 5q^2 + 0.15q^3$, find the:

(a) Total cost function of the firm.

(b) Firm's average cost function and average variable cost function.

(c) Marginal cost function of the firm.

Solution

(a) Remember, from Equation 4.2 the total cost of a firm is the sum of total fixed cost and total variable cost, mathematically written as $TC = TFC + TVC$.

$TVC = 100q - 5q^2 + 0.15q^3$ and $TFC = 200$

Therefore, the total cost function of the firm is

$TC = 200 + 100q - 5q^2 + 0.15q^3$

(b) To determine the average cost and average variable cost of the firm, divide the respective total cost and total variable cost functions with the level of output (q).

$$AC = \frac{TC}{q} = \frac{200}{q} + \frac{100q}{q} - \frac{5q^2}{q} + \frac{0.15q^3}{q}$$

$$AC = \frac{200}{q} + 100 - 5q + 0.15q^2$$

and

$$AVC = \frac{TVC}{q} = \frac{100q}{q} - \frac{5q^2}{q} + \frac{0.15q^3}{q}$$

$$AVC = 100 - 5q + 0.15q^2$$

(c) To determine the marginal cost, find the first derivative of the total cost function with respect to output q. Recall equation 4.10,

$$MC = \frac{\partial(TC)}{\partial q} = 100 - 10q + 0.45q^2$$



Activity 4.1

Recall and use knowledge of short-run cost of production to fill the table below. Share your work with your classmates.

Units of output (Q)	Total Fixed cost (TFC)	Total Vari-able cost (TVC)	Total cost (TC)	Average variable cost (AVC)	Average fixed cost (AFC)	Average cost (AC)	Margin-al cost (MC)
0	50	0					
1	50	20					
2	50	25					
3	50	30					
4	50	40					
5	50	60					
6	50	80					

Long-run cost-output relations

Recall that long-run is a period that is long enough to make all inputs variable. That is, there is no distinction between fixed and variable costs. In the long-run, a firm can increase or decrease its output according to the existing demand by increasing or decreasing all the factors of production. The firm is able to expand the scale of operation in the long-run by purchasing larger quantities of all inputs, which implies that all factors are variable.

Changing the scale of production is one way of changing the production in the long-run. In this regard, the long-run cost-output relationship will be influenced by the returns to scale. The other way, which is beyond the scope of this book is substitutability between factor inputs.

In the long-run, a firm has a number of alternatives with regard to the scale of its operation. For each scale of production or plant size, the firm has a separate short-run average cost curve. Hence, the long-run average cost curve is composed of a series of short-run average cost curves, as presented in Figure 4.5. In the long-run, a firm will examine which short-run average cost curve to operate with in producing a given quantity of output at a minimum possible cost. A firm may decide to operate at SATC1, SATC2 or SATC3. Thus, Figure 4.5 shows how costs vary with the scale, that is, the size of a firm's operation.

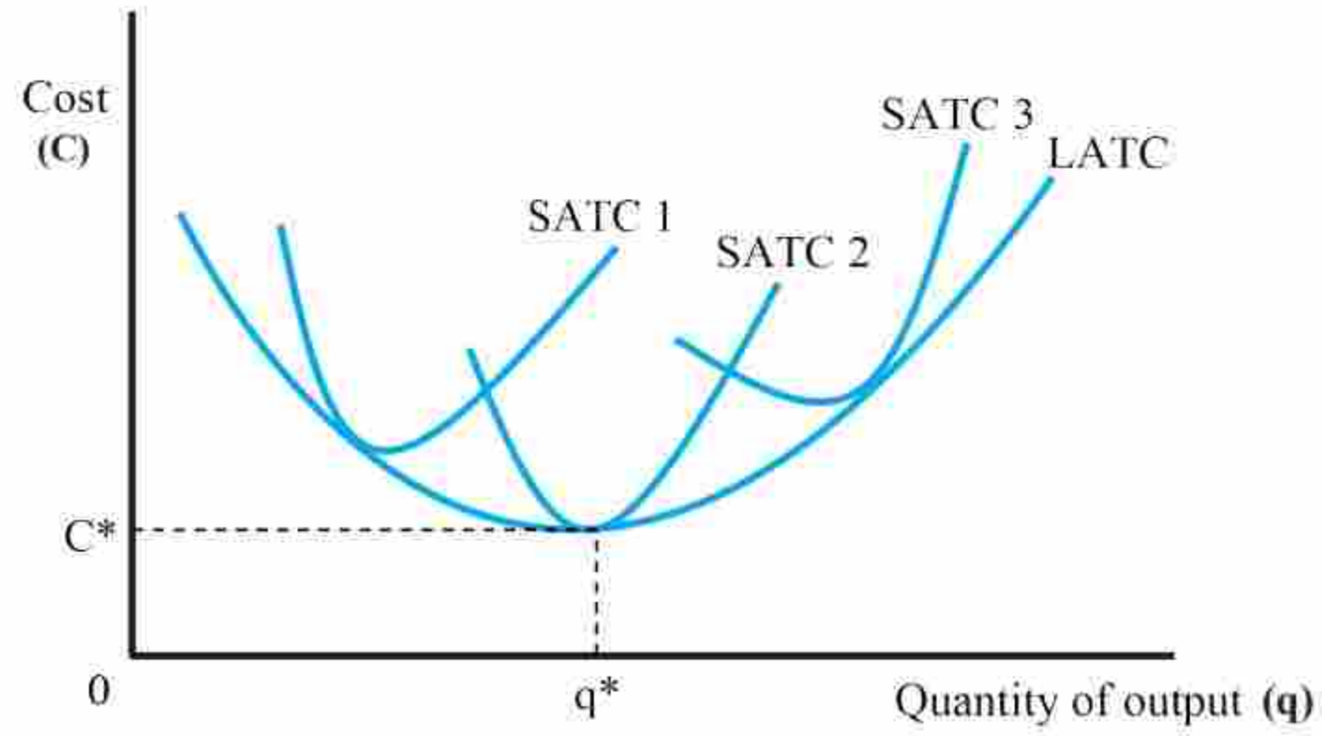


Figure 4.5: *Long-run average cost curve*

The long-run ATC curve in Figure 4.5 is a U-shape because of existence of **economies** and **diseconomies** of scale in production. As explained in Chapter Three, economies of scale are advantages enjoyed by a firm when engaged in large scale production that lead to decrease in cost per unit of output produced.

Diseconomies of scale are disadvantages faced by a firm when engaged in large scale production that leads to an increase in the unit cost of outputs produced by a firm. Thus, in Figure 4.5, when long-run average cost declines as the quantity of output increases, the firm gains from economies of scale. When the long-run average cost does not vary with the level of output produced, the firm earns constant returns to scale. When the long-run average cost increases, the firm faces diseconomies of scale.

Optimum size of the firm

A firm is said to be at an optimum point of production over the long-run when its long-run average total cost (LATC) is at its lowest level (minimum), as presented in Figure 4.6. To determine its optimal output, the firm has to consider demand as well as cost.

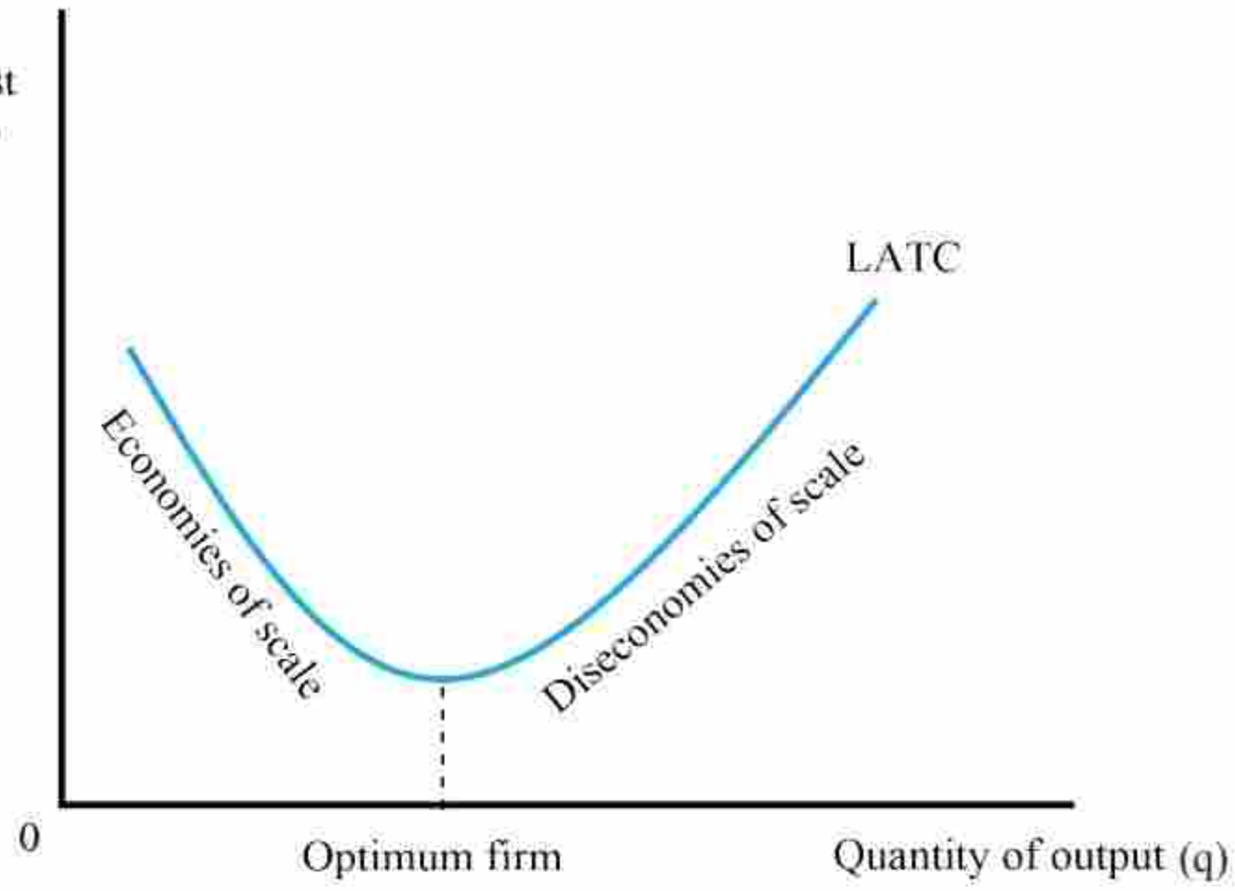


Figure 4.6: *Optimum firm*

As shown in Figure 4.6, an optimal firm is considered as a firm that has set up a plant with the lowest possible cost and is operating it at the lowest point of the long-run average cost curve of production. At this point, resources are mostly efficiently utilised.

The long-run average cost of the firm is influenced by various economies and diseconomies of scale. Such economies and diseconomies are determined by various technical, managerial, financial and marketing factors.

The optimal size of the firm is achieved when all the internal economies such as division of labour, use of specialised machinery, managerial advantages of large scale production are being fully enjoyed by the firm.

Exercise 4.3

1. Explain the relationship between total cost, average total cost, and marginal cost.
2. Differentiate between short-run and long-run average total cost curves.
3. Fill the following table:

Term	Definition	Mathematical description
Fixed costs		
Variable costs		
Total cost		
Average fixed cost		
Average variable cost		
Marginal cost		

4. Draw a diagram depicting the relationship between marginal cost and marginal product curves of a firm. Explain shapes of the two curves.
5. Given a total cost function, $TC = 64 + 74q - 33q^2 + 17q^3$, derive;
   (a) Average total cost and average variable cost functions.
   (b) Marginal cost function of the firm.
   (c) The firm's total cost, average total cost, average variable cost and marginal cost when it produces 16 units of output.

Revenue analysis and profit maximisation of a firm

The main objective of most firms is to maximise profit. Profit is the difference between a firm's total revenue and total costs. The revenue of a firm is the amount of money that it receives from selling goods or services. If a firm gets Tshs 40,000 from sale of 20 pairs of socks it produced, then the amount Tshs 40,000 is the **total revenue (TR)** of the firm.

The total revenue of a firm is obtained by multiplying the price (p) per unit of the good and quantity of the good sold (q).

$$\text{Total revenue} = \text{Price} \times \text{Quantity}$$

$$TR = p \times q \quad \ldots\ldots\ldots\

Example 4.3:

The demand function equation faced by a firm is provided by $P = 35,000 - 5q$, determine:

(a) Marginal revenue equation.

(b) Price and quantity at which marginal revenue is zero.

(c) The price elasticity of demand of a firm using the results obtained in item (b) and interpret the results.

Solution

(a) To obtain the marginal revenue equation, first find the total revenue function by multiplying the given demand equation by q, since $TR = q \times p$. Therefore,

$$TR = (35,000 - 5q) \times q$$

$$TR = 35,000q - 5q^2$$

The marginal revenue equation is given by the first derivative of the total revenue function as:

$$MR = \frac{\partial(TR)}{\partial q} = 35,000 - 10q$$

$$MR = 35,000 - 10q$$

(b) To obtain the price and quantity at which the marginal revenue is zero, set the marginal revenue function equal to zero. Thus:

$$35,000 - 10q = 0$$

$$35,000 = 10q$$

$$q = 3,500 \text{ units}$$

Substitute the value of q into the demand equation to obtain the price associated with the quantity (3,500).

$$P = 35,000 - 5(3,500)$$

$$P = 35,000 - 17,500$$

$$P = Tshs 17,500$$

Therefore, when quantity produced is 3,500 units and price is Tshs 17,500 per unit sold, the marginal revenue is zero.

(c) To obtain the price elasticity of demand of a firm, recall that:

$$e_p^d = \left| \frac{\partial Q}{\partial P} \times \frac{P}{Q} \right|$$

and the given demand function is $P = 35,000 - 5q$

The $\frac{\partial Q}{\partial P}$ is obtained by rearranging the demand function in terms of quantity demanded as:

$$5q = 35,000 - P$$

$$q = 7,000 - \frac{1}{5}P$$

from the demand function equation, the coefficient of P is $-\frac{1}{5}$ which means that $\frac{\partial Q}{\partial P} = \frac{1}{5}$. Remember, in (b) the quantity equals to 3,500 and price equals to Tshs 17,500 when the total revenue is at maximum.

$$e_p^d = \left| \frac{\partial Q}{\partial P} \times \frac{P}{Q} \right|$$

$$e_p^d = \left| -\frac{1}{5} \times \frac{17,500}{3,500} \right|$$

$$e_p^d = |-1| = 1$$

Therefore, price elasticity of demand of a firm is equal to 1. This is unitary elasticity since the coefficient of elasticity equals to one, which means that a one percentage change in price leads to a one percentage change in the quantity demanded.

Example 4.4:

Mr Kasewe is running a barbershop nearby a school. The demand function for barber services is $q = 135 - 3P$.

(a) Determine the total revenue function.
(b) At what quantity and price of barber services is the revenue is maximised?
(c) Find the value of maximum total revenue.

Solution

(a) To determine the total revenue function, rewrite the demand function in terms of price as a function of quantity demand.

$$3P = 135 - q$$

$$P = 45 - \frac{q}{3}$$

Since total revenue equals price times quantity,

$$TR = \left(45 - \frac{q}{3}\right) \times q$$

$$TR = 45q - \frac{q^2}{3}$$

The total revenue function of Mr Kasewe's barbershop is given as

$$TR = 45q - \frac{q^2}{3}.$$

(b) To obtain the quantity and price at which Mr Kasewe's barbershop revenue is maximised, first, determine the first derivative of the total revenue function.

$$\frac{\partial(TR)}{\partial q} = 45 - \frac{2}{3}q$$

Equate the first derivative of the total revenue to zero, to obtain the quantity sold at which the total revenue is maximised.

$$45 - \frac{2}{3}q = 0$$

$$45 = \frac{2}{3}q$$

$$q = 67.5$$

Substitute the value of q into the demand equation

$$P = 45 - \frac{q}{3}$$

$$P = 45 - \frac{67.5}{3}$$

$$P = 22.5$$

Therefore, Mr Kasewe's barbershop maximises revenue at 67.5 units and price of Tshs 22.5.

(c) To obtain the value of maximum total revenue, substitute the value of revenue maximising quantity (q=67.5) into the total revenue function obtained in item (a).

$$TR = 45q - \frac{q^2}{3}$$

$$TR = 45(67.5) - \frac{(67.5)^2}{3}$$

$$TR = 3,037.5 - \frac{4,556.25}{3}$$

$$TR = 3,037.5 - 1,518.75$$

$$TR = 1,518.75$$

Therefore, maximum total revenue (TR) is Tshs 1,1518.75.

Example 4.5:

Mkude and Ambakise Company Limited dealing with rice cultivation in Morogoro, had kept the following records: price of labour is Tshs 1,000, price of capital is Tshs 10,000, quantity of labour 50 units and quantity of capital 100 units. Ambakise, who acts as a financial advisor to the company, revealed that Tshs 500,000 could be obtained by the company if the money was invested in shares. During that period, however, capital depreciated by Tshs 2,000. Total output of the undertaking was 420 bags of rice and price per each bag was Tshs 10,000. Obtain:

(a) Total cost of production
(b) Total implicit cost
(c) Total explicit cost
(d) Total revenue

(e) Economic profit
(f) Accounting profit

Solution

(a) To obtain the total cost of production recall that:

$$TC = TFC + TVC$$

But, TVC = Units of labour × Price of labour

Thus, given the units of labours = 50 units and price of labour = Tshs 1,000

$$TVC = 50 \times 1,000$$

$$TVC = 50,000$$

Then, TFC = Units of capital × Price of capital

Given the units of capital = 100 units and price of capital = Tshs 10,000

$$TFC = 100 \times 10,000$$

$$TVC = 1,000,000$$

Therefore, $TC = 1,000,000 + 50,000$

$$TC = \text{Tshs } 1,050,000$$

(b) Total implicit cost, which is the amount that could be earned by the fixed capital in alternative uses is given as:

Total implicit cost =Tshs 500,000

(c) Total explicit cost, which is the actual payment in the form of money is obtained as:

$$\text{Explicit cost} = TFC + TVC + \text{Depreciation}$$

$$\text{Explicit cost} = 1,000,000 + 50,000 + 2,000$$

$$\text{Explicit cost} = 1,052,000$$

(d) Total revenue of the firm is obtained as follows:

$$TR = \text{Price} \times \text{Quantity}$$

$$TR = 420 \times 10,000$$

$$TR = \text{Tshs}\, 4,200,000$$

(e) Economic profit

$$\text{Economic profit} = 4,200,000 - (500,000 + 1,052,000)$$

$$\text{Economic profit} = \text{Tshs}\ 2,648,000$$

(f) Accounting profit

$$\text{Accounting profit} = \text{Total revenue} - \text{Explicit cost}$$

$$\text{Accounting profit} = 4,200,000 - 1,052,000$$

$$\text{Accounting profit} = \text{Tshs}\ 3,148,000$$

Rule for profit maximisation by a firm

A firm will maximise profit when the difference between its total revenue (TR) and total cost (TC) is at maximum. A firm will minimise loss when the difference between total cost and total revenue is at minimum.

Mathematically:

$$\pi = TR - TC \quad \ldots\ldots\ldots\

Implying that,

$$\frac{\partial TR}{\partial q} = \frac{\partial TC}{\partial q} \quad \ldots\ldots\ldots\

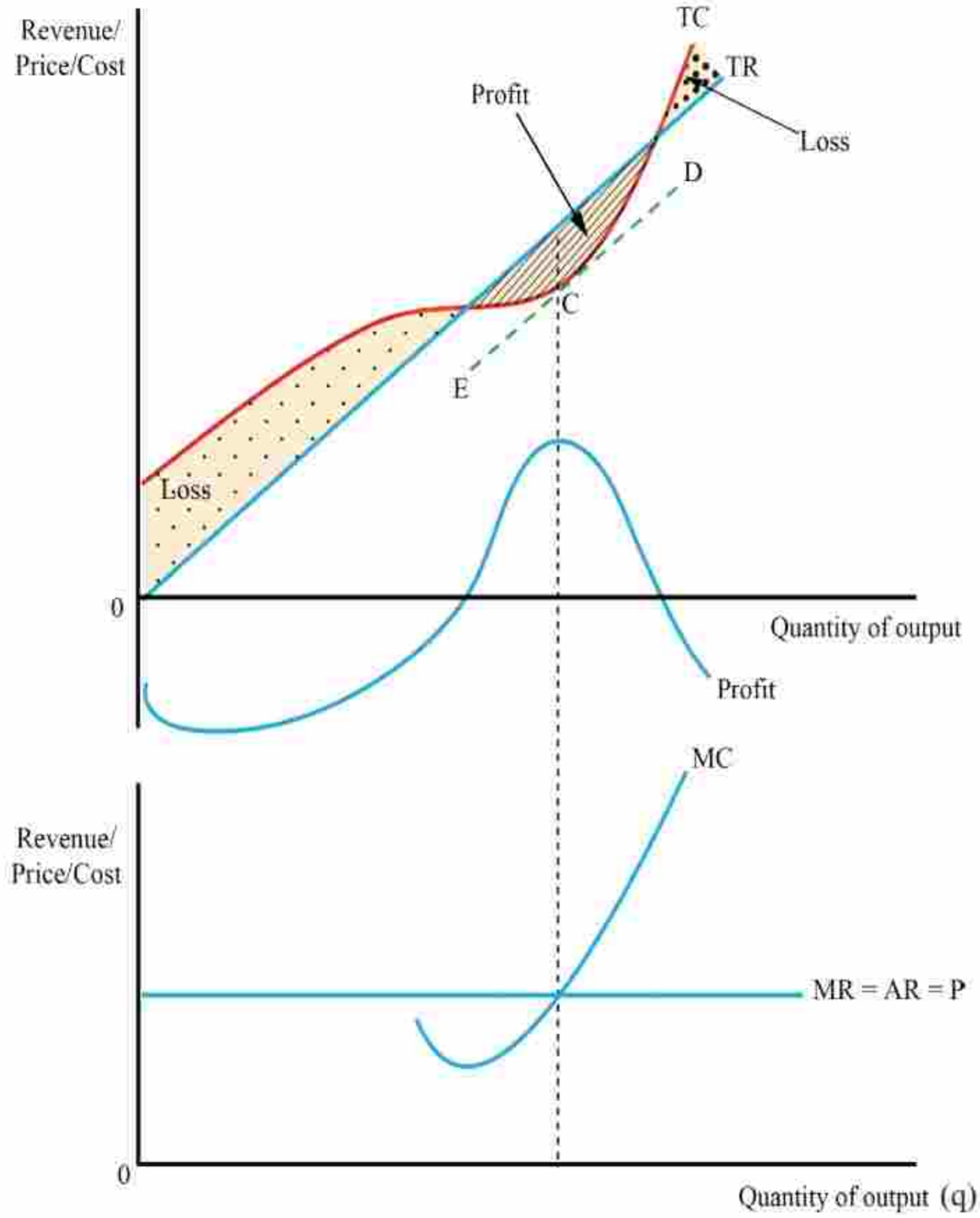


Figure 4.7: *Profit maximisation curve of a perfectly competitive firm*

Remember, Figure 4.7 shows that the slope of the line ED, which is tangent to the TC curve, at C is equal to the slope of TR. This emphasises that it is at C only where the firm will maximise profit.

The first and second order conditions can also be elaborated graphically. In Figure 4.8, the MC = MR at q_1 and q_2. The question is at which point profit is maximised. At point q_1, the MC is still decreasing and it is less than MR. So, any additional output will generate more revenue because the addition to cost is less than the addition to revenue. Therefore, q_1 is not the equilibrium level of output of the firm.

The firm will continue to increase output as long as $MC < MR$ until it reaches point q_2 where $MC = MR$. Any additional output will lead marginal cost to be greater than marginal revenue (MC > MR), and so the firm will opt to remain at q_2 where

it attains its equilibrium. At the point of equilibrium, whether a firm gets profit or loss depends on the price and cost of production.

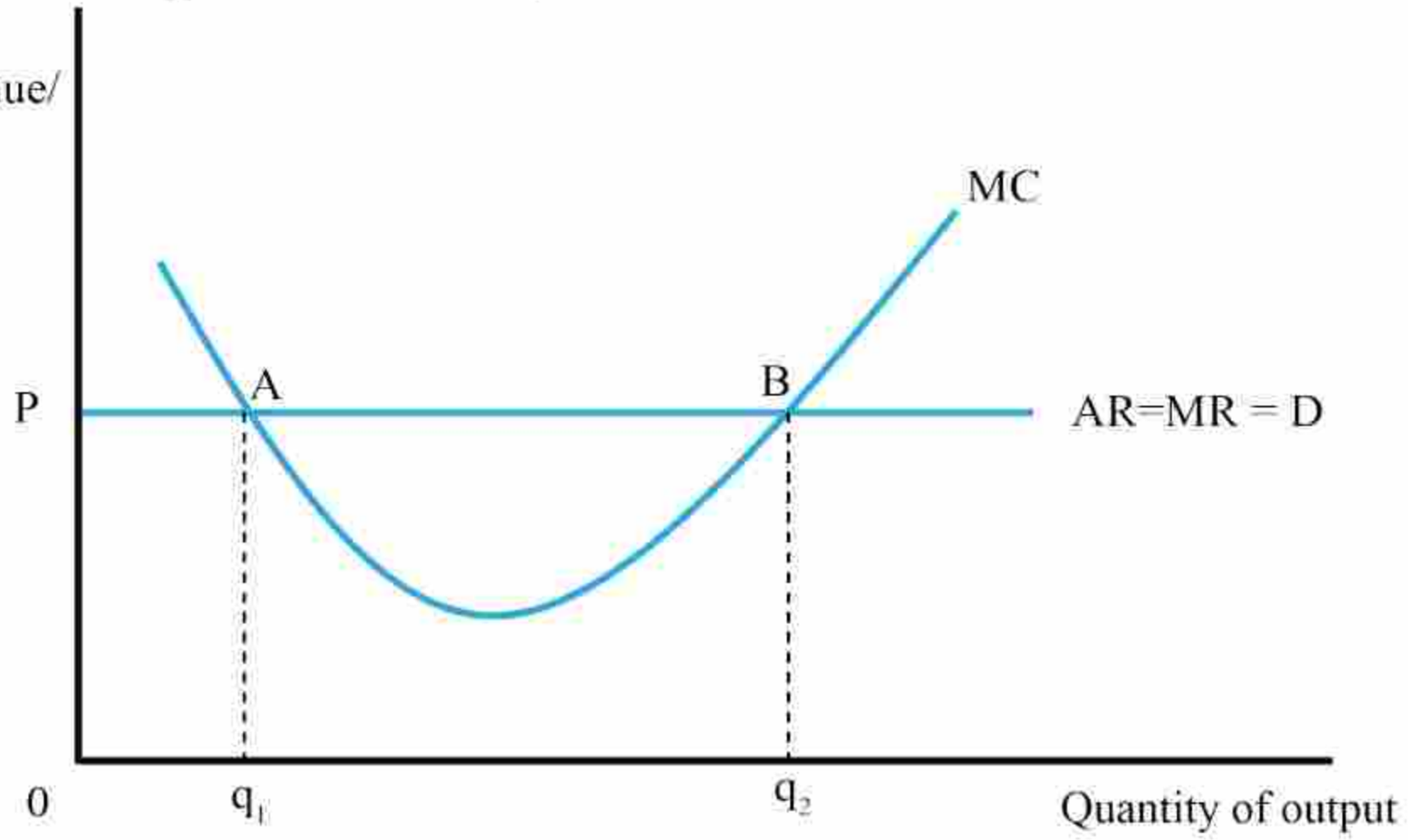


Figure 4.8: *Equilibrium of the firm*

Example 4.6:

A coconut oil firm has the following revenue and total cost function:

$$TR = 640q - 4q^2$$

$$TC = 3600 + 100q + 6q^2$$

Find the level of output at which the firm will maximise profit.

Solution

Profit maximising output of a firm is obtained when marginal revenue equals marginal cost (MR = MC).

Take the first derivative of the total revenue function to obtain the MR function,

$$TR = 640q - 4q^2$$

$$MR = \frac{\partial TR}{\partial q} = \frac{\partial\left(640q - 4q^2\right)}{\partial q}$$

$$MR = \frac{\partial TR}{\partial q} = 640 - 8q$$

Take the first derivative of the total cost function to obtain the MC function,

$$TC = 3600 + 100q + 6q^2$$

$$MC = \frac{\partial TC}{\partial q} = \frac{\partial\left(3600 + 100q + 6q^2\right)}{\partial q}$$

$$MC = \frac{\partial TC}{\partial q} = 100 + 12q$$

Now, in equilibrium, MR = MC,

$$100 + 12q = 640 - 8q$$

$$640 - 100 = 12q + 8q$$

$$540 = 20q$$

$$q = 27$$

The profit maximising level of output is 27 units.

Exercise 4.4

1. Distinguish between a firm's profit and its revenue. Explain which among the two is maximised by the firms.
2. Suppose the firms demand for a given product is $q = 300 - 3P$. Where Q is quantity demanded of the product and P is price of the product per unit.
   (a) Determine the output and price at which the firm will maximise revenue.
   (b) Obtain the price elasticity of demand of a firm and interpret the results.
3. Suppose a firm faces a cost function $C = 6 + 3q + q^3$
   (a) What is the firm's fixed cost?
   (b) Find the firm's average cost and marginal cost.



Activity 4.2

In a group of three students, visit the library and use the website to understand conditions for a firm to maximise profit.

(a) Elaborate the conditions using a mathematical derivation.

(b) Present the conditions graphically.

Share your work with the rest of the groups in class.

Chapter Summary

1. A firm is an entity that uses factors of production to produce goods and/or services.
2. An industry comprises firms producing similar products.
3. A business oriented firm always aims at maximising profit. Profit is maximised when the difference between total revenue and total cost of the firm is at maximum level.
4. Total costs of a firm can be divided into fixed costs and variable costs. Fixed costs do not vary with the quantity of output produced by the firm. However, the variable costs changes when the quantity of output produced changes.
5. A firm's cost behaviour can be analysed by drawing of average total cost and marginal cost. As output increases, the average cost initially falls and eventually increases as output continues to increase. Marginal cost curve always cuts the average total cost curve and the average variable cost curve from below at the minimum of both curves.
6. Economies and diseconomies of scale explain why the long-run average cost curve is U-shaped.
7. A firm maximises profit if it is in equilibrium, and the equilibrium occurs when the marginal cost equals the marginal revenue.

Revision exercise

1. Explain the following terms:
   (a) Implicit cost
   (b) Explicit cost
   (c) Private cost
   (d) Social cost
   (e) Accounting cost
2. Explain how economic costs differ from accounting costs.
3. The following information was extracted from the books of accounts of Kakakuona Road Construction Company. Price of labour input Tshs. 10,000. Units of labour required are 100 units. Price of capital used is Tshs. 100,000. Quantity of capital required is150 units. Total length of road constructed for the inputs used above is 800 kilometres. Moreover, the general Manager revealed that his company had foregone some interest of Tshs. 50,000, which would have been obtained by the company if the money was deposited in the bank during the period. Assuming that the firm paid Tshs. 200,000 price per km constructed. Find;
   (a) Total cost of production
   (b) Implicit cost
   (c) Explicit cost
   (d) Total revenue
   (e) Profit
4. What does it mean to say that an individual firm is "small relative to the industry?"
5. Explain why when MP is at maximum MC is at minimum. At this point, when output is increased what happens to MC and MP respectively? Illustrate by using a graph.
6. With fixed costs of Tshs 400,000, a firm has average total costs of Tshs 3,000 and average variable costs of Tshs 2,500. What is its output?
7. If you know that with 8 units of output, average fixed cost is Tshs 12,500 and average variable cost is Tshs 81,250, what will be the total cost at this output level?
8. Suppose that a firm produces 200,000 units a year and sells all of them for Tshs 10 each. The explicit costs of production are Tshs 1,500,000 and the

implicit costs of production are Tshs 300,000. What is the firm's accounting profit?

9. Study the table carefully, and answer questions which follow:

OUTPUT	TVC	AVC	AFC	TFC	TC	MC
0	0				120,000	
1,000	150,000					
2,000	290,000					
3,000	420,000					
4,000	540,000					
5,000	640,000					
6,000	720,000					
7,000	795,000					
8,000	875,000					
9,000	975,000					
10,000	1,095,000					

Fill in the blanks.

(a) Plot the AC, AVC, AFC and MC curves.

(b) Explain the shape of the average cost curve.

10. Show that the average variable cost (AVC) of a firm is a reciprocal of the average product of labour.

11. A firm's total variable cost is given as $TVC = 320q - 10q^2 + q^3$. Will a firm produce the product if the price of the product is Tshs 40?

12. Suppose the revenue and total cost of a firm are given by the equation:

$TR = 60q$ and $C = 10 + 5q^2$

Find the profit maximising output and total profit of a firm.

13. Sketch the graph of the short-run cost curves and name the curves.

(a) Explain why the total cost and total fixed cost curves originate at a common point.

(b) Provide reasons for a constant gap between total cost curve and total variable cost curve.

(c) Explain why the total variable cost curve starts from the origin.

14. A firm's total revenue and total cost functions are given as: $TR = 10q$ and $TC = 1500 + 2q + 0.01q^2$ respectively. Determine the profit maximising quantity of output.

15. The total cost equation for the production of spectacles at Boma's factory is given as:

$C = 1200 + 500 - 15q^2 + q^3$ where, C = Cost, measured in Tanzanian shillings and q = Quantity measured in pairs of glasses.

(a) Compute the total and average costs at the output quantity of 10 and 11 pairs of glasses.

(b) What is the marginal cost of the 10^{th} pair?

16. Given that $AVC = 10 + 20q$ and fixed cost = 100. Determine:

(a) Total cost function

(b) Marginal cost function

(c) Profit function if the price per unit output is Tshs 15 and quantity sold is q units.

Chapter Five

The market

Introduction

You might have noticed that there are different types of markets in an economy. Some markets are dominated by a single or few sellers and many buyers. Others have a single or few buyers and many sellers. The difference in number of sellers and buyers gives the different definitions of what a market is and different types of markets that may exist in an economy. In this chapter, you will learn the concept of market from the economics point of view, classification of markets according to its type and structures, and determination of price and quantity of a good or service in the market. The competencies developed will enable you to determine the efficient level of price for various market structures.

Concept of a market

The word 'market' in economics means a mechanism through which buyers and sellers interact to exchange goods and services. The market exists when two or more individuals are willing to enter into an exchange transaction, regardless of place or time. Therefore, the market is where buyers and sellers are in contact to exchange goods and services thereby determining the price and quantity of the good bought and sold.

Traditionally, markets have been identified by physical locations. For example, Kariakoo and Tandale markets in Dar es Salaam, Memorial market in Kilimanjaro, Kingalu market in Morogoro, Mwanjelwa market in Mbeya, Makoroboi market in Mwanza, Kemondo market in Kagera and Kibirizi market in Kigoma and Mwanakwerekwe and Darajani markiti Market in Zanzibar. However, due to development and advancement in information and communication technologies (ICT), a market is not necessarily a physical place. Given the ICT, trade transactions may take place by digital means. For example, some of the modern

forms of markets are such as Alibaba, Amazon, Kikuu, and Kupatana. Figure 5.1 presents different types of digital markets that take place through E-mails, mobile applications, social media, and websites.

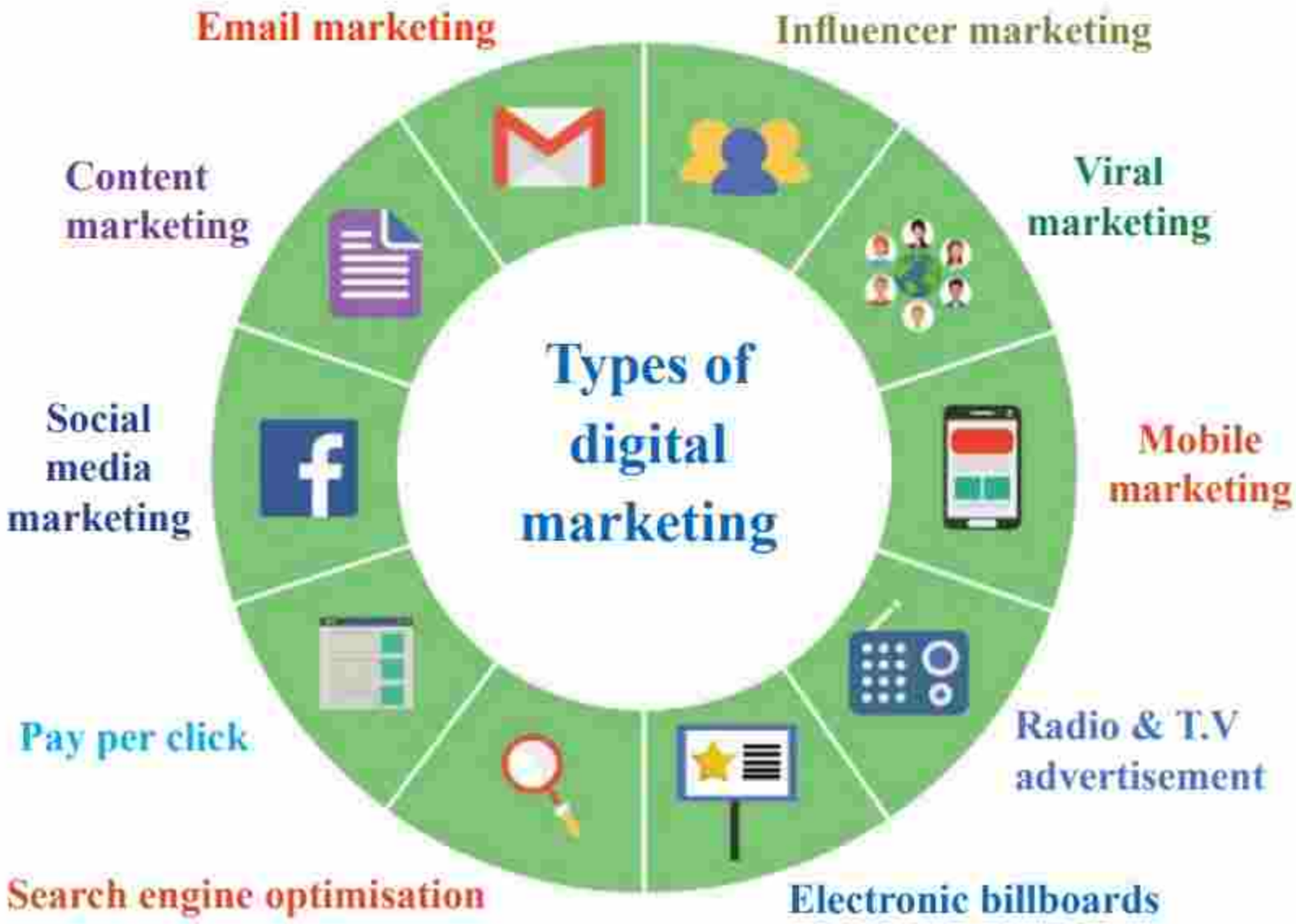


Figure 5.1: *Digital marketing*

Essentials of a market

Existence of a market is influenced by various factors, among others, presence of a good or service to be exchanged in the market at a specific price; and existence of buyers and sellers who exchange or trade the goods and services in the market. It is essential to have the location of a market (not necessarily a physical place) where buyers and sellers can interact to exchange goods and services. Likewise, there must exist a contact between the buyers and sellers. The contact could include physical or means of communication media (telephones and websites).

Determinants of market size

The size of markets for various commodities is quite variable. Some markets may cover the whole country or even the whole world while others can be small and cover a small village or a street. The size of a market also depends on the type and the nature of a commodity. For example, goods that are durable or easily portable tend to have a wider market than goods that are easily perishable, bulky, or heavy to carry.

The size of the market also depends on nature of demand for a commodity. Commodities that have universal demand and general consumption such as fuel (diesel and petrol) and minerals (gold and silver) have a wide market.

Similarly, well-developed means of transport and communication systems enable goods and services to be transported to various places in the world. This leads to expansion of the market. In the contrary, if means of transport and communication are not well developed, transportation of goods and services becomes difficult, leading to existence of a small market size.

The size of the market may also be affected by government policies. On one hand, policy and institutional factors can restrict growth of markets through prohibitive laws, duties, tariffs and quotas. On the other hand, some policies and institutional factors such as provision of subsidies can promote consumption of certain goods and services that lead to expansion of the size of the market. In addition, stability of a currency induces people to have confidence in it and thus, expand the market. One example is the dollar (USD) of the United States of America (USA), whose stability has made it one of the major hard currencies in the world.

Functions of a market

A market performs various functions but mainly, it facilitates pricing and allocation of goods and services. Selling and buying of goods and services facilitated by the market establish prices at which consumers buy and producers sell goods as well as services.

Besides, price signals from the market guide allocation of resources in the economy. A rise in the price of a good due to an increase in its demand will induce the producer(s) of the good to increase its supply to earn more profits. Eventually, the increase in supply of the good will lead to a fall in its price in the market.

In addition, the market gives a signal to producers either to expand or to reduce the scale of its activity over time. When the size of the market for a particular commodity is large enough, it encourages firms to scale-up (increase) production. However, expansion of production depends on nature of the commodity. Commodities of specific brands or that are special tend to have a small market with few buyers who are willing to pay a higher than normal market price.

Markets also act as a source of social relations. They facilitate development of close relationships between buyers and sellers. Frequent contacts between a buyer and a seller may lead to creation of mutual relationship and form a sense of community (local or international) to promote peace and security.

Classification of markets

Markets are usually classified according to their types and structure. The following are some major classifications of markets.

Classification of markets according to type

A market is classified according to goods traded (exchanged). Markets that involve buying and selling of final consumer goods and services are referred to as product or commodity markets. Examples include markets for cereals, clothes or health services.

There are also markets for factors of production, which are referred to as factor markets, for example, a market for labour or a market for land. There are financial markets in which money and other financial assets such as shares (stocks), treasury bills market and bond are traded.

Classification of a market according to structure

A market structure reflects the nature and degree of competition in the market. Determining the market structure requires establishment of number and size of firms in the market, nature of the goods traded, market shares of the largest firms, extent of product differentiation, entry and exit conditions and cost of transport. Taken together, the forementioned factors provide a useful picture of degree of efficiency that exists in a market.

Market structures are classified into two categories: perfectly competitive and imperfectly market structures. The latter includes monopoly, monopolistic competitive and oligopoly markets.

Perfectly competitive market (PCM)

This is a benchmark market structure. All other market structures are compared against perfectly market structure. A pure or perfectly competitive market is rare in the real world. It is important in economics to understand its main features that help in examining industries with characteristics related to pure competition. Specific examples of markets that are characterised by perfect competition include

agricultural commodity markets, 'pikipiki' or motorcycle transport market and foreign exchange markets.

Characteristics of perfect competitive market

The following are characteristics of a perfect competitive market:

Many buyers and sellers: A perfect competitive market (PCM) is an ideal market with many buyers and sellers (firms). Thus, any individual action by a single buyer or seller in a perfectly competitive market has an insignificant impact on the market price. Hence, in PCM, the market price is determined by forces of demand and supply, which implies that both buyers and sellers are price takers.

Homogeneous products: In a perfect competitive market, goods or services sold by firms and bought by households are largely homogenous. Thus, all firms produce same standard product in the market, and thus, high competition exists between firms.

No barriers to entry: A perfect competitive market has no restrictions on entry or exit of a firm to or from the market. For instance, anyone is allowed to start selling maize. If the existing maize sellers are no longer interested in the business, they can leave because there is no barrier to entry or exit from the market.

High competition: Because there are many sellers of the same standard product in the market, high competition exists between firms. As a result, there occurs proper and effective utilisation of resources (no wastage of resources) by the firms. It is assumed that sellers and buyers have perfect knowledge of the product sold in the market. Hence, advertisement has no role to play in a perfectly competitive market structure.

No transport cost: It is assumed that there are no transport costs incurred by the firms operating under perfect competitive market. This is because goods under PCM are charged the same price everywhere. Thus, if transport costs are added to the price, homogeneous commodities produced under this market would have different prices.

Price and output determination under perfect competitive market

A firm is said to be optimising the use of its resources when it maximises profit or conversely, minimises its costs of production. Therefore, at the optimal level of production, a firm has no desire to change its level of output because it will be maximising profit at that level.

Individual firms in a competitive market regime cannot influence the price of goods. They are price takers in that they accept the price determined by market forces of demand and supply.

Suppliers will supply the quantity that consumers will buy in the market and there will be no excess supply. Because of competition, an excess supply would reduce the price of the goods. In turn, excess demand for a good would cause the price of the goods or services to rise to clear the market. If price is above or below the equilibrium price, the forces of demand and supply will pull it down or push it up to the equilibrium situation.

Figure 5.2 shows the industry price is P, which is determined by the forces of demand and supply at a point where the two curves intersect. At the equilibrium point (E), an individual firm cannot change the price, which is for the whole industry (market), because there are no incentives to do so, given their size. Since every individual firm is atomistic, all firms take price as given. Therefore, line P on the individual demand curve is the demand curve for a firm, showing that the firm can produce any quantity of output that it is capable of producing at price P.

In addition, because the firm sells every unit of its products at the same price, the price of a good is equal to the firm's average revenue and marginal revenue. Thus, P = AR = MR.

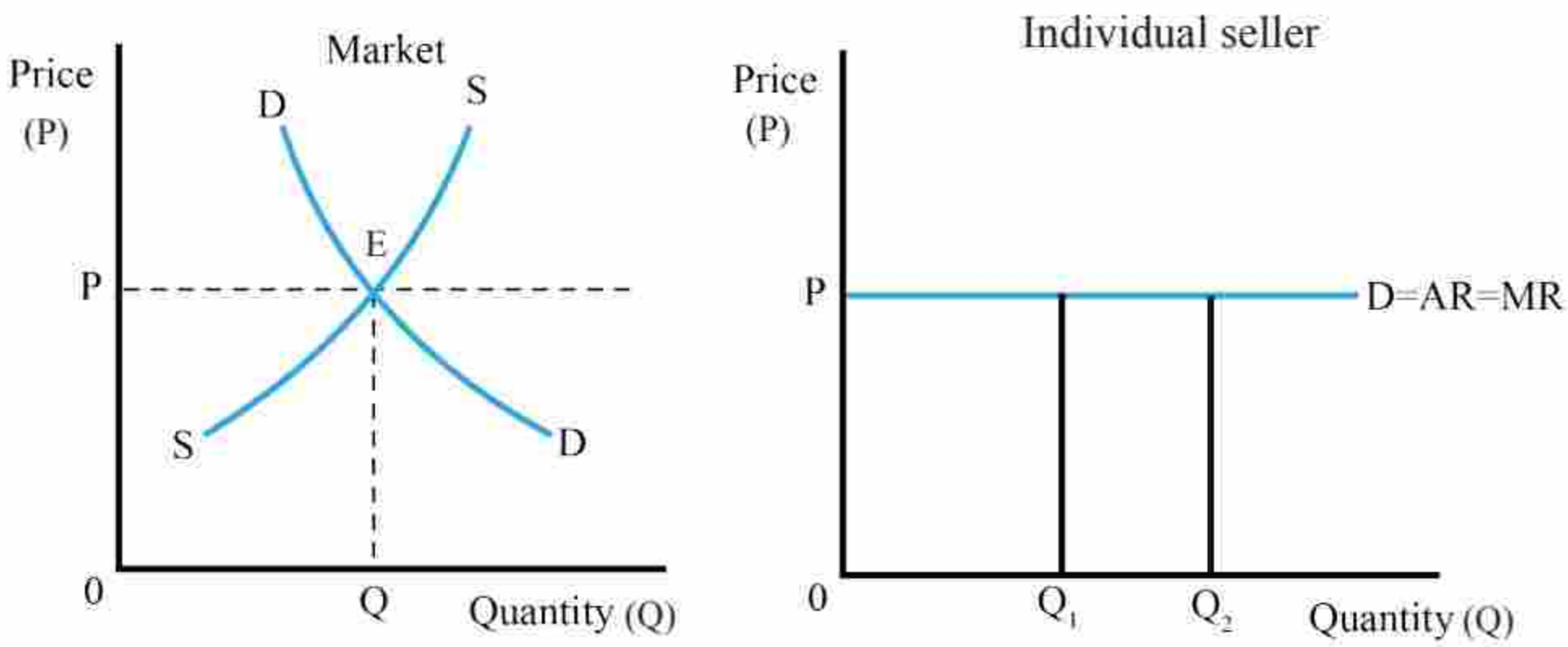


Figure 5.2: *Demand curve under perfect competitive market*

Profit maximisation and cost minimisation under perfect competitive market

One of the main objective of all firms in any perfectly competitive market economy is to maximise profit. These firms are faced with two key decisions:

(a) Maximising profit in the short-run by deciding the quantity of output to produce; and

(b) Maximising profit by deciding whether to stay, enter or exit the market in the long-run.

Condition for short-run profit maximisation

"One may ask, Can a firm in a competitive market earn profit?"

In the short-run, a firm may earn super normal profit, normal profit or loss. Profit is the difference between total revenue and total cost of production. The profit is calculated as total revenue minus total costs whereby total costs is sum of fixed costs and variable costs. A firm may earn super normal profit in the short-run when average revenue (AR) is greater than average total costs (ATC). However, in the long-run, a firm earns normal profit since average revenue (AR) of a firm is equal to average total costs (ATC).

Supernormal profit

A firm in a perfectly competitive market economy maximises profit when:

(a) $MC = MR$, since for a perfectly competitive firm, $MR = P$. Thus, the profit maximising condition can be re-written as $P = MC$;

(b) Marginal cost (MC) curve cuts the marginal revenue (MR) curve from below; and

(c) $P > ATC$

In the short-run, a firm in a perfectly competitive market can get super normal profit when AR is greater than ATC. Now, because the firm sells every unit of a good it produces at the same price, it means that AR = MR = P. Hence, to earn super normal profit, the price should be greater than average total cost, that is, $P > ATC$ at the equilibrium, and MC curve must cut ATC curve from below.

In Figure 5.3, when the firm is in equilibrium at point A, where MC is equal to MR, that is, MC = MR, P > ATC and output of the firm is Q*. The firm's revenue (area 0Q*AP) is greater than the firm's costs (area 0Q*BC). Therefore, the firm earns super normal profit as shown by the shaded area (CBAP).

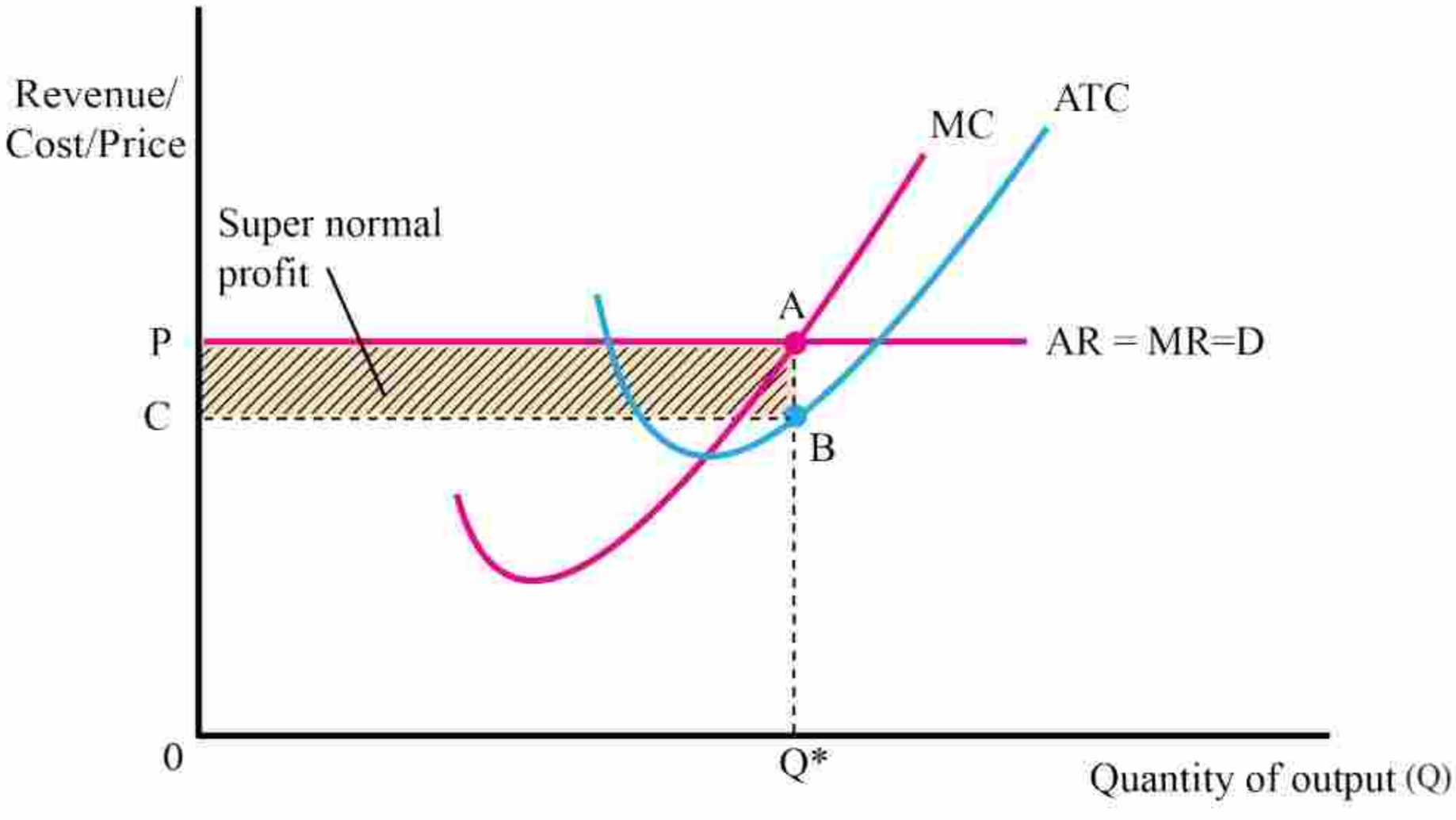


Figure 5.3: *Short-run supernormal profit under a perfect competition market*

Normal profit

A firm earns normal profit when its total revenue is equal to its total cost. Thus average revenue is equal to average cost, which implies that normal profit is part of the total cost. Now, because normal profit is included in the total cost of a firm, the firms' earnings over and above the total cost are called supernormal profit. Hence, when the firm earns just normal profit, it means that supernormal profit is zero for the firm.

A firm in a perfectly competitive market earns normal profit when:

(a) MC = MR;

(b) MC curve cuts the MR and ATC curves from below, which means that the MC curve must be on its rising portion; and

(c) P = ATC.

In Figure 5.4, the firm is in equilibrium at point A where MC = MR = P = ATC. That is, the firm's total revenue (area 0QAP) is equal to the firm's total cost (area 0QAP). Therefore, the firm earns a normal profit.

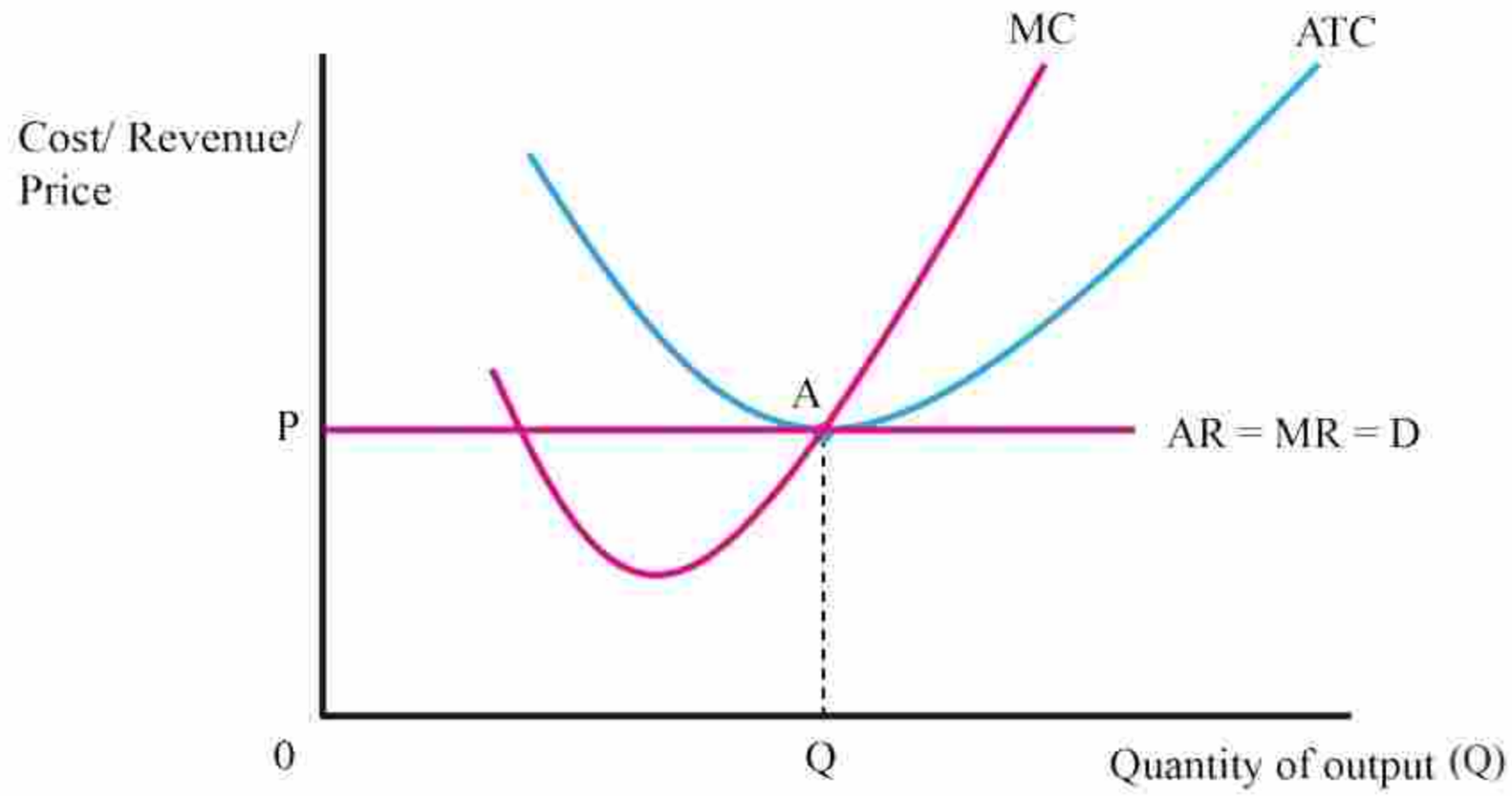


Figure 5.4: *Short-run normal profit under a perfect competition market*

Loss

A firm incurs loss when its total revenue is less than its total costs. The condition for loss making in perfect competition is as follows:

(a) MC = MR (Equilibrium point);

(b) P < ATC; and

(c) MC curve must cut MR and ATC curves from below, which shows that it must be on its rising portion.

In Figure 5.5, at point A where MC = MR, price is less than average total cost (ATC), that is, P < ATC. The firm's revenue (area 0QAP) is less than the firm's total costs (area 0QBC). Therefore, the firm is incurring loss (area PABC) as shown in the shaded part.

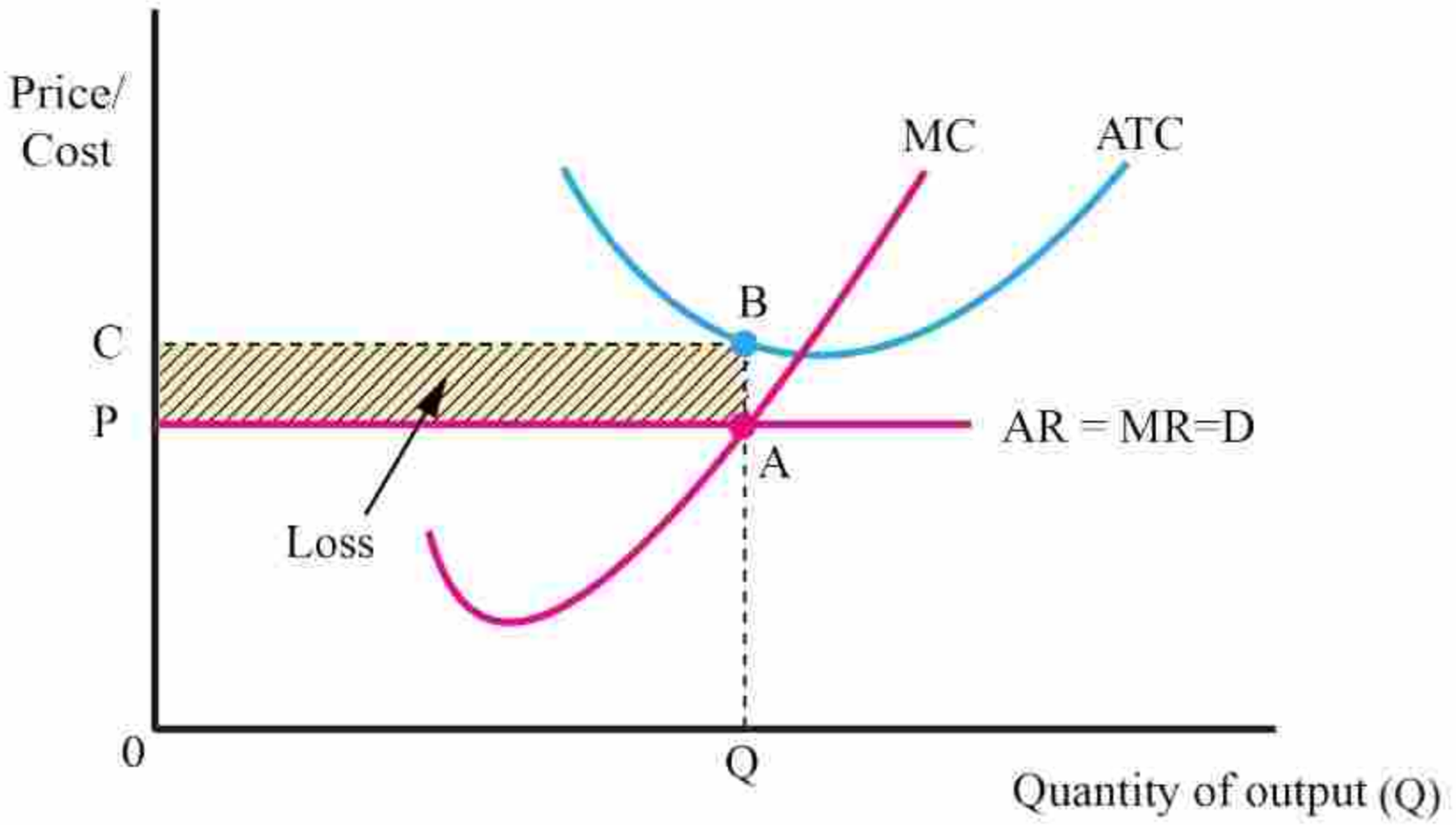


Figure 5.5: *Short-run loss under perfect competition market*

Break-even point

Break-even point of a firm occurs when the firm is able to just cover its total cost of production, which includes both fixed and variable costs. Figure 5.6 shows the break-even point as point D. At this point, P_1 = ATC = MC, which means that at the break-even point, a firm earns just normal profit.

Shut-down point

A firm will shut down if it cannot cover its variable costs. Hence, at the shut-down point, a firm is able to cover only its variable costs such as salaries, cost of raw materials, and other business expenses. A firm cannot avoid the fixed costs even when it does not produce any output. However, it can avoid variable costs by not producing any output. If the price falls below the minimum AVC, it means that it cannot cover the variable costs, which it can then avoid by not producing any output. Hence, if the price falls below the minimum AVC, it will decide to shut-down thereby not to produce any output, and hence, its loss will be only the fixed costs, which are unavoidable.

Figure 5.6 shows that at point E, the shut-down point is the level of output and price at which the firm just covers its total variable cost. If the price of the product (P_2) is less than the minimum average variable cost (AVC), the firm will shut-down to minimise loss because when P_2 is less than minimum AVC, each additional unit produced would increase the loss. At the shut-down point (E), the

firm's economic loss equals its total fixed costs. Thus, over the short-run, if the price is smaller than the minimum AVC, then the firm should shut-down. If the price is greater than AVC, the firm should continue to produce.

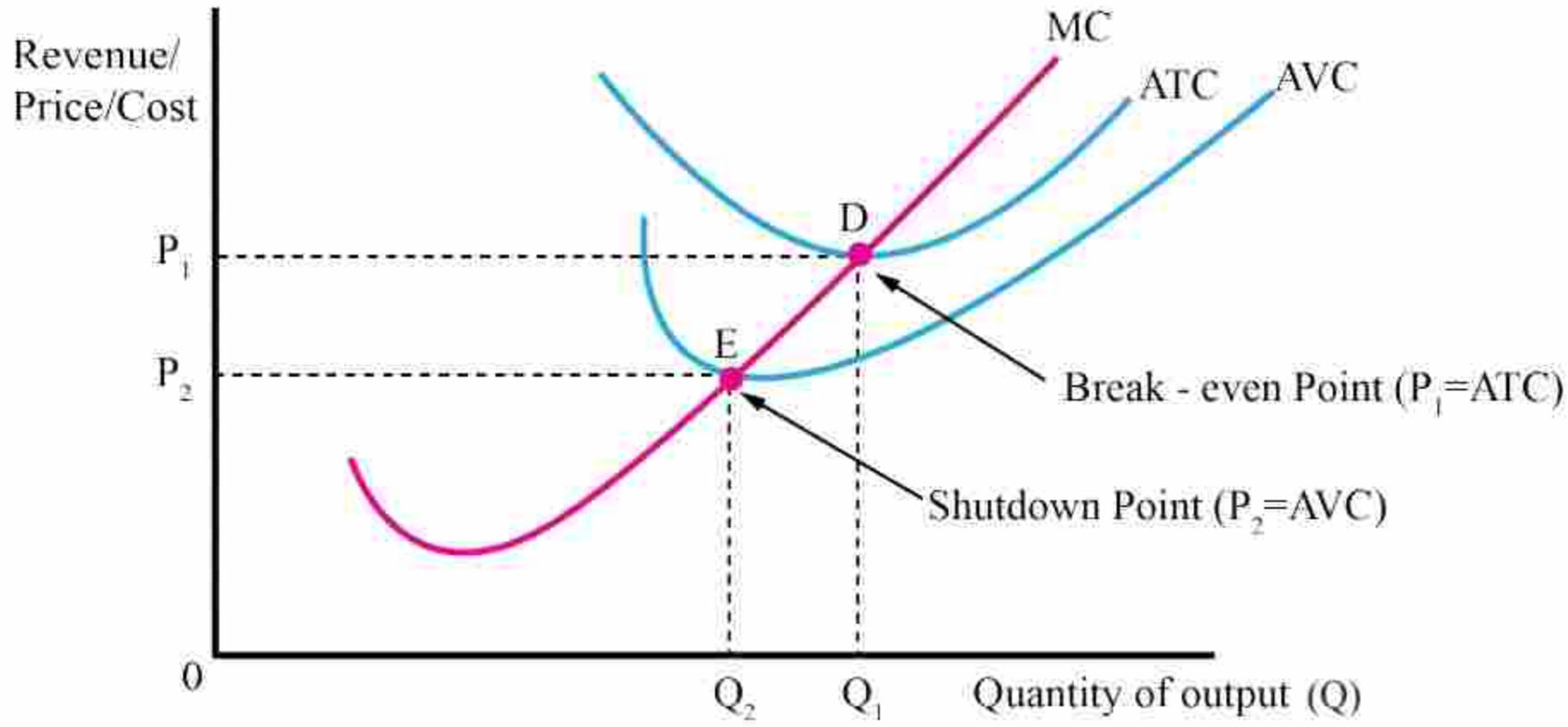


Figure 5.6: *Break-even and shut-down point for a perfectly competitive firm*

Conversely, price and MC are compared to find profit maximising or loss minimising output. This leads to the following conditions:

(a) If $P > AVC$ firms produce in the short-run (normal profit or supernormal profit);

(b) If $AVC \leq P \leq ATC$ firm still produces in the short-run but if the economic situation does not improve, it will exit the market in the long-run [that is to say, although it is incurring loss (negative profit), the loss is smaller than total fixed costs]; and

(c) If $P < AVC$, a firm will shut down straightaway (since the loss is larger than total fixed costs) and hence, the firm is better off by shutting down and incurring loss of only the fixed costs.

Example:

Let a firm's fixed cost be Tshs 1,000 and its minimum AVC be Tshs 555. If the market price is Tshs 550, which is less than minimum AVC, the firm will incur a loss of Tshs 5 for each unit of output it produces. It follows that if the firm produces only one unit of output, its total loss is Tshs 5 plus Tshs 1,000, which is the fixed costs. If instead the firm decides to shut down, its loss will be only Tshs

1,000, as the variable costs will be zero. Thus, the decision to shut down will be a loss minimisation strategy of the firm.

If the market price is Tshs 600, the firm will gain Tshs 5 more by producing each unit of output. If the firm produces one unit, its loss will be the fixed cost less the variable cost, that is, Tshs 1,000 – Tshs 5 = Tshs 995. Therefore, the firm will be better off by producing one unit and not shutting down because the incurred loss (Tshs 995) will be lower than if it were to shut down (Tshs 1,000).

When the market price is higher than the minimum AVC, then the optimal level of output will be at a point where MR = P = MC.

Short-run supply curve under perfect competition

In a perfectly competitive market, the short-run supply curve is the part of marginal cost curve above the minimum average variable cost. It shows how the firm's profit maximising output varies as the price varies, other things remaining the same. The supply curve is based on the marginal cost and the so-called shut-down decision. In Figure 5.7, the supply curve at prices less than P_1 is the vertical line $0P_1$ since at any price below P_1, the firm would have shut-down. Hence, positive output Q starts from where the firm's MR and MC intersect at point E with equilibrium output Q_1. If the price rises from $0P_1$ to $0P_2$, the output produced increases to Q_2. Hence, the portion of the firms' short-run supply curve along which Q greater than 0 starts at point E upwards.

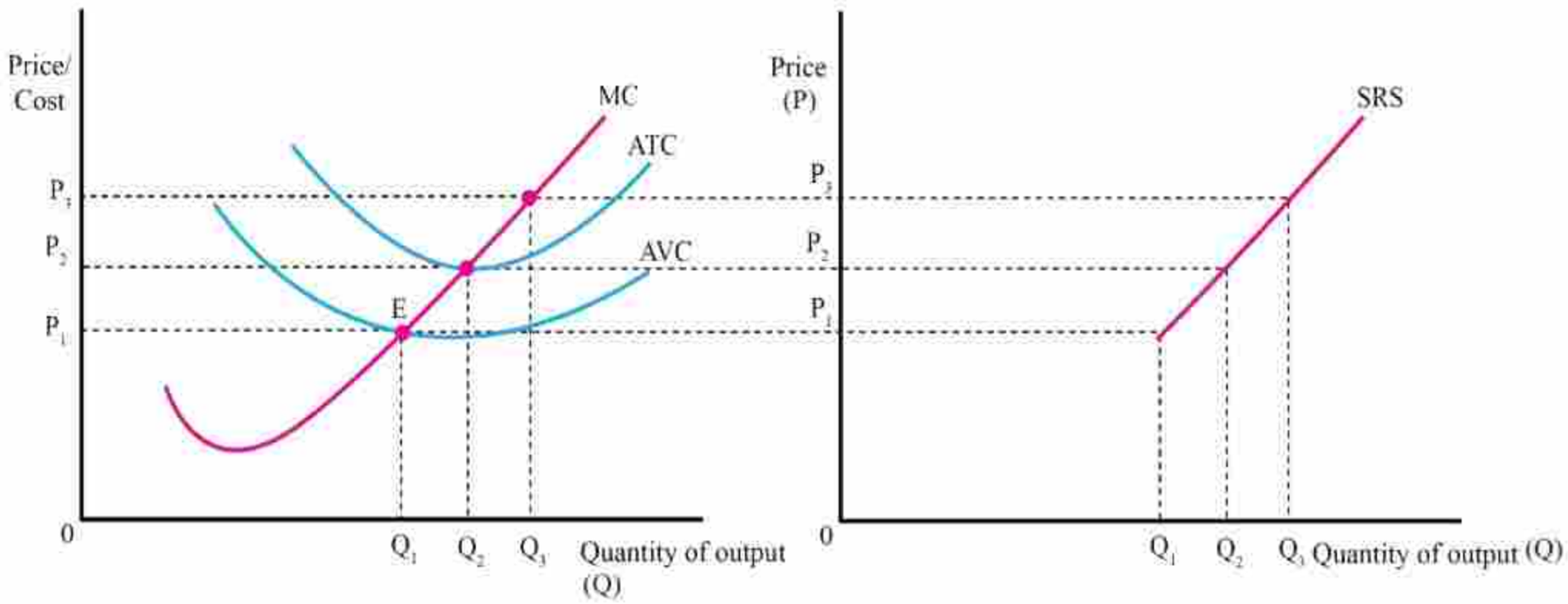


Figure 5.7: *Short-run supply curve of a perfect competition market*

Long-run equilibrium of a firm under perfectly competitive markets

In the long-run, firms in a perfect competition earn normal profit. This is because in the short-run, some firms earn supernormal profit, which attract other firms to join the industry and share the profit that after some time becomes normal. The transition from supernormal profit to normal profit results from the assumption of free entry and free exit from the industry. However, firms that are making loss in the short-run have to stop production in the long-run, in which all factors of production are variable.

Condition for long-run profit maximisation

(a) The firm must be in the equilibrium, that is, long-run marginal cost (LMC) = long-run marginal revenue (LMR).
(b) LMC curve cuts LMR curve from below.
(c) Price = long-run average cost (LATC).

In the long-run, firms in a perfectly competitive market earn normal profit. Normal profit is earned when the firm's total costs equal total revenue.

In Figure 5.8, the LMC = LMR = LAR = LATC at point A. The firm's total revenue is area 0QAP and the firm's total cost is area 0QAP .

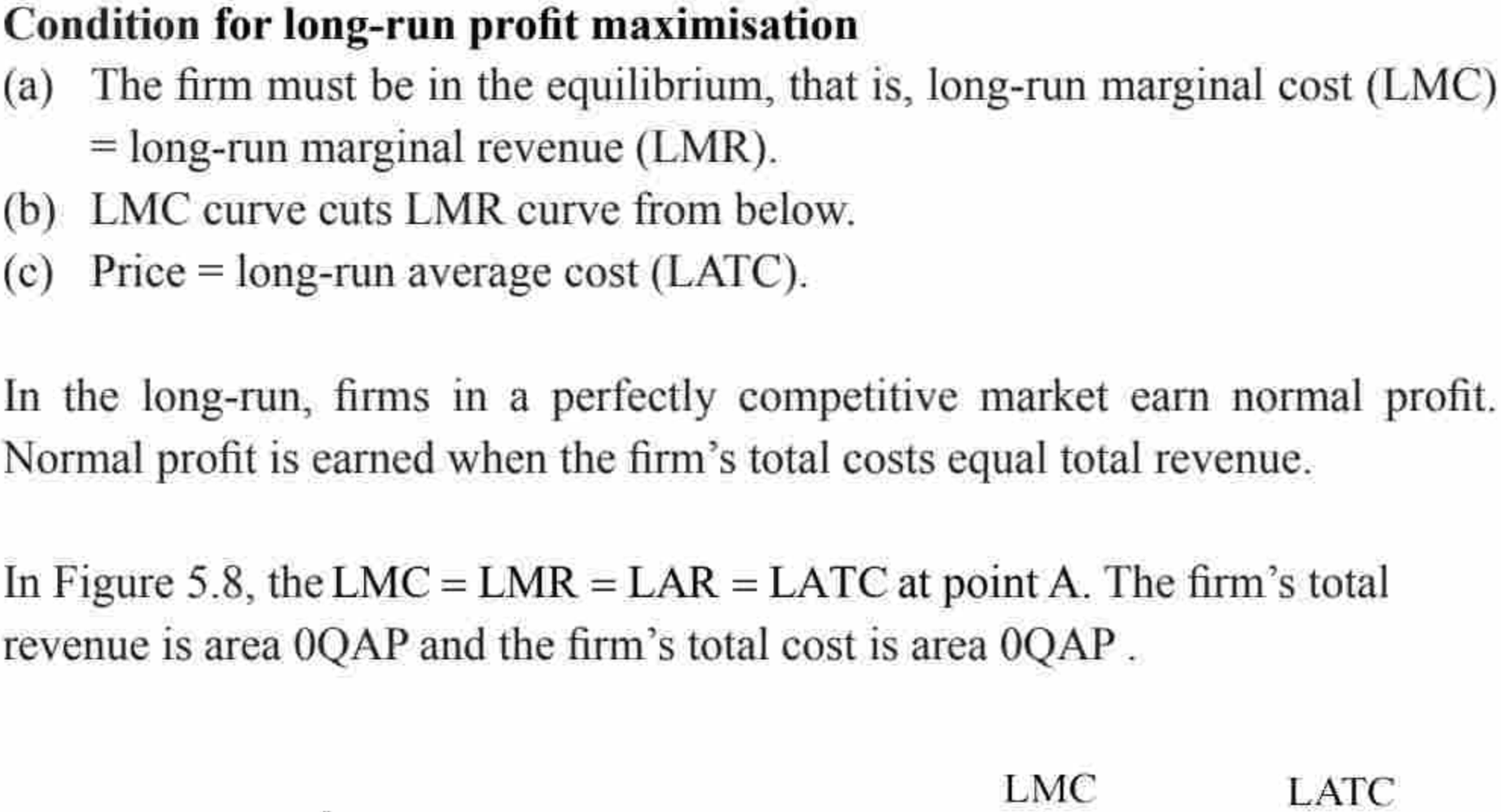


Figure 5.8: *Long-run profit maximisation under perfect competition*

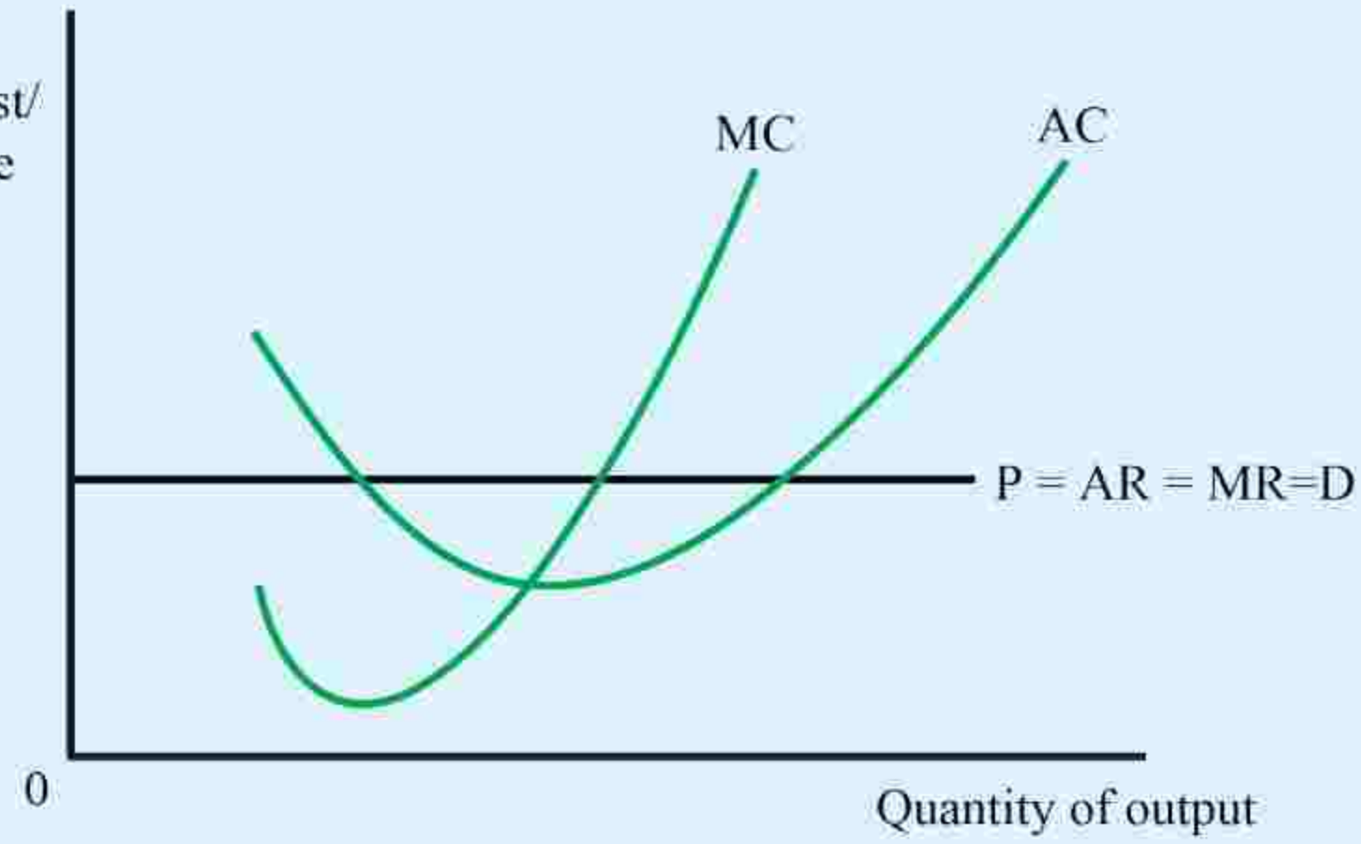
In the long-run, firms produce at minimum average cost. Moreover, there is no excess capacity, costs and prices are at the minimum and as such, they are fair, and stable to afford a good standard of living to consumers. In addition, there is reduction in the unemployment rate due to perfect mobility of factors of production. There is no shortage of goods and services due to existence of many sellers in the market, which cause existence of equilibrium between demand for and supply of goods or services.

However, a perfect competitive market lacks varieties of goods and services because commodities produced by firms are homogeneous, which limit consumers' choice. Unemployment also is likely to occur due to elimination of inefficient firms in the market or business. Additionally, certain goods such as public goods (for example, roads or utilities) may not be provided.

Furthermore, the assumptions of perfectly competitive markets are criticised for being unrealistic as they present a market situation that does not exist in practice. For example, consumers are assumed to have perfect knowledge of the product, whereas in reality, they are ignorant with regard to quality and price of commodities in the market. Lastly, the competitive markets discourage investments because the normal profit earned in the long-run may not incentivise firms to undertake innovation, invention and research to improve products and expand outputs.

Exercise 5.1

1. What is meant by market structures and why do they exist?
2. Distinguish between short-run and long-run curves for a perfectly competitive firm.
3. Study the figure and answer the following questions:



(a) Explain whether the firm makes loss or profit. Give reasons for your answer.
(b) Shade the area that shows profit or loss.
(c) Identify the points at which the firm is at equilibrium

4. Explain the difference between a competitive firm's break-even point and its shut-down point. Describe conditions under which each would occur and illustrate them graphically.

Imperfect market structures

Imperfect market structures include monopoly, monopolistic competition and oligopoly markets. These market structures arise from violation of assumptions of the perfect competitive market such as imposition of barriers to entry and exit of firms in the market and reduction in number of buyers in the market.

Monopoly market

Monopoly is a market for a good or service with no close substitute and only one dominant seller in the market. A monopoly exists because a stable supplier in the market is protected from competition by barriers that prevent entry of new firms. In practice, a pure monopoly does not exist, although public utilities such as water, electricity and natural gas services resemble forms of monopoly. A typical example of a monopoly in Tanzania is the Tanzania Electricity Supply Company Limited (TANESCO). Similarly, Tanzania would be a monopoly in supply of Tanzanite if it could control production and supply of the mineral.

Characteristics of monopoly market

Single seller of the product: In a monopoly market, there is a single seller with no close substitute for the commodity in the market.

Barriers in entry: In a monopoly market, there are restrictions of entry. Such restrictions can be of forms like economical, legal or institutional.

No close substitute: A monopolist sells a product, which does not have close substitute. Thus, cross elasticity of demand for such product is either zero or very small. Also, the price elasticity of demand for the monopolist product is less than one. Hence, in the monopoly market, the monopolist faces a downward sloping demand curve.

Price maker: Since there is only one firm selling the product, it becomes the price maker for the whole industry. The consumers have to accept the price set by the firm as there are no other sellers or close substitute of the good.

A monopoly is in a position to earn supernormal profits in the long-run because there is lack of competition in the market. In addition, demand for the firm's product is the market demand curve since it is the only firm in the market and the producer is a price maker.

However, in the monopoly market, consumers' choice is limited. The consumers lack freedom of choice because monopoly produces a commodity that has no close substitute.

Factors leading to emergence of a monopoly market

Several factors lead to the existence of a monopoly market. Some of the most important factors are the following:

Patent or copyright laws: One of the important sources of monopoly is that a firm may legally possess a patent or a copyright, which prevents others from producing the same product or use a particular production process. Usually, patents and copyrights are granted for a certain period.

Control over essential raw materials: Another source of monopoly is control by a particular firm over essential raw materials or inputs used in the production of a particular commodity. For example, the Organisation of Petroleum Exporting

Countries (OPEC) commands world monopoly power over the production and supply of petroleum products.

Grant of exclusive legal right: Another important source of monopoly is the grant of franchise (legal rights) by the government to a firm. A firm may be granted the exclusive legal right to produce a given product or service in a particular region or area. For example, TANESCO has exclusive legal rights to produce and supply grid electricity in Tanzania.

Natural Monopoly (Economies of scale): Significant economies of scale over a wide range of output can endow monopoly on an economic entity. When significant economies of scale are present, the LAC of producing a product continues to decline over a wide range of output and reaches a minimum at an output that is large enough for a single firm to meet the entire market demand at price that is profitable. Such situation may stifle competition from other firms. For example, production of iron and steel requires high investment in infrastructure. If a firm is to establish an iron and steel manufacturing plant for the entire market in Tanzania, it would be difficult for another firm to enter.

The entry barriers of new competitors: New competitors may be blocked or the rival one may be eliminated by aggressive cutthroat competition and aggressive innovations by the monopolists. One of the familiar technique is aggressive price-cutting designed to pre-empt competitors.

Government owned and regulated monopolies: There may exist firms, such as postal services, water and sewage-systems, which are solely owned by the government.

Nature of monopoly's demand and revenue curves

In a monopoly, the demand curve slopes downward from left to right, as shown in Figure 5.9. Hence, the monopoly's demand curve differs from that of the perfectly competitive firm whereby the perfectly competitive firm's demand curve is horizontal (perfectly elastic). Since the monopolist is the only seller in the market, the monopoly's demand curve is the market demand curve of the product. The monopolist faces a downward sloping demand curve.

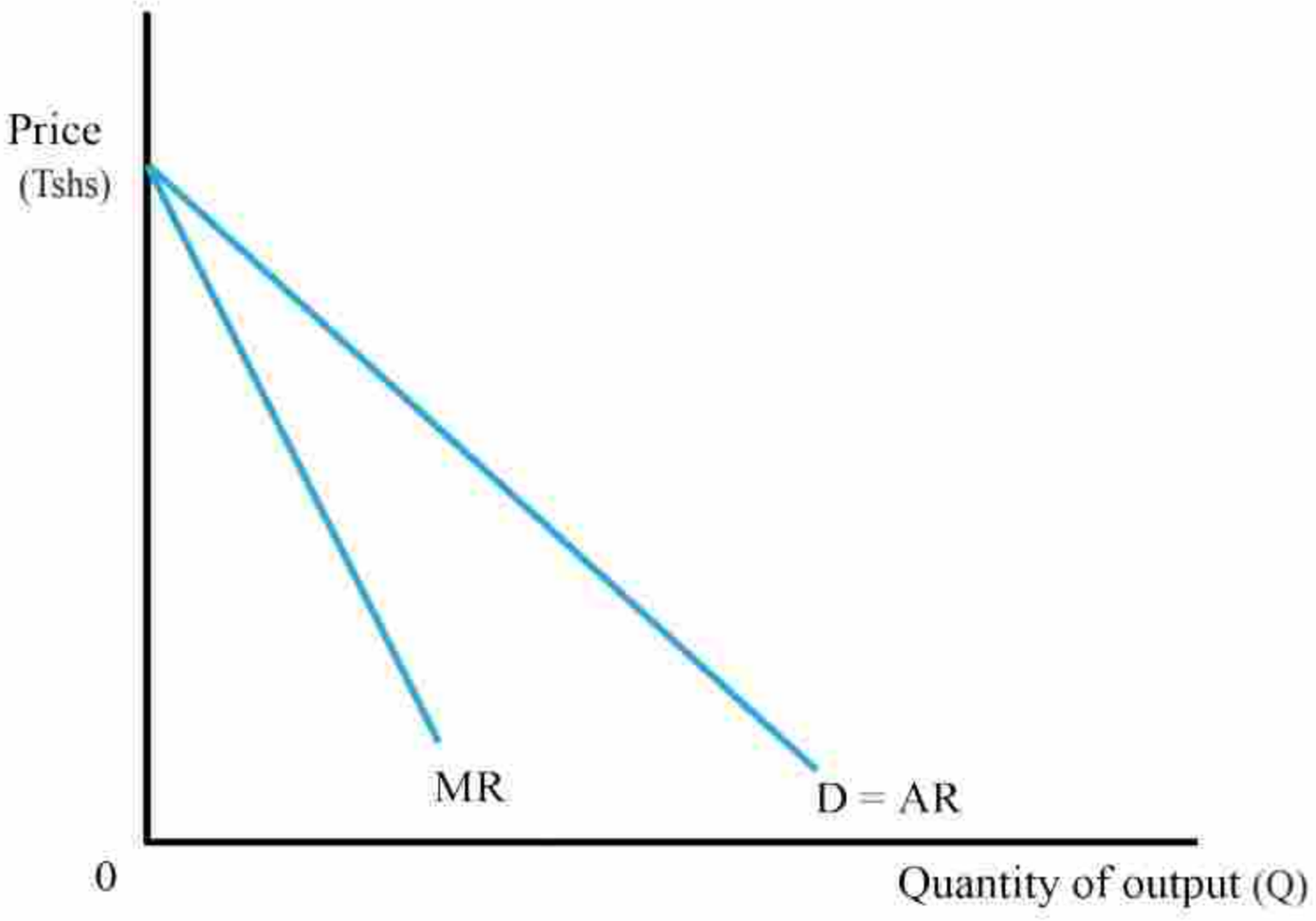


Figure 5.9: *Monopoly demand curve*

Figure 5.9 also shows the relationship between AR and MR of a firm. The AR curve shows how the quantity sold affects the price of the good, and the MR curve shows how the firm's revenue changes when the quantity increases by one unit. Thus, both curves slope downwards from left to right. However, the MR curve lies below the AR curve.

Furthermore, the slope of MR is twice the slope of AR such that:

$$\text{if } P = a - mQ,$$

$$\text{and, } TR = P \times Q$$

$$\text{then, } TR = (a - mQ) \times Q$$

$$TR = aQ - mQ^2$$

But recall that the $AR = \frac{TR}{Q}$ thus:

$$AR = \frac{aQ - mQ^2}{Q} = a - mQ$$

Then, $MR = \frac{\Delta TR}{\Delta Q}$, thus:

$$MR = \frac{\Delta\left(aQ - mQ^2\right)}{\Delta Q} = a - 2mQ.$$

The slope of AR = – m and the slope of MR = –2m, thus the slope of the MR is twice the slope of AR.

Price and output determination under monopoly market

Condition for profit maximisation under a monopoly market in the short-run:

(a) A firm must be at equilibrium such that MC = MR;

(b) Marginal cost curve must cut the marginal revenue curve from below; and

(c) Price (P) must be greater than ATC.

In the short-run, a monopoly only makes supernormal profit. Figure 5.10 illustrates how a monopoly maximises profit by choosing the quantity of output at which the MR equals the MC. The firm is in equilibrium at point A where MC = MR and the produced quantity is equal to Q. At the equilibrium output (Q), the AR is at B and ATC is at E whereby the AR is greater than the ATC. This means that the shaded area PBEC represent super normal profit earned by a monopolist. The distance from C to E shows the quantity (in units) sold by the firm, which is equal to 0Q.

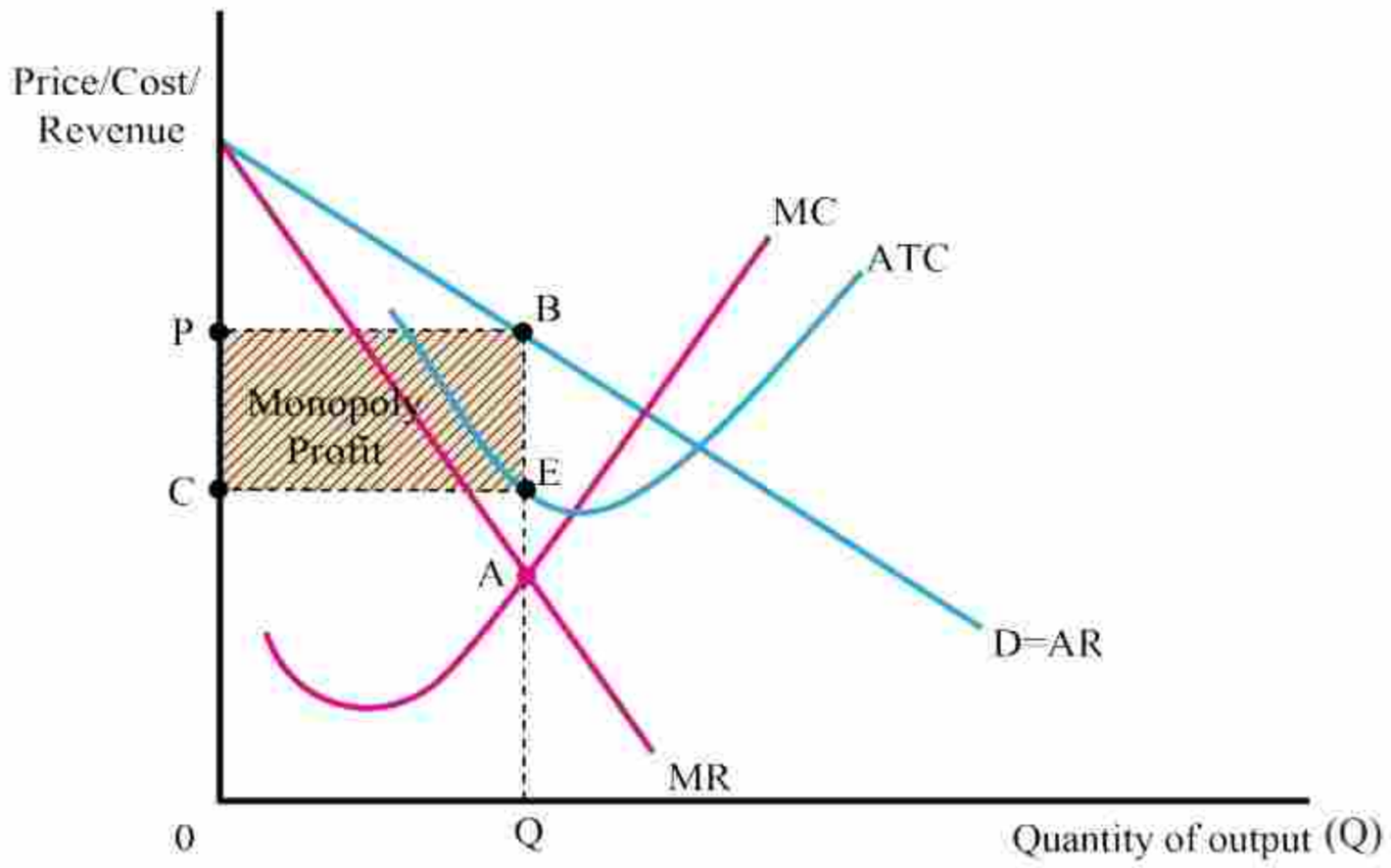


Figure 5.10: *Short-run profit maximisation under monopoly market*

Long-run profit maximisation under monopoly market

In the long-run, the firm is in equilibrium where LMC = LMR. This equilibrium is shown in Figure 5.11 at point E where the monopoly firm produces quantity Q. At the equilibrium, the LAR is at D and LATC is at E such that D > E, which means that the monopoly firm earns super normal profit in the long-run. This profit is presented in Figure 5.11 by the area PCED.

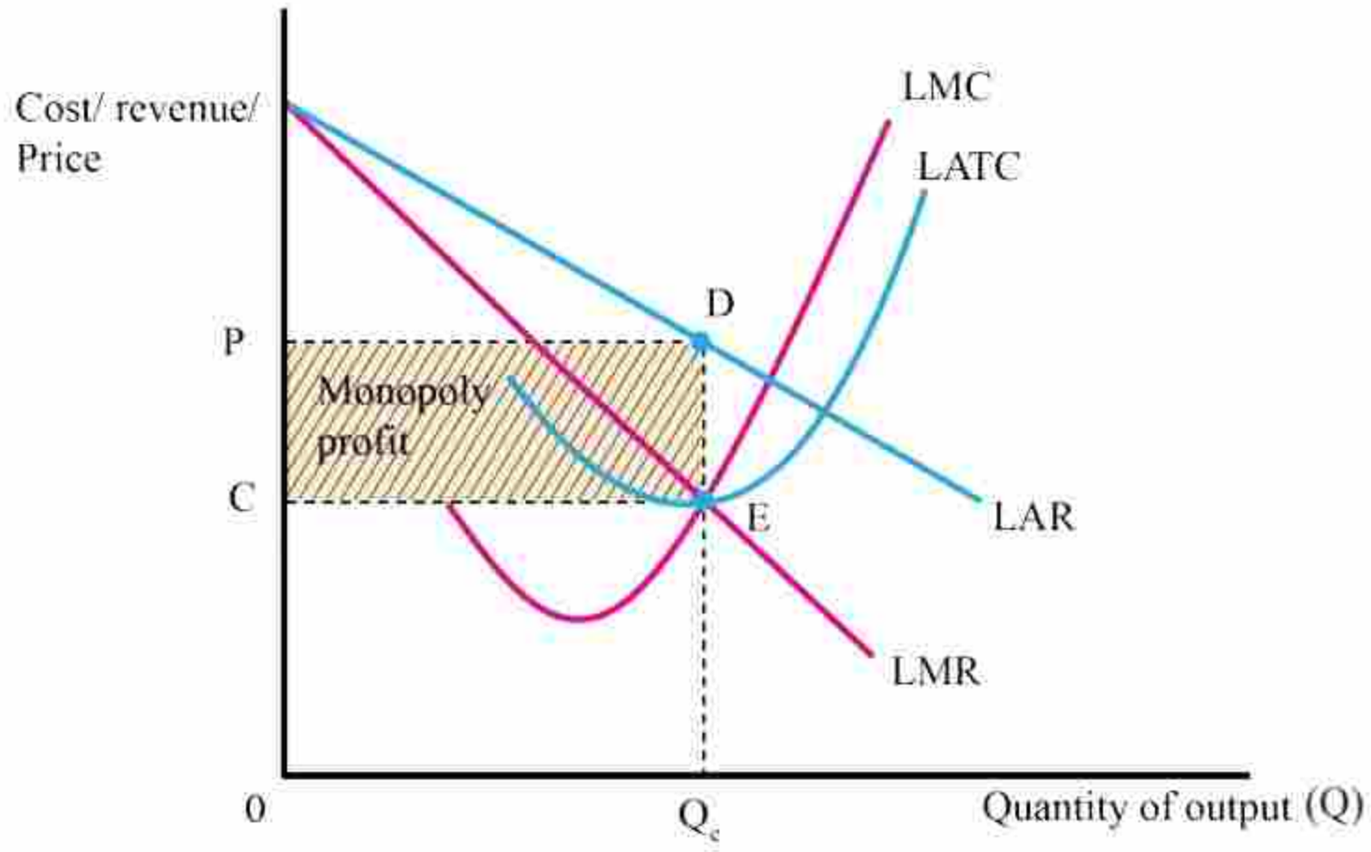


Figure 5.11: *long-run profit maximisation under monopoly*

A monopoly firm earns super normal profit over the short-run and long-run period because it is a sole producer in the market and hence, it has market power. However, the monopoly firm can temporarily make loss because it cannot influence market demand.

Loss under monopoly market

A monopoly firm can also make loss in the short-run. In Figure 5.12, the firm is in equilibrium at point E where MR = MC and produces quantity Q_e, which is sold at price P. At the equilibrium point, the AR is at A and ATC is at B such that B > A, which means that the monopoly firm incurs loss, which is equal to the shaded area CPAB.

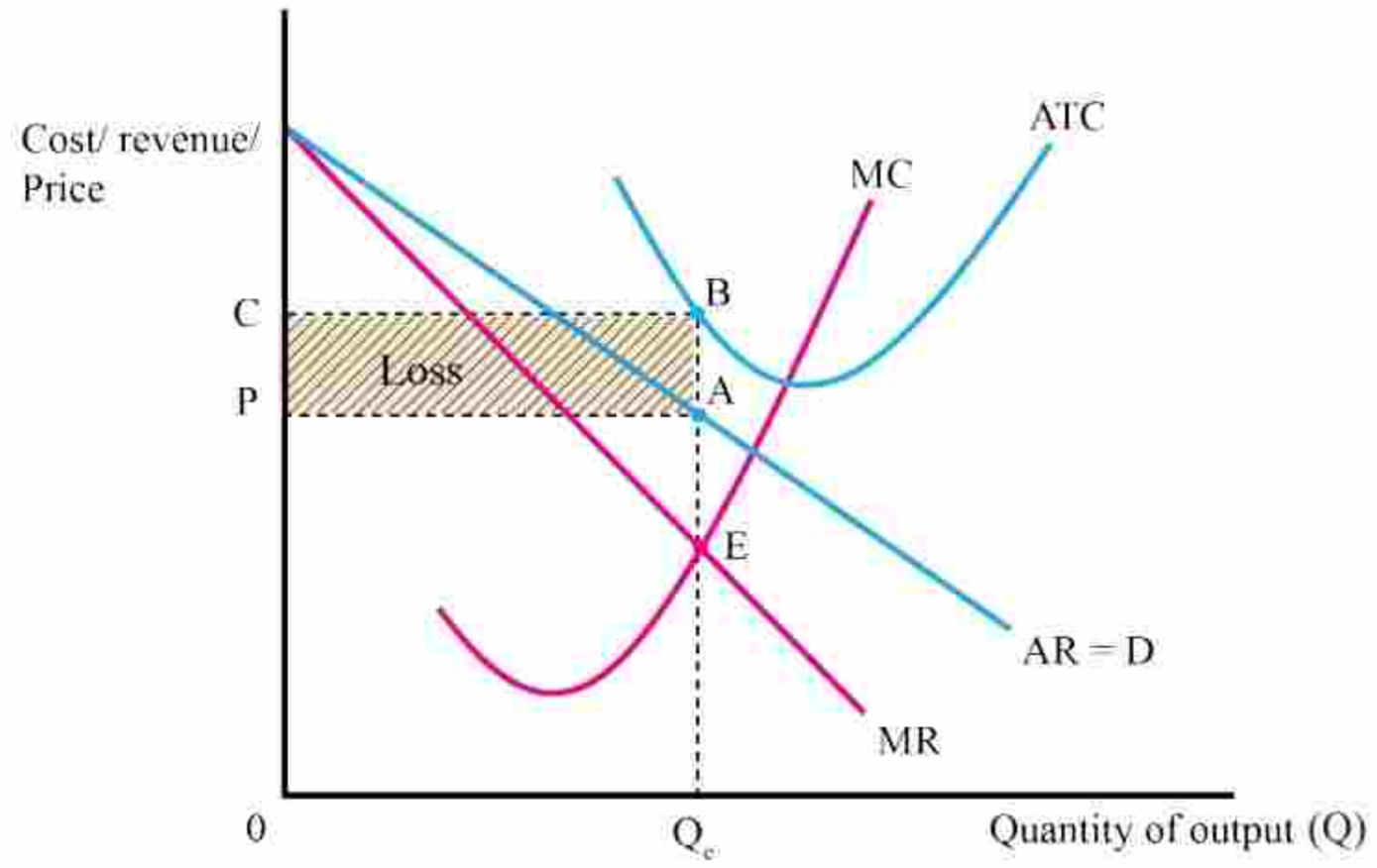


Figure 5.12: *Loss under monopoly market*

Price discrimination under monopoly market

A monopolist may decide to charge different prices to different buyers for the same good. Price discrimination is a rational decision for a monopolist, whose objective is to maximise profit. Since, a monopolist is able to increase profit by charging different prices to different customers, as it charges a price closer to a customer's willingness to pay. For example, TANESCO charges different prices of electricity it supplies to its wide range of customers.

Degrees of price discrimination

Price discrimination can be classified into three types, namely, first degree price discrimination, second degree price discrimination and third degree price discrimination.

First degree price discrimination: It is sometimes referred to as perfect price discrimination because sellers charge each buyer the maximum price he or she is willing to pay with the objective of maximising profits. This type of price discrimination occurs when a monopolist is able to charge each separate unit of a commodity sold at a different price. Thus, sellers charge each buyer the highest price a buyer is willing to pay for a given quantity of goods.

Second degree price discrimination: This is price discrimination, which occurs when a monopolist is able to charge different prices for different quantities of goods purchased by buyers. Under the second degree price discrimination, the monopolist may charge lower price for larger quantities and higher price for smaller quantities sold to buyers.

Third degree price discrimination: This occurs when a seller divides the market into two or more sub-markets and charges a different price in each sub-market, depending on the demand condition of the buyers. A good example is bus fair in public transport whereby they charge low bus fair to children and high bus fair to adults. But also, in concerts or football matches, they charge high price for very important person (VIP) tickets and low price for normal tickets.

Conditions necessary for price discrimination

The following are conditions necessary for price discrimination to be successful:

A firm must operate in imperfect competition, particularly monopoly: The firm must be a price maker with a downward sloping demand curve.

A firm must be able to separate markets and prevent resale: For example, if a firm has segmented cinema entry tickets into adults and children, then it must be able to stop an adult using a child's ticket. It should not be possible to transfer any unit of the commodity from a cheaper market to an expensive market.

Different consumer groups must have different elasticity of demand: In order to segment the market, a seller has to know categories of consumers, whose demand is price elastic and consumers, whose demand is price inelastic. Thus, consumers, whose demand is price elastic would be charged a lower price, whereas consumers, whose demand is price inelastic would be charged a higher price.

Consumers should be unaware of price charged in other markets: Buyers in different markets should not be aware of the price charged in other markets as they will be motivated to buy from the market with the lowest price.

To maximise profits, a firm sets output and price where MR = MC. If there are two submarkets with different elasticities of demand, the firm will realise profits by setting different prices depending on the slope of the demand curve. Therefore, for a segment of the population, whose price elasticity of demand (e_p^d) is inelastic, the price charged would be higher. On the other hand, for a segment of the population, whose demand is elastic, the price charged would be lower. These segments are illustrated in Figure 5.13 whereby profit is maximised at a point where MR = MC. Because demand is inelastic in market A, it leads to a higher price being set. In market B, demand is price elastic and hence, the profit-maximising price is lower.

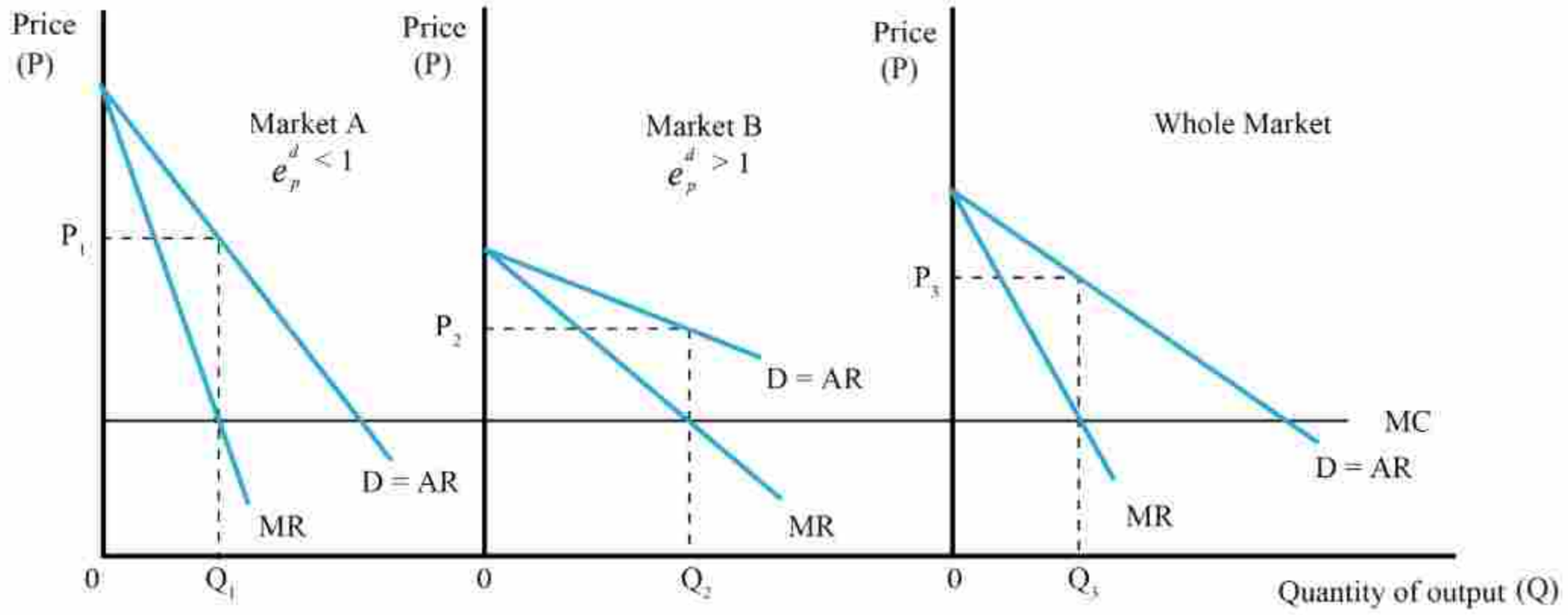


Figure 5.13: *Price discrimination in a monopoly market*

Through price discrimination, firms are able to increase revenue, which helps them to stay in business, otherwise, they would incur losses. Price discrimination results in some consumers paying higher prices. Such higher prices may cause allocative inefficiency if the charged price is greater than the marginal cost. In addition, price discrimination may lead to consumption of all quantities in the market resulting in a decrease in the consumer surplus. With no consumer surplus, the price of the good would tend to rise. Furthermore, there may be administration costs incurred in separating the markets, which would be incorporated into the prices of goods and services. Lastly, price discrimination may result in a situation where poor people end up paying higher prices, while the rich pay lower prices.

Advantages of a monopoly market

A monopoly market enhances profit generation for projects with high initial costs. If a good is consumed by a large proportion of the population and the initial cost of its production is very high, price discrimination will enable an increase in revenue. Such projects include electricity generation, water supply, airways and railway transport.

Moreover, in the monopoly market, there is no wasteful competition. This is because as far as there is only one seller or producer, there will be no duplication of resources, and the monopolist can allocate resources more efficiently to earn maximum profits. Similarly, it encourages innovation in the long-run because the earning of supernormal profit is an incentive for innovations and invention as well as a source of resources for funding research for such ventures.

Lastly, monopoly gains economies of scale, which may lower prices thereby benefit consumers. That is to say, through the economies of scale, the monopolist may be able to lower average total cost (ATC) thereby lowering the price of the good.

Disadvantages of a monopoly market

The monopoly market has the following disadvantages:

Exploitation of consumers: Monopolists could reduce consumers' utility by setting high price. A monopolist determines the price of the good it sells in two ways: One, by choosing the quantity of output to produce. Thus, it can choose a very small quantity, which will lead to inflating the price of the good and creating artificial scarcity of the good. The other is by choosing the price, which if it is very high will result in producing a small quantity. However, the guiding objective will be to maximise its profit, which will be determined at a point where MR = MC.

Inefficiency in production: Inefficiency in production arises because there is lack of competition in the market. The price it charges for the good produced by the monopolist is higher than that under perfect competition, which is higher than total average costs. This implies the monopolist could lower price of a good and increase the quantity produced. Because there is no close substitute or competition for the good, the monopolist may reduce the quality of its product. In addition, it may increase the cost, for example, through unnecessary advertisement.

Consumers lack freedom of choice: A monopolist produces a commodity, which has no close substitutes and hence, limits consumers' choice. If consumers' choice is limited, then the welfare of the society tends to be compromised.

Monopoly may lead to unemployment: A policy to restrict output leads to decrease in demand for inputs. The decrease in demand for the inputs leads to unemployment of labour and other factors of production.

Control and regulation of monopoly

There are several ways of controlling monopoly such as the following:

Adoption of price control: Government authorities may regulate monopoly through price control measures. For instance, the government can put a price

ceiling (maximum price) in order to prevent a monopolist from inflating the price uncontrollably. For example, in Tanzania, the prices of electricity and water are regulated by EWURA.

Adoption of anti-monopoly laws: Through this method, the government can pass laws that prevent establishment of monopolistic associations. The law may also prevent monopoly practices in the country.

Taxation of monopolist profit: The taxes work to reduce the profit margins of the monopolists.

Restricting a monopoly: Another method, which can be used to regulate monopoly is to restrict expansion of large firms. The government's antitrust policy should be such that big firms are restricted to expand further. This may be achieved by restricting their access to scarce materials.

Allow entry of new firms: The government could reduce barriers to entry. Other new firms could be allowed to enter into the monopolistic industry. The government could ensure that barriers to entry by outside firms into the monopolised industry are maintained at minimum.

Formation of monopoly commission: The government could form a monopoly commission to report on evils resulting from working of the monopolies. The report could aim at preventing evils from happening rather than punishing the monopolies after they have done such evils.



Activity 5.1

1. Visit the school library, use the website and other journals on economics to prepare a table showing similarities and differences between perfectly competitive markets (PCM) and monopoly markets.
2. Choose two monopoly markets operating within the East African countries, known to practice price discrimination. Explain why the two monopolists have chosen to discriminate price.

Monopolistic competition market

Monopolistic competition is a market structure with a large number of firms selling products that are similar but not identical. In monopolistic market, every firm has

a monopoly power over the product it produces, though the products are similar. Examples of monopolistic competition can be found in publishing, clothing, and shoes industries, among others. In the clothing and footwear industries, for example, the designer's name distinguishes products.

Characteristics of a monopolistic competitive market

The following are some features of a monopolistic competitive market:

Many sellers: In this kind of a market, there are many sellers competing for the same group of consumers.

Products are slightly differentiated: Products are differentiated in terms of brands, packages, colour and other minor differences but they all serve the same purpose.

Relatively free entry and exit: There are no barriers to entry into and exit from the market.

Firms are price makers: Firms have some monopoly power over their respective products. Each firm may have a tiny 'monopoly' because of differentiation of their products thereby forming sub markets.

There is stiff competition: There is competition among firms or producers since their products are similar.

Advertisement is practiced: Unlike a monopoly market and PCM, advertisement is practiced in monopolistic competition in order to increase the share of the market because products are slightly differentiated and there is competition.

The demand curve for firms under monopolistic competition is elastic, unlike pure monopoly: The demand curve of the firms in monopolistic competition firms is elastic because of substitutability of commodities sold in the market.

Price and output determination under monopolistic competition market

A monopolistic competitive market does not face a perfectly elastic demand due to product differentiation. Each differentiated product forms a sub-market and hence, has a downward sloping curve in accordance with the law of demand. The technical condition of production of firms in monopolistic competition firms allows it to become a price maker with a downward sloping demand curve as presented in Figure 5.14.

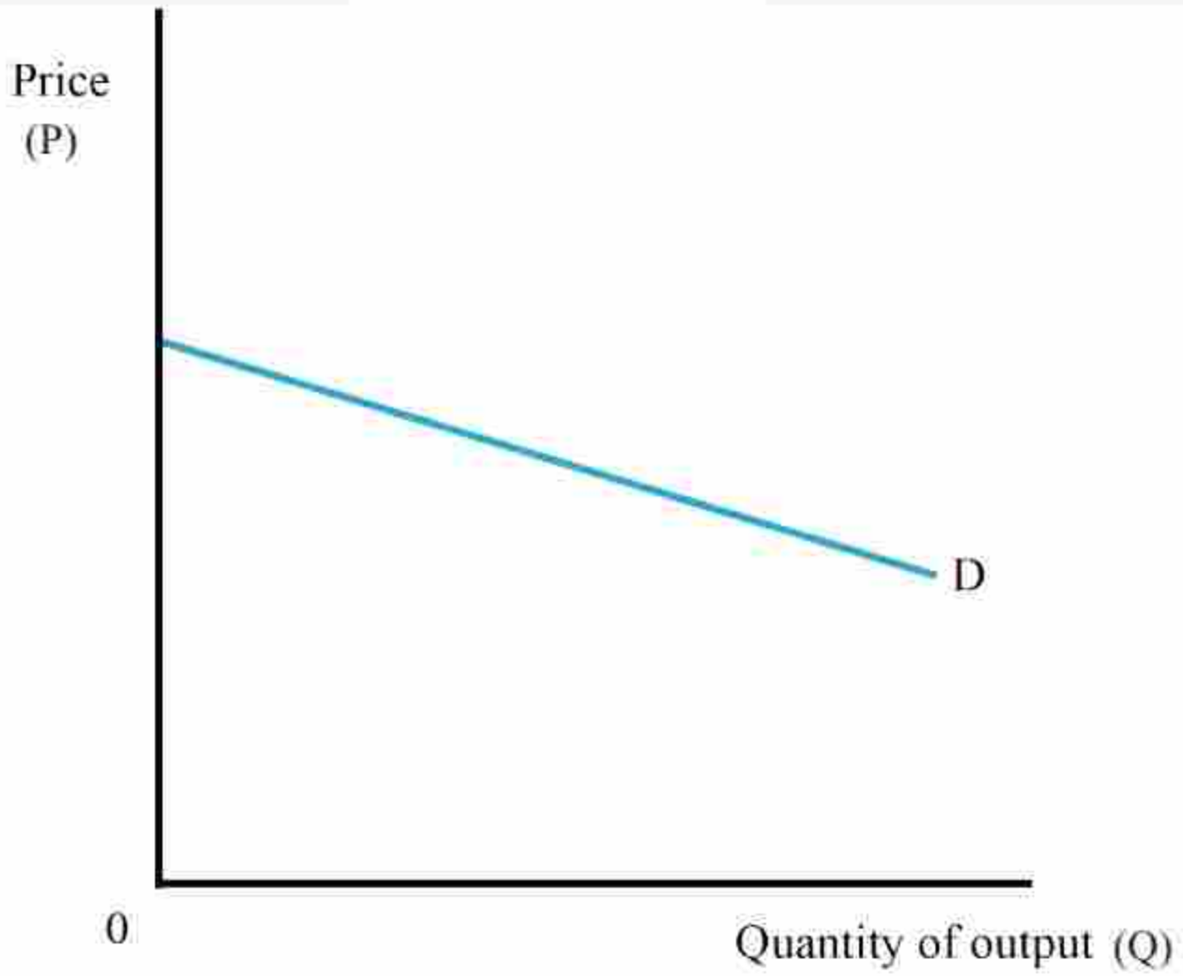


Figure 5.14: *Monopolistic competition market demand curve*

Under monopolistic competition, a good is produced in large quantity and has higher quality than in monopoly due to competition. Consumers have wide choice because the produced products differ under monopolistic competition. Products produced by monopolistic competitive markets may differ in size, brands, taste, weight and colour. In a monopolistic competitive market, there is no price discrimination and hence, the industry treats all consumers indiscriminately. The resources used in production are mobile across firms. For instance, if labour is paid lowly in one firm, it is free to move to another firm that offers a high wage.

However, under the monopolistic competition, there may exist inefficient use of resources due to competition, which lead to production of multiple products. There may be no need to produce multiple products, since few would suffice. Hence, production of multiple products results in wastage of resources. Firms may spend a large part of their resources in advertisement and sales promotion due to competition that exists in the market. In the long-run, there is a possibility of loss making by small firms due to increased competition for profit, given unrestricted entry into the market.

Profit maximisation under monopolistic competition

In the short-run, firms under the monopolistic competition can make supernormal profit, normal profit or loss.

Condition for supernormal profit under monopolistic competition

A firm maximises profit when:

(a) MC = MR;

(b) Marginal cost curve must cut marginal revenue curve from below; and

(c) Price must be greater than average cost P > ATC.

In Figure 5.15, the firm is in equilibrium when producing at point E where MC=MR, at output Q_e and price P_1. At A, P > ATC, which implies that the firm is making supernormal profit presented by the shaded area CBAP.

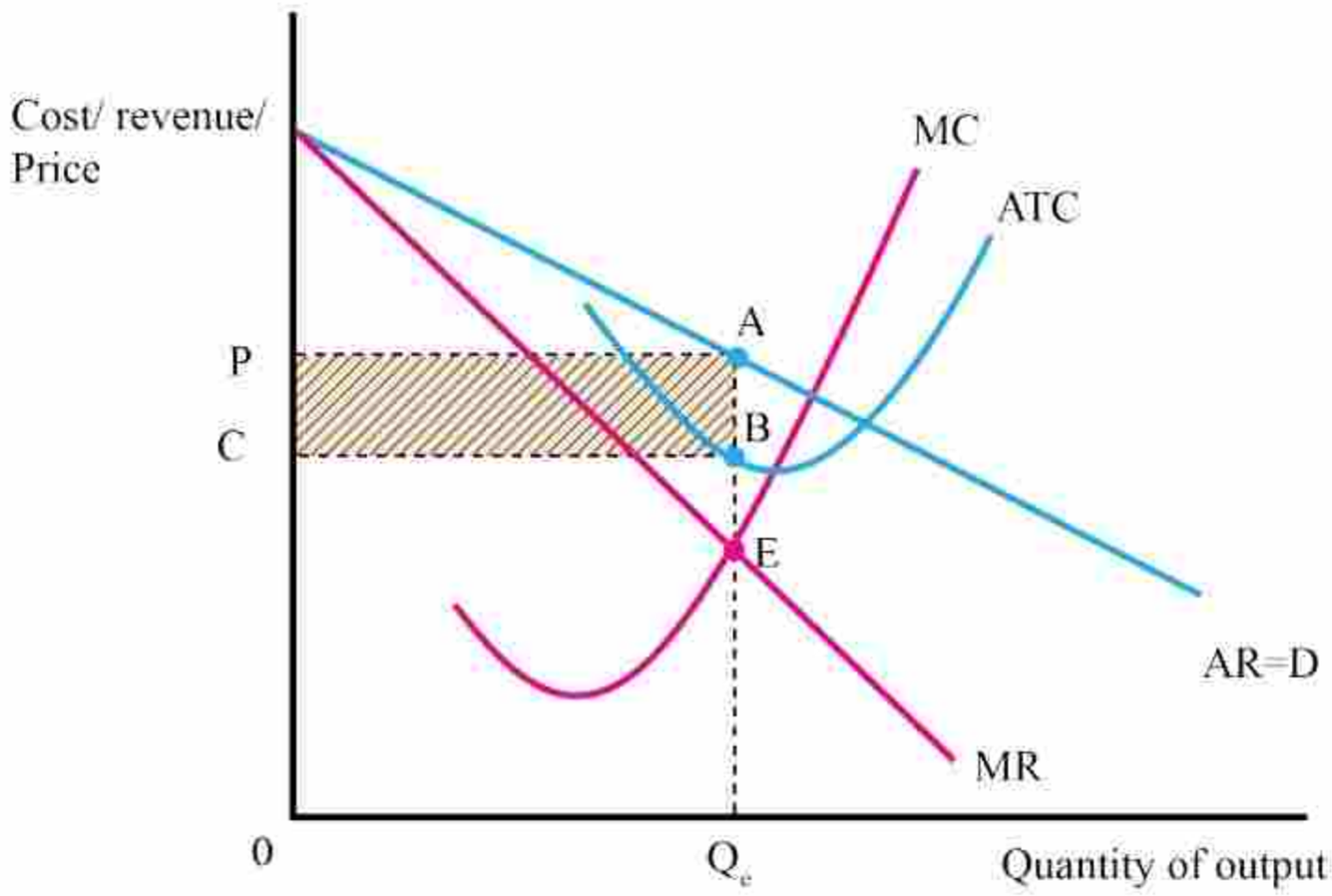


Figure 5.15: *Supernormal profit under monopolistic competition*

Normal profit under monopolistic competition

The firm under monopolistic competition will earn normal profit under the following conditions:

(a) MC = MR;

(b) MC curve cuts MR curve from below; and

(c) P = ATC.

Figure 5.16, shows normal profit for a firm under monopolistic competition.

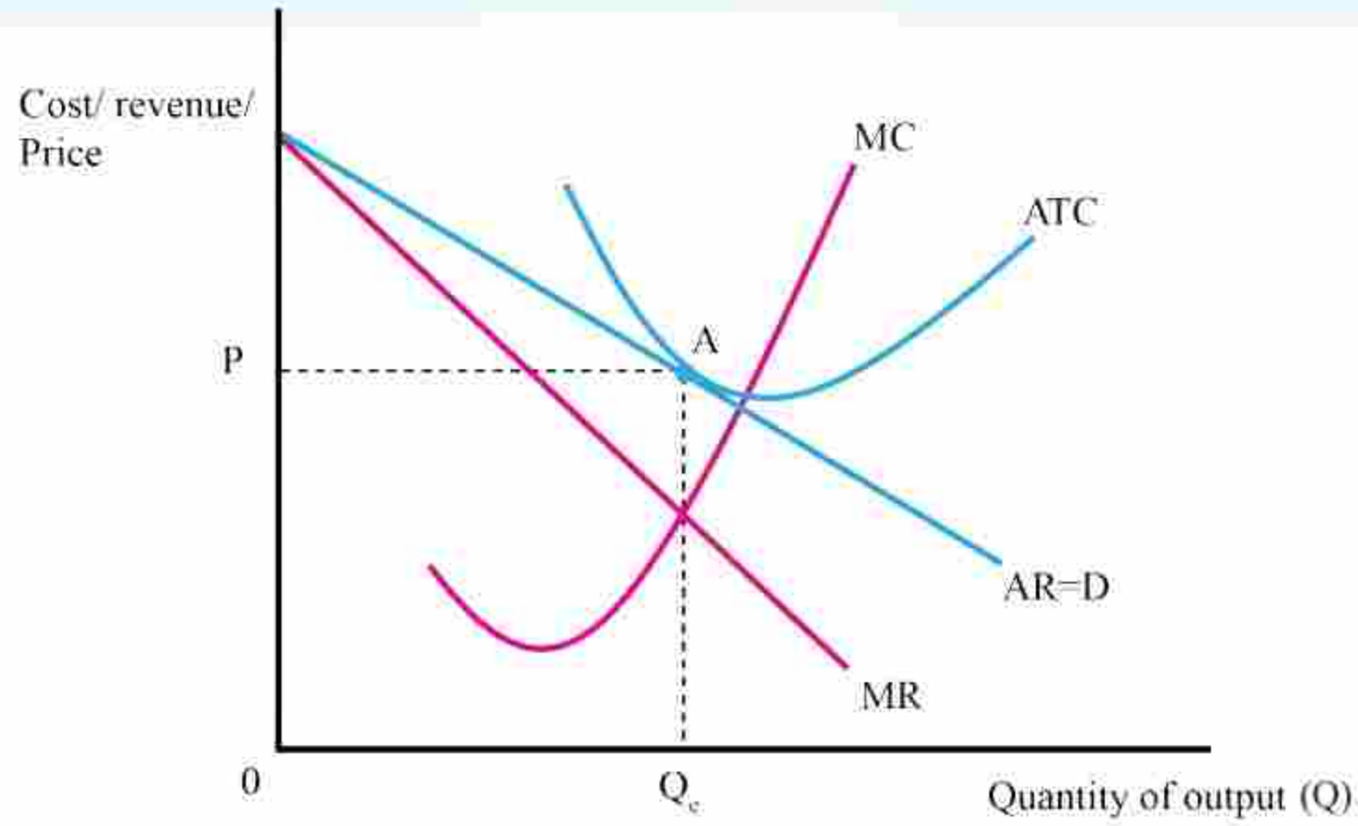


Figure 5.16: *Normal profit under monopolistic competition market*

In the short-run, a firm in monopolistic competition may sometimes earn normal profit at a point where AR = ATC and equilibrium output Q_e.

Loss in monopolistic competition

A firm under monopolistic competition incurs loss under the following conditions:

(a) MC = MR;
(b) MC curve cuts MR curve from below; and
(c) P < ATC.

In the short-run, firms in monopolistic competition may sometimes incur loss when AR is less than ATC. In Figure 5.17, the firm is in equilibrium when producing the quantity of output (Q_e) at point E, where MR = MC. However, ATC is at B, whereas AR is at A, which means that ATC > AR. As a result, the firm is making loss, which is depicted by the shaded area PABC.

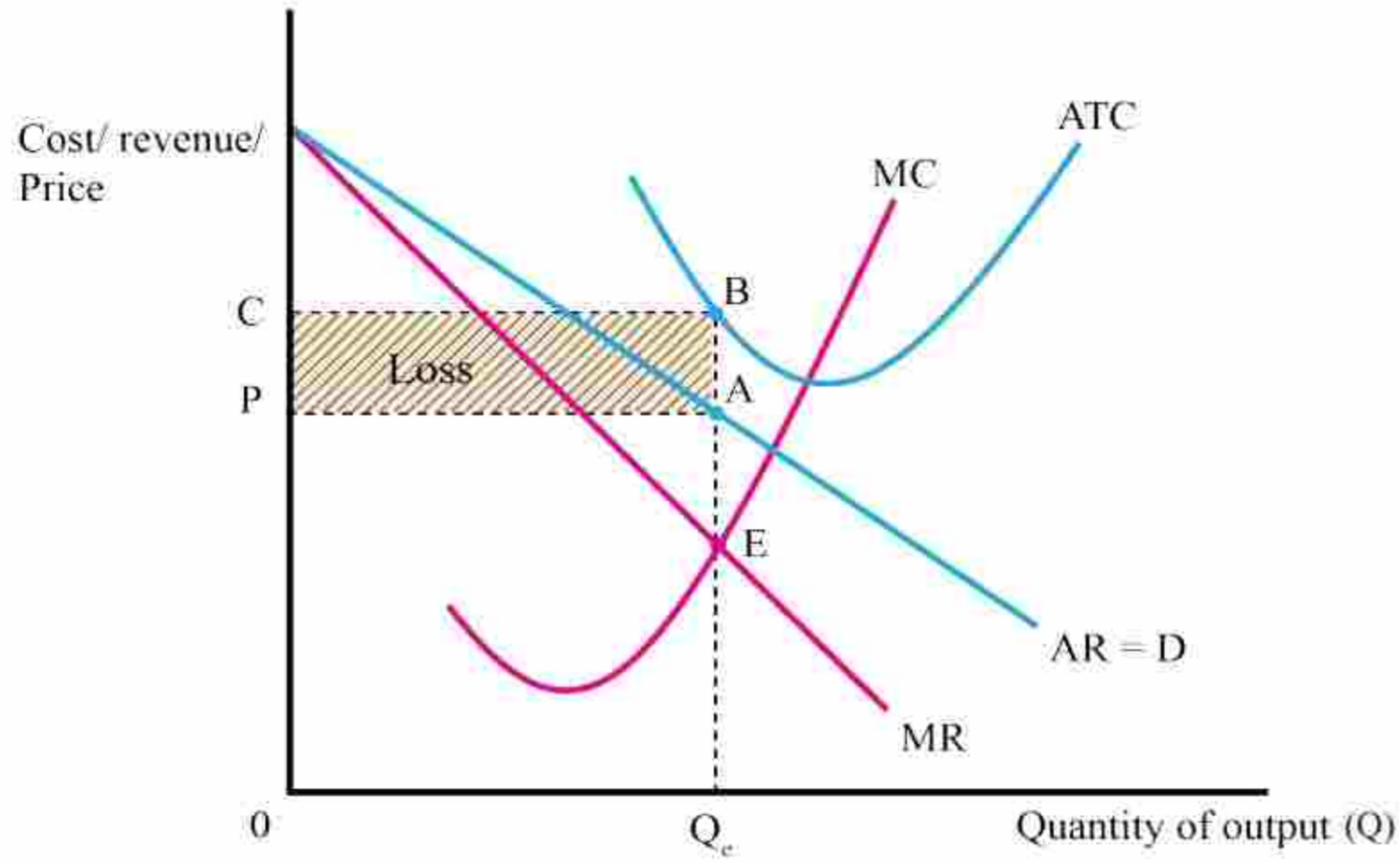


Figure 5.17: *Loss under monopolistic competition market*

Long-run profit maximisation under monopolistic competition

In the long-run, monopolistic competitive firms make normal profit because there is no restriction to entry into the industry for a firm seeking to take advantage of supernormal profit. In addition, in the short-run, a monopolistic firm can survive even if it incurs loss as long as it can cover its variable costs. In the long-run, a loss-making firm has to stop production because all factors will be now variable.

Condition for long-run profit maximisation

The following conditions must hold for a monopolistic competition market to maximise profit in the long-run:

(a) LMC = LMR;
(b) P = LATC; and
(c) LMC curve cuts LMR curve from below.

In the long-run, firms in monopolistic competition are in equilibrium at Q_e where the long-run marginal cost (LMC) equals the long-run marginal revenue (LMR), as shown in Figure 5.18. At A, LAR equal the LATC and the firm is making normal profit.

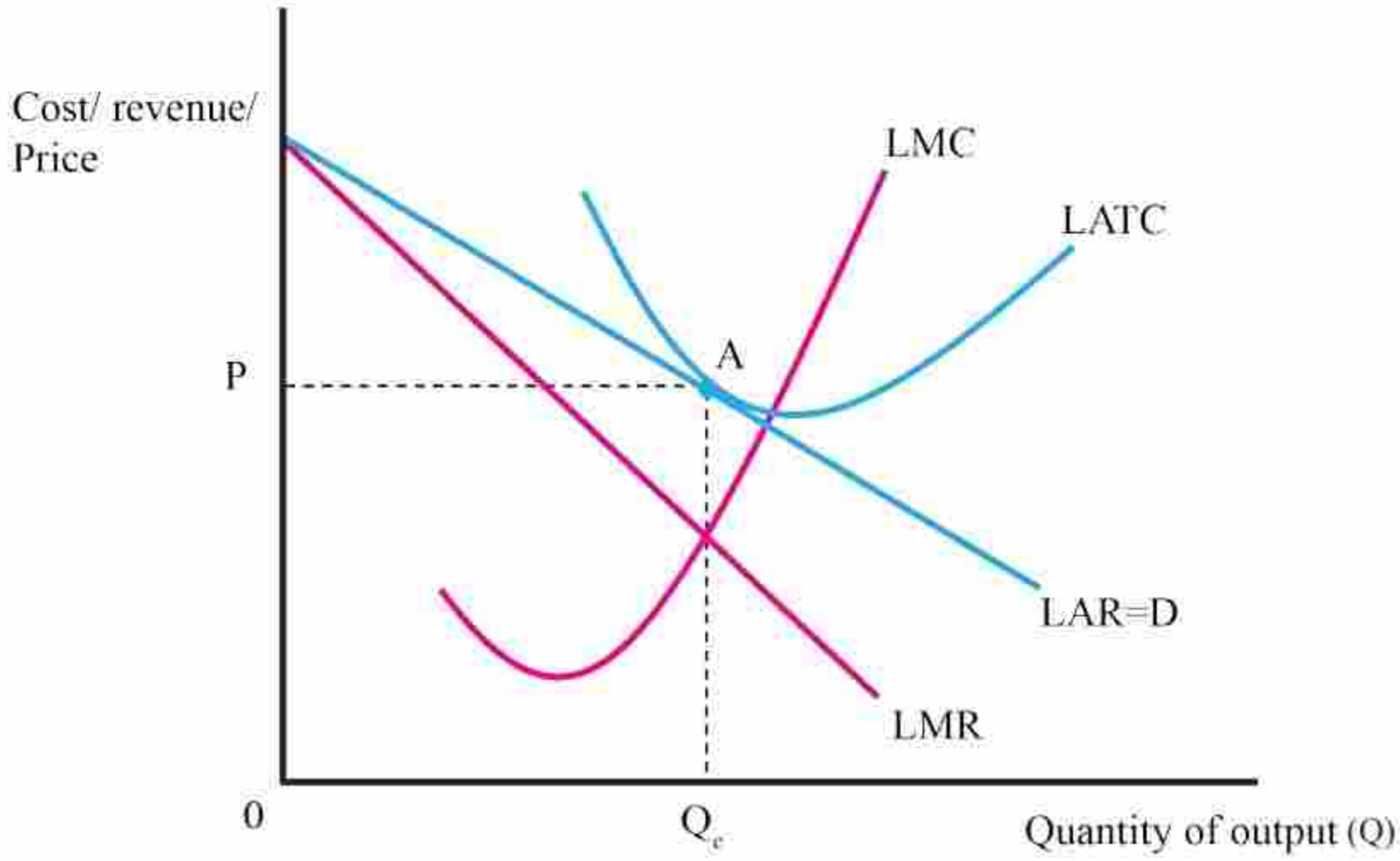


Figure 5.18: *Long-run normal profit under monopolistic competition*

Exercise 5.2

1. Explain advantages and disadvantages of monopolistic competition.
2. Distinguish between a monopoly and monopolistic competitive market. Present not less than five (5) points and support them with relevant examples.
3. What distinguishes monopolistic competition from perfectly competitive market structures?
4. Describe similarities between monopolistic competition and perfect competition market structures.

Oligopoly market

Oligopoly market is a market structure dominated by few large firms. In an oligopolistic market, firms produce and sell homogenous and/or differentiated products. Examples of domestic oligopoly markets are Cement industry (Twiga Cement, Simba Cement, Tanga Cement, Dangote Cement, Kilimanjaro Cement Company Limited); and the telecommunications industry (TTCL, Vodacom, Airtel, Halotel, Zantel, and Tigo). Internationally, there are automobiles industry (TOYOTA, BMW, NISSAN, SUBARU, MERCEDES, FORD and MITSUBISHI). One of the difference between an oligopoly and perfect competition is that because there is only a small number of firms in the industry, the action of any one of the firms will have a large effect on other firms. Therefore, firms in an oligopolistic market must act strategically. That is to say, before a firm takes any

action, it has to consider the possible reactions of its rivals. An oligopoly market could be of two types, namely:

(a) A perfect or pure oligopoly market, which involves few sellers producing homogeneous or identical products such as cement, cable wires or dry batteries manufacturers and telecommunications companies. In this type, the pricing policy of one firm affects other firms that produce the same good.

(b) Imperfect or differentiated oligopoly market involves few sellers who produce differentiated products, for example, automobile companies.

Characteristics of an oligopoly market

The following are characteristics of oligopoly market:

Industry dominated by small number of large firms: There are few firms in the industry, each with a substantial market share.

High barriers to entry: In an oligopoly market, entry of a new firm is difficult, and relatively few sellers dominate the market. Barriers may be in form of ownership of technology and high initial capital to start the firm. For example, the oil industries own extraction and refining technology, and if any country wants to own the technology, it has to pay a huge amount of money, which is difficult.

Products could be differentiated or homogeneous: In an oligopoly, the products produced and sold in the market may be homogeneous or differentiated.

Non–price competition: There is price stability within the market. In oligopoly, market prices are relatively rigid. They do not change frequently unless when there are price wars. When one firm increases price, the other firms will not follow. Hence, the firm that has increased the price will lose some of its customers thereby forcing it to reduce the price back to the original level. There is no need for a firm to reduce price with the objective of increasing sales because all the other firms will follow the same strategy, resulting in lowering profit for all the firms in the industry.

Potential for collusion: Firms can collude instead of competing against each other. In this case, firms can jointly agree to set price and quantities that maximise the sum of their profits. For instance, the Organisation of Petroleum Exporting Countries (OPEC) brings together oil producing countries and agree on the price and quantity of oil to be sold in the world market.

Supernormal profits: When oligopoly firms collude, they earn supernormal profits because it is easier for them to form cartels and decide on price to charge and the quantity to produce. These firms can increase price to earn supernormal profit or they can cut production to make the commodity scarce in the market for the price to rise.

High degree of interdependence between firms: In oligopoly market, the actions or decisions of one firm affects the decision of others. If one firm changes the price or quantity of the good, then other firms will have to react to the changes.

Nature of oligopoly's demand and revenue curves

Given the interdependence of firms in the oligopoly market, the firms do not have a definite demand curve as it may keep shifting as the rival firms change their prices. According to an economist Paul Sweezy in 1939, an ordinary demand curve cannot determine the price and output in an oligopoly market. Thus, he introduced a **kinked demand curve**. The logic behind the kinked demand curve is that, each firm in an oligopolistic market believes that rivals will match its price reduction but ignore its decision to increase the price. This mixed response of competitors will produce a demand curve for the firm that is kinked, as shown in Figure 5.19, at point E.

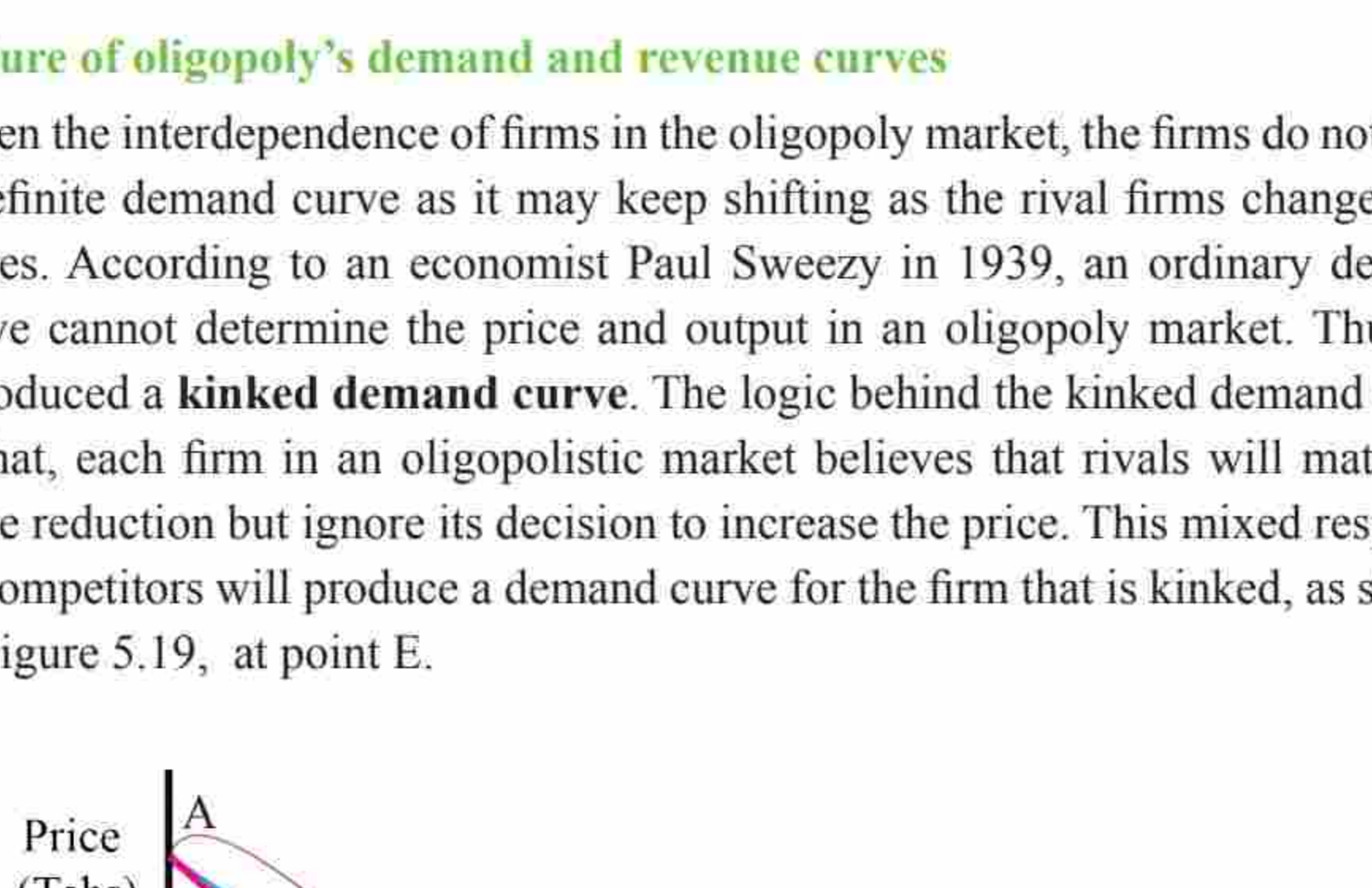


Figure 5.19: *Kinked demand curve under oligopoly market*

Presence of a kinked demand curve causes the MR curve to be discontinuous at the level of output corresponding to the kink. The marginal revenue has two segments. Segment AB corresponds to the elastic part of demand curve while segment CD corresponds to the inelastic part of demand. Small fluctuations in cost occur within the gap BC without affecting the level of price.

The kinked demand curve explains how an oligopolistic firm might have to accept price stability in the market. In Figure 5.19, the equilibrium market price is Tshs 400 and the firm produces 4 units of output.

Now assume that if a firm raises its price above Tshs 400 (let us say to Tshs 600), its competitors will not follow. As a result, many of its customers would switch to its competitors' products. In this case, its revenue and hence, profits would decline. The demand curve would be relatively elastic above the price of Tshs 400. Now suppose that the firm reduces its price below Tshs 400 (let us say to Tshs 200), its competitors will also reduce their prices to prevent it from taking away their customers. As a result, with a price cut, the firm would find its market share unchanged, whereas its revenue would decline. Its demand curve is, therefore, inelastic below price of Tshs 400. Hence, the kink at point E. In this situation, the firm will not want to change its price unless it is forced to do so by huge changes in its costs. Thus, there is price stability in an oligopolistic industry.

Weakness of kinked demand curve

The kinked demand curve helps to explain why oligopoly prices are stable even without collusion among firms. However, it does not explain how the firm sets the existing price (for example, the price of Tshs 400 in the Figure 5.19). It simply explains that once an equilibrium price is established, it is unlikely to change. There is no empirical evidence to support presence of the point where the demand curve is kinked. Despite the rivalrous nature of the firms in oligopoly market, the interdependent oligopolistic firm still make profit. The firms do not compete on prices and instead, they use non-price measures through products development and advertisement. Examples of non-price measures include free delivery and installation, long opening hours (especially in petrol stations), after sale customer services and huge expenditure on advertisement. Thus, we considered to introduce other models of oligopoly that explain the behaviour of oligopolistic firms and one of those models is the collusion model.

Collusion model

The collusion model states that if a small number of firms that make up an oligopoly are able to get together and make price and output decisions jointly, the resulting outcome would be the same as that under a monopoly. When a group of firms get together to make joint price and output decisions, it is called a **cartel.** The best example of a cartel is the Organisation of Petroleum Exporting Countries (OPEC), which constitutes 13 countries that meet to determine oil production levels and prices to charge.

Collusion is an agreement among rival firms not to engage in price competition or other kinds of competitive activities that could reduce their individual profits. They agree to work together and behave like monopoly. As pointed out, a cartel is a collection of colluding oligopoly firms who agree on price and output levels. Price and output decisions of a collusive oligopoly are the same as those of a monopolist. As shown in Figure 5.20, the whole industry equates MC with MR, charges the price P. At the profit maximising output Q_e, the firm earns supernormal profit presented by the shaded area CBAP.

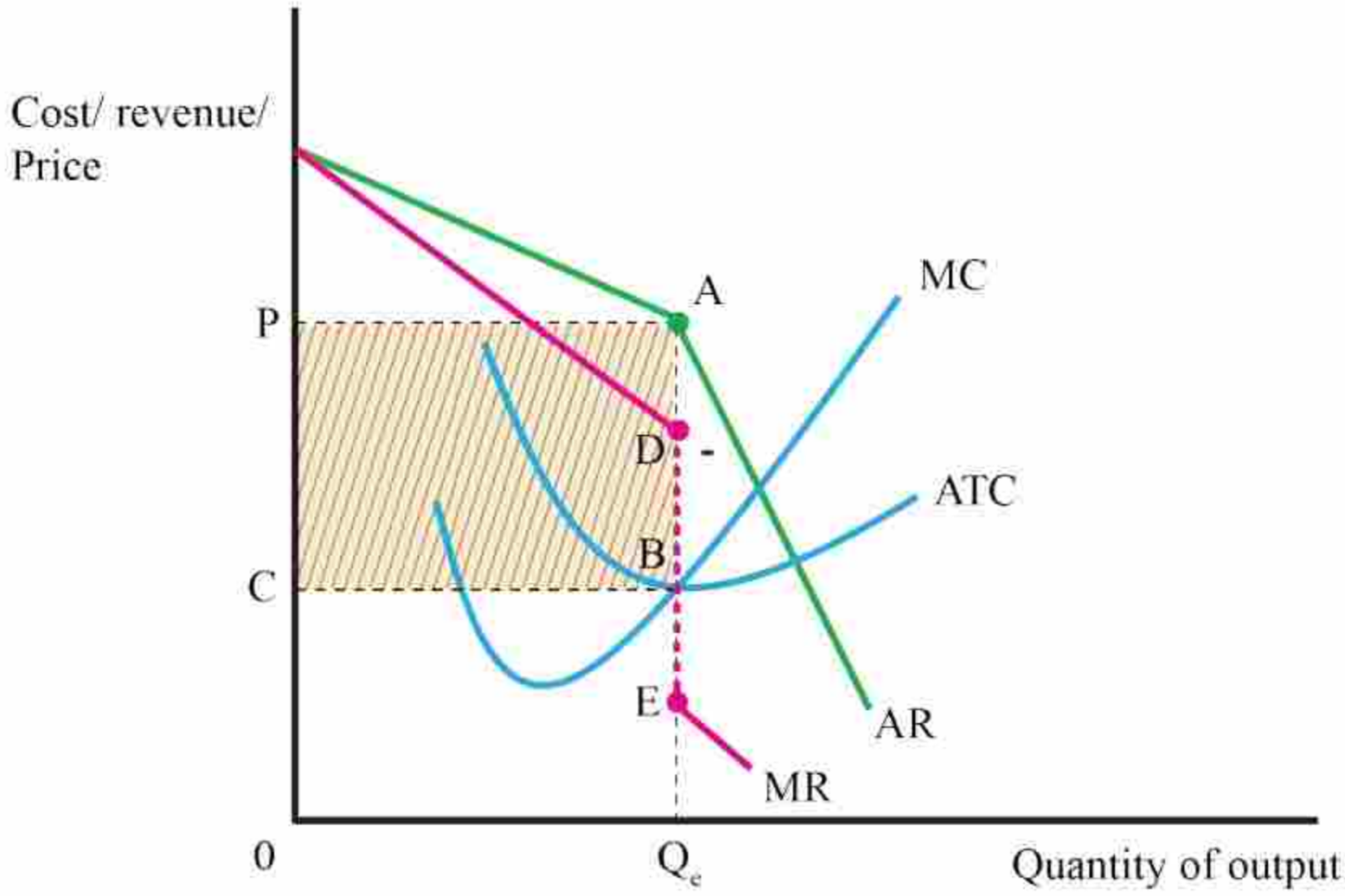


Figure 5.20: *Profit maximisation under oligopoly market*

Supernormal profit is earned when the price of a good is greater than the average total cost of the firm, that is, (P > ATC) or when total revenue is larger than total cost, that is, (TR > TC).

Collusion has certain advantages to an individual oligopoly firm. These advantages are the followwing:

(a) There is relative certainty about the demand curve;

(b) There is no fear from price cutting; and

(c) Supernormal profits are maintained at the highest possible level.

For a cartel to maximise its industry's joint profits, there must be:

(a) Strict agreements on the price of the good and size of the market each firm would supply.

(b) Production quotas for ensuring that the output for the whole industry does not exceed the total output by individual firms requiring that each firm equates MR with its MC. Once a member cheats by secretly increasing output and selling it at the fixed cartel price, other members of the cartel may retaliate thereby leading to price wars that can leave the entire industry in disarray. If most of the firms decide to increase their output in that way, the cartel would collapse because the high price may not be maintained as the excess supply would drive down the price implied by the agreed upon quotas.

This has been the experience among the OPEC oil-producing countries in the recent years. Attempts to agree on a restricted output quota for each country in order to push up oil prices have often broken down because some member countries decide to cheat.

In order for cartels to be successful, several conditions must hold. These conditions include the following:

Demand for the cartel's product must be inelastic: The reason why OPEC can be successful is that there are few substitutes for oil. If it cuts production and raises the price, consumers have little option but to pay high prices. If, for example, cabbage producers set up a cartel, it will not be very successful. As soon as the cabbage cartel raises prices of cabbage, consumers will just switch to other vegetables.

Members of the cartel must not cheat: This is a hard condition to maintain as firms have a strong incentive to cheat. For example, suppose that OPEC cuts production and raises prices. A single member country, call it country X, could decide to increase its production, offer a low price, and make large profits, as everyone would rush to buy oil from Country X. Country X would, therefore,

increase its market share by cheating, at the expense of other countries that stick to the agreement. Other countries might have similar thoughts, and if a great number of the cartel members decide to cheat, the cartel agreement will collapse.

Disadvantages of cartels

Cartel causes inefficiency in production because they raise price above the competitive level and reduce output. Consumers choose to pay high prices for some or all of the cartelised products that they desire and thus, foregoing the product, or they pay the cartel price thereby unknowingly transfer wealth to the cartel. Cartel creates monopolies. Monopolies are inefficient because they supply low output at higher prices.

Protection of inefficient firms. Cartels protect their members from full exposure to market forces, reducing pressure on them to control cost and to innovate. The protection affects efficiency of the market economy.



Activity 5.2

Visit the library or use the website to read and understand the different types of markets.

(a) Create a table with 5 columns and 5 rows. On the first row from the second column horizontally write the market structures. On the second row the first column vertically, write features like number of sellers and buyers, nature of commodity, seller's entry barriers and knowledge of the market situation.

(b) Fill in the blanks the information that you have extracted from your readings.

(c) Present your work to the rest of the class.

Chapter Summary

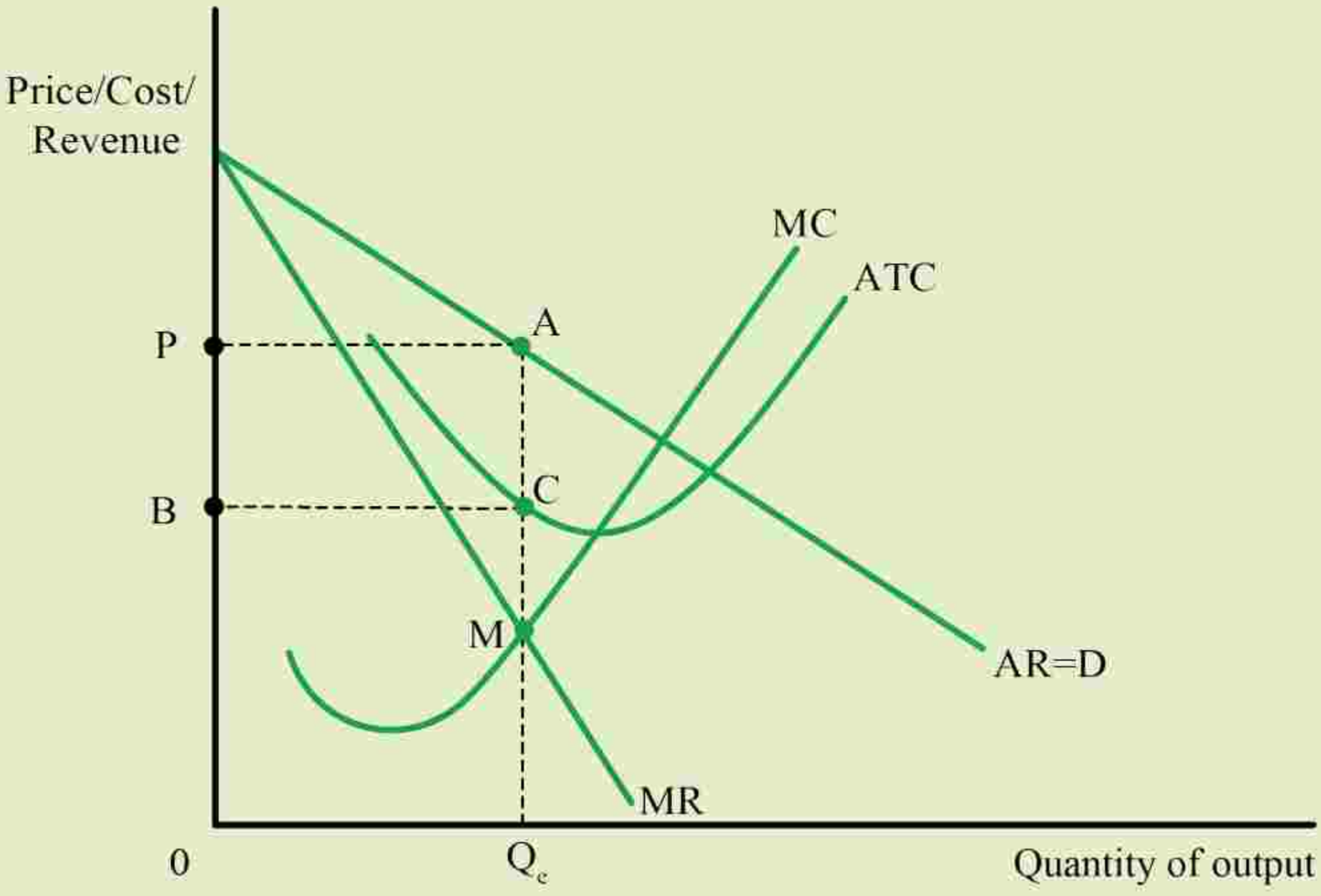
1. Market involves buyers and sellers interacting but it may not necessarily be through physical contact, following emergence of digital markets.
2. Markets can be classified according to what good or service is sold and bought in the market and its structure or situation.
3. Markets can be classified into goods markets, factor markets or financial markets.

4. Markets are structured according to their characteristics such as number of buyers and sellers, entry and exit into the market, if products are homogeneous or differentiated, the mobility of factors of production as well as whether the information about the market is perfect or imperfect.
5. A perfectly competitive market (PCM) is a market that has many buyers and sellers with the traded products being homogeneous. The price and quantity determination under PCM is through market forces of demand and supply. The supply curve of the competitive firm is determined by its marginal cost.
6. Monopoly market is a market in which there is a single seller and the product it sells has no close substitute. The monopoly market can adopt price discrimination (charging different prices basing on geographical location or income of the consumers).
7. Monopolistic competition is a market structure with a large number of firms selling products that are similar but not identical. In monopolistic market, every firm has a monopoly over the product it produces though the products are similar.
8. Oligopoly market is a structure dominated by few large firms selling differentiated or homogeneous products. Firms under oligopoly market can collude and agree on price and quantity of the good or service.
9. Demand curve for oligopoly firm is a kinked demand curve in that an increase in the price by the firm will lower the quantity demanded of the product and a decrease in price by the firm will be responded to by the same decrease in price by other firms.

Revision exercise

1. Present and discuss the assumptions behind the traditional model of perfectly competitive markets.
2. What are causes and measures for controlling monopoly market?
3. Explain the role that the number of firms and barriers to entry play in determining how real-world oligopolistic industries behave.
4. Briefly explain the concept of a Kinked demand curve under oligopoly.

5. Discuss if the mobile networks in Tanzania have features of oligopolistic markets. Give reasons for your answer.
6. Describe differences and similarities between a monopoly and monopolistic market.
7. Explain sources of monopoly.
8. Describe the concept of collusion in oligopolistic firms.
9. Explain how a monopoly causes inefficiency for the society.
10. Study the diagram below and attempt the following questions.



(a) Identify the equilibrium output and price.
(b) Identify total revenue, total cost and profit.
(c) Does the firm operate under the perfectly competitive market? Elaborate your answer.
(d) Show the relationship between AR and MR.

11. A firm faces a demand function given as $Q = 40 - 0.02P$. The fixed costs for the firm are Tshs 8,000, the marginal cost of producing a unit of output is Tshs 4. Suppose a firm is operating under a perfectly competitive market, what will be the level of profit earned by this firm?

12. Mpalaki Company Limited is the supplier of fish in Kazinde market. If the demand curve of their product is given by equation:

$Q = 250 - 0.2P$

Where Q is quantity of fish demanded and P is price of fish per unit.

(a) Calculate the output price that will maximise the firm's profit from the sale of fish.

(b) Obtain the total revenue of the firm.

(c) Determine the average revenue of the firm.

13. Given that:

$TC = Q^3 - 29.5Q^2 + 115Q + 120$ and $AR = 65 - 1.5Q$

(a) Derive the profit equation in terms of Q.

(b) Determine the level of output that maximises profit given that the second order conditions for profit maximisation hold.

14. TANESCO is the sole supplier of electricity in Tanzania. Two separate groups of consumers, namely, commercial users and domestic users, consume this good. The company charges different prices to these two groups. Assume that the major goal of the company is to maximise profit.

(a) How should the company allocate its total output of electricity between the two groups of consumers?

(b) Which group is likely to be charged a higher price? Provide reasons for your answer.

(c) Elaborate on necessary conditions for the company to charge different prices for the same commodity.

15. Graphically, explain how a monopoly firm maximises profit.

16. Show that MC=MR by using a profit equation.
Where, MC is marginal cost and MR is marginal revenue.

Chapter Six

Market failure

Introduction

Think about a village or household near a factory that can very easily be affected by its emission. Are these costs incurred by the neighbourhood compensated? Moreover, think about those people using private vehicles such as cars and motorcycles, which cause congestion and pollution in towns and cities. It is of interest to know, are the households or the villages compensated for the emissions from the factory? Also, is pollution and congestion included in the price of the factory and service provider? Are effects accounted for in the price? Lastly, what if the government was not responsible for constructing roads and providing necessary utilities such as clean and safe water to the society, would these services be provided sufficiently? In this chapter, you will learn about the concept of market failure, causes of market failure and various ways to correct market failure. The competencies developed will enable you to recommend effective economic policies that can be used when market mechanism does not produce efficient outcomes.

Meaning of market failure

As explained in Chapter Two, through forces of demand and supply, the market can attain an equilibrium, which is an efficient outcome. However, there are some situations when market forces do not lead to efficient outcomes. In economic theory, these situations are termed as market failure, situations where a market fails to allocate resources in the most efficient manner. Thus, the failure of the market is its inability to capture the entire social costs or social benefits of the produced goods and services. Hence, the society's scarce resources are not allocated optimally to maximise social welfare.

Causes of market failure

The failure of market mechanism to generate optimal outcome is commonly attributed to two sources. One is the case of public goods whereby the market fails to price the goods to be supplied, which gives no incentives to private producers to produce it, despite their importance to society. Two is the case of externalities whereby the competitive market fails to account for all costs and benefits of producing a good. As a result, private producers do not become obliged to incur some costs to produce the goods and consumers are not obliged to pay to enjoy such benefits. Those costs or benefits are referred to as external costs or external benefits to production of the goods, since they are external to decisions of economic agents and forces of demand and supply in the market.

Externalities

The prerequisite for forces of demand and supply is to guide markets in the society's interest. Market transactions should have no effects commonly referred to as external effects on members of society who are not a party to the transaction made. These are side effects that affect individuals or firms that are not involved in the transaction. An externality is referred to as the effect of an economic activity on an uninvolved third party such that economic agents causing the effect do not take into account effects caused in their economic decisions.

An externality is an indirect cost or benefit that arises from production or consumption decisions that falls on a firm or an individual other than the involved producer or consumer. Hence, it is a situation where an agent "external" to a transaction is materially affected and is not fully compensated for the effect. The parties involved in the economic transaction do not bear these indirect costs or enjoy the benefits exclusively. Hence, the society carries the burden of bearing the costs, on one hand, or enjoying the benefits inclusively, on the other hand.

Thus, externalities are considered to result from goods or services involved in a given transaction to be priced incorrectly by the market. Because the price of the good or the service involved in the transaction is not right, the welfare of an individual or a firm is directly affected by an action of another agent in the economy. Suppose that a firm produces a product and sells it in the market, but part of the cost of producing the product is not borne by the firm. For example, a

firm may dump freely industrial waste products from its productive activities.The waste products may contain hazardous chemicals that affect the quality of water consumed by the community and the lives of creatures in the river, such as fish. Dumping of waste materials into the river will negatively affect the community, which is a third party to the transaction. The hazardous chemicals will poison the fish, and the water will be harmful to humankind, other species and the environment if not treated. Such an indirect cost to society, which is not borne by the parties involved in the transaction are called a negative externality. On the other hand, the third party may reap benefits from some economic transactions, which are called positive externalities. For example, a community may benefit from education of individuals in some ways that it does not pay for it.

Basic concepts in economic analysis of externalities

Marginal private costs (MPC)

Marginal private costs (MPC) are costs of producing an additional unit of a good or service that is borne by the producer of that good or service. The MPC include both explicit and implicit costs incurred by a firm in the production of a good and a service in a market transaction. In production, private costs include costs that are incurred on labour, raw material and machinery, among others. The private costs do not take into account any costs that are not priced by the market, such as the negative effects or harm caused by dumping freely the waste into the nearby river.

Marginal external cost (MEC)

Marginal external cost (MEC) is the cost of producing an additional unit of a good or service, which is borne by all people in a community. The MEC is an external cost at the margin that the producer or the consumer of a good or service should be incurring but does not because the market forces fail to include it in the price of the good or service. Hence, external cost is the cost passed to the whole society that result from an economic transaction whereby some individuals in the society are not directly involved. An example of such costs is air pollution caused by motor vehicles. The whole society incurs the cost of air pollution and not just the producers or users of the motor transport.

Marginal social cost (MSC)

Marginal social cost (MSC) is the cost of producing an additional unit of a good or service that the entire society incurs. Social costs include both private costs and external costs to society arising from production or consumption of a good

or a service. Generally, social costs of a productive or consumption activity sums up all costs, regardless of who incurs them.

$$MSC = MPC + MEC$$

$$MEC > 0 \rightarrow MSC > MPC$$

An example of private versus social cost may be deduced from the action of individuals driving private cars in a city thereby causing traffic congestion and pollution. Whereas the individuals incur private costs, such as fuel costs, the whole society incurs both private costs and external costs from the decision of individuals to drive cars that cause congestion and air pollution. The sum of all private costs and the external cost to society is referred to as social costs.

Marginal private benefits (MPB)

Marginal private benefit (MPB) is the benefit from an additional unit of a good or service that a consumer or producer of the good or service receives. To a consumer, private benefit is satisfaction (utility) from consuming a given good and to a producer, the private benefit is the profit received from selling the produced good.

Marginal external benefits (MEB)

Marginal external benefit (MEB) is the benefit realised from an additional unit of a good or service that people other than the consumer or the producer of that good or service enjoy. The benefits are realised by an individual or producer who does not pay anything to reap them. Examples of external benefits include health benefits to society, which result from individual vaccination or benefits from education of individuals in a community. Society may benefit from either or both vaccination and education actions in terms of consumption through improved health (or knowledge) or in terms of production through increased productivity. Consumers may benefit from a beautiful garden in the neighbourhood by getting extra satisfaction from improvement in the air quality caused by the garden. For producers, pollination of an orchard by bees from a nearby hive, for example, gives freely to the producer of apples an essential input. Thus, apple producers benefit in terms of getting free input for producing apples. A beekeeper could also benefit from the nectar from the orchard in production of honey.

Marginal social benefits (MSB)

Marginal social benefit (MSB) is the marginal benefit, which result from suming marginal private benefits and marginal external benefits. Social benefit is the

total benefit to society that arises from production or consumption of a good or service. Social benefit includes all private benefits plus external benefits of production or consumption. If a good has external benefits, then the social benefit will be greater than the private benefit.

$$MSB = MPB + MEB$$

$$MEB > 0 \rightarrow MSB > MPB$$

Types of externalities

There are two types of externalities namely, negative externality and positive externality. Both externalities are explained as follows:

Negative externality

Negative externalities are costs imposed on a third party in a society because the market fails to price correctly actions of some producers or consumers. Examples of negative externalities are pollutants created by firms and carbon monoxide gas emitted by motor vehicles, which are harmful to humankind. In unregulated markets, producers do not take responsibility for the negative externality they cause, which affects people and could affect production thereby reducing the welfare of society. For example, due to negative externality that vehicles create, the social cost of producing a vehicle exceeds the private cost that the producer incurs. However, the negative externality is not reflected in the market price.

Negative externalities can be analysed by using supply and demand curves. In Figure 6.1, market equilibrium output is at Q_e (where Demand = Supply). However, that level of output is socially inefficient because at Q_e, external costs of vehicles lead to greater social costs that are higher than the private cost incurred by a firm. Therefore, to attain efficiency in this market, a firm has to produce at Q_s where marginal private benefit (MPB) = marginal social cost (MSC).

The government attempts to reduce a negative externality by imposing taxes on fuel and motor vehicles in order to reduce the number of private vehicles on the road. Alternatively, the government can provide subsides in order to promote use of public transport that would reduce emissions from vehicles.

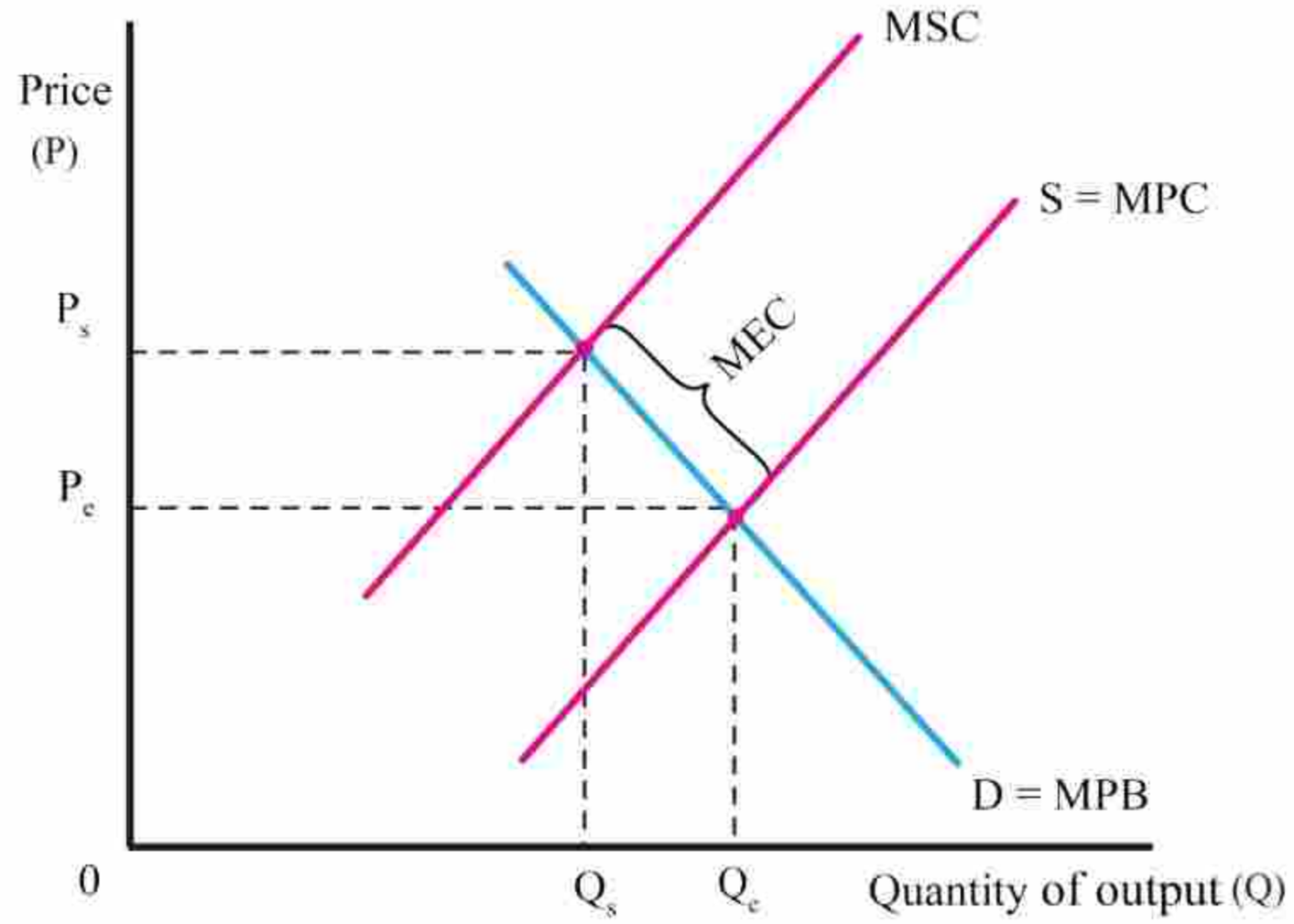


Figure 6.1: *Negative externality*

Example 6.1:

Consider an inverse demand function $P^d = 24 - Q$ and an inverse supply function given as $MPC = 2 + Q$. The Marginal external cost is given as $MEC = Q - 2$.

(a) Determine the market equilibrium price and output.
(b) Find the economically efficient output.
(c) Compute the MEC at equilibrium level.
(d) Graphically illustrate the marginal private cost (MPC), demand curve, marginal external cost (MEC) and marginal social cost (MSC) functions.

Solution

(a) The market equilibrium is obtained where the quantity demanded equals to the quantity supplied. Since, the quantity supplied is depicted by marginal private cost (MPC), then

$$P^s = MPC = 2 + Q$$

At equilibrium,

$$Q^d = Q^s$$
$$24 - P = P - 2$$

$$2P = 26$$

$$P = 13$$

To obtain the equilibrium quantity, substitute the value of price (P =13) into the supply function

$$13 = 2 + Q$$

$$Q^e = 11$$

Therefore, the equilibrium price is Tshs 13 and equilibrium quantity is 11 units.

(b) The economically efficient output is at a point where the marginal social cost (MSC) is equal to demand. Since the MSC is the sum of marginal private cost (MPC) and marginal external cost (MEC), therefore:

$$MSC = MPC + MEC$$

$$MSC = (2 + Q) + (Q - 2) = 2Q$$

and demand $P^d = 24 - Q$.

At socially efficient price and output:

$$MSC = MPB$$

Thus, $2Q = 24 - Q \rightarrow Q^* = 8$ and $P^* = 16$.

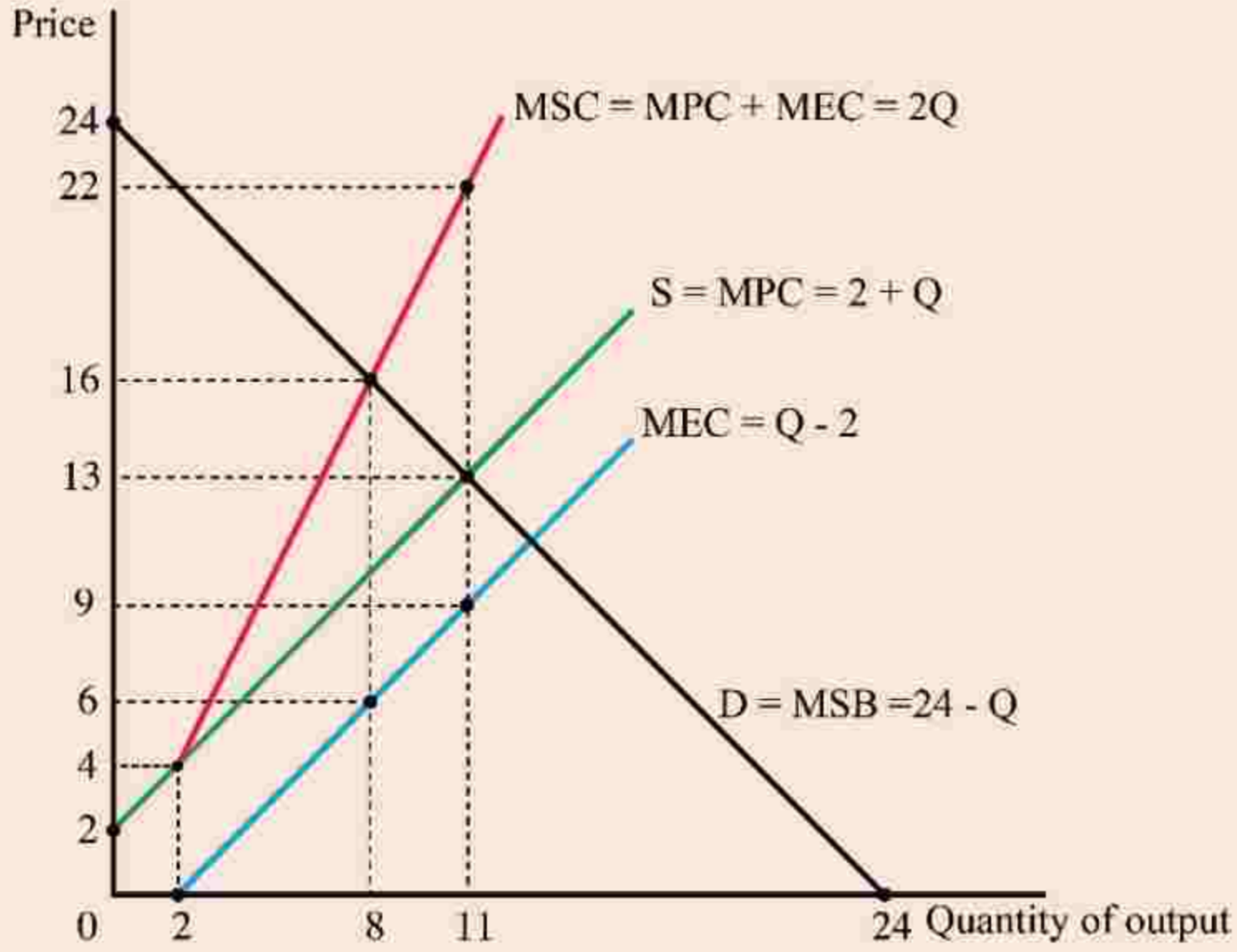
(c) To compute the MEC at equilibrium substitute the value of Qe = 11 at equilibrium into the MEC function $MEC = Q - 2$ such that:

$$MEC = 11 - 2$$

$$MEC = 9$$

Therefore, the value of MEC at equilibrium is 9.

(d) Graphical illustration of the marginal private cost (MPC), demand curve, marginal external cost (MEC) and marginal social cost (MSC) functions.



Positive externality

If an activity causes benefits to a third party, it is termed a positive externality. Education, for example, has a positive externality. In addition to private gains to the individual beneficiary such as good employment and income, education has a positive externality to society that benefits from education of the individual. When a large part of the population is educated, the society gets professionals in various fields such as lawyers, doctors and teachers. Moreover, highly educated people in the society are more likely to participate in projects that aid in solving local problems. Another externality of education is that it provides the society with educated leaders who have necessary knowledge and skills to make appropriate decisions for benefits of citizens.

An example of externality is when a firm decides to spend money on research and development (R&D), and the benefits of innovation spread to other enterprises. For example, research efforts could lead to a new marketing strategy that will attract more customers. Thus, if other firms adopt legally the new marketing strategy, they will also use the strategy to increase customers.

Figure 6.7 demonstrates a positive externality. In the figure, market equilibrium quantity is at Q_e, where the quantity demanded equals to quantity supplied.

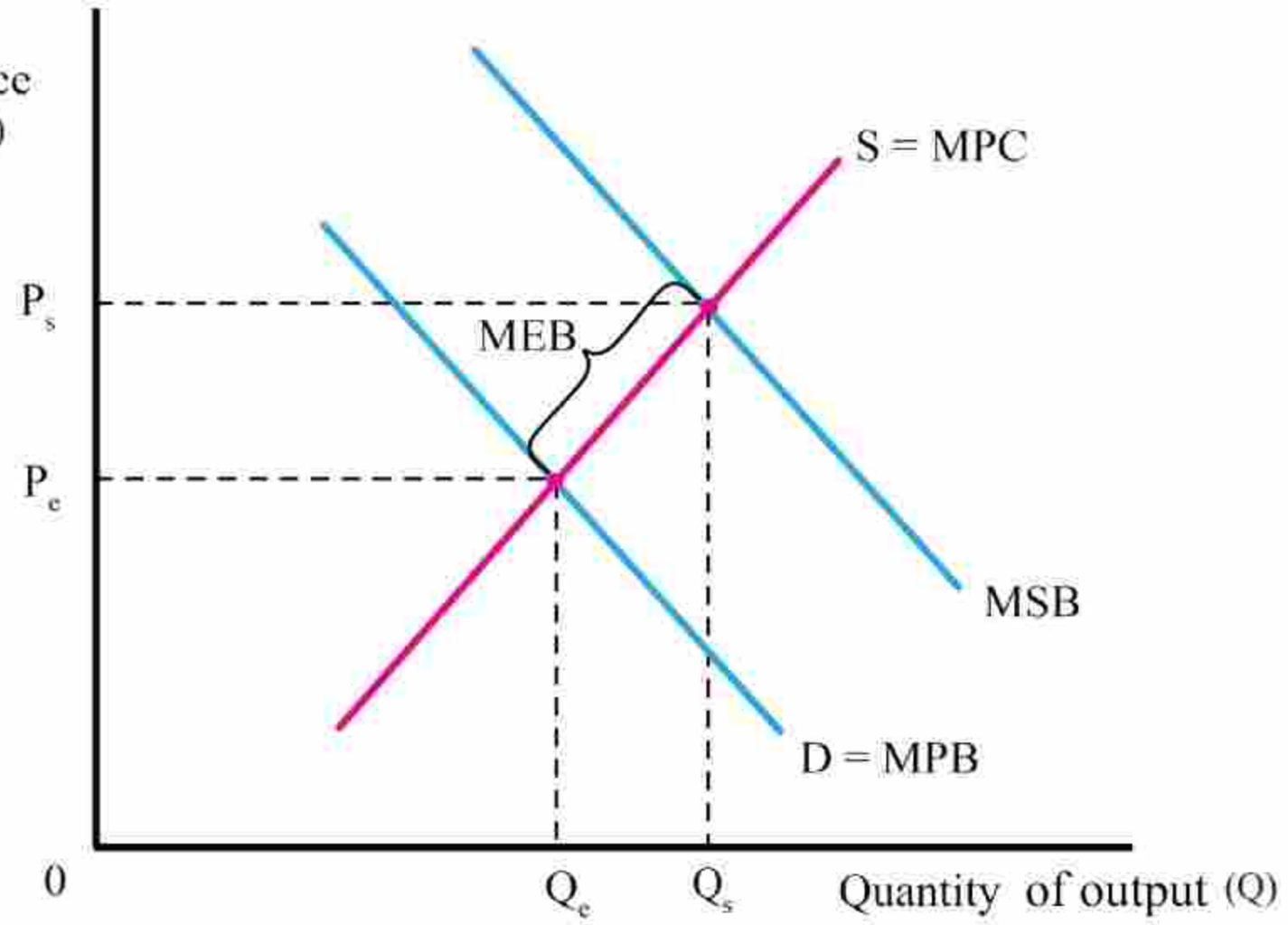


Figure 6.2: *Positive externality*

However, the level of output is socially inefficient because at Q_e, social benefits are larger than private benefits. Therefore, there is under-consumption of the good. In this case, social efficiency is at Q_s where the marginal social benefit (MSB) is equal to the marginal private cost (MPC).

Exercise 6.1

1. Even though the concept of an invisible hand states that markets can automatically result in an efficient price and quantity of a good or service, in practice, we frequently observe interventions by the government.
   (a) Explain the reasons for government intervention?
   (b) What are forms of government intervention?
2. Distinguish the following concepts:
   (a) Marginal social cost and marginal external cost.
   (b) Marginal private cost and marginal private benefits.
3. Explain the difference between negative and positive externalities.

Activity 6.1

Use a case study of any factory in your area. Present and briefly explain types of externalities that exist and propose remedies for controlling them. Share your work with your fellow classmates.

Concept of public goods

Provision of public goods is another source of market failure. In economics, a public good has two characteristics: it is non-excludable and non-rival. A good is excludable if those who do not pay the market price to buy it cannot consume the good. A good is a rival one if its consumption reduces the quantity available for others to consume. Thus, non-excludable goods are goods for which there is no mechanism of excluding an individual from consuming the good once it has been availed. Hence, such goods are available free of charge to anyone to consume. An example of such goods is the air that we breathe. No one can exclude anyone from breathing the air. Another example of non-excludable goods include shared goods such as street lights or waterbodies such as rivers and lakes. It is difficult to exclude people from use of goods that are publicly shared, even if they do not pay for their use.

For the non-rival goods, their consumption by one person does not diminish the amount available for consumption by other people. This means that consumption of the good by one person does not preclude or diminish the benefit from another consumer who chooses to consume the good. For example, your breathing of air does not diminish the amount or quality of air that is available for your colleagues to breath. A television broadcast is another example of a non-rival good. Watching of a specific television channel by one person does not deny other people to benefit from watching the same channel.

By using the two characteristics of goods, the goods can be categorised into four different types, namely, public goods, private goods, club goods and common goods. Table 6.1 shows types of goods according to their characteristics.

Table 6.1: *Types of goods*

	Excludable	Non-excludable
Rival	**Private goods** Food stuff, tap water	**Common goods** Fisheries, forest, clean water
Non-rival	**Club goods/ merit goods** Cable television, education	**Public goods** Radio broadcast, national defence, street lights

Basic concepts in economic analysis of public goods

The following are basic concepts in economic analysis of public goods:

Public goods

Public goods are freely available for consumption because they possess characteristics of non-rivalry and non-excludability. Because of these two characteristics, public goods usually are not provided through the free market system because it is difficult for people to pay a price directly when they know they can get them freely. Even more difficult would be the supply because the price mechanism does not work for public goods. Hence, private firms cannot adequately provide public goods and charge for them. Thus, public goods are supplied and financed mainly by the government through taxes. Consider streetlights as an example. If streetlights are provided, no road user can be excluded from benefiting from the lights. It is also not possible to charge people who walk under strretlights. Walking under streetlights does not make the lights dimmer. Hence, use of streetlights by some people does not diminish the amount available for other road users. Therefore, streetlights are non-excludable and non-rival, that is to say, they are pure public goods. A public good is for society (the public) as a whole.

Private goods

A private good has characteristics that are opposite to those of a public good. Because private goods are excludable and rival, they are supplied through the market. A bottle of soda, for example, is a private good. Its owner can exclude others from using it and once it has been consumed, it cannot be used again.

Common goods (or commons)

Like public goods, common goods are not excludable. They are available free of charge to anyone who wishes to consume them. For example, fresh water, grazing land and wildlife. However, unlike public goods, common goods are rival because one person's use of the good reduces other people's use. Trees in the forest are rivalrous in nature. When one person clears a forest, there are fewer trees for other people to benefit. Grazing land is also rival at some point as it reaches the limit of cattle stock that can graze on land. Because common goods are a common property, they tend to be overused.

Club goods and merit goods

Club goods are non-rival but excludable, as Table 6.1 shows. For example, a television-broadcasting firm may prevent people from accessing certain channels by requiring them to pay for the channels. However, once the service is provided it becomes non-rival. Thus, watching the television channel by one person does not diminish the quality or amount available for other people to benefit from watching that channel.

Merit goods, like club goods, are excludable but non-rival. They could be provided through a free market mechanism, but they would be under provided, for example, education. Education is important to any nation because educated people tend to create positive externalities to society. If education was not provided at all, there would still be private schools for those who could afford it. However, because market only takes into account private costs and benefits, there would be under provision of education for everyone because private costs may be prohibitive to some people to afford. The market system does not take into account external benefits that may arise to society.

Public goods and the market failure

Existence of public goods provide an important example of market failure. Markets cannot supply public goods efficiently. Presence of public goods gives rise to a free-rider problem. Free ride means to enjoy benefits from provision of a public good and shirks to pay for the good. Free riders do not pay for the good be it a public or private good, knowing that others will bear the cost of supplying the good. Non-excludability and non-rivalry characteristics of public goods allow existence of free riders. Inability of the market to charge price for public goods prevents private firms from supplying these goods. Thus, dependence on the

market to supply the public good would lead to lack or shortfall in its supply. It is for this reason provision of public goods requires central government or local government interventions.

Exercise 6.2

1. Explain the free rider problem? Why does the free rider problem induce governments to provide public goods?
2. Define and give examples of public goods.
3. Explain what is meant by a good being non-excludable and non-rival.
4. Why public goods should always be provided by the government?



Activity 6.2

Think of one problem of negative externality that you have experienced or witnessed in Tanzania. Using a concept of public goods, mention possible causes of the problem.

Ways of controlling market failure

The government is concerned with negative externalities because of its associated costs imposed on society. In addition, the imposed costs cause inefficient allocation of resources in the market. Thus, the government, firms and individuals find ways for controlling externalities. The following are some ways of solving the problem of externalities so as to facilitate existence of efficient allocation of resources in the market.

Emission charges

Emission charges are charged to minimise environmental pollution. The charges aim at punishing the polluter for external cost of pollution it imposes to the society. Hence, they provide an incentive for polluters to seek new technologies that are less polluting. Emission charges allow the polluting firm to pollute as much as it is able to pay for the pollution. The more pollution a firm creates, the more it pays in emission charges.

To reduce emission charges, polluting firms work to find the best way to reduce pollution, for example, acquiring the best technology that creates minimum pollution.

In Figure 6.3, market equilibrium output is at Q_e (where Demand = Supply). However, this level of output is socially inefficient because at Q_e, external costs lead to greater social cost than private cost incurred by a firm. To attain efficiency in the market, a firm has to produce at Q_s where marginal social benefit (MSB) = marginal social cost (MSC) after imposition of emission charge.

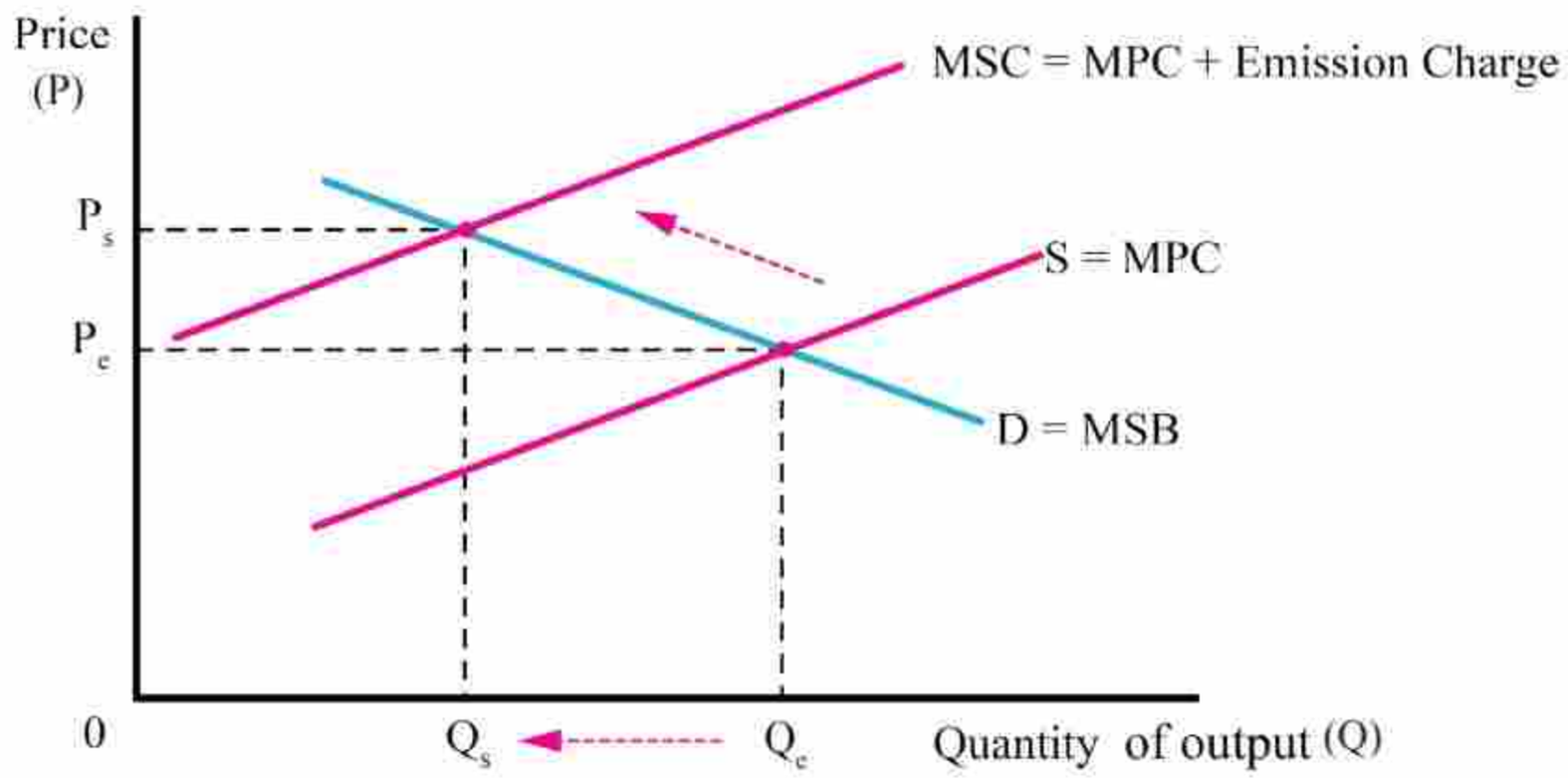


Figure 6.3: *Emission charges*

Pigouvian taxation

Pigouvian Tax is tax imposed on economic activities that generate negative externalities. In the case of negative externalities, the social cost of an activity is greater than the private cost of the activity. In such a case, the market outcome is inefficient because it leads to overproduction of a good. Taxes make it highly expensive for firms to produce the good or service generating a negative externality thereby providing an incentive to produce less of it. In theory, a Pigouvian tax is equal to cost of externality. Theoretically, taxes work in the same way as emission charges. Practically, an emission charge is imposed per unit of pollution, while tax is imposed per unit of a good or service produced. The main aim of the tax is to increase the firm's private marginal cost (PMC) until it equates with social marginal cost curve (SMC). This will result in a socially efficient output level. In Figure 6.4, the market equilibrium output is at Q_e. However, at this level of output, external costs lead to greater social cost than the private cost incurred

by a firm. To attain efficiency in the market, a firm has to produce at Q_s after imposition of tax.

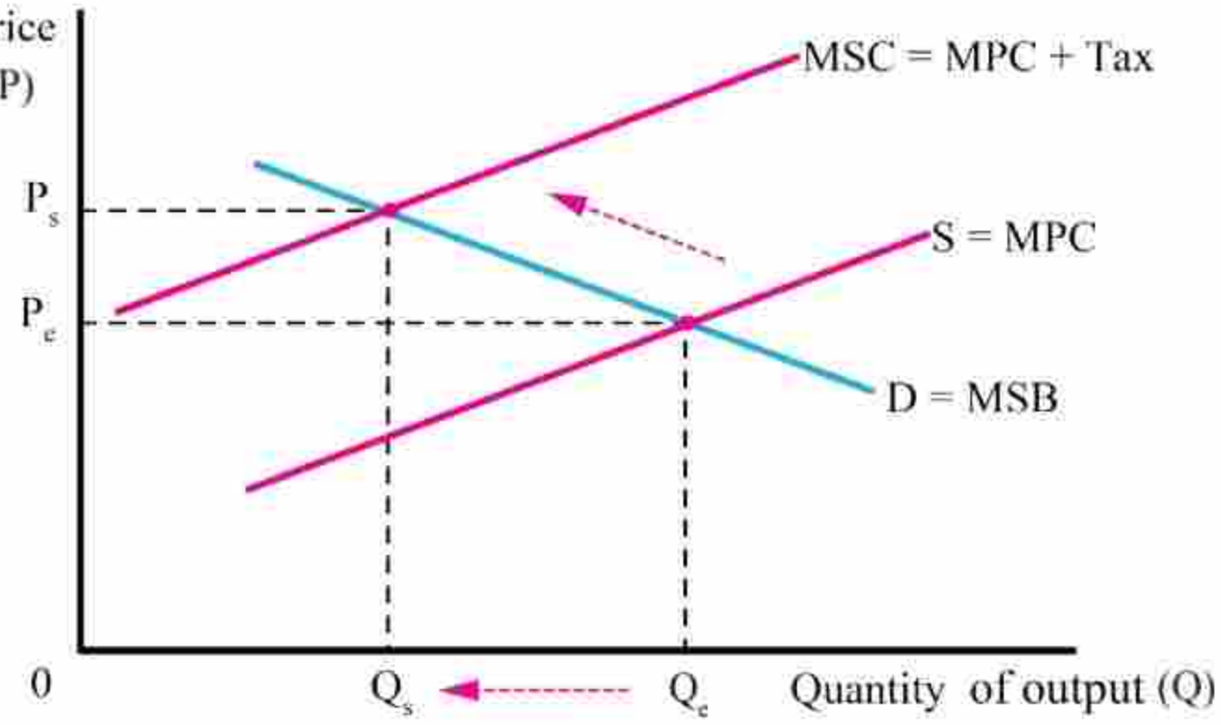


Figure 6.4: *Pigouvian taxation*

Example 6.2

Suppose the government wants to impose the Pigouvian tax that will induce the market to produce an economically efficient amount of output, Q_s. Use information in Example 6.1 to find the tax level that the government should impose to correct the problem of externality.

The emission charge or tax is the difference between what consumers pay for the good and what producers receive for the good, which reflect the tax per unit.

That is, what producers receive for the good is $MPC = 2 + Q$

$Q = Q_s = 8$ units, which is an efficient output

$MPC = 2 + 8$

$MPC = 10$

What consumers pay for the good is $MSC = 2Q$

$MSC = 2(8)$

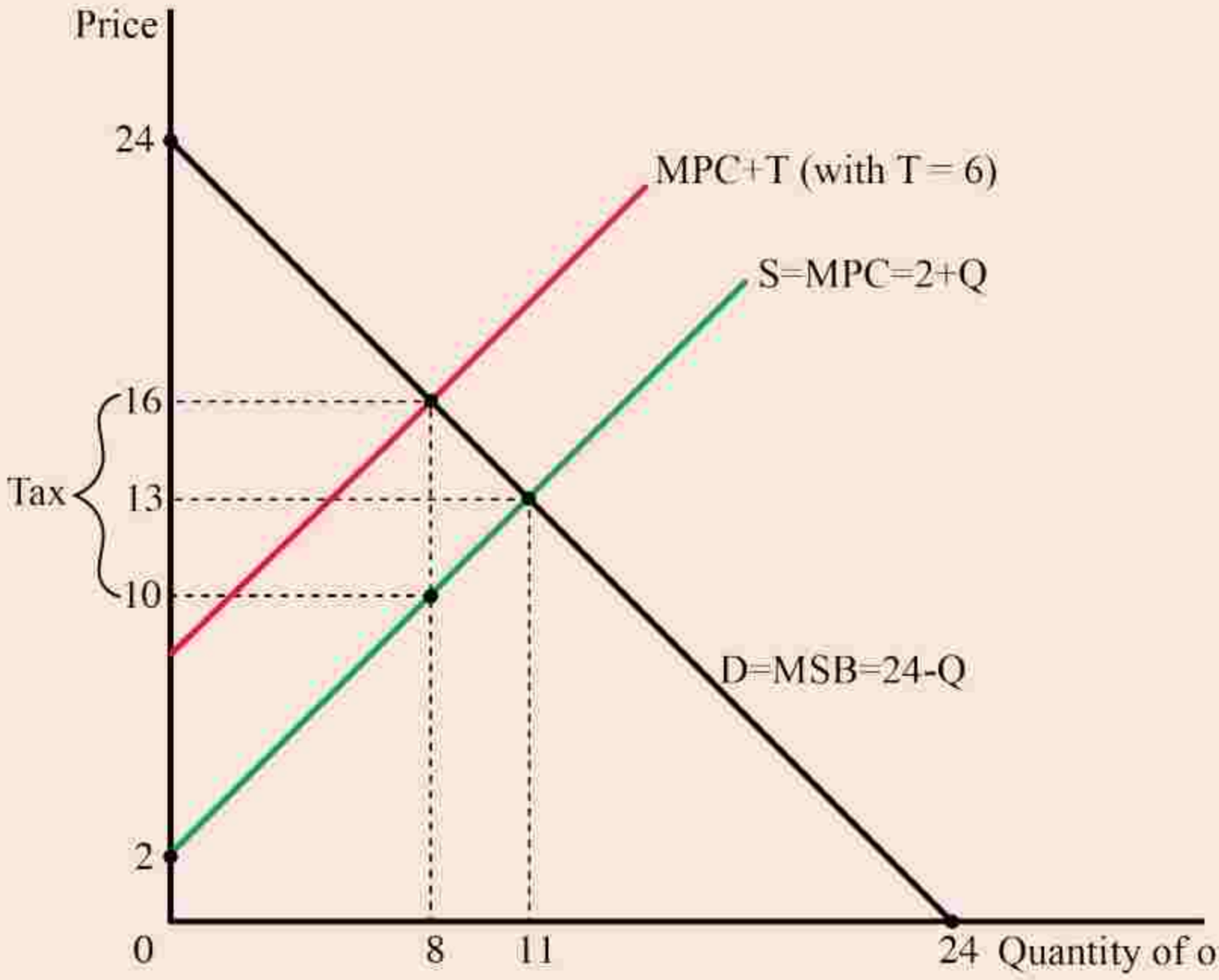
$MSC = 16$

Thus

Tax per unit = Tshs 16 – Tshs 10 = Tshs 6 .

Therefore, the government should charge Tshs 6 per unit to induce the market to produce the economically efficient output.

Graphically,



Subsidies

Government subsidy is a payment that effectively reduces the cost of producing a given good or service. Such subsidies provide an incentive for firms to increase production of goods that give positive externalities. The government can decide to provide subsidies to stimulate certain economic activities and encourage production of goods or services that generate spill over benefits. Subsidies are commonly used to increase consumption of goods with positive externalities. The main aim of the subsidy is to reduce the firm's marginal private cost (MPC) until it equates with marginal social cost curve (MSC). This will result in a socially efficient output level as shown in Figure 6.5.

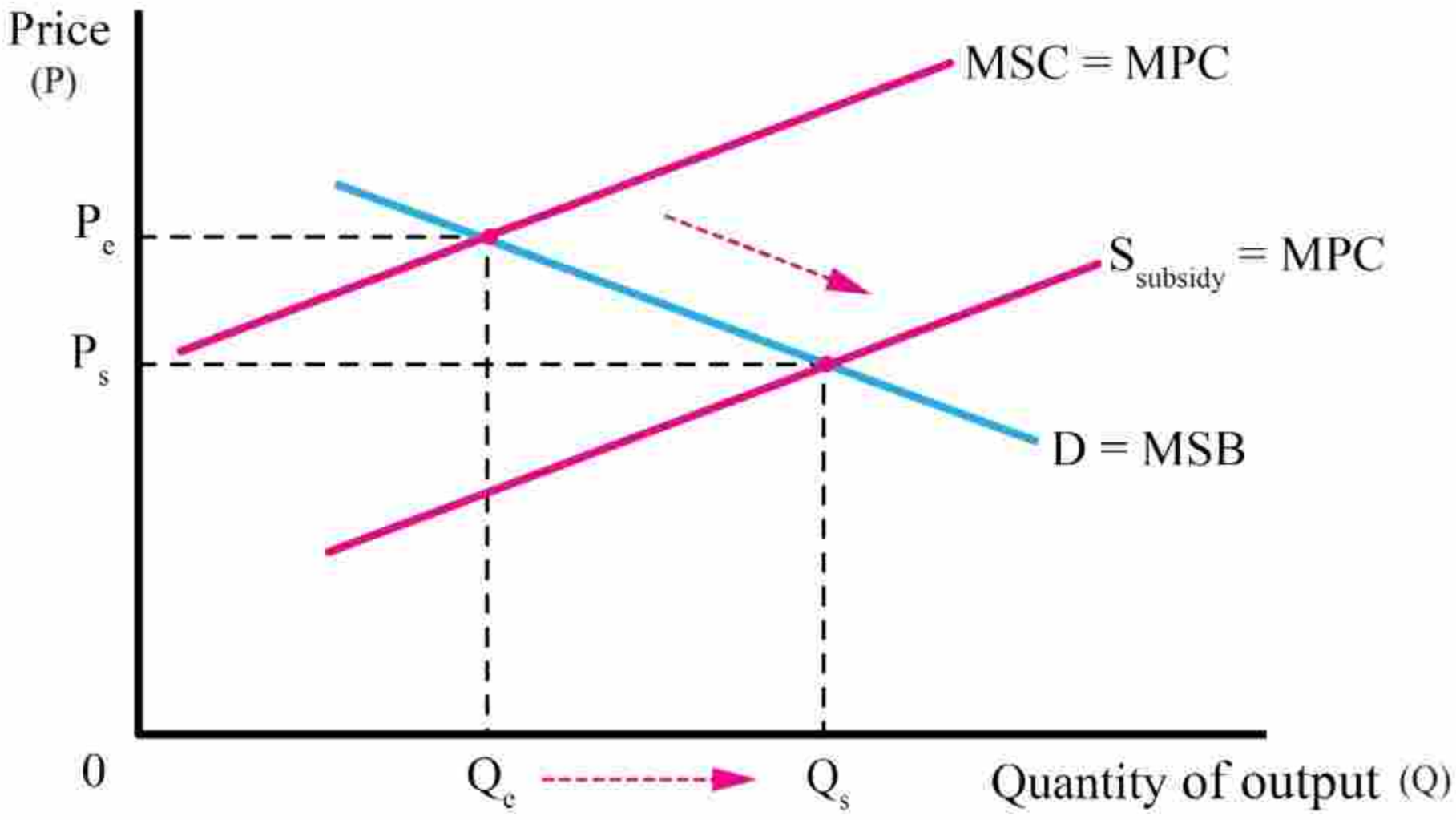


Figure 6.5: *Subsidy*

Regulations

The government can directly regulate externality activities by either using the command or direct control policies. Governments would, for example, impose standards, targets, process requirements, or outright bans to regulate certain economic behaviours. In Moshi municipality, for example, people who litter streets are penalised because littering imposes cost to the municipality and other road users. Thus, by using regulations, the government aims to protect the environment and individuals that live in Moshi municipality that would otherwise suffer from the negative effects of littering.

Marketable or tradeable permits

To address the problem of negative externalities, the government may use a permit system to reduce effects associated with negative externalities. The government, for example would introduce the right to pollute and set a goal in terms of the desired pollution level. Then the government would issue permits and allow permit holders to trade them. In total, the permits issued to firms would equal the desired pollution level. Firms that are able to reduce production of externalities would be able to sell permits to firms that are unable to make such reductions and are willing to pay for the permits.

Market solution

According to the Coase theorem, a market can provide an efficient solution to an externality problem. Given a suitable assignment of property rights, private bargaining between individuals can correct externality problems and lead to efficient outcomes. The efficient solution to an externality does not depend on which party is assigned the property rights (polluter or victim) as long as someone is assigned those rights. In the context of the Coase theorem, if property rights are well assigned, individuals in conflict caused by an externality problem can bargain or negotiate and come into an agreement on allocation of resources without incurring social costs. For example, Mwasha owns a poultry farm at his house in Kazinikazi Street. The poultry farm produces noise and smell, which are externalities that disturb neighbours. If the benefits of keeping the poultry farm exceeds costs of the externality problem of smell and noise, Mwasha will compensate the neighbours and they will find an agreement in which Mwasha will keep the poultry farm. Otherwise, if the costs of Mwasha keeping the poultry farm exceeds the benefits, then the neighbours could pay Mwasha to stop operating the poultry farm that disturbs them.

Activity 6.3

1. From tomorrow morning, observe all your activities for the whole day, then:
   (a) Identify positive and negative externalities.
   (b) Suppose the government decides to pay for positive externalities and tax for negative externalities, will you be the net gainer or net looser? Explain.

2. With reference to activity 6.1
   (a) How can the government intervene a market system to solve the problem?
   (b) How can the problem be solved without government intervention?

Chapter summary

1. One of the reasons governments intervene in the market is summed up in the concept "market failure."
2. Market failure is a situation by which a market fails to allocate resources efficiently. There are two sources of market failure: externalities and public goods.
3. Externalities refer to effects of a decision by consumers/producers that has an impact on a third party who is not involved in consumption/production decision when the resulting effects are not compensated for.
4. Negative externalities are external costs incurred by the society when producers and/or consumers do not take into account the cost to society as a result of their decisions. Examples of negative externalities include traffic congestion, smoking in public areas and destruction of wildlife.
5. Positive externalities or external benefits are beneficial effects on third parties as a result of a decision by another party, for example, societal benefits that are associated with vaccinations, public transport, education and attractive gardens.
6. Public goods have two characteristics (in consumption): non-rival and non-excludable.
7. A market system fails to supply public goods. Since people cannot be excluded from enjoying benefits of a public good, individuals may withhold paying for the good assuming that others will pay for it. This problem is referred to as free-rider problem.
8. Measures to correct market failure include government interventions, through direct provision of public goods and service. For externalities, they include use of taxes (for example, a Pigouvian tax), subsidies, emission charges, regulations, tradeable permit, and the market solution.
9. Pigouvian tax is tax charged to a firm that causes an externality in order to cover the marginal external cost of pollution. The tax reduces pollution through the increase of the firm's cost of production and reduction of output.
10. Market solutions to externalities as advocated by Coase theorem that if property rights are well assigned, individuals facing the externality problem

could enter into negotiations and bring about an efficient outcome to the externality problem, without any government interventions.

Revision Exercise

1. Elaborate examples of a positive externality and negative externality.
2. Give examples of a private good, a public good and a club good, and classify each of the good as either a rival or non-rival good and whether an excludable or non-excludable good.
3. What is a common good? Is it a rival or non-rival good? Is it an excludable or non-excludable good? Give examples.
4. What will happen when a common good is used with and without government intervention? Will the consumer use too little or too much of the good? Explain why?
5. What type of a good is Fishery? Is it a rival or non-rival good? Is it an excludable or non-excludable good?
6. Define a club good, and:
   (a) Categorise whether it is excludable or non-excludable and rival or non-rival.
   (b) Give examples of club goods.
7. Suppose a cement factory is located on Rufiji River bank. The marginal private cost (MPC) of producing bags of cement (in Tsh. per ton) is given by the function: $MPC = 1,000 + 0.5Q$, where Q are tons of bags of cement produced. In addition to this marginal private cost, an external cost is incurred. Each ton of cement produced emits pollutant that flow into the river that cause damage valued at Tshs 1,000. The demand for cement by the society is given by $P^{d} = 3,000 - 0.5Q$.
   (a) Find profit maximising output of the factory.
   (b) Find output of the factory that maximises social net benefits.
   (c) Explain why the socially efficient output of cement is lower than the private profit-maximising output level.
   (d) Draw a diagram showing the relationship between marginal private cost (MPC), demand curve, marginal external cost (MEC) and marginal social cost (MSC) functions.

8. With reference to characteristics of goods discussed in this chapter, what is the difference between national defence and a bottled water? Can the private market provide an efficient amount of each of these goods on its own? Explain.
9. Consider the case of river pollution by industries, which affect water users downstream. The marginal external cost function of affected people is given by $MEC = 5Q$. The marginal private cost of the firm is given by $MPC = 2Q + 3$.
   (a) Calculate the price for a competitive market if the quantity $Q = 10$.
   (b) Calculate the socially desirable price if the quantity $Q = 10$.
   (c) Why does the price in part (b) differ from that obtained in part (a)?
10. The demand function for Takata Soap Company is $Q = 1,000 - 2Q$ where Q is the number of units of soap produced. The marginal private cost of producing Q is constant at Tshs 500 and marginal external cost of producing an extra unit of Takata soap is Tshs 200.
   (a) How many units of Takata soap will the company produce in a competitive market?
   (b) What is an efficient production level?
   (c) Suggest a Pigouvian tax that wold be charged to obtain an efficient output level.
   (d) How much revenue would the tax generate?
11. Compare and contrast different approaches for correcting negative externalities.
12. Mzee Matata loves watching football games on his local public TV station, but he never contribute to support the station during its fund-raising drives.
   (a) What name do economists have for Mzee Matata? Why?
   (b) How can the government address the behaviour of people like Mzee Matata?

Glossary

Art

a practical application of knowledge that teaches how to do things using the knowledge that an individual has acquired

Bonds

a certificate of indebtedness

Budget deficit

shortage of tax revenue from government spending

Cartel

a group of firms working in unity to promote their own interest

Coase theorem

the proposition that private parties can solve the problem of externalities on their own by bargain without cost over allocation of resources

Collective bargaining

the method used by unions and firms agree on terms of employment

Complementary goods

commodities whose consumption depends on each other. If two commodities are complementary, an increase in the price of one good leads to a decrease in demand of the other, and a decrease in price of one good leads to an increase in demand for the other. Particular examples of such goods are tooth paste and tooth brush

Cost

the value of something a seller must forego to produce a good

Currency

coins and paper bills in hands of the public

Desire

something one would like to have even if he or she cannot afford

Discrimination

a situation of offering different opportunities to individuals based on their personal characteristics such as race, age, sex, ethnic group and income level

Dividend

amount of money paid (typically annually) by a company to its shareholders out of the company's earnings

Efficiency

property of resource allocation that maximises total surplus received by all members in the society

Elasticity

measure of responsiveness of quantity demanded or quantity supplied to a change in one of its determinants

Equilibrium

state of balance between opposing forces. An object is at equilibrium if it is in a state of rest and has no tendency to change

Equilibrium price

the price of a commodity determined in the market by the intersection of a supply curve and demand curve. Price that maximises the firm's profit

Equilibrium quantity

the quantity of a commodity determined in the market by the intersection of a supply curve and demand curve. Quantity that maximises the firm's profit

Exports

goods and services that are produced in the domestic economy and sold for consumption in other countries

Foreign Direct Investment (FDIs)

investment by foreign nationals in the domestic economy Particular examples of foreign investments are in gold mines and cement industries

Free rider

a consumer who consumes a commodity without paying for it

Gross Domestic Product (GDP)

the market value of all final goods and services produced domestically (within a country) in a given period of time

Human capital

knowledge and skills that a worker acquires through education and job trainings

Imports

goods and services that are produced abroad but consumed in the domestic economy

Incentives

anything which bears influence on behaviour or decision of an individual or firm

Inflation

an increase in the general price level in the economy

Labour force

a total number of working age population both employed and unemployed in a country

Liquidity

the ease with which an asset can be transformed into the economy's medium of exchange

Money

an asset in an economy that people often use to buy goods and services from other people locally or internationally

Natural resources

the inputs used in production of goods and services that are provided by nature such as land, rivers, and mineral deposits

Price taker

a seller or buyer of a good that is unable to affect price at which the good sells by changing the amount it sells or buys, respectively

Productivity

the quantity of goods and services produced from each unit of labour input

Profit

difference obtained between total revenue and total cost

Purchasing

action of acquiring goods and services by buying them

Purchasing power

quantity of goods and services a consumer can buy with a fixed amount of income

Rational choice

a choice whereby an economic agent allocates efficiently and effectively available resources to satisfy his or her wants to the possible limit

Rational individuals

are individuals that are systematic and consistent to do things that will enable them to achieve declared objectives

Scarcity

limited nature of society's resources

Slope

change in the y value divided by the corresponding change in the x value of a curve

Spillovers

external benefits or cost for which no compensation is made. They are also called externalities

Substitute goods

alternative goods that give the same level of satisfaction. Typically, commodities that are substitutes serve the same purpose such that an increase in price of one of them leads to an increase in the demand of the other, and a decrease in the price of one of them leads to a decrease in the demand of the other. For example, charcoal and gas

Super normal profit

excess profit earned by a firm over its maximum return that would keep the firm in business

Transaction

an agreement between a buyer and a seller to exchange a good or a service for money

Treasury bills

a short-term financial instrument yielding no interest but issued at a discount on its redemption price.

Union

a worker's association that bargains with employers over wages and working conditions of workers

Utility

satisfaction realised from consumption of a good or service

X-axis

the horizontal axis in a graph

Y-axis

the vertical axis in a graph

Bibliography

Duran, D. C., Gogan, L. M., Artene, A., & Duran, V. (2015). The components of sustainable development-a possible approach. *Procedia Economics and Finance*, *26*, 806-811.

Field, B. C., & Field, M. K. (2016). *Environmental Economics, An Introduction, (7th ed)*. New York: McGraw-Hill .

Jhingan, M. L.(2012). Principles of Economic, (4th ed). Delhi:Vrinda Publications (P) LTD.

Jain, R. K., Cui, Z. C., & Domen, J. K. (2016). Environmental impacts of mining. *Environmental impact of mining and mineral processing*, 53-157.

Kolmar, M. (2017). *Principles of Microeconomics An Integrative Approach.* Switzerland: Springer International Publishing AG.

Lobley, D. T. (1976). *Applied Economics, (3rd ed)*. London: W. H. Allen & Co., Ltd.

Libecap, G. D. (2014). Addressing global environmental externalities: Transaction costs considerations. *Journal of Economic Literature*, *52*(2), 424-79.

Mankiw, G. N. (2019). *Principles of Economics (8th ed)*. United States of America: Cengage Learning.

McConnell, C. R., & Brue, S. L. (1990). *Micro-Economics*. United States of America: McGraw-Hill, Inc.

Pindyck, R. S., & Rubinfeld, D. L. (2013). *Microeconomics, (8th ed)*. United States of America: Pearson Publishing Company.

Pindyck, R. S., & Rubinfeld, D. L. (1989). *Microeconomics*. New York: Macmillan Publishing Company.

Spencer, M. H. (1986). *Contemporary Economics, (6th ed)*. New York: Worth Publishers, Inc.

Roland, G. (2016). *Development Economics (2nd ed)*. New York: Routledge: pg 55-80.

Wang, L. K., & Pereira, N. C. (Eds.). (2012). *Air and Noise Pollution Control: Volume 1* (Vol. 1). Springer Science & Business Media.

Ziolo, M., Jednak, S., Savić, G., & Kragulj, D. (2020). *Link between energy efficiency and sustainable economic and financial development in OECD Countries*. Energies, 13(22), 5898.

Index

A

accounting costs 183
accounting profit 184, 185, 208, 209
arc elasticity of demand. 66
average product 156, 161, 164, 173, 191, 192, 193, 216

B

bargaining theory of wage 137
breakeven point 227

C

capital vi, 6, 10, 16, 36, 96, 107, 120-122, 135, 136, 141-147, 149, 152-158, 162, 165, 166, 169, 172, 179, 182, 183, 184, 186, 207, 208, 215, 249
capitalist economic system 39
cartel 252, 253, 254
change in demand 56, 72, 116
change in quantity demanded 56, 116
change in supply 92, 95, 100
change in the quantity supplied 92
choice 29, 31, 32, 36, 41, 45, 86, 231, 233, 241, 244
classical theory 145
classification of markets 221
comforts 8
competitive demand 60
competitive supply 95, 96
complementary demand 60, 61
composite demand 61
consumer goods 5
consumption 3, 12
cost Function 185
costs of production 22, 129, 168, 180, 181, 182, 215, 222
cross elasticity of demand 81

D

deductive method 28
demand curve 50, 51, 115, 255
demand function 55
demand schedule 48
demerit goods 7
derived demand 61
diseconomies of scale 167, 170, 173
distribution 2, 13, 19, 22, 24, 28, 29, 36, 86, 108, 141
durable goods 6
dynamic theory of profit 148, 150

E

economic agents 13, 14, 16, 121
economic analysis 23, 28
economic costs 182, 183, 184, 215
economic goods 4
economic laws 26, 27
economic problems 1
economic profit 184, 185
economic resources 2
economic systems 28, 37
economic terminologies 2
economies of scale 152, 153, 167, 169, 179, 200, 234, 241
elasticity of demand 116, 117, 139
entrepreneurship 10, 16, 120, 141, 147, 154, 172, 174, 182, 183
environmental economics x, 259
exceptions of the law of demand 53
exchange 2, 3, 4, 5, 7, 9, 13, 122, 218, 219, 222
explicit cost 182, 215
externalities

F

factors of production 120, 165, 172
firm vi, x, 19, 21, 75, 85, 87, 88, 92, 103, 124, 126, 133, 137, 139, 140, 143, 144, 148, 151-153, 156, 157, 158, 164-178, 180, 189, 191-193, 196, 197, 198, 200-217, 222-239, 243, 245-253-257
free goods 3, 4, 5, 11, 31, 45

G

goods 2-8, 10-19, 29, 31, 36, 38, 45, 54, 55, 58, 59, 60-64, 75, 78, 80-87, 91, 95, 96, 102, 103, 107, 108, 111, 113, 115, 117, 121, 124, 126, 130, 131, 132, 142-148, 151-154, 172, 176, 177, 179, 182, 184, 202, 203, 214, 218-222, 223, 231, 238, 240, 254

H

human wants 5, 9, 11, 16, 29, 151, 172

I

implicit costs 182
income elasticity 75
inductive method 28
inputs 22, 93, 102, 113, 115, 120, 122, 130, 148, 153, 154, 156, 162, 163, 165, 166, 168, 172, 177, 179, 184, 186, 187, 191, 200, 215, 233, 241

J

joint supply 95

L

labour 10, 16, 19, 62, 63, 86, 94, 96, 102, 107, 120-122, 125- 132, 134,-142, 149, 152-158, 161-169, 172-187, 191-193, 196, 207, 208, 215, 216, 221, 241, 244
land viii, 4, 10, 13, 16, 96,

107, 120-124, 141-144, 149, 155, 162, 172, 182, 183, 184, 186, 221
law of demand 47, 48, 54, 63, 88, 110, 115, 124, 243
law of diminishing returns 161
law of supply 88
linear cost function 196
loanable fund theory of interest 145, 146
localisation of industries 180, 181
location of industries 178
long run cost-output relations 200
long-run period 165, 185, 186
luxuries 8
M
macroeconomics 19, 20
marginal product vi, 135, 136, 156, 158, 161, 162, 164, 166, 172, 173, 174, 191, 195, 202
marginal productivity theory of wage 135
marginal revenue 203- 207, 211, 212, 214, 230, 236, 245, 247, 257
market equilibrium 103, 104
market structure 221, 222, 242
merit goods 7
microeconomics i, x, 18, 20
mobility of labour 127
modern theory of rent 122, 124
modern theory of wages 142
monopolistic competition 242, 244
monopoly market 232, 255
N
nature of economics 1, 24
necessary good 74
necessities 8, 54
normal profit 183, 225, 230, 245, 246
normative economics 26
O
oligopoly market 248, 255
opportunity cost 30, 31, 183
P
perfectly competitive markets 221, 230, 231, 242, 255
perishable goods 3, 6
piece rate system 132
point elasticity of demand 66
pollution 23, 171, 181, 184, 185
positive economics 25
price ceiling 111
price controls 109
price discrimination 85, 238, 240
price elasticity of demand 64, 65, 67, 68, 69, 71, 72, 73, 74, 75, 77, 81, 82, 83, 85, 86, 87, 109, 116, 204, 205, 213, 239
price elasticity of supply 96, 97
price floor 110
price mechanism 107, 116
private cost 184, 215
private goods 3, 6, 7
producer goods 3, 6
production v, vi, viii, 2, 7, 10-23, 28, 32, 34, 36, 38, 45, 61, 87, 96, 102, 103, 107, 110, 113, 114, 120-126, 128-131, 133-137, 139, 141-145, 147-158, 161-192-208, 212, 214-217, 220, 221, 222, 224, 227, 230-234, 240-244, 247, 250, 252, 253, 255
production function 153-156, 158, 162, 165, 172, 174
production possibilities curve 32
profit sharing system 133
Q
Quasi-rent 124
R
rational choices 17, 22, 32
rent theory of profit 148, 149
residual claimant theory of wages 141
returns to scale 165
revenue analysis 202
risk bearing theory of profit 148, 150
S
scale of preference 30
scale of production 139, 151, 165, 167
scarcity 9, 12, 16, 28, 31, 36, 45, 121, 123, 143, 241
scarcity 29, 177
short run period 124, 155, 156, 162
short run supply curve 229
shut down point 227
social cost 184, 215
socialist economic system 37
stages of production 163
subsistence theory of wages 134
supernormal profit 224, 245
supply function 91, 92
supply schedule 88
T
time rate system 132
total product viii, 156, 161, 163, 174
trade union 137, 138, 139, 141
transfer earnings 124
transition period 43
U
uncertainty theory of profit 148, 150
utility 10, 12
W
wage differentials 141
wage fund theory 136
wage theory of profit 149
wants 2, 8, 9, 11, 12, 16, 17, 18, 28, 29, 30, 36, 249
wealth vi, 1, 2, 9- 15, 23, 29, 45, 142, 143, 254

Microeconomics for Advanced Secondary Schools Student's Book Form Five

Property of the Government of the United Republic of Tanzania, Not for Sale

