

BASICS OF ECONOMICS

Introduction

Dear learner! microeconomics deals with the behavior of individual economic units. Any individual or entity such as consumers, workers, firms etc that play a role in the functioning of our economy considered as an economic unit. Microeconomics explains how and why these economic units make economic decisions. It also explains how consumers and firms buy out puts and sale inputs and how their choices are affected by changing prices/costs and incomes/revenues. Consumers, workers, firms etc are interested to know causes of price and output instability as well as unemployment. Moreover, most of our issues and problems are related to economic matters. In general, an individual who does not understand basic economic principles will not appreciate and evaluate public issues that are most of them are related with economics. This unit is divided in to two sections. In the first section of the unit, you will see the *definitions and the central aims of economics*. The second section focuses on *the resource availability and its impact on the production and consumption of goods and services*.

Objectives

After successful completion of this unit, you will be able to:

- define what economics is and its purpose;
- explain the central aim of economics;
- explain how the problem of scarcity is the basic problem in any economy;
- discuss the problem of scarcity of factors of production and opportunity cost

1. The Concept of Economics

Overview

Dear learner! the nature and scope of the science of economics have dramatically grown and it has become very wide and vast. In this section, we will define economics from different angles like

scarcity, unlimited wants, choices etc. You will also try to understand the nature and scope of economics.

Objectives

After completing this section, you will be able to:

- Define the term economics;
- Explain the major divisions of Economics;
- Discuss the central aim of economics;
- Explain the method of economic analysis; and
- Explain the basic economic problems in terms of unlimited wants and scarce resources.

1.1 Definition of Economics

**Dear learner, what does it mean by Economics?
(Use the space left below to write your response)**

Economics is the study of efficient allocation of resources in order to attain the maximum fulfillment of unlimited human wants or needs. It is also defined as the study of how people make choices to cope with scarcity. Economics studies how people choose to use scarce resources to produce various commodities, how people consume goods and services and how they trade.

In the above definition, efficiency, unlimited human wants, scarce resources and choices are the key phrases. Thus, it is important to look at each one of them thoroughly.

Unlimited wants and scarce resources

Human beings want food, clothes, shelter and other variety of goods and services for their survival. These human wants are unlimited and increase from time to time; however, economic resources that include Land, Labor, Capital and Entrepreneurship are scarce by their nature. Although using these economic resources society produces what it wants, it will not be able to produce and consume all the goods and services it wants. Thus, economics describes various sets of tools that enable societies to use their scarce resources efficiently in order to achieve the highest possible standard of living.

1.2 The Central Aim of Economics

Dear learner! What is the aim of economics?

(Use the space left below to write your response)

Dear learner! As we have discussed above human wants are unlimited while they live in a world of scarce resources. *Scarcity* refers to a physical condition where the quantity desired of a particular resource exceeds the quantity available. Therefore, since the available resources are scarce, the ability of every society to produce goods and services are limited. The need to balance unlimited wants with limited resources has raised the question of efficient utilization of scarce resources. Therefore, the central aim of economics is the efficient use of the scarce resources by minimizing loss so as to get the maximum possible satisfaction. In the absence of scarcity, there will be no need of economizing. Therefore, the foundation of economics lies on the concepts of scarcity and choice (unlimited human wants). The field of economics bases itself on two fundamental facts. The following are facts that provide foundations for the field of economics:

1.3 Opportunity cost and making choice

Dear learner! What do you know about an opportunity cost?

(Use the space left below to write your response)

Opportunity cost is the single most important concept for making optimizing choices. An opportunity cost of any action is the best alternative forgone. The real opportunity cost of an action is measured in goods and services forgone, not in monetary units.

To make economic choices in an activity both marginal costs (the additional opportunity cost that can be incurred) and marginal benefits (the additional benefit that can be obtained) are considered. If marginal benefit exceeds the marginal cost, an individual prefers increasing the activity. Whenever the opportunity cost of an activity increases, individuals substitute other activities in its place. Thus, changes in marginal cost (opportunity cost) and marginal benefits change the incentives people face and change their actions.

1.4 Major Economic Problems Caused by Scarcity

Dear learner! As we discussed above many of our economic problems are raised due to scarce resources and unlimited human wants. This scarcity of resources created three major problems that every society faces. These economic problems are what, how and for whom to produce.

What to produce: It refers to those goods and services and the quantity of each that the economy should produce. Since resources are scarce or limited, an economy can not produce as much of every good and service as desired by all members of society. For this reason, more of one good or service means less of others. Therefore, every society must choose exactly which goods and services to produce and in what quantities. In other words, what to produce refers to the problem of allocation of scarce resource between their alternative uses.

How to produce: It refers to the choice of the combination of factors and the particular technique to use in producing a good or service. Different techniques of production can be used to produce goods and services. Even if resources are generally scarce, some resources may be relatively abundant than others in a country. For instance, in Ethiopia labor is relatively abundant than capital. If the country uses more of labor and less of capital it minimizes cost of production.

For whom to produce: This question refers to how the total output produced is to be divided among different consumers. In every economy, due to scarcity no nation is capable of satisfying all the needs of its society. As a result, the nation has to choose how to distribute the output. For example, in market economy the distribution of goods and services depends on the distribution of money income. That means, those who have more income can enjoy more of the goods and services and those who have less income can enjoy less of the goods and services.

1.5 Methods of Economic Analysis

Dear learner! How do economists solve the basic economic problems?

(Use the space left below to write your response)

In order to solve the basic economic problems (What, how and for whom to produce) economists design **policies** based up on **principles** or **theories** .This principles or theories can be derived from

facts. Economic theories/analysis are drawn from facts through induction (from particular to general) and deduction (from general to particular) methods.

1.6 Theories and Economic Models

Like other sciences economics is concerned with the explanation and prediction of observed phenomenon. Explanation and prediction of phenomenon is based on the use of theories.

Theory: A theory is developed to explain observed phenomenon based on a set of basic rules and assumptions. It is a framework that helps us to understand the relationship between cause and effect. It is simplification of an actual relationship and a hypothesis that has been successfully tested. It can be true in general or on average and is often subject to exceptions because of individual differences. For example, according to the theory of demand, other things remaining constant, when price of a product increases, the quantity demanded of the product will decrease. This is generally true. However, there may be some exceptional individuals who may not like to buy cheaper products and decide to stop buying any quantity of a product when its price decreases.

Dear learner! As we have discussed above, the objective of a theory is to predict and explain the cause of phenomena we observe. Thus, it simplifies, generalizes, predicts and explains the event. For instance, the theory of demand helps us to predict by how much the quantity demanded of a product will increase if its price falls by a certain amount and explains the reasons for such negative relationship.

Using application of statistical tools and econometric techniques, we can construct models from economic theories.

Economic Models: It is a skeletal and rough representation of the actual economy. In other words, it is a simplified representation of the real situation that is achieved by a set of meaningful and consistent assumption. A model can be represented by using graphs, mathematical equations, computer programs etc.

We use models in order to simplify our complex real world and to minimize the costs we incur while obtaining information. We evaluate models based on its assumptions, generality, simplicity and how well it predicts its phenomenon.

1.7 Microeconomics and Macroeconomics

Dear learner! Can you define microeconomics and macroeconomics?

(Use the space left below to write your response)

Economics is categorized on two broad categories. These are microeconomics and macroeconomics. Let us see each of these as follows:

Microeconomics: It is a branch of economic analysis of the economic behavior of individual decision making units such as individuals, households, business firms, industries or prices of different goods and services. In other words, microeconomics is concerned with the 'elements' of economic activity, the firm and the consumer. It refers to the study of economic motives and behavior of individual consumers and producers and the principals involved in organizing and operating the individual firms or industries.

Macroeconomics: It is a branch of economic analysis that examines the economy as a whole or its basic sub-divisions or aggregates such as the government, household and business sector. It is the study of the economy as a whole; of total saving, investment and employment in the system. It deals with the great aggregates and averages of the system rather than with particular parts of it and attempts to define these aggregates in a useful manner and to examine how they are related and determined.

1.8 Positive and Normative Economics

Positive economics deals with specific statements that are capable of verification by reference to the facts about economic behavior. It deals with facts or relationships which can be proven or disproven.

Examples of positive economic statements are:

- The 2000 fiscal year deficit of Ethiopia exceeded \$5 billion.
- When the value of Birr falls, imported products into our country become more expensive.
- If investment rises, national income will increase.

A **normative economics** is someone's opinion or value judgment about an economic issue. Such a statement can never be proven. It has a moral or ethical aspect and goes beyond a science can say.

Examples of normative economic statements are:

- The government should raise taxes and lower government spending to reduce the budget deficit.
- We need to try to lower the value of Birr in order to discourage the importation of foreign goods into this country.
- Families with income below birr \$3,500 per year should be exempted from income taxes.

Summary

Economics is the study of efficient allocation of scarce resources in order to attain the maximum fulfillment of unlimited human wants. *Scarcity* refers to a physical condition where the quantity desired of a particular resource exceeds the quantity available. This scarcity of resources created three major problems that every society faces. These economic problems are what, how and for whom to produce.

Moreover, since the available resources are scarce, the ability of every society to produce goods and services are limited. The need to balance unlimited wants with limited resources has raised the question of efficient utilization of scarce resources. Therefore, the central aim of economics is the efficient use of the scarce resources by minimizing loss so as to get the maximum possible satisfaction

A **theory** is a framework that helps us to understand cause and effect relationships. It is a simplification of actual relationships. It is developed to explain observed phenomenon based on a set of basic rules and assumptions.

Economics is categorized on two broad categories. These are **microeconomics** and **macroeconomics**. Microeconomics is a branch of economic analysis of the economic behavior of individual decision making units such as individuals, households, business firms etc while Macroeconomics is the study of the economy as a whole.

Positive economics and **normative** economics are the two approaches in economic analysis. Positive economics deals with specific statements that are capable of verification by reference to the facts about economic behavior. In other words, it is an economic analysis that provides statements about “what is”, “what was” or “what will be ” rather than “what should be”. Normative Economics is someone’s opinion or value judgment about an economic issue. In other words, it is concerned with “what ought to be” or “what should be” done about the economy.

UNIT TWO

THEORY OF CONSUMER BEHAVIOR AND DEMAND

Introduction

The theory of consumer choice lies on the assumption of the consumer being rational to maximize level of satisfaction. The consumer makes choices by comparing bundle of goods. There are two approaches to analyze consumer's decision making process. These are, the cardinal and ordinal utility approaches.

Objectives

After completing this unit, you will be able to:

- explain the concept of cardinal and ordinal utility approach;
- state effects of changes in money income and price on equilibrium;
- identify the difference between the indifference curve and indifference map;
- discuss the properties of indifference curve;
- derive the budget line
- describe and show slusky's decomposition;
- state and explain consumer and producer surplus;
- derive the market demand;

2. Consumer Preferences and Choices

Overview

Dear learner, in this section you will see how consumers allocate their limited income among different number of goods and services Moreover, you will learn how consumers allocation decisions determine quantity demand of goods and services.

After completing this section, you will be able to:

- Describe the theory of consumer preference an choice
- Explain the assumptions of consumer preference

2.1 Consumer Preference

Dear learner, given any two consumption bundles (groups of goods) available for purchase, how a consumer compares the goods? Does he prefer one good to another, or does he indifferent between the two groups.

Given any two consumption bundles, the consumer can either decide that one of consumption bundles is strictly better than the other, or decide that he is indifferent between the two bundles.

Strict preference: Given any two consumption bundles (X_1, X_2) and (Y_1, Y_2) , if $(X_1, X_2) > (Y_1, Y_2)$ or if he chooses (X_1, X_2) when (Y_1, Y_2) is available the consumer definitely wants the X-bundle than Y.

Weak preference: Given any two consumption bundles (X_1, X_2) and (Y_1, Y_2) , if the consumer is indifferent between the two commodity bundles or if $(X_1, X_2) \geq (Y_1, Y_2)$, the consumer would be equally satisfied if he consumes (X_1, X_2) or (Y_1, Y_2) .

Completeness: For any two commodity bundles X and Y, a consumer will prefer X to Y, Y to X or will be indifferent between the two.

Transitivity: It means that if a consumer prefers basket A to basket B and to basket C, then the consumer also prefers A to C.

More is better than less: Consumers always prefer more of any good to less and they are never satisfied or satiated. However, bad goods are not desirable and consumers will always prefer less of them.

2.2 Utility

Dear learner economists use the term utility to describe the satisfaction or enjoyment derived from the consumption of a good or service.

Definition

Utility is the level of satisfaction that is obtained by consuming a commodity or undertaking an activity.

In defining strict preference, we said that given any two consumption bundles (X_1, X_2) and (Y_1, Y_2) , the consumer definitely wants the X bundle than the Y bundle if $(X_1, X_2) > (Y_1, Y_2)$. This means, the consumer preferred bundle (X_1, X_2) to bundle (Y_1, Y_2) if and only if the utility (X_1, X_2) is larger than the utility of (Y_1, Y_2) .

The concept of utility is characterized with the following properties:

- ‘Utility’ and ‘Usefulness’ are not synonymous. For example, paintings by Picasso may be useless functionally but offer great utility to art lovers.
- Utility is subjective. The utility of a product will vary from person to person. That means, the utility that two individuals derive from consuming the same level of a product may not be the same. For example, no-smokers do not derive any utility from cigarettes.
- The utility of a product can be different at different places and time. For example, the utility that we get from meat during fasting is not the same as any time else.

A Consumer considers the following points to get maximum utility or level of satisfaction:

- How much satisfaction he gets from buying and then consuming an extra unit of a good or service.
- The price he pays to get the good.
- The satisfaction he gets from consuming alternative products.
- The prices of alternative goods and services.

Dear learners, how do you measure the satisfaction level (Utility) that you get from goods and services?

2.3 Approaches to measure Utility

There are two major approaches of measuring utility. These are Cardinal and ordinal approaches. This sub unit is divided into two Sections. In Section one the Cardinal utility approach will be discussed while in Section two the concept of ordinal Utility will be addressed.

The Cardinal Utility theory

Overview

Dear learner, Utility maximization theories are important to deal with consumer behavior. Thus, in this section, you will learn about the Cardinal Utility theory. Neo classical economists argued that utility is measurable like weight, height, temperature and they suggested a unit of measurement of satisfaction called utils. A util is a cardinal number like 1,2,3 etc simply attached to utility. Hence, utility can be quantitatively measured.

Objectives

After completing your study on this section, you will be able to:

- explain the concept of cardinal approach;
- explain the assumptions of the cardinal approach; and
- Determine utility maximization of the consumer.

2.4 Assumptions of Cardinal Utility theory

1. **Rationality of Consumers.** The main objective of the consumer is to maximize his/her satisfaction given his/her limited budget or income. Thus, in order to maximize his/her satisfaction, the consumer has to be rational.
2. **Utility is Cardinally Measurable.** According to this approach, the utility or satisfaction of each commodity is measurable. Money is the most convenient measurement of utility. In other words, the monetary unit that the consumer is prepared to pay for another unit of commodity measures utility or satisfaction.
3. **Constant Marginal Utility of Money.** According to assumption number two, money is the most convenient measurement of utility. However, if the marginal utility of money changes with the level of income (wealth) of the consumer, then money can not be considered as a measurement of utility.

4. **Limited Money Income.** The consumer has limited money income to spend on the goods and services he/she chooses to consume.
5. **Diminishing Marginal Utility (DMU).** The utility derived from each from each successive units of a commodity diminishes. In other words, the marginal utility of a commodity diminishes as the consumer acquires larger quantities of it.
6. The total utility of a basket of goods depends on the quantities of the individual commodities.

If there are n commodities in the bundle with quantities, X_1, X_2, \dots, X_n the total utility is given by:

$$TU = f(X_1, X_2, \dots, X_n)$$

2.5 Total and Marginal Utility

Definitions

Total Utility (TU): It refers to the total amount of satisfaction a consumer gets from consuming or possessing some specific quantities of a commodity at a particular time. As the consumer consumes more of a good per time period, his/her total utility increases. However, there is a saturation point for that commodity in which the consumer will not be capable of enjoying any greater satisfaction from it.

Marginal Utility (MU): It refers to the additional utility obtained from consuming an additional unit of a commodity. In other words, marginal utility is the change in total utility resulting from the consumption of one or more unit of a product per unit of time. Graphically, it is the slope of total utility.

Mathematically, the formula for marginal utility is:

$$MU = \frac{\Delta TU}{\Delta Q} \quad \text{Where,} \quad \Delta TU \text{ is the change in Total Utility, and,}$$

ΔQ is change in the amount of product consumed.

2.6 Law of diminishing marginal Utility (LDMU)

Dear learner, is the utility you get from consumption of the first orange is the same as the second orange?

The utility that a consumer gets by consuming a commodity for the first time is not the same as the consumption of the good for the second, third, fourth, etc.

The Law of Diminishing Marginal Utility States that as the quantity consumed of a commodity increases per unit of time, the utility derived from each successive unit decreases, consumption of all other commodities remaining constant.

The LDMU is best explained by the MU curve that is derived from the relationship between the TU and total quantity consumed.

Equilibrium of a consumer

A consumer that maximizes utility reaches his/her equilibrium position when allocation of his/her expenditure is such that the last birr spent on each commodity yields the same utility.

For example, if the consumer consumes a bundle of n commodities i.e x_1, x_2, \dots, x_n , he/she would be in equilibrium or utility is maximized if and only if:

$$\frac{MU_{x_1}}{P_{x_1}} = \frac{MU_{x_2}}{P_{x_2}} = \dots = \frac{MU_{x_n}}{P_{x_n}} = MU_m \text{ Where: } MU_m \text{ –marginal utility of money}$$

Diagrammatically,

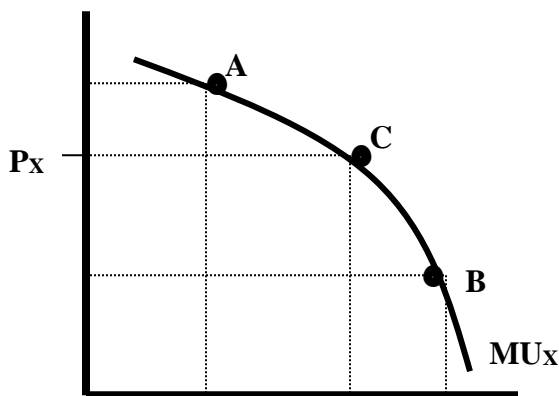


Figure 2.2 marginal utility of a consumer

Note that: at any point above point C like point A where $MUX > Px$, it pays the consumer to consume more. At any point below point C like point B where $MUX < Px$ the consumer consumes less of X. However, at point C where $MUX = Px$ the consumer is at equilibrium.

Table 2.2 Utility schedule for a single commodity

Quantity of Orange	Total utility	Marginal utility	Marginal utility per Birr (price=2 birr)	Marginal utility of money
0	0	-	-	1
1	6	6	3	1
2	10	4	2	1
3	12	2	1	1
4	13	1	0.5	1
5	13	0	0	1
6	11	-2	-1	1

For consumption level lower than three quantities of oranges, since the marginal utility of orange is higher than the price, the consumer can increase his/her utility by consuming more quantities of oranges. On the other hand, for quantities higher than three, since the marginal utility of orange is lower than the price, the consumer can increase his/her utility by reducing its consumption of oranges.

Mathematically, the equilibrium condition of a consumer that consumes a single good X occurs when the marginal utility of X is equal to its market price.

$$MU_x = P_x$$

Proof

The utility function is:

$$U = f(X)$$

If the consumer buys commodity X, then his expenditure will be $Q_x P_x$. Thus, the consumer wants to maximize the difference between his/her utility and expenditure

$$\text{Max}(U - Q_x P_x)$$

The necessary condition for maximization is equating the derivative of a function with zero.

Thus, X

$$\frac{dU}{dQ_x} - \frac{d(Q_x P_x)}{dQ_x} = 0$$

$$\frac{dU}{dQ_x} - P_x = 0 \Rightarrow MU_x = P_x$$

Table 2.3 Utility schedule for two commodities

Orange, Price=2birr				Banana, Price=4birr			
Quantity	TU	MU	MU/P	Quantity	TU	MU	MU/P
0	0	-	-	0	0	-	-
1	6	6	3	1	6	6	6
2	10	4	2	2	22	16	4
3	12	2	1	3	32	12	3
4	13	1	0.5	4	40	8	2
5	13	0	0	5	45	5	1.85

6	11	-2	-1	6	48	3	0.75
----------	-----------	-----------	-----------	----------	-----------	----------	-------------

Dear learner, as we discussed earlier, utility is maximized when the condition of marginal utility of one commodity divided by its market price is equal to the marginal utility of the other

commodity divided by its market price **MU i.e.** $\frac{MU_1}{P_1} = \frac{MU_2}{P_2}$

Thus, the consumer will be at equilibrium when he consumes 2 quantities of Orange and 4

quantities of banana, because $\frac{MU_{orange}}{P_{orange}} = \frac{MU_{banana}}{P_{banana}} = \frac{4}{2} = \frac{8}{4} = 2$

Limitation of the Cardinalist approach

The Cardinalist approach involves the following three weaknesses:

1. The assumption of cardinal utility is doubtful because utility may not be quantified.
2. Utility can not be measured absolutely (objectively). The satisfaction obtained from different commodities can not be measured objectively.
3. The assumption of constant MU of money is unrealistic because as income increases, the marginal utility of money changes.

The Ordinal Utility Theory

Overview

Dear learner, in the previous section, we have discussed one of the approach for measurement of utility that is cardinal utility approach. In this section, we will discuss the second approach that is the ordinal utility approach.

In the ordinal utility approach, utility cannot be measured absolutely but different consumption bundles are ranked according to preferences. The concept is based on the fact that it may not be possible for consumers to express the utility of various commodities they consume in absolute terms, like, 1 util, 2 util, or 3 util, but it is always possible for the consumers to express the utility in relative terms. It is practically possible for the consumers to rank commodities in the order of their preference as $1^{st} 2^{nd} 3^{rd}$ and so on.

Objectives

After completing this section you will be able to:

- Describe the assumption of ordinal utility approach
- Define an indifference curve and a budget line
- Identify the properties of indifference curve;
- Derive an income-consumption curve and price-consumption curve.
- Derive the income and substitution effects.
- Distinguish the difference between substitution and income effects

2.7 Assumptions of Ordinal Utility theory

Dear learner, like the previous approach, this approach is based on the following assumptions:

1. The Consumers are rational-they aim at maximizing their satisfaction or utility given their income and market prices.
2. Utility is ordinal, i.e. utility is not absolutely (cardinally) measurable. Consumers are required only to order or rank their preference for various bundles of commodities.
3. Diminishing Marginal Rate of Substitution (MRS): The marginal rate of substitution is the rate at which a consumer is willing to substitute one commodity (x) for another commodity (y) so that his total satisfaction remains the same. When a consumer continues to substitute X for Y the rate goes decreasing and it is the slope of the Indifference curve.
4. The total utility of the consumer depends on the quantities of the commodities consumed, i.e., $U=f(X_1, X_2, \dots, X_n)$
5. Preferences are transitive or consistent:
 - It is transitive in the senses that if the consumer prefers market basket X to market basket Y, and prefers Y to Z, and then the consumer also prefers X to Z.
 - When we said consistent it means that If market basket X is greater than market basket Y ($X > Y$) then Y not greater than X ($Y \not > X$).

The ordinal utility approach is expressed or explained with the help of indifference curves. An indifference curve is a concept used to represent an ordinal measure of the tastes and preferences of the consumer and to show how he/she maximizes utility in spending income. Since it uses ICs to study the consumer's behavior, the ordinal utility theory is also known as the Indifference Curve Analysis.

2.8 Indifference Set, Curve and Map

Dear learner, can you give your own definition on the indifference curve?

(You can use the space left below to write your response)

Indifference Set/ Schedule: It is a combination of goods for which the consumer is indifferent, preferring none of any others. It shows the various combinations of goods from which the consumer derives the same level of utility.

Table 2.4 Indifference Schedule

Bundle (Combination)	A	B	C	D
Orange(X)	1	2	4	7
Banana (Y)	10	6	3	1

Each combination of good X and Y gives the consumer equal level of total utility. Thus, the individual is indifferent whether he consumes combination A, B, C or D.

Indifference Curves: an indifference curve shows the various combinations of two goods that provide the consumer the same level of utility or satisfaction. It is the locus of points (particular combinations or bundles of good), which yield the same utility (level of satisfaction) to the consumer, so that the consumer is indifferent as to the particular combination he/she consumes.

By transforming the above indifference schedule into graphical representation, we get an indifference curve.

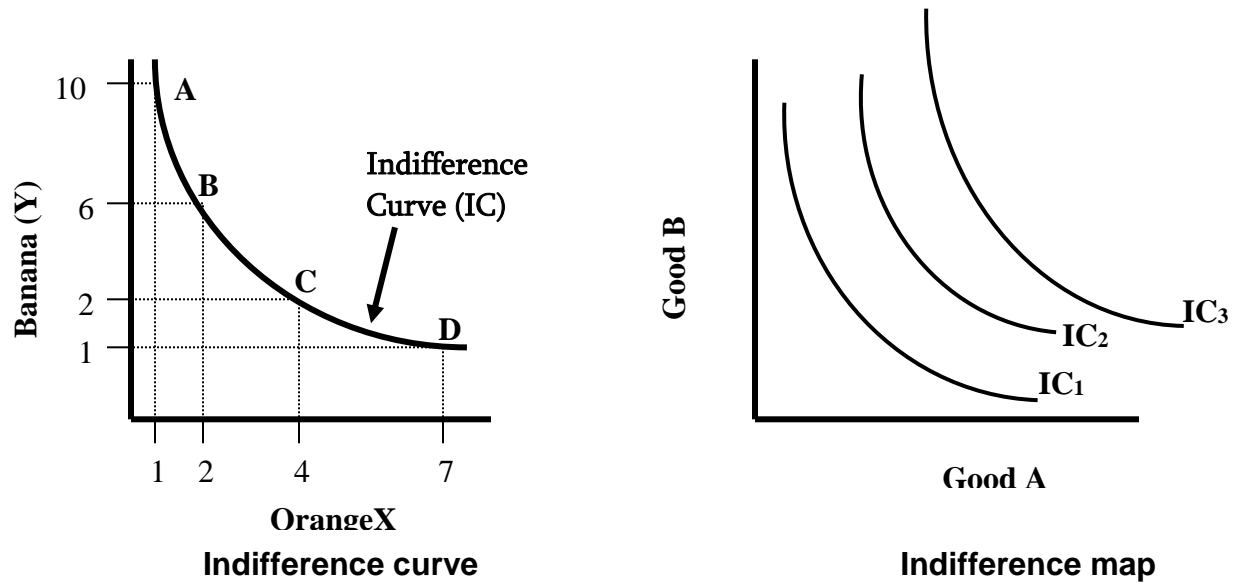


Fig2.4 indifference curves and indifference map.

Indifference Map: To describe a person's preferences for all combinations potato and meat, we can graph a set of indifference curves called an indifference map. In other words it is the entire set of indifference curves is known as an indifference map, which reflects the entire set of tastes and preferences of the consumer. A higher indifference curve refers to a higher level of satisfaction and a lower indifference curve shows lesser satisfaction. IC2 reflects higher level of utility than that of IC1. Any consumer has lots of indifference curves, not just one.

Properties of Indifference Curves:

Indifference curves have certain unique characteristics with which their foundation is based.

1. Indifference curves have negative slope (downward sloping to the right). Indifference curves are negatively sloped because the consumption level of one commodity can be increased only by reducing the consumption level of the other commodity. That means,

if the quantity of one commodity increases with the quantity of the other remaining constant, the total utility of the consumer increases. On the other hand, if the quantity of one commodity decreases with the quantity of the other remaining constant, the total utility of the consumer reduces. Hence, in order to keep the utility of the consumer constant, as the quantity of one commodity is increased, the quantity of the other must be decreased.

2. Indifference curves do not intersect each other. Intersection between two indifference curves is inconsistent with the reflection of indifference curves. If they did, the point of their intersection would mean two different levels of satisfaction, which is impossible.
3. A higher Indifference curve is always preferred to a lower one. The further away from the origin an indifferent curve lies, the higher the level of utility it denotes: baskets of goods on a higher indifference curve are preferred by the rational consumer, because they contain more of the two commodities than the lower ones.
4. Indifference curves are convex to the origin. This implies that the slope of an indifference curve decreases (in absolute terms) as we move along the curve from the left downwards to the right. This assumption implies that the commodities can substitute one another at any point on an indifference curve, but are not perfect substitutes.

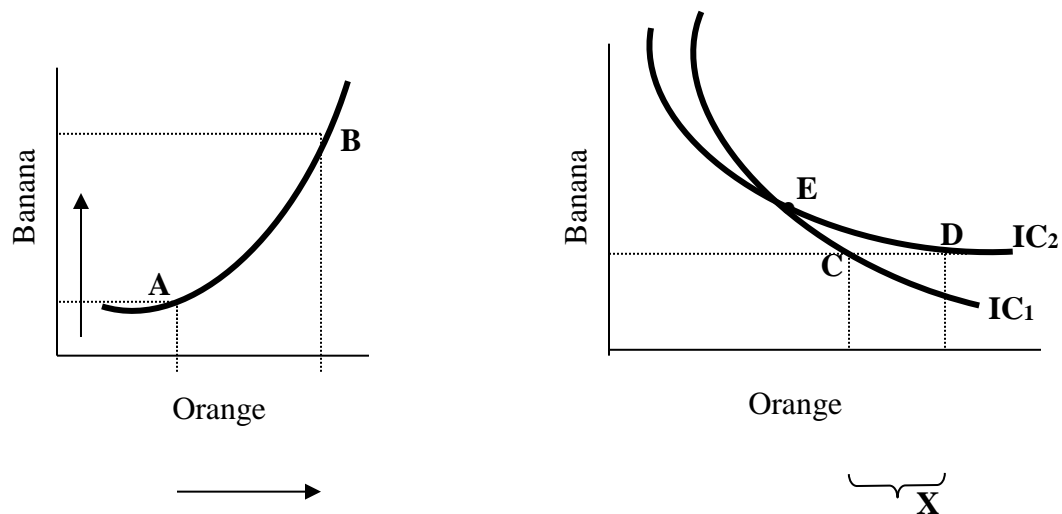


Fig.2.5 positively sloped and intersected indifference curves

Dear learner, as we discussed earlier, Indifference curves cannot intersect each other. If they did, the consumer would be indifferent between C and E, (Right panel of figure 2.6) since both are on indifference curve one (IC1). Similarly, the consumer would be indifferent between points D and E, since they are on the same indifference curve, IC2. By transitivity, the consumer must also be indifferent between C and D. However, a rational consumer would prefer D to C because he/she can have more Orange at point D (more Orange by an amount of X).

2.9 The Marginal rate of substitution (MRS)

Dear learner, how do you perceive the concept of marginal rate of substitution?

(You can use the space left below to write your response)

Dear learner, to quantify the amount of one good that a consumer will give up to obtain more of another, we often use marginal rate of substitution as a measurement (MRS).

Definition: Marginal rate of substitution of X for Y is defined as the number of units of commodity Y that must be given up in exchange for an extra unit of commodity of X so that the consumer maintains the same level of satisfaction.

$$MRS_{x,y} = \frac{\text{Number of units of Y given up}}{\text{Number of units of X gained}}$$

It is the negative of the slope of an indifference curve at any point of any two commodities such as X and Y, and is given by the slope of the tangent at that point:

i.e., Slope of indifference curve

$$\frac{\Delta y}{\Delta x} = MRS_{x,y}$$

In other words, MRS refers to the amount of one commodity that an individual is willing to give up to get an additional unit of another good while maintaining the same level of satisfaction or

remaining on the same indifference curve. The diminishing slope of the indifference curve means the willingness to substitute X for Y diminishes as one move down the curve.

Note that ($MRS_{X,Y}$) measures the downward vertical distance (the amount of y that the individual is willing to give up) per unit of horizontal distance (i.e. per additional unit of x required) to remain on the same indifference curve. That is, $MRS_{X,Y} = -\frac{\Delta Y}{\Delta X}$ because of the reduction in Y, MRS is negative. However, we multiply by negative one and express $MRS_{X,Y}$ as a positive value.

The rationale behind the convexity, that is, diminishing MRS, is that a consumer's subjective willingness to substitute A for B (or B for A) will depend on the amounts of B and A he/she possesses.

Table 2.5 level of consumption of good X and Y

Bundle (Combination)	A	B	C	D
Orange(X)	1	2	4	7
Banana (Y)	10	6	3	1

$$MRS_{X,Y}(\text{between points A and B}) = \frac{\Delta Y}{\Delta X} = \frac{4}{1} = 4$$

Dear learner, in the above case the consumer is willing to forgo 4 units of Banana to obtain 1 more unit of Orange. If the consumer moves from point B to point C, he is willing to give up only 2 units of Banana(Y) to obtain 1 unit of Orange (X), so the MRS is 2($\Delta Y/\Delta X = 4/2$). Having still less of Banana and more of Orange at point D, the consumer is willing to give up only 1 unit of Banana so as to obtain 3 units of Orange. In this case, the MRS falls to $\frac{1}{3}$. In general, as the amount of Y increases, the marginal utility of additional units of Y decreases. Similarly, as the quantity of X

decreases, its marginal utility increases. In addition, the MRS decreases as one move downwards to the right.

Marginal Utility and Marginal rate of Substitution

Dear learner, it is also possible to show the derivation of the MRS using MU concepts. The $MRS_{x,y}$ is related to the MU_x and the MU_y is:

$$MRS_{x,y} = \frac{MU_x}{MU_y}$$

Proof:

Suppose the utility function for two commodities X and Y is defined as:

$$U = f(X, Y)$$

Since utility is constant on the same indifference curve:

$$U = f(X, Y) = C$$

The total differential of the utility function is:

$$dU = \frac{\partial U}{\partial X} dX + \frac{\partial U}{\partial Y} dY = 0$$

$$MU_x dX + MU_y dY = 0$$

$$\frac{MU_x}{MU_y} = -\frac{dY}{dX} = MRS_{x,y}$$

Or,

$$\frac{MU_y}{MU_x} = -\frac{dX}{dY} = MRS_{y,x}$$

Example

Suppose a consumer's utility function is given by $U = 5\frac{X^4}{Y^{-2}}$. Compute the $MRS_{X,Y}$.

$$MRS_{x,y} = \frac{MU_x}{MU_y}$$

$$MU_x = \frac{dU}{dX} \quad \text{and} \quad MU_y = \frac{dU}{dY}$$

Therefore, $MU_x = 4(X^{4-1}Y^2) = 4(X^3Y^2)$ and $MU_y = 2(X^4Y^{2-1}) = 2X^4Y$

$$MRS_{x,y} = \frac{MU_x}{MU_y} = \frac{4X^3Y^2}{2X^4Y} = 2\frac{Y}{X}$$

Special Indifference Curves

Dear learner, convexity or down ward sloping is among the characteristics of indifference curve and this shape of indifference curve is for most goods. In this situation, we assume that two commodities such as x and y can substitute one another to a certain extent but are not perfect substitutes. However, the shape of the indifference curve will be different if commodities have some other unique relationship such as perfect substitution or complementary.

Here, are some of the ways in which indifference curves/maps might be used to reflect preferences for three special cases.

I. Perfect substitutes: If two commodities are perfect substitutes (if they are essentially the same), the indifference curve becomes a straight line with a negative slope. MRS for perfect substitutes is constant. (Panel a)

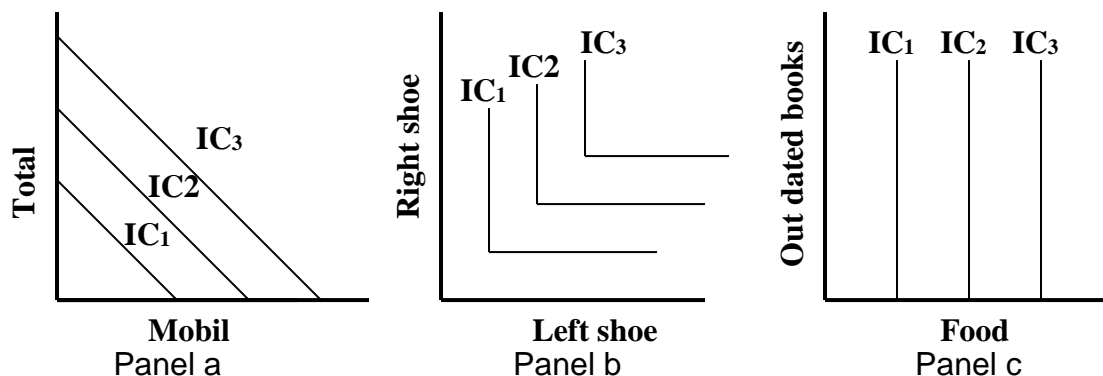


Fig.2.6 Special cases of indifference curves

II. Perfect complements: If two commodities are perfect complements the indifference curve takes the shape of a right angle. Suppose that an individual prefers to consume left shoes (on the horizontal axis) and right shoes on the vertical axis in pairs. For example, if an individual has two

pairs of shoes, additional right or left shoes provide no more utility for him/her. MRS for perfect complements is zero (both MRS_{XY} and MRS_{YX} is the same, i.e. zero).

III.A useless good: Panel C in the above figure shows an individual's indifference curve for food (on the horizontal axis) and an out-dated book, a useless good, (on the vertical axis). Since they are totally useless, increasing purchases of out-dated books does not increase utility. This person enjoys a higher level of utility only by getting additional food consumption. For example, the vertical indifference curve IC_2 shows that utility will be IC_2 as long as this person has some units of food no matter how many out dated books he/she has.

2.10 The Budget Line or the Price line

**Dear learner, how do you explain the idea of the budget line?
(You can use the space left below to write your response)**

Dear learner, indifference curves only tell us about the consumer's preferences for any two goods but they can not tell us which combinations of the two goods will be chosen or bought.

In reality, the consumer is constrained by his/her money income and prices of the two commodities. Therefore, in addition to consumer preferences, we need to know the consumer's income and prices of the goods. In other words, individual choices are also affected by budget constraints that limit people's ability to consume in light of prices they must pay for various goods and services. Whether or not a particular indifference curve is attainable depends on the consumer's money income and on commodity prices. A consumer while maximizing utility is constrained by the amount of income and prices of goods that must be paid. This constraint is often presented with the help of the budget line constructing by alternative purchase possibilities of two goods. Therefore, before we discuss consumer's equilibrium, it is better to understand his/ her budget line.

The budget line is a line or graph indicating different combinations of two goods that a consumer can buy with a given income at a given prices. In other words, the budget line shows the market basket that the consumer can purchase, given the consumer's income and prevailing market prices.

Assumptions for the use of the budget line

In order to draw the budget line facing the consumer, we consider the following assumptions:

1. there are only two goods, X and Y, bought in quantities X and Y;
2. each consumer is confronted with market determined prices, P_x and P_y , of good X and good Y respectively; and
3. the consumer has a known and fixed money income (M).

By assuming that the consumer spends all his/her income on two goods (X and Y), we can express the budget constraint as:

$$M = P_x X + P_y Y \quad \text{Where, } P_x = \text{price of good X}$$

$P_y = \text{price of good Y}$

$X = \text{quantity of good X}$

$Y = \text{quantity of good Y}$

$M = \text{consumer's money income}$

This means that the amount of money spent on X plus the amount spent on Y equals the consumer's money income.

Suppose for example a household with 30 Birr per day to spend on banana(X) at 5 Birr each and Orange(Y) at 2 Birr each. That is, $P_x = 5, P_y = 2, M = 30 \text{ birr}$.

Therefore, our budget line equation will be:

$$5X + 2Y = 30$$

Table 2.6 Alternative purchase possibilities of the two goods

Consumption Alternatives	A	B	C	D	E	F
Kgs of banana (X)	0	1	2	3	4	6
Kgs of Orange(Y)	15	12.5	10	7.5	5	0
Total Expenditure	30	30	30	30	30	30

At alternative A, the consumer is using all of his /her income for good Y. Mathematically it is the y-intercept (0, 15). And at alternative F, the consumer is spending all his income for good X. mathematically; it is the x-intercept (6, 0). We may present the income constraint graphically by the budget line whose equation is derived from the budget equation.

$$M = P_x X + P_y Y$$

$$M - XP_x = YP_y$$

By rearranging the above equation we can derive the general equation of a budget line,

$$Y = \frac{M}{P_y} - \frac{P_x}{P_y} X$$

$$\frac{M}{P_y} = \text{Vertical Intercept (Y-intercept), when } X=0.$$

$$-\frac{P_x}{P_y} = \text{slope of the budget line (the ratio of the prices of the two goods)}$$

The horizontal intercept (i.e., the maximum amount of X the individual can consume or purchase given his income) is given by:

$$\frac{M}{P_y} - \frac{P_x}{P_y} X = 0$$

$$\frac{M}{P_Y} = \frac{P_X}{P_Y} X$$

$$X = \frac{M}{P_X}$$

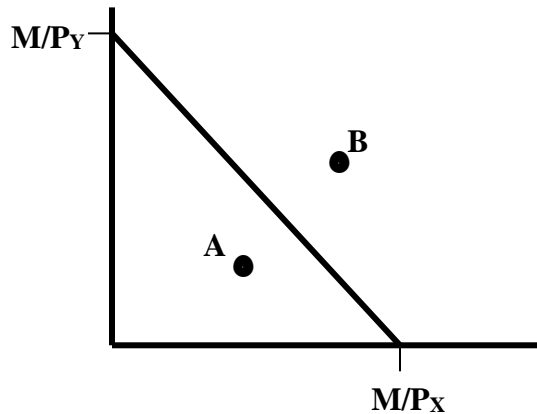


Fig.2.7 Derivation of the Budget Line

Therefore, the budget line is the locus of combinations or bundle of goods that can be purchased if the entire money income is spent.

2.11 Optimum of the Consumer

Dear learner, a rational consumer seeks to maximize his utility or satisfaction by spending his or her income. It maximizes the utility by trying to attain the highest possible indifference curve, given the budget line. This occurs where an indifference curve is tangent to the budget line so that the slope of the indifference curve (MRS_{XY}) is equal to the slope of the budget line (P_X / P_Y).

Thus, the condition for utility maximization, consumer optimization, or consumer equilibrium occurs where the consumer spends all income (i.e. he/she is on the budget line) and the slope of the indifference curve equals to the slope of the budget line $MRS_{XY} = P_X / P_Y$.

The preferences of the consumer (what he/she wishes) are indicated by the indifference curve and the budget line specifies the different combinations of X and Y the consumer can purchase with the limited income. Therefore, the consumer tries to obtain the highest possible satisfaction with in his budget line.

However, the consumer cannot purchase any bundle lying above and to the right of the budget line. Because Indifference curves above the region of the budget line are beyond the reach of the consumer and are irrelevant for equilibrium consideration. The question then arises as to which combinations of X and Y the rational consumer will purchase.

Graphically, the consumer optimum or equilibrium is depicted as follows:

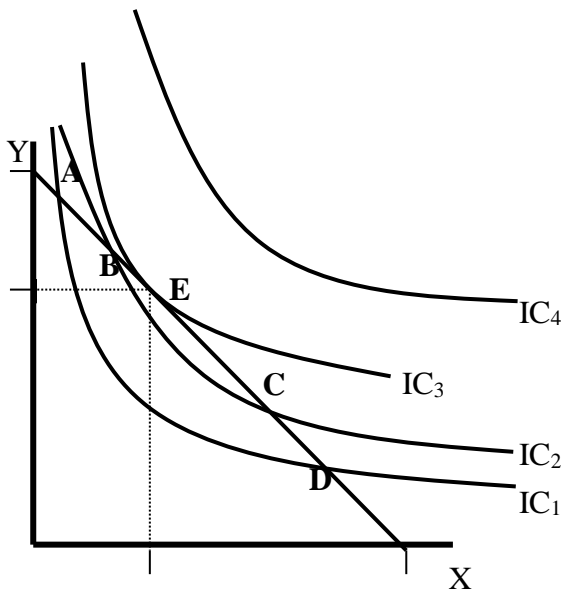


Figure2.14 Consumer equilibrium

At point 'A' on the budget line, the consumer gets IC₁ level of satisfaction. When he/she moves down to point 'B' by reallocating his total income in favor of X he/she derives greater level of satisfaction that is indicated by IC₂. Thus, point 'B' is preferred to point 'A'. Moving further down to point 'E', the consumer obtains the greatest level of satisfaction (IC₃) relative to other indifference curves.

Therefore, point 'E' (which represents combination X and Y) is the most preferred position by the consumer since he/she attains the highest level of satisfaction within his/her reach and point 'E' is known as the point of consumer equilibrium (or consumer optimum). This equilibrium occurs at the point of tangency between the highest possible indifference curve and the budget line. Put differently, equilibrium is established at the point where the slope of the budget line is equal to the slope of the indifference curve.

Mathematically, consumer optimum (equilibrium) is attained at the point where:

$$MRS_{XY} = \frac{P_X}{P_Y}, \text{ But we know } \frac{MU_X}{P_X} = \frac{MU_Y}{P_Y} = \dots\dots MU_X P_Y = MU_Y P_X \dots\dots, \frac{MU_X}{MU_Y} = \frac{P_X}{P_Y}$$

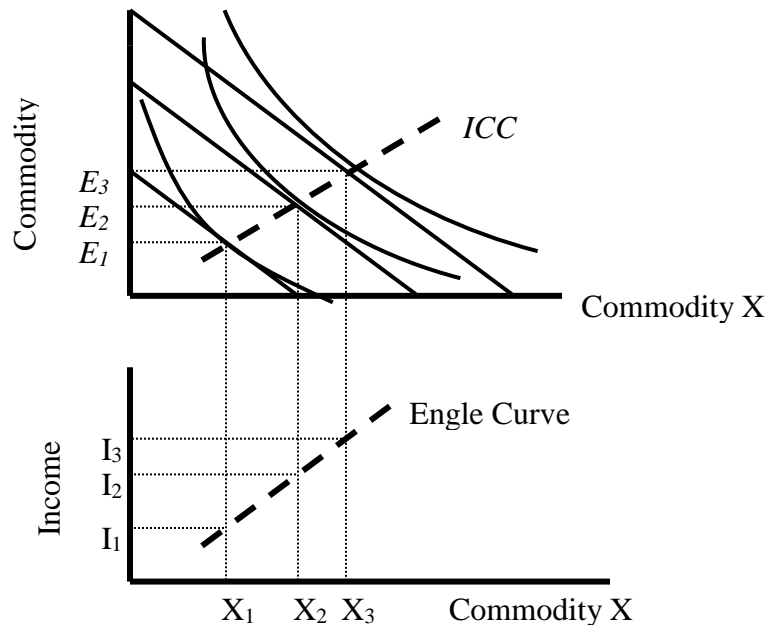
2.12 Effects of Changes in Income and Prices on Consumer

Dear learner, let us now analyze the effect of changes in consumer's income and the price of the good that are the two important determinants of quantity demanded (or also consumer equilibrium). Let us first consider the effect of change in income on the equilibrium of the consumer all other things remaining constant.

A. Changes In Income: Income Consumption Curve and the Engel Curve

In our previous discussion, we noted that an increase in the consumer's income (all other things held constant) results in an upward parallel shift of the budget line. This allows the consumer to buy more of the two goods. And when the consumer's income falls, *ceteris paribus*, the budget line shifts downward, remaining parallel to the original one.

If we connect all of the points representing equilibrium market baskets corresponding to all possible levels of money income, the resulting curve is called the ***Income consumption curve (ICC)*** or ***Income expansion curve (IEC)***. The Income Consumption Curve is a curve joining the points of consumer optimum (equilibrium) as income changes (*ceteris paribus*). Or, it is the locus of consumer equilibrium points resulting when only the consumer's income varies.



From the Income Consumption Curve we can derive the Engle Curve. The Engle curve is named after Ernest Engel, the German statistician who first derived family budgets and expenditure positions.

Figure 2.15 the income –consumption and the Engle curves

The Engle Curve is the relationship between the equilibrium quantity purchased of a good and the level of income. It shows the equilibrium (utility maximizing) quantities of a commodity, which a consumer will purchase at various levels of income; (ceteris paribus) per unit of time.

In relation to the shape of the income-consumption and Engle curves goods can be categorized as normal (superior) and inferior goods. Thus, commodities are said to be normal, when the income consumption curve and its Engle curve are positively sloped; meaning that more of the goods are purchased at higher levels of income. On the other hand, commodities are said to be inferior when the income consumption curve and Engle curve is negatively sloped, i.e. their purchase decreases when income increases.

B. Changes in Price: Price Consumption Curve (PCC) and Individual Demand Curve

We now look at the second factor that affects the equilibrium of the consumer that is price of the goods. The effect of price on the consumption of good is even more important to economists than the effect of changes in income. Here, we hold money income constant and let price change to analyze the effect on consumer behavior.

In our earlier previous discussion, we have seen that an increase in the price of good X, for example, increases the absolute value of the slope of the budget line, but it does not affect the vertical (Y) intercept of the line. Thus, the change in the price of x will result in out ward shift of the budget line that makes the consumer to buy more of good x. If we connect all the points representing equilibrium market baskets corresponding to each price of good X we get a curve called *price-consumption curve*.

The price-consumption curve is the locus of the utility-maximizing combinations of products that result from variations in the price of one commodity when other product prices, the money income and other factors are held constant.

We can derive the demand curve of an individual for a commodity from the price consumption curve. Below is an illustration of deriving the demand curve when price of commodity X decreases from P_{x_1} to P_{x_2} to P_{x_3} .

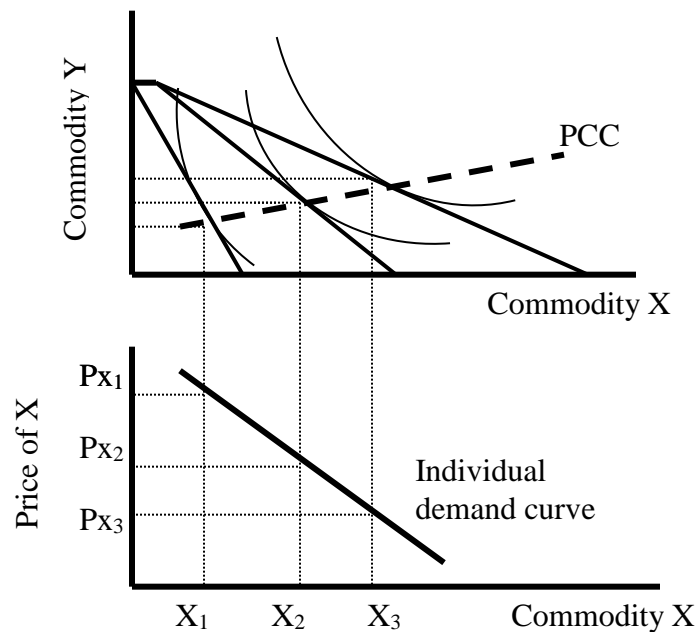


Figure 2.17 the PPC and derivation of the demand curve

Mathematical derivation of equilibrium

Suppose that the consumer consumes two commodities X and Y given their prices by spending level of money income M. Thus, the objective of the consumer is maximizing his utility function subject to his limited income and market prices. In utility maximization, the function that represents the objective that the consumer tries to achieve is called the objective function and the constraint that the consumer faces is represented by the constraint function.

The maximization problem will be formulated as follows:

$$\text{Maximize } U = f(X, Y)$$

$$\text{Subject to } P_X X + P_Y Y = M$$

We can rewrite the constraint as follows:

$$M - P_X X + P_Y Y = 0 \quad \text{or} \quad P_X X + P_Y Y - M = 0$$

Multiplying the constraint by Lagrange multiplier λ

$$\lambda(M - P_x X + P_y Y) = 0$$

Forming a composite function gives as the Lagrange function:

$$\ell = U(X, Y) + \lambda(M - P_x X + P_y Y)$$

Or,
$$\ell = U(X, Y) - \lambda(P_x X + P_y Y - M)$$

The first order condition requires that the partial derivatives of the Lagrange function with respect to the two goods and the langrage multiplier be zero.

$$\frac{\partial \ell}{\partial X} = \frac{\partial U}{\partial X} - \lambda P_x = 0; \frac{\partial \ell}{\partial Y} = \frac{\partial U}{\partial Y} - \lambda P_y = 0 \text{ and } \frac{\partial \ell}{\partial \lambda} = -(P_x X + P_y Y - M) = 0$$

From the above equations we obtain:

$$\frac{\partial U}{\partial X} = \lambda P_x \quad \text{and} \quad \frac{\partial U}{\partial Y} = \lambda P_y$$

$$\frac{\partial U}{\partial X} = MU_x \quad \text{and} \quad \frac{\partial U}{\partial Y} = MU_y$$

Therefore, substituting and solving for λ we get the equilibrium condition:

$$\lambda = \frac{MU_x}{P_x} = \frac{MU_y}{P_y}$$

By rearranging we get:

$$\frac{MU_x}{MU_y} = \frac{P_x}{P_y}$$

The second order condition for maximum requires that the second order partial derivatives of the Lagrange function with respect to the two goods must be negative.

$$\frac{\partial^2 \ell}{\partial X^2} = \frac{\partial^2 U}{\partial X^2} < 0 \quad \text{and} \quad \frac{\partial^2 \ell}{\partial Y^2} = \frac{\partial^2 U}{\partial Y^2} < 0$$

Example

A consumer consuming two commodities X and Y has the following utility function $U = XY + 2X$. If the price of the two commodities are 4 and 2 respectively and his/her budget is birr 60.

- a) Find the quantities of good X and Y which will maximize utility.
- b) Find the $MRS_{X,Y}$ at optimum.

Solution

The Lagrange equation will be written as follows:

$$\ell = XY + 2X + \lambda(60 - 4X - 2Y)$$

$$\frac{\partial \ell}{\partial X} = Y + 2 - \lambda 4 = 0 \dots\dots\dots (1)$$

$$\frac{\partial \ell}{\partial Y} = X - \lambda 2 = 0 \dots\dots\dots (2)$$

$$\frac{\partial \ell}{\partial \lambda} = 60 - 4X - 2Y = 0 \dots\dots\dots (3)$$

From equation (1) we get $Y + 2 = 4\lambda$ and from equation (2) we get $X = 2\lambda$. Thus, we can get that $X = \frac{Y + 2}{2}$ and equation (2) gives as $\lambda = \frac{1}{2} X$.

By substituting $X = \frac{Y + 2}{2}$ in to equation (2) we get $Y = 14$ and $X = 8$.

$$\begin{aligned} MRS_{X,Y} &= \frac{MU_X}{MU_Y} \\ &= \frac{Y + 2}{X} \end{aligned}$$

After inserting the optimum value of $Y=14$ and $X=8$ we get 2 which equals to the price ratio of the two goods ($\frac{P_x}{P_y} = \frac{4}{2} = 2$).

2.13 Income and Substitution Effects

Dear learner, we now turn to a more complete analysis of why demand curves slope downward. In our previous discussion we have noted that there are two effects of a price change. If price falls (rises), the good becomes cheaper (more expensive) relative to other goods; and consumers substitute toward (away from) the good. This is the substitution effect. Also, as price falls (rises), the consumer's purchasing power increases (decreases). Since the set of consumption opportunities increases (decreases) as price changes, the consumer changes the mix of his or her consumption bundle. This effect is called the income effect. Let us analyze each effect in turn, and then combine the two in order to see why demand is assumed to be downward sloping.

- **Let us Consider the case of a price-decline:**

First a decrease in price increases the consumer's real income (purchasing power), thus enhancing the ability to buy more goods and services to some extent. Second, a decrease in the price of a commodity induces some consumers (the consumer) to substitute it for others, which are now relatively expensive (higher price) commodities. The 1st effect is known as the income effect, and the 2nd effect is known as the substitution effect. The combined effect of the two is known as the total effect (net effect).

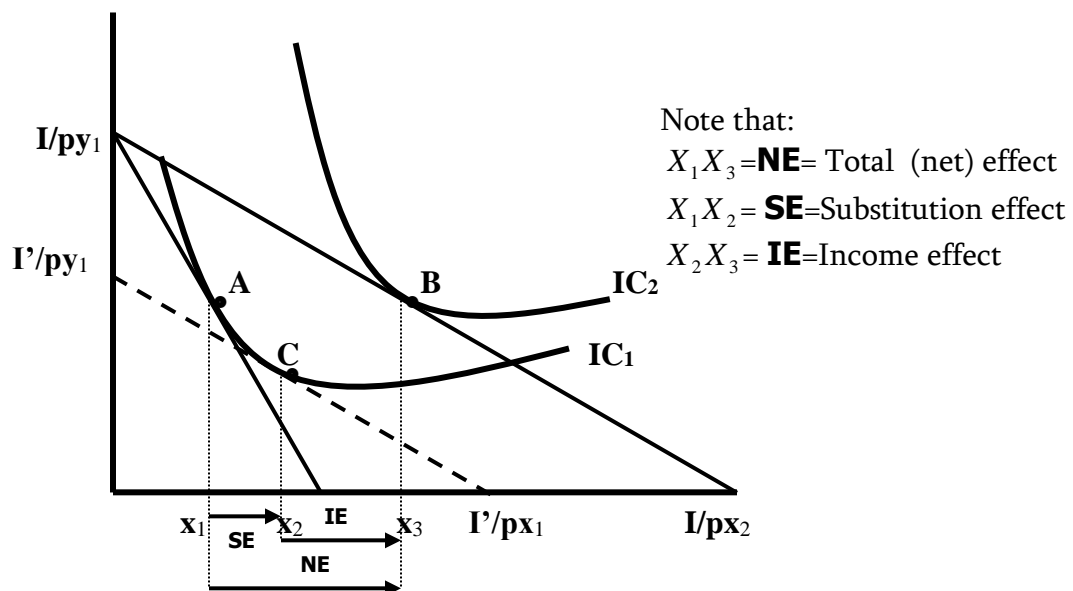


Figure 2.18 Income and Substitution effect for a normal good

Suppose initially the income of the consumer is I_1 , price of good Y is P_{y_1} , and Price of good X is P_{x_1} , we have the budget line with y-intercept $\frac{I}{P_{y_1}}$ and X-intercept $\frac{I}{P_{x_1}}$. The consumer's equilibrium is point A that indicates the point of tangency between the budget line and indifference curve IC_1 . As a result of a decrease in the price of X from P_{x_1} to P_{x_2} the budget line shifts outward with y-intercept $\frac{I}{P_{y_1}}$ & X-Intercept $\frac{I}{P_{x_2}}$. The consumer's new equilibrium will be on point B.

The total change in the quantity purchased of commodity X from the 1st equilibrium point at A to the second equilibrium point at B shows the Net effect or total effect of the price decline (change).

The total effect of the price change can be conceptually decomposed into the substitution effect and income effect.

The Substitution Effect

The substitution effect refers to the change in the quantity demanded of a Commodity resulting exclusively from a change in its price when the consumer's real income is held constant; thereby restricting the consumer's reaction to the price change to a movement along the original indifference curve. The decline in the price of X results in an increase in the consumer's real income, as evidenced by the movement to a higher indifference curve even though money income remains fixed.

Now, imagine that we decrease the consumer's income by an amount just sufficient to return to the same level of satisfaction enjoyed before the price decline. Graphically, this is accomplished by drawing a fictitious (imaginary) line of attainable combinations with a slope corresponding to new ratio of the product price $\frac{P_{x_2}}{P_{y_1}}$ so that it is just tangent to the original indifference curve IC_1 .

The point of tangency is the imaginary point **C** (imaginary equilibrium). The movement from point **A** to the imaginary intermediate equilibrium at point **C**, which shows increase in consumption of X from X_1 to X_2 is the substitution effect. In other words, the effect of a decrease in price encourages the consumer to increase consumption of X than Y .

The Income Effect

The income effect may be defined as the change in the quantity demanded of a commodity exclusively associated with a change in real income. The income effect is determined by observing the change in the quantity demanded of a commodity that is associated solely with the change in the consumer's real income.

In figure 2.18, letting the consumer's real income rise from its imaginary level (defined by the line of attainable combinations tangent to point **C**) back to its true level (defined by the line of attainable combinations tangent to point **B**) gives the income effect. Thus, the income effect is indicated by the movement from the imaginary equilibrium at point **C** to the actual new equilibrium at point **B**, the increase in the quantity of X purchased from X_2 to X_3 is the income effect.

The income effect of a change in the price of good shows the change in quantity demanded via change in real income, while the relative price ratio remains constant. This movement does not involve any change in prices; the price ratio is the same in budget line 1 as in budget line 2. It is due to a change in total satisfaction and such a change is a movement from one indifference curve to another.

When we look at both the substitution and income effects, the magnitude of the substitution effect is greater than that of the income effect. The reason is that:

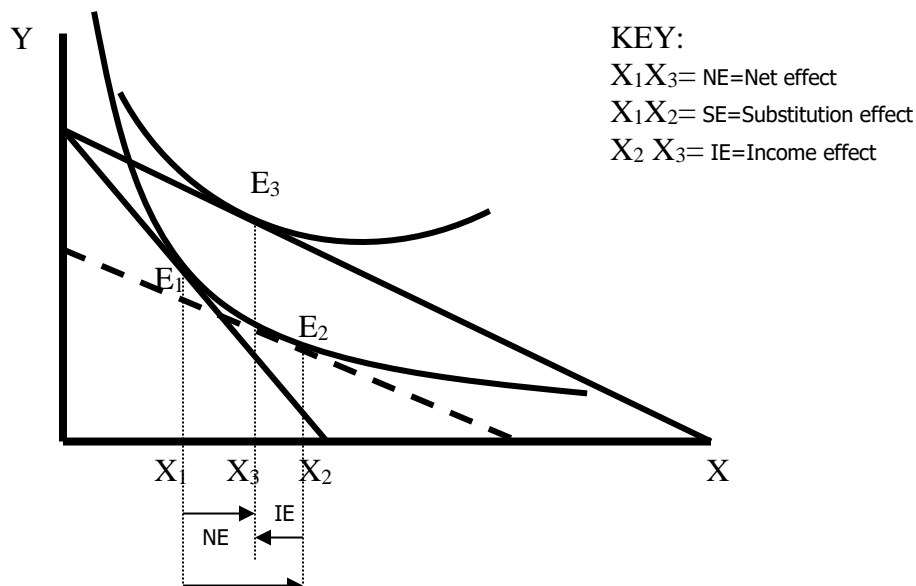
- Most goods have suitable substitutes and when the price of good falls, the quantity of the good purchased is likely to increase very much as consumers substitute the now cheaper good for others.

- Spending only a small fraction of his /her income, i.e. with the consumers purchasing many goods and spending only a small fraction of their income on any one good, the income effect of a price change of any one good is likely to be small.

Usually, the income and substitution effects reinforce one another i.e. they operate in the same direction. The substitution effect is always negative. i.e. if the price of a good X increases and real income is held constant, there will always be a decrease in the consumption of good X, and vice versa. This result follows from the fact that indifference curves have negative slopes. However, the income effect is not predictable from the theory alone. In most cases, one would expect that increases in real income would result in increases in consumption of a good. This is the case for so called Normal goods.

In short in the case of normal goods, the income effect and the substitution effect operate in the same direction –they reinforce each other. But not all goods are normal. Some goods are called inferior goods because the income effect is the opposite (of that of a normal good) for them-they operate in opposite direction. For an inferior good, a decrease in the price of the commodity causes the consumer to buy more of it (the substitution effect), but at the same time the higher real income of the consumer tends to cause him to reduce consumption of the commodity (the income effect). We usually observe that the substitution effect still is the more powerful of the two; even though the income effect works counter to the substitution effect, it does not override it. Hence, the demand curve for inferior goods is still negatively sloped.

Let us consider the following diagram that shows the income, substitution and net effect for an **inferior commodity** in the case of a decline in the price of good X.



IC₁ IC₂

Figure 2.19 Income, Substitution, and Net effect for an inferior commodity

In very rare occasions, a good may be so strongly inferior that the income effect actually overrides the substitute effect. Such an occurrence means that a decline in the price of a good would lead to a decline in the quantity demanded and that a rise in price will induce an increase in quantity demanded. In other words, price and quantity move in the same direction. The name given to such a unique situation is Giffen paradox; and it constitutes an exception to the Law of demand. That is for Giffen goods the income effect (which decreases the quantity demanded) is so strong that it offsets the substitution effect (which increases the quantity demanded), with the result that the quantity demanded is directly related to the price, at least over some range of variation of price.

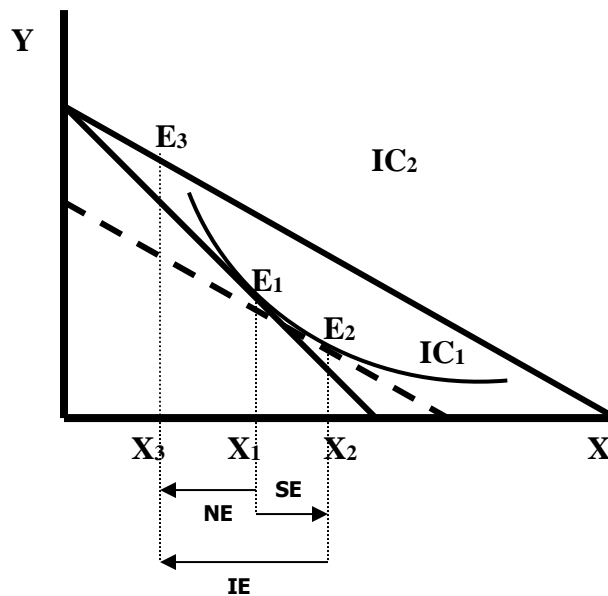


Figure 2.20 Income, Substitution and net effects for a Giffen good, When there is a price decline.

2.14 The Slutsky Equation

Dear learner, can you explain Slutsky equation ?

(You can use the space left below to write your response)

Dear learner as we have discussed earlier, when the price of a good decreases, there will be two effects in consumption. The change in relative prices makes the consumer to consume more of the cheaper good—substitution effect. The increase in purchasing power due to the lower price may increase or decrease consumption of the good—income effect. Generally, the Slutsky equation says that the total change in demand is the sum of the substitution effect and the income effect.

Numerical Example

Suppose that the consumer has a demand function for good X is given by

$$X = 20 + MP_X^{-2}$$

Originally his income is \$ 200 per month and the price of the good is 5 per kilogram.

Therefore, his demand for good X will be $20 + \frac{200}{5^2} = 28$ per month.

Suppose that the price of the good falls to 4 per kilogram. Therefore, the new demand at the new price will be: $20 + \frac{200}{4^2} = 32.5$ per month.

Thus, the total change in demand is 4.5 that is 32.5-28.

When the price falls the purchasing power of the consumer changes. Hence, in order to make the original consumption of good X, the consumer adjusts his income. This can be calculated as follows:

$$M^1 = P_1^1 X + P_2 Y$$

$$M = P_1 X + P_2 Y$$

Subtracting the second equation from the first gives: 4-

$$M^1 - M = X[P_1' - P_1]$$

$$\Delta M = X\Delta P_1$$

Therefore, new income to make the original consumption affordable when price falls to 4 is:

$$\Delta M = X\Delta P_1$$

$$\Delta M = 28 * [4 - 5] = -28$$

Hence, the level of income necessary to keep purchasing power constant is

$$M^1 = M + \Delta M = 200 - 28 = 172$$

The consumer's new demand at the new price and income will be :

$$X(4, 172) = 20 + \frac{172}{4^2} = 30.75$$

Therefore, the substitution effect will be:

$$\Delta X = X(4, 172) - X(5, 200) = 30.75 - 28 = 2.75$$

The income effect will be:

$$X(4, 200) - X(4, 172) = 32.5 - 30.75 = 1.75$$

Since the result we obtained is positive we can conclude that the good is a normal good.

2.15 The Consumer Surplus

Dear learner, how do you explain the concept of consumer surplus?

(You can use the space left below to write your response)

While consumers purchase goods and services, they often pay less than what they are willing to pay. Thus, the difference between what they are willing to pay and what they actually paid is considered as their surplus.

Therefore, *consumer surplus* is the difference between what a consumer is willing to pay and what he actually pays.

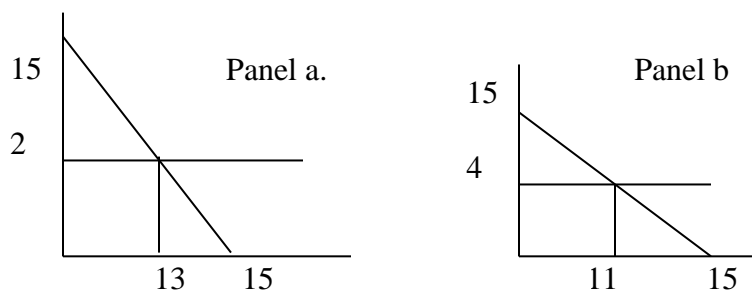
Numerical Example

Suppose the demand function of a consumer is given by: $Q = 15 - p$

- Compute the consumer surplus when the price of the good is 2
- Compute the consumer surplus when the price of the good is 4
- Compute the change in consumer surplus when the price changes from 2 to 4.

Solution

When Price is zero the demand for quantity purchased will be 15 and when the demand for quantity is put to zero then the price level will be 15. And finally, when we insert the given price level 2 in the demand equation we get the level of quantity demanded that is 13. Hence, we can easily compute the area of the triangle that is found above the given price level that is 2.



In panel a, the area of the triangle above price that is the consumer surplus is $\frac{1}{2} * 13 * 13 = 84.5$ and in panel b the consumer surplus is 60.5. Therefore, due to a change in the price level the consumer surplus will be $84.5 - 60.5 = 24.5$

Summary

Consumers given their income and prices of the commodities, they spend their income so as to attain the highest possible satisfaction or utility from commodities. **Utility** is thus the satisfaction obtained from the consumption of a good. The maximization of utility is referred to as the axiom of utility maximization. To attain this utility maximization objective, the consumer must be able to compare the utility of the various baskets of goods, which they can buy with their income. In order to explain the comparison of these commodities we have two approaches. These are: **cardinal approach and the Ordinal approach.**

Cardinalist believed that cardinal numbers could be used to express the utility derived from the consumption of a commodity while ordinalists believed that utility is not measureable, but is an ordinal magnitude. The main ordinal theories are the **indifference curves approach** and the **revealed preferences hypothesis**. These approaches are also known as the indifference curve theories .

An **indifference curve** is the locus of points which provide the same level of satisfaction to the consumer. The slope of an indifference curve shows the marginal rate of substitution between goods.

The consumer aims at maximization of utility, given his/her income and market prices of the commodities available for consumption. Therefore, to determine the equilibrium of the consumer, we have to bring together the indifference map and the budget line facing the consumer on the same diagram.

There are two effects of a price change. If price falls (rises), the good becomes cheaper (more expensive) relative to other goods; and consumers substitute toward (away from) the good. This is **the substitution effect**. Also, as price falls (rises), the consumer's purchasing power increases (decreases). Since the set of consumption opportunities increases (decreases) as price changes, the consumer changes the mix of his or her consumption bundle. This effect is called **the income effect**.

UNIT THREE

THE THEORY OF PRODUCTION

Objectives

After successful completion of this unit, you will be able to

- Define production and production function
- Differentiate short run and long run, and fixed and variable inputs
- Know the concepts of short run production and efficiency
- Know the concepts of long run production, laws of returns to scale and how to determine least cost production process.

3.1 Introduction: Definition and basic concepts

Dear learner, how do you define production? What is needed to undertake production? Attempt to answer these questions and then read the following paragraphs to be sure on the answer. _____

To an economist production means creation of utility for sales. Alternatively, production may be defined as the act of creating those goods/services which have exchange value for sale (not for personal consumption). Raw materials yield less satisfaction to the consumer by themselves. In order to get utility from raw materials, first they must be transformed into output. However, transforming raw materials into final products require factor inputs such as land, labor, and capital and entrepreneurial ability.

Thus, no production (transforming raw material into output) can take place without the use of inputs.

Fixed Vs variable inputs

In economics, inputs can be classified as fixed & variable. *Fixed inputs* are those inputs whose quantity cannot readily be changed when market conditions indicate that an immediate change in

output is required. In fact no input is ever absolutely fixed, but may be fixed during an immediate requirement. For example, if the demand for Beer shoots up suddenly in a week, the brewery factories can not plant additional machinery over a night to respond to the increased demand. It takes long time to buy new machineries, to plant them and use for production. Thus, the quantity of machinery is fixed for some times such as a week. Buildings, machineries and managerial personnel are examples of fixed inputs because their quantity can not be manipulated easily in short time periods.

Variable inputs, on the other hand, are those inputs whose quantity can be changed almost instantaneously in response to desired changes in output. That is, their quantity can easily be diminished when the market demand for the product decreases and vice versa. The best example of variable input is unskilled labor.

In our previous example, if the brewery factory had idle machinery before the market demand shot up, the factory can easily and immediately respond to the market condition by hiring laborers.

Short run Vs. long run

Dear learner, what do you expect about the definitions of short run and long run? _____

In economics, **short run** refers to that period of time in which the quantity of at least one input is fixed. For example, if it requires a firm one year to change the quantities of all the inputs, those time periods below one year are considered as short run. Thus, short run is that time period which is not sufficient to change the quantities of all inputs, so that at least one input remains fixed. One thing to be noted here is that short run periods of different firms have different duration. Some firms can change the quantity of all their inputs within a month while it takes more than a year to change the quantity of all inputs for another type of firms. For example, the time required to change the quantities of inputs in an automobile factory is not equal with that of flour factory. The later takes relatively shorter time. **Long run** is that time period (planning horizon) which is sufficient to change the quantities of all inputs. Thus there is no fixed input in the long -run.

3.2 Production in the short run: Production with one variable input

Production with one variable input (while the others are fixed) is obviously a short run phenomenon because there is no fixed input in the long run.

Assumption of short run production analysis

In order to simplify the analysis of short run production, the classical economist assumed the following:

1. *Perfect divisibility of inputs and outputs*

This assumption implies that factor inputs and outputs are so divisible that one can hire, for example a fraction of labor, a fraction of manager and we can produce a fraction of output, such as a fraction of automobile.

2. *Limited substitution between inputs*

Factor inputs can substitute each other up to a certain point, beyond which they can not substitute each other

3. *Constant technology*

They assumed that level of technology of production is constant in the short run.

Suppose a firm that uses two inputs: Capital (which is a fixed input) and labor (which is variable input). Given the assumptions of short run production, the firm can increase output only by increasing the amount of labor it uses.

Hence, its production function is

$$Q = f(L) \quad K - \text{being constant}$$

Where Q is the quantity of production (Output)

L is the quantity of labor used, which is variable, and

K is the quantity of capital (which is fixed)

The production function shows different levels of output that the firm can obtain by efficiently utilizing different units of labor and the fixed capital. In the above short run production function, the quantity of capital is fixed. Thus output can change only when the amount of labor used for production changes. Hence, Q is a function of L only in the short run.

3.3 Total product, marginal product and average product

Dear learner, do you remember what is meant by total utility and marginal utility from the previous chapter discussion? Then what do you suggest about total product, marginal and average product?_____

Total product: is the total amount of output that can be produced by efficiently utilizing a specific combination of labor and capital. The total product curve, thus, represents various levels of output that can be obtained from efficient utilization of various combinations of the variable input, and the fixed input. It shows the output produced for different amounts of the variable input, labor.

Dear learner, do you think that output can always be increased by increasing the variable input while there is a fixed input?

Any ways, increasing the variable input (while some other inputs are fixed) can increase the total product only up to a certain point. Initially, as we combine more and more units of the variable input with the fixed input output continues to increase. But eventually, increasing the unit of the variable input may not help output increase. Even as we employ more and more unit of the variable input beyond the carrying capacity of a fixed input, out put may tends to decline. Thus increasing the variable input can increase the level of output only up to a certain point, beyond which the total product tends to fall as more and more of the variable input is utilized. This tells us what shape a total product curve assumes. The shape of the total variable curve is nearly S-shape (see fig 2.1 Panel A)

Marginal Product (MP)

The marginal product of variable input is the addition to the total product attributable to the addition of one unit of the variable input to the production process, other inputs being constant (fixed). Before deciding whether to hire one more worker, a manager wants to determine how much this extra worker ($\Delta L = 1$) will increase output, Δq . The change in total output resulting from using this additional worker (holding other inputs constant) is the marginal product of the worker. If output changes by Δq when the number of workers (variable input) changes by ΔL , the change in output per worker or marginal product of the variable input, denoted as MP_L is found as

$$MP_L = \frac{\Delta Q}{\Delta L} \text{ or } MPL = \frac{dTP}{dL}$$

Thus, MPL measures the slope of the total product curve at a given point. In the short run, the MP of the variable input first increases reaches its maximum and then tends to decrease to the extent of being negative. That is, as we continue to combine more and more of the variable inputs with the fixed input, the marginal product of the variable input increases initially and then declines.

Average Product (AP)

The AP of an input is the ratio of total output to the number of variable inputs.

$$AP_{labour} = \frac{\text{total product}}{\text{number of } L} = \frac{TP}{L}$$

The average product of labor first increases with the number of labor (i.e. TP increases faster than the increase in labor), and eventually it declines.

Graphing the short run production curves

The following figures shows how the TP , MP and AP of the variable (labor) input vary with the number of the variable input.

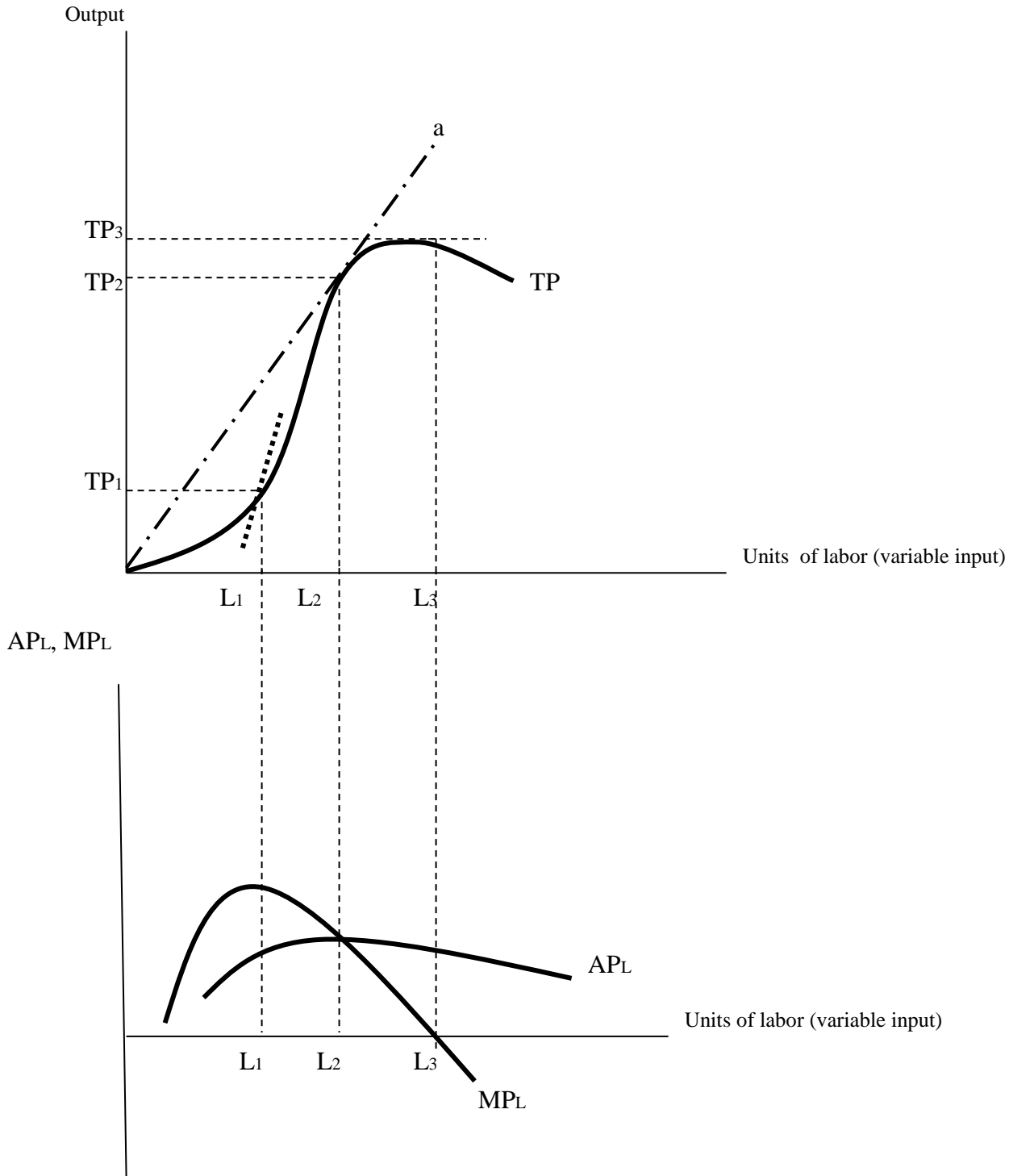


Fig 3.1 Total product, average product and marginal product curves: As the number of the labor hired increases (capital being fixed), the TP curve first rises, reaches its maximum when L_3 amount of labor is employed, beyond which it tends to decline. Assuming that this short run production curve represents a certain car manufacturing industry, it implies that L_3 numbers of workers are

required to efficiently run the machineries. If the numbers of workers fall below L_3 , the machine is not fully operating, resulting in a fall in TP below TP_3 . On the other hand, increasing the number of workers above L_3 will do nothing for the production process because only L_3 number of workers can efficiently run the machine. Increasing the number of workers above L_3 , rather results in lower total product because it results in over crowded and unfavorable working environment.

Marginal product curve increases until L_1 number of labor reaches its maximum at L_1 , and then it tends to fall. The MPL is zero at L_3 (when the TP is maximal); beyond which its value assumes zero indicating that each additional worker above L_3 tends to create over crowded working condition and reduces the total product. Thus, in the short run (where some inputs are fixed), the marginal product of successive units of labor hired increases initially, but not continuously, resulting in the limit to the total production. Geometrically, the MP curve measures the slope of the TP. The slope of the TP curve increases (MP increases) up to L_1 , it decreases from L_1 to L_3 and it becomes negative beyond L_3 .

The average product curve increases up to L_2 , beyond which it continuously declines. The AP curve can be measured by the slope of rays originating from the origin to a point on the TP curve. For example, the APL at L_2 is the ratio of TP_2 to L_2 . This is identical to the slope of ray a.

The relationship between AP and MP of the variable input

The relationship between MPL and APL can be stated as follows:

- For all number of workers (Labor) below L_2 , MPL lies above APL.
- At L_2 , MPL and APL are equal.
- Beyond L_2 , MPL lies below the APL

Thus, the MPL curve passes through the maximum of the APL curve from above. This relationship between APL and MPL can be shown algebraically as follows:

Suppose the production function is given as

$$TP = f(L), K \text{ -being constant}$$

Given the total product function,

$$MPL = \frac{dTP}{dL} = \frac{df(L)}{dL} \quad \text{and} \quad APL = \frac{TP}{L} = \frac{f(L)}{L}$$

To determine the relationship between APL and MPL, consider the slope of the APL function.

$$\text{Slope of APL} = \frac{dAPL}{dL} = \frac{d\left(\frac{f(L)}{L}\right)}{dL} = \frac{\frac{df(L)}{dL} \cdot L - \frac{dL}{dL} \cdot f(L)}{L^2} \text{----- (quotient rule of differentiation)}$$

$$\text{Slope of APL} = \frac{\frac{df(L)}{dL} \cdot L}{L^2} - \frac{f(L)}{L^2} \text{----- (note that } \frac{a+b}{c} = \frac{a}{c} + \frac{b}{c} \text{)}$$

$$= \frac{\frac{df(L)}{dL}}{L} - \frac{f(L)}{L}$$

$$\text{Slope of APL} = \frac{MPL - APL}{L}, \text{ because } \frac{df(L)}{dL} = MPL \text{ and } \frac{f(L)}{L} = APL$$

Now – when $MPL > APL$, Slope of APL is positive (APL rises)

- When $MPL = APL$, Slope of APL is zero (APL is at its maximum).
- When $MPL < APL$, Slope of APL is negative (APL falls)

The law of diminishing marginal returns (LDMR): short –run law of production

The LDMR states that as the use of an input increases in equal increments (with other inputs being fixed), a point will eventually be reached at which the resulting additions to output decreases. When the labor input is small (and capital is fixed), extra labor adds considerably to output, often because workers get the chance to specialize in one or few tasks. Eventually, however, the LDMR operates: when the number of workers increases further, some workers will inevitably become ineffective and the MPL falls (this happens when the number of workers exceeds L1 in fig 2.1)

Note that the LDMR operates (MP of successive units of labor decreases) not because highly qualified laborers are hired first and the least qualified last. Diminishing marginal returns results from limitations on the use of other fixed inputs (e.g. machinery), not from decline in worker quality.

The LDMR applies to a given production technology (when the level of technology is fixed). Over time, however, technological improvements in the production process may allow the entire total product curve shift upward, so that more output can be produced with the same input.

3.5 Efficient Region of Production in the short-run

Dear learner, we are now not in a position to determine the specific number of the variable input (labor) that the firm should employ because this depends on several other factors than the productivity of labor such as the price of labor, the structure of input and output markets, the demand for output, etc. However, it is possible to determine ranges over which the variable input (labor) be employed.

To do best with this, let's refer back to fig 2.1 and divide it into three ranges called stages of production.

- Stage I – ranges from the origin to the point of equality of the APL and MPL.
- Stage II – starts from the point of equality of MPL and APL and ends at a point where MP is equal to zero.
- Stage III – covers the range of labor over which the MPL is negative.

Now, which stage of production is efficient and preferable?

To answer the question, let us follow elimination method.

Obviously, a firm should not operate in stage III because in this stage additional units of variable input are contributing negatively to the total product (MP of the variable input is negative) because of overcrowded working environment i.e., the fixed input is over utilized.

Stage I is also not an efficient region of production though the MP of variable input is positive. The reason is that the variable input (the number of workers) is too small to efficiently run the fixed input; so that the fixed input is under-utilized (not efficiently utilized)

Thus, the efficient region of production is stage II. At this stage additional inputs are contributing positively to the total product and MP of successive units of variable input is declining (indicating that the fixed input is being optimally used). Hence, the efficient region of production is over that range of employment of variable input where the marginal product of the variable input is declining but positive.

3.6 Long run Production: Production with two variable inputs

Dear learner, we have completed our analysis of the short-run production function in which the firm uses one variable input (labor) and one fixed input (capital). Now we turn to the long run analysis of production. Remember that long run is a period of time (planning horizon) which is sufficient for the firm to change the quantity of all inputs. For the sake of simplicity, assume that the firm uses two inputs (labor and capital) and both are variable.

The firm can now produce its output in a variety of ways by combining different amounts of labor and capital. With both factors variable, a firm can usually produce a given level of output by using a great deal of labor and very little capital or a great deal of capital and very little labor or moderate amount of both. In this section, we will see how a firm can choose among combinations of labor and capital that generate the same output. To do so, we make the use of isoquant. So it is necessary to first see what is meant by isoquants and their properties. Dear learner, what is an isoquant?

Isoquants

An isoquant is a curve that shows all possible efficient combinations of inputs that can yield equal level of output. If both labor and capital are variable inputs, the production function will have the following form.

$$Q = f(L, K)$$

Given this production function, the equation of an isoquant, where output is held constant at q is

$$q = f(L, K)$$

Thus, isoquants show the flexibility that firms have when making production decision: they can usually obtain a particular output (q) by substituting one input for the other.

Isoquant maps: when a number of isoquants are combined in a single graph, we call the graph an isoquant map. An isoquant map is another way of describing a production function. Each isoquant represents a different level of output and the level of outputs increases as we move up and to the right. The following figure shows isoquants and isoquant map.

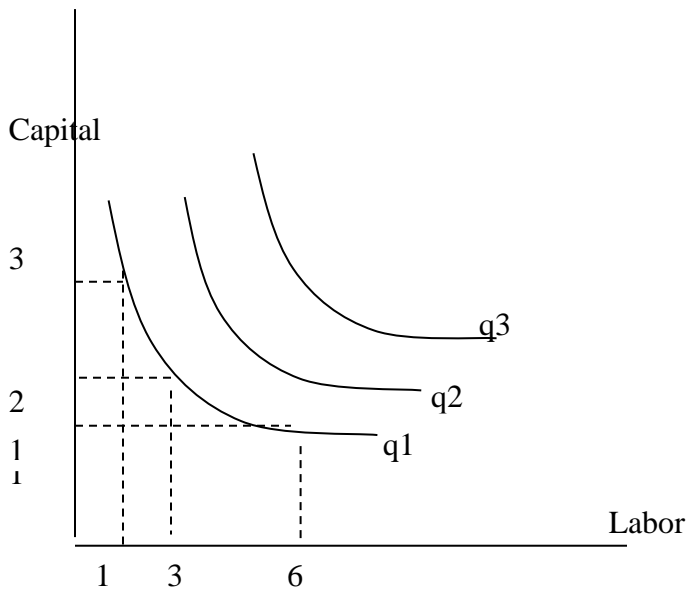


Fig 3.2 Isoquant and isoquant map. Isoquants show the fact that long run production process is very flexible. A firm can produce q_1 level of output by using either 3 capital and 1 labor or 2 capital and 3 labor or 1 capital and 6 labor or any other combination of labor and labor on the curve. The set of isoquant curves q_1 q_2 & q_3 are called isoquant map.

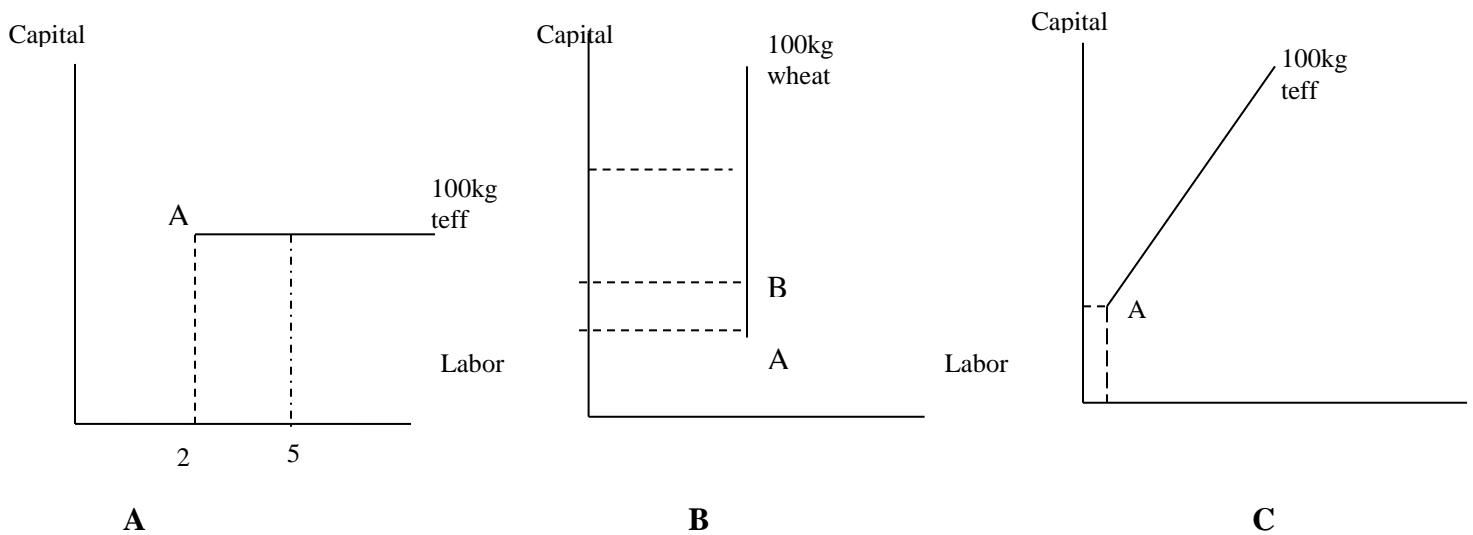
Properties of isoquants

Dear learner, do you remember the properties of a well behaved indifference curve? Use this place to enumerate them.-----

Isoquants have most of the same properties as indifference curves. The biggest difference between them is that output is constant along an isoquant where as indifference curves hold utility constant. Most of the properties of isoquants, results from the word '*efficient*' in its definition.

1. Isoquants slope down ward. Because isoquants denote efficient combination of inputs that yield the same output, isoquants always have negative slope. Isoquants can never be horizontal, vertical or upward sloping. If for example, isoquants have to assume zero slopes (horizontal line) only one point on the isoquant is efficient. See the following figures.

Fig 3.3



A

An isoquant can never be horizontal. In this figure, the firm can produce 100kg of teff by using either of the following alternatives: 4 capital and 2 labor, 4 capital and 5 labor or any other combination of labor and capital along the curve. Obviously, only the first alternative is efficient as it uses the least possible combination of inputs. Thus, all points, except A, are inefficient and not part of

B

In this figure, a firm can produce 100kg of wheat by using any combination of labor and capital along the isoquant. But only point A is efficient. For example, point B shows the same number of labor as point A, but higher capital. Thus point B is inefficient because it shows higher combination of inputs. Thus, isoquants can never be vertical line

C

In this figure, all points above point A utilize higher combination of both inputs to produce the same output (100 kg coffee). Point A shows the least combination of inputs that can yield 100 kg coffees. Thus all other points are inefficient and not part of the isoquants.

Thus, efficiency requires that isoquants must be negatively sloped. As employment of one factor increases, the employment of the other factor must decrease to produce the same quantity efficiently.

2. The further an isoquant lays away from the origin, the greater the level of output it denotes. Higher isoquants (isoquants further from the origin) denote higher combination of inputs. The more inputs used, more outputs should be obtained if the firm is producing efficiently. Thus efficiency requires that higher isoquants must denote higher level of output.

3. Isoquants do not cross each other. This is because such intersections are inconsistent with the definition of isoquants.

Consider the following figure.

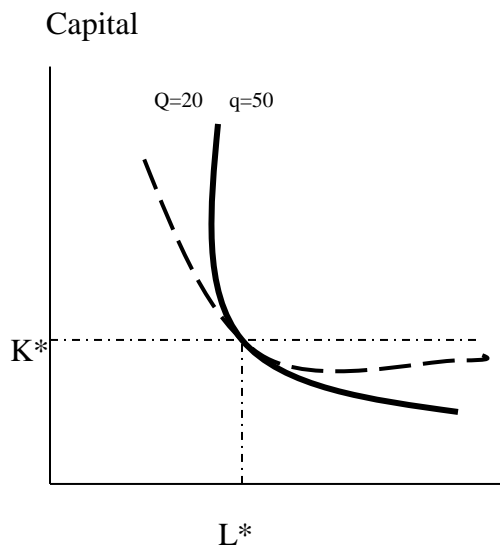


Fig 3.4 Efficiency requires that isoquants do not cross each other, because the point of their intersection implies that there is inefficiency at this point.

This figure shows that the firm can produce at either output level (20 or 50) with the same combination of labor and capital (L^* and K^*). The firm must be producing inefficiently if it produces $q = 20$, because it could produce $q = 50$ by the same combination of labor and capital (L^* and K^*). Thus, efficiency requires that isoquants do not cross each other.

4. Isoquants must be thin. If isoquants are thick, some points on the isoquant will become inefficient. Consider the following isoquant.

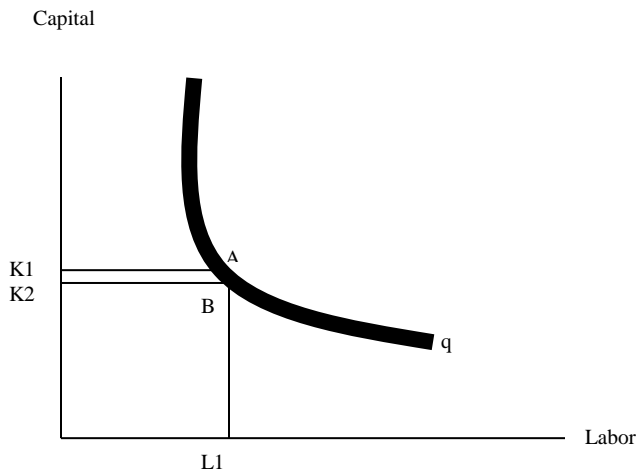


Fig.3.5: Iso quants can never be thick. Points A and B are on the same iso quant. But point A denotes higher amount of capital and the same amount of labor as point B. Hence point A denotes inefficient combination of inputs and thus it lies out of the iso quant. The iso-quant should be thin if point A is to be excluded from the iso quant.

Shape of isoquants

Isoquants can have different shapes (curvature) depending on the degree to which factor inputs can substitute each other.

1-Linear isoquants

Isoquants would be linear when labor and capital are perfect substitutes for each other. In this case the slope of an iso quant is constant. As a result, the same output can be produced with only capital or only labor or an infinite combination of both. Graphically,

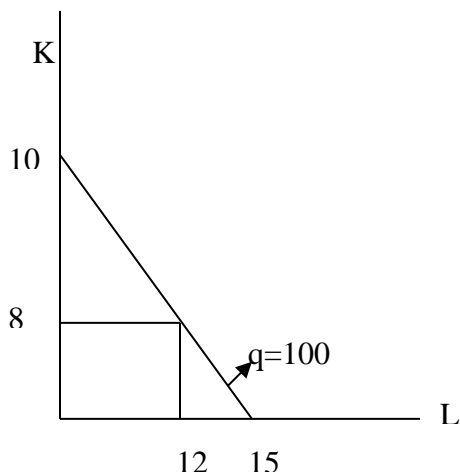


Fig.3.6 linear isoquant. Capital and labor can perfectly substitute each other so that the same output ($q=100$) can be produced by using either 10k or 8K and 12L or 15L or an infinite combinations of both inputs.

2. Input output isoquant

It is also called Leontief isoquant. This assumes strict complementarities or zero substitutability of factors of production. In this case, it is impossible to make any substitution among inputs. Each level of output requires a specific combination of labor and capital: Additional output cannot be obtained unless more capital and labor are added in specific proportions. As a result, the isoquants are L-shaped. See following figure

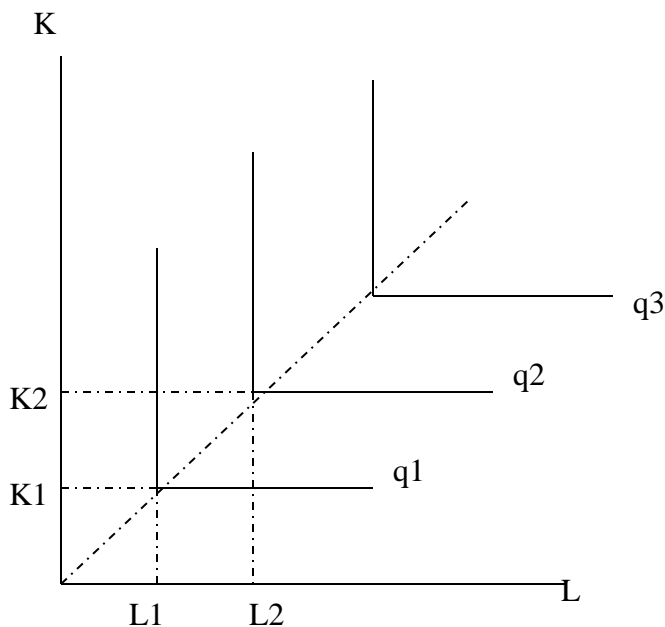


Fig. 2.7 L-shaped isoquant. When isoquants are L-shaped, there is only one efficient way of producing a given level of output: Only one combination of labor and capital can be used to produce a given level of output. To produce q_1 level of output there is only one efficient combination of labor and capital (L_1 and K_1). Output cannot be increased by keeping one factor (say labor) constant and increasing the other (capital). To increase output (say from q_1 to q_2) both factor inputs should be increased by equal proportion.

3. Kinked isoquants

This assumes limited substitution between inputs. Inputs can substitute each other only at some points. Thus, the isoquant is kinked and there are only a few alternative combinations of inputs to produce a given level of output. These isoquants are also called linear programming isoquants or activity analysis isoquants. See the figure below.

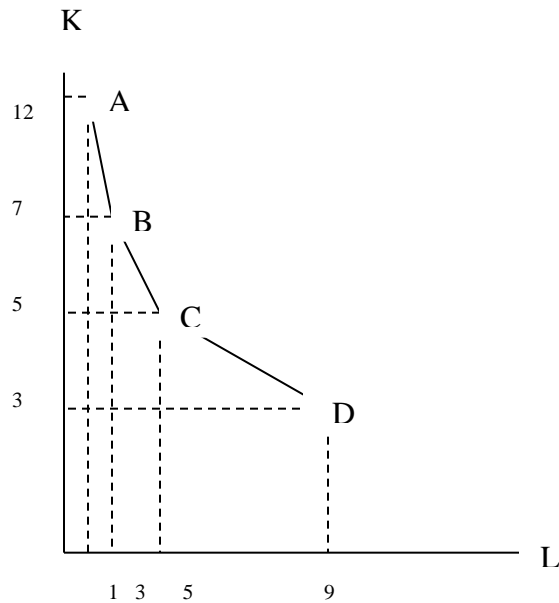


Fig. 3.8 kinked isoquant in this case labor and capital can substitute each other only at some point at the kink (A, B, C, and D). Thus, there are only four alternative processes of producing $q=100$ out put.

4. Smooth, convex isoquants

This shape of isoquant assumes continuous substitution of capital and labor over a certain range, beyond which factors cannot substitute each other. Basically, kinked isoquants are more realistic: There is often limited (not infinite) method of producing a given level of output. However, traditional economic theory mostly adopted the continuous isoquants because they are mathematically simple to handle by the simple rule of calculus, and they are approximation of the more realistic isoquants (the kinked isoquants). From now on we use the smooth and convex isoquants to analyze the long run production.

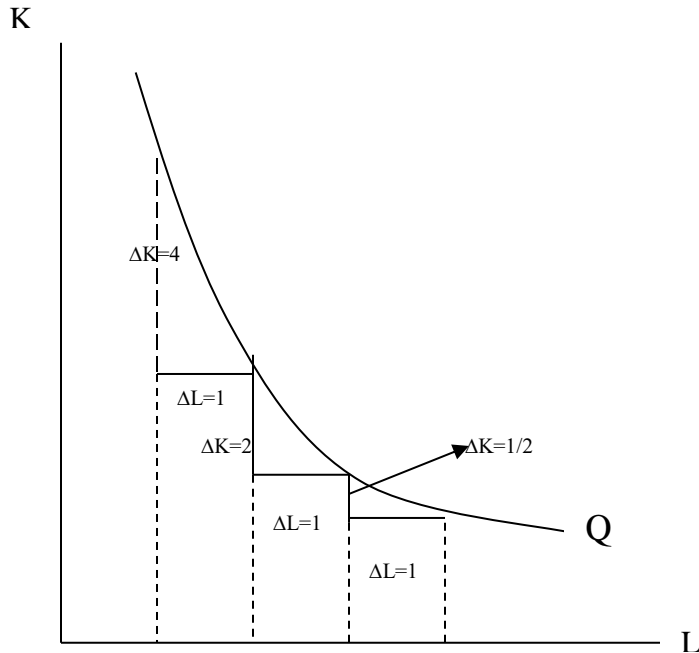


Fig: 3.9 the smooth and convex isoquant. This type of isoquant is the limiting case of the kinked isoquant when the number kink is infinite. The slope of the iso quant decrease as we move from the top (left) to the right (bottom) along the isoquant. This indicates that the amount by which the quantity of one input (capital) can be reduced when one extra unit of another inputs(labor) is used (so that out put remains constant) decreases as more of the latter input (labor) is used.

The slope of an isoquant: marginal rate of technical substitution (MRTS)

The slope of an isoquant ($-\Delta K/\Delta L$) indicates how the quantity of one input can be traded off against the quantity of the other, while out put is held constant. The absolute value of the slope of an isoquant is called marginal rate of technical substitution (MRTS). The MRTS shows the amount by which the quantity of one input can be reduced when one extra unit of another input is used, so that output remains constant. MRTS of labor for capital, denoted as $MRTS_{L,K}$ shows the amount by which the input of capital can be reduced when one extra unit of labor is used, so that output remains constant.

This is analogous to the marginal rate of substitution (MRS) in consumer theory.

$MRTS_{L,K}$ decreases as the firm continues to substitute labor for capital (or as more of labor is used). In fig.2.9 to increase the amount of labor from 1 to 2, the firm reduces 4 units of capital

($\Delta K=4$), to increase labor from 2 to 3, the firm reduce 2 unit of capital ($\Delta K=2$), and so on. Hence, the firm reduces lower and lower number of capital for the successive one unit of labor. Dear learner, why does this happen?

The reason is that when the number of capital is large and that of labor is low, the productivity of capital is relatively lower and that of labor is higher (due to the law of diminishing marginal returns). Thus, at this point relatively large amount of capital is required to replace one unit of labor (or one unit of labor can replace relatively large amount of capital). As the employment of labor increases and that of capital decreases (as we move down ward along the isoquant), quite the reverse will happen. That is, productivity of capital increases and that of labor decreases. Hence, the amount of capital that needs to be reduced increase when one extra labor is used decreases. The fact that the slope of an isoquant is decreasing makes an isoquant convex to the origin.

MRTS L, K (the slope of isoquant) can also be given by the ratio of marginal products of factors. That is,

$$MRTS_{L,K} = -\frac{\Delta K}{\Delta L} = \frac{MPL}{MPK}$$

This can be shown algebraically as follows:

Let the production function is given as:

$$q = f(L, K)$$

Where q- is output

L- is unit of labor employed

K-is the amount of capital employed.

Given this production function, the equation of a specific isoquant can be obtained by equating the production function with a given level of output, say \bar{q} .

$$\bar{q} = f(L, K) = \bar{q}$$

Total differential of \bar{q} measures the total change in \bar{q} that happens as a result of a simultaneous change in L and K. i.e,

$$dq = \frac{\partial q}{\partial L} .dL + \frac{\partial f}{\partial k} .dk = d\bar{q}$$

But since \bar{q} is constant, $d\bar{q}$ is zero ($d\bar{q}=0$)

$$\text{So, } \frac{\partial q}{\partial L} \cdot dL + \frac{\partial q}{\partial k} dk = 0$$

$$\left(\text{But, } \frac{\partial q}{\partial L} = MPL \text{ and } \frac{\partial q}{\partial k} = MPk \right)$$

Thus, the above equation can be written as:

$$MPL \cdot dL + MPK \cdot dk = 0$$

$$\Rightarrow \frac{MPL}{MPk} = \frac{-dK}{dL}$$

There fore, the slope of an isoquant can be given as the ratio of marginal products of inputs.

Elasticity of substitution

MRTS as a measure of the degree of substitutability of factors has a serious defect. It depends on the units of measurement of factors. A better measure of the ease of factor substitution is provided by the elasticity of substitution, δ . The elasticity of substitution is defined as

$$\begin{aligned} \delta &= \frac{\% \Delta \frac{K}{L}}{\% \Delta MRTS} = \frac{\% \Delta \frac{K}{L}}{\% \frac{MPL}{MPK}} \\ &= \frac{\left(\frac{K}{L}\right)}{\frac{L}{K}} \div \frac{\left(\frac{MPL}{MPK}\right)}{\left(\frac{MPL}{MPK}\right)} \end{aligned}$$

The elasticity of substitution is a pure number independent of the units of measurement of K and L, since the numerator and the denominator are measured in the same units and be cancelled.

Factor intensity

A process of production can be labor intensive or capital intensive or neutral process. A process of production is called labor intensive if it uses many labors and relatively few capitals. If it uses

many capitals and relatively few labor it is called capital intensive technology. On the other hand, if the process uses equal proportion of both it is called neutral technology. The factor intensity of any process is measured by the slope of the line through the origin representing the particular process. Thus, the factor intensity is the capital-labor ratio. The higher the capital-labor ratio is the higher the capital intensity but the lower the capital-labor ratio is the higher labor intensity of the process.

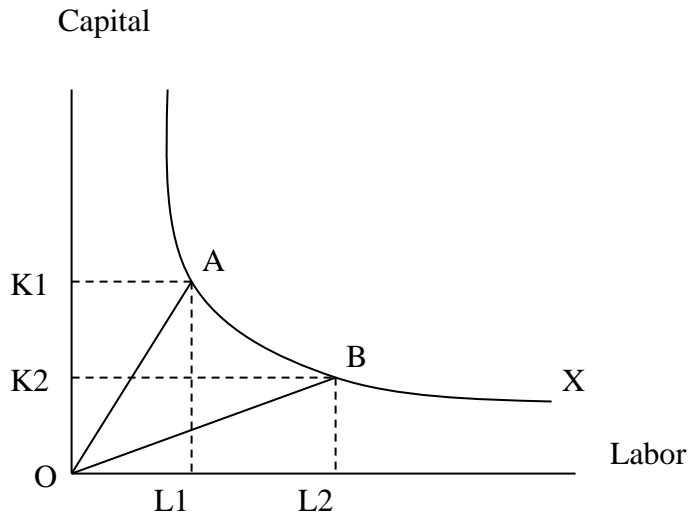


Fig 3.10 Process A uses k_1 and L_1 units of labor and capital to produce x amount of output. The factor intensity of this process can be measured by the slope of OA , which equals $AL_1/OL_1 =$

$$\frac{OK_1}{OL_1} = \frac{K_1}{L_1}$$

Similarly, factor intensity of process B is given by $\frac{K_2}{L_2}$

Since $\frac{K_1}{L_1} > \frac{K_2}{L_2}$, process A is more capital intensive than process B or B is more labor intensive than A. The upper part of the isoquant includes more capital intensive processes and the lower part, labor intensive techniques.

Now let's illustrate the above concepts with the most popular and applicable form of production function, **Cobb-Douglas** production function

The Cobb- Douglas production function is of the form

$$x = b_0 L^{b_1} K^{b_2}$$

From this production function

$$\begin{aligned}
 1. \text{MP}_L &= \frac{2X}{2L} = b_1 b_0 L^{b_1-1} K^{b_2} \\
 &= b_0 b_1 \frac{L^{b_1}}{L} K^{b_2} = b_0 \frac{X}{L} \\
 &= b_0 \cdot \text{APL}
 \end{aligned}$$

$$\begin{aligned}
 \text{MP}_K &= \frac{2K}{2L} = b_2 b_0 L^{b_1} K^{b_2-1} \\
 &= b_0 \frac{X}{L} = b_0 \cdot \text{AP}_K
 \end{aligned}$$

2. Marginal rate of technical substitution

$$(\text{MRT}_{LK}) = \frac{\text{MPL}}{\text{MPK}} = \frac{b_1 \frac{X}{L}}{b_2 \frac{X}{K}} = \frac{b_1}{b_2} \cdot \frac{K}{L}$$

3. The elasticity of substitution

$$\begin{aligned}
 \delta &= \frac{d\left(\frac{k}{l}\right) \frac{l}{k}}{d \frac{\text{MPL}}{\text{MPK}} \frac{\text{MPK}}{\text{MPL}}} \\
 &= \frac{d\left(\frac{k}{l}\right) / k/l}{d\left(\frac{b_1 k}{b_2 l}\right) / b_1 K / b_2 l} = 1
 \end{aligned}$$

4. Factor intensity is measured by the ratio b_1/b_2 . The higher the ratio, the more labor intensive the technique. Similarly, the lower the ratio $\frac{b_1}{b_2}$ the more capital intensive the technique

5. The efficiency of production. This is measured by the coefficient b_0 . Obviously it is clear that if two firms have the same K , L , b_1 and b_2 and still produce different quantities of output, the difference could be due to the superior organization and entrepreneurship of one of the firms,

which results in different efficiencies. The more efficient firm will have a larger b_0 than the less efficient one.

3.8 The efficient region of production: long run

In principle the marginal product of a factor may assume any value, positive, zero or negative. However, the basic production theory concentrates only on the efficient part of the production function, i.e. over the range of output over which the marginal product of factors are positive and declining. In the short run production function efficient region of production prevails in stage two (stage II), where $MP_L > 0$, but $\frac{\partial MP_L}{\partial L} < 0$.

Similarly, efficient region of production in the long run prevails when the marginal product of all variable inputs is positive but decreasing. Graphically this can be represented by the negatively sloped part of an isoquant. The locus of points of isoquants where the marginal products of factors are zero form the ridge lines. The upper ridge line implies that the MP of capital is zero. MP_k is negative for all points above the upper ridge line and positive for points below the ridge line. The lower ridge line implies that the MPL is zero. For all points below the lower ridge line the MPL is negative and positive for points above the line. Production techniques are technically efficient inside the ridge lines symbolically; in the long run efficient production region can be illustrated as:

$$MP_L > 0, \text{ but } \frac{\partial MP_L}{\partial L} < 0$$

$$MP_k > 0, \text{ but } \frac{\partial MP_k}{\partial K} < 0$$

Graphically, efficient region of production is shown as follow:

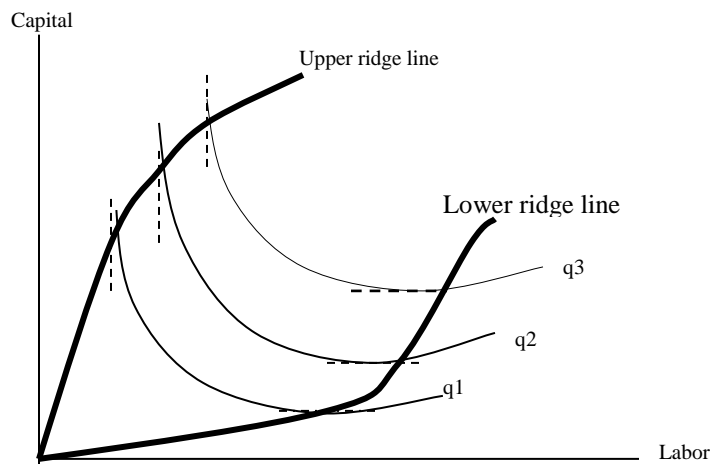


Fig 3.11: Thus efficient region of production is defined by the range of isoquants over which they are convex to the origin.

The long run law of production: The law of returns to scale

The laws of production describe the technically possible ways of increasing the level of production. Output may increase in various ways. In the long run output can be increased by changing all factors of production. This long run analysis of production is called *Law of returns to scale*.

In the short run output may be increased by using more of the variable factor, while capital (and possibly other factors as well) are kept constant. The expansion of output with one factor (at least) constant is described by the *law of variable proportion or the law of (eventually) diminishing returns of the variable factor*.

Now let's have a deep examination of law of returns to scale.

3.9 Laws of returns to scale: long run analysis of production

In the long run all inputs are variable. Expansion of output may be achieved by varying all factors of production by the same proportion or by different proportions.

The traditional theory of production concentrates on the first case, i.e. the study of output as all inputs change by the same proportion. The term returns to scale refers to the change in output as all factors change by the same proportion. Suppose initially the production function is

$$X_0 = f(L, K)$$

If we increase all factors by the same proportion t , we clearly obtain a new level of output X^* where,

$$X^* = f(tL, tK)$$

- If X^* increases by the same proportion t or if $X^* = tX_0$, we say that there is constant returns to scale.
- If X^* increases less than proportionally with the increase in the factors (or if X^* increases by a proportion less than t), we have decreasing returns to scale.
- If X^* increases more than proportionally with the increase in the factors (by a more than t proportion), we have increasing returns to scale.

Returns to scale and homogeneity of production function

Suppose we increase both factors of the function $X_0 = f(L, K)$ by the same proportion ' t ', and we get the new level of output $X = f(tL, tK)$

If t can be factored out (that is, may be taken out of the brackets as a common factors), then the new level of output X^* can be expressed as a function of t (to any power V) and the initial level of output, and the production function is said to be homogeneous.

$$X^* = t^V f(L, K) \text{ or } X^* = t^V X_0$$

If t can not be factored out, the production function is non-homogeneous. Thus, a homogeneous function is a function such that if each of the input is multiplied by t , then t can be completely factored out of the function. The power V of t is called degree of homogeneity of the function and is measure of returns to scale.

If $V=1$, we have constant returns to scale. This production function is some times called linear homogeneous

If $V < 1$, decreasing return to scale prevails

If $V > 1$, increasing return to scale prevails

For a Cobb-Douglas production function

$$X = b_0 L^{b_1} K^{b_2}, V = b_1 + b_2 \text{ and it is a measure of returns to scale.}$$

Proof: Let L and K increase by t . The new level of output is

$$X^* = b_0 (tL)^{b_1} (tk)^{b_2}$$

$$X^* = b_0 t^{b_1} l^{b_1} t^{b_2} k^{b_2}$$

$$X^* = b_0 L^{b_1} K^{b_2} t^{b_1+b_2}$$

$$X^* = X (t^{b_1+b_2})$$

$$\text{Thus } V = b_1+b_2$$

Exercise

1. Which of the following production function is/are homogeneous?

A. $q = L+K$

B. $q = 10L +K$

C. $q = L^\alpha L^\alpha K^\beta +K$

D. $q = L^2+K$

2. Suppose the production function is

$$q = L^{\frac{3}{4}} K^{\frac{1}{4}}$$

A. what is APL, holding capital fixed?

B. Calculate MPL & MPK?

C. Does this production function have increasing, constant or decreasing returns to scale?

For a homogeneous production function the returns to scale may be represented graphically in an easy way. Before explaining the graphical representation of the returns to scale it is useful to introduce the concept of product line and isocline.

3.10 Product Lines

A product line shows the (physical) movement from one isoquant to another as we change the employment of both factors or a single factor, and it describes the technically possible alternative paths of expanding output. Which path actually chosen by the firm will depend on the prices of factors.

The product line (curve) passes through the origin if all factors are variable. If only one factor is variable (the other being constant) the product line is a straight line parallel to the axis of the variable factor. For such product lines the K/L ratio diminishes along the product line.

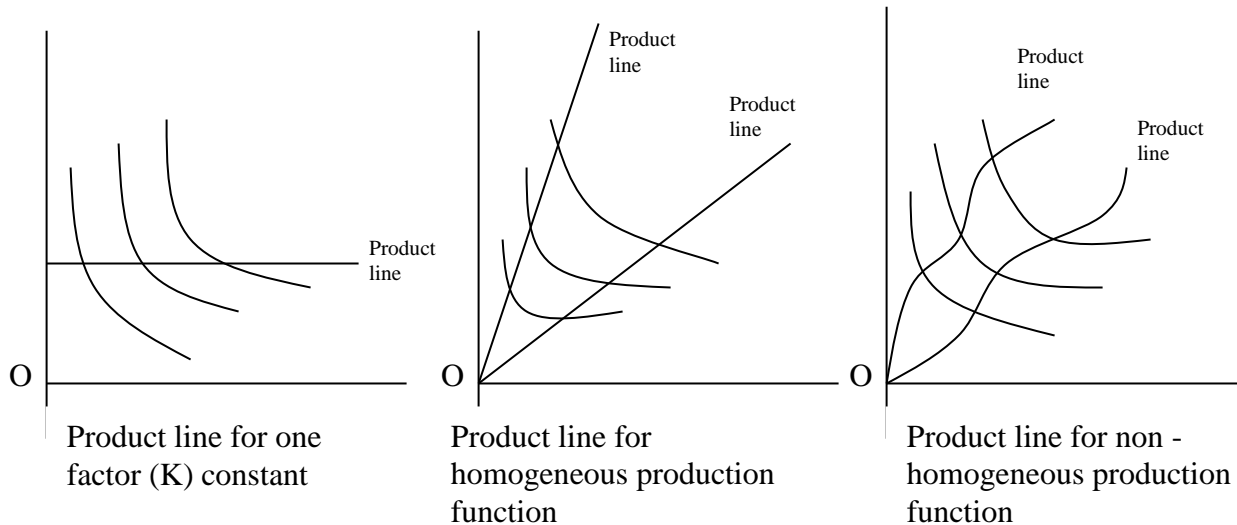


Fig 3.12

A product line along which the MRTS of factors is constant is called an isocline. So an isocline is the locus of points of different isoquants at which MRTS of factors is constant. If the production function is homogeneous the isoclines are straight lines through the origin. Along any one isocline the K/L ratio is also constant (as is the MRTS of the factors). But K/L ratio and the MRTS are different for different isoclines.

If the production is non homogeneous the isocline will not be straight line, but their shape will be twiddle. The K/L ratio changes along each isocline and on different isoclines.

3.11 Graphical presentation of returns to scale for homogeneous production function

The returns to scale may be shown graphically by the distance (on an isocline) between successive “multiple level-of-output” isoquants, i.e. isoquants that show levels of output which are multiple of some base level such as X , $2X$, $3X$ etc.

Constant returns to scale

Along any isocline the distance between successive multiple- isoquant is *constant*. Doubling the factor inputs doubles the level of initial output; trebling inputs trebles output, and so on.

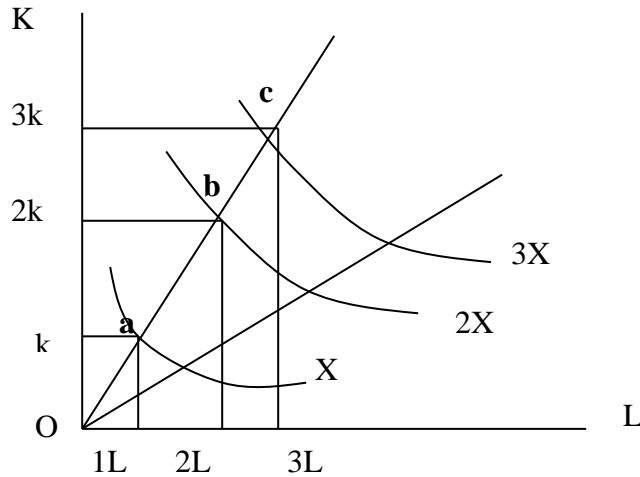


Fig 3.13 Constant returns to scale: $oa = ab = bc$

Decreasing returns to scale

Here, the distance between consecutive multiple- isoquants increases. By doubling inputs output increases by less than twice of its original level.

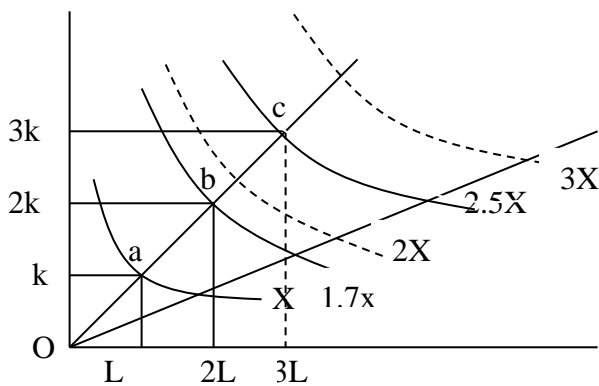


Fig 3.14 Constant returns to scale: $oa < ab < bc$

Increasing returns to scale

The distance between consecutive multiple isoquants decrease, by doubling the inputs, output is more than doubled.

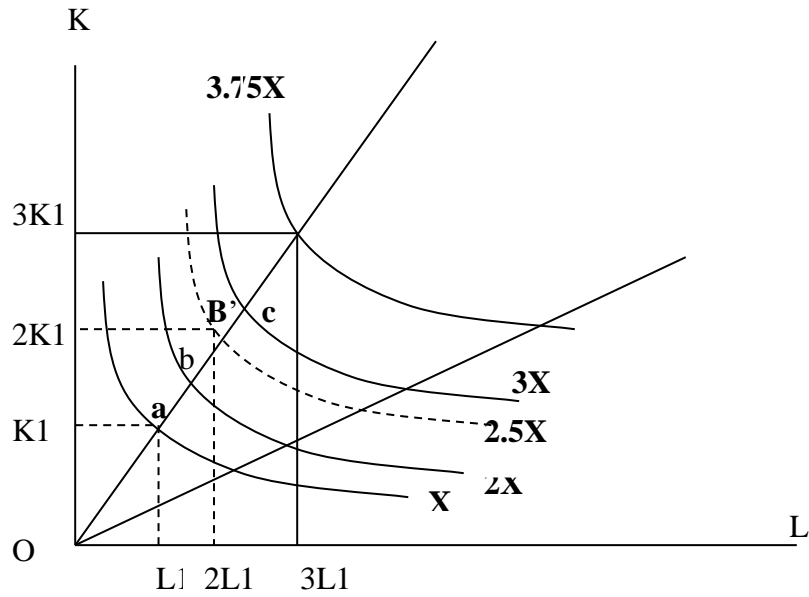


Fig 3.15 Doubling K and L leads to B' which lies above an isoquant denoting $2X$ (i.e., $2.5X$), and trebling K & L results in an isoquant which lies above $3X$ (i.e., $3.75X$) and so on.

Returns to scale are usually assumed to be the same every where on the production surface i.e., the same along all the expansion product lines. All processes are assumed to show the same returns to scale over all ranges of output. Either constant returns to scale every where, or decreasing returns every where, or increasing returns everywhere. However, the technological conditions of production may be such that returns to scale may vary over different ranges of output. Over some range we may have constant returns to scale, while over another range we may have increasing or decreasing returns to scale.

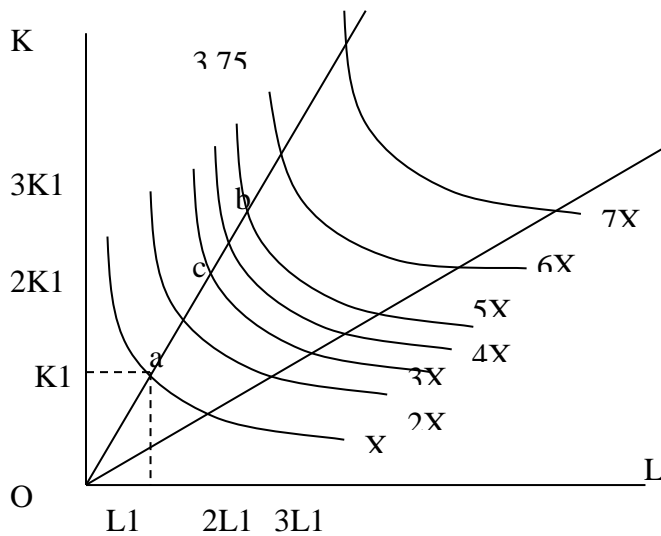


Fig3.16 Up to point C, increasing returns to scale prevails in the firm, from C to B constant returns to scale prevails, and beyond B decreasing returns to scale prevails.

Causes of increasing returns to scale

Technical and /or managerial indivisibility. Mostly, processes of production can be doubled but it may not be possible to half them. When the production system expands, workers will specialize in one extreme and their productivity increases.

Causes of decreasing returns to scale: The most common causes are ‘diminishing returns to management’. If we expand the out put beyond optimum, the top management personnel will be over burdened and the productivity of additional unit of the variable inputs decline eventually. E.g., doubling fishing fleet may not double fish catch.

3.12 Technological process and production function

Technological improvement (progress) makes factors of production more productive or it makes production system more efficient; so that the firm will get higher output from the same combinations of labor and capital than before. Graphically, this can be shown by **upward** movement of the total product curve (indicating higher output level can be achieved from the same input) and **down** ward movement of isoquant denoting lower combinations of factors of production can produce equal level of output. See the figures

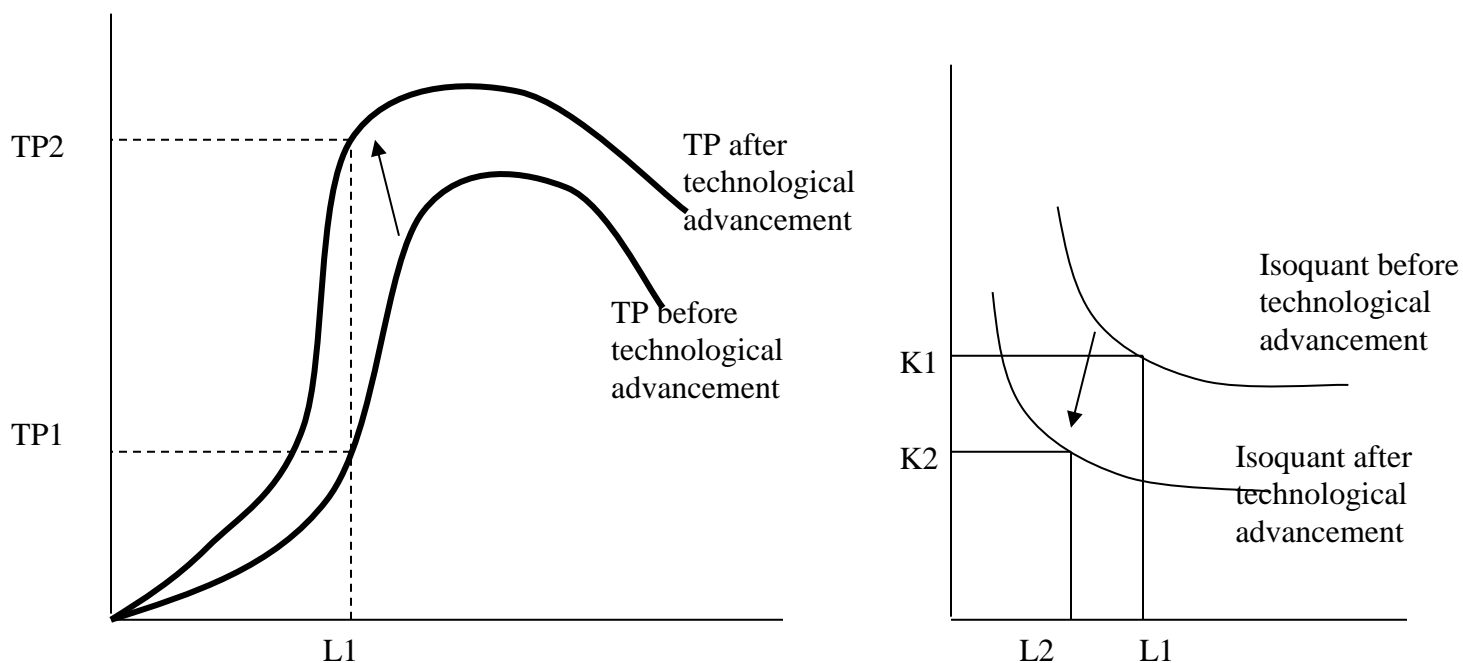


Fig 3.17 technological progress shifts the TP curve up ward &the isoquant down ward

3.13 Equilibrium of the firm: Choice of optimal combination of factors of production

Dear learner, in our previous discussion we have said that an isoquant denotes efficient combination of labor and capital required to produce a given level of out put. But, this does not mean that the monetary cost of producing a given level of out put is constant along an isoquant. That is, though different combinations of labor and capital on a given isoquant yield the same level of out put, the cost of these different combinations of labor and capital could differ because the prices of the inputs can differ. Thus, isoquant shows only technically efficient combinations of inputs, not economically efficient combinations. Technical efficiency takes in to account the physical quantity of inputs where as economic efficiency goes beyond technical efficiency and seeks to find the least cost (in monetary terms) combination of inputs among the various technically efficient combinations. Hence, technical efficiency is a necessary condition, but not a sufficient condition for economic efficiency. To determine the economically efficient input combinations we need to have the prices of inputs.

To determine the economically efficient input combination, the following simplifying assumptions hold true:

Assumptions

1. The goal of the firm is maximization of profit (Π) where $\Pi = R - C$
Where Π -Profit, R-revenue and C-is cost outlay.
2. The price of the product is given and it is equal to \bar{P}_x .
3. The prices of inputs are given (constant).Price of a unit of labor is \bar{w} and that of capital is \bar{r} .

Now before we go to the discussion of optimal input combination (or economically efficient combination), we need to know the isocost line, because optimal input is defined by the tangency of the isoquant and isocost line.

Isocost line

Dear learner, do you remember what the budget line denotes?

Isocost lines have most of the same properties as that of budget lines, an isocost line is the locus points denoting all combination of factors that a firm can purchase with a given monetary outlay, given prices of factors.

Suppose the firm has C amount of cost out lay (budget) and prices of labor and capital are w and r respectively. The equation of the firm's isocost line is given as:

$$C = rK + wL \quad , \text{where } K \text{ and } L \text{ are quantities of capital and labor respectively.}$$

Given the cost outlay C , the maximum amounts of capital and labor that the firm can purchase are equal to $\frac{C}{r}$ and $\frac{C}{w}$ respectively. The straight line that connects these points is the iso-cost line. See the following figure:

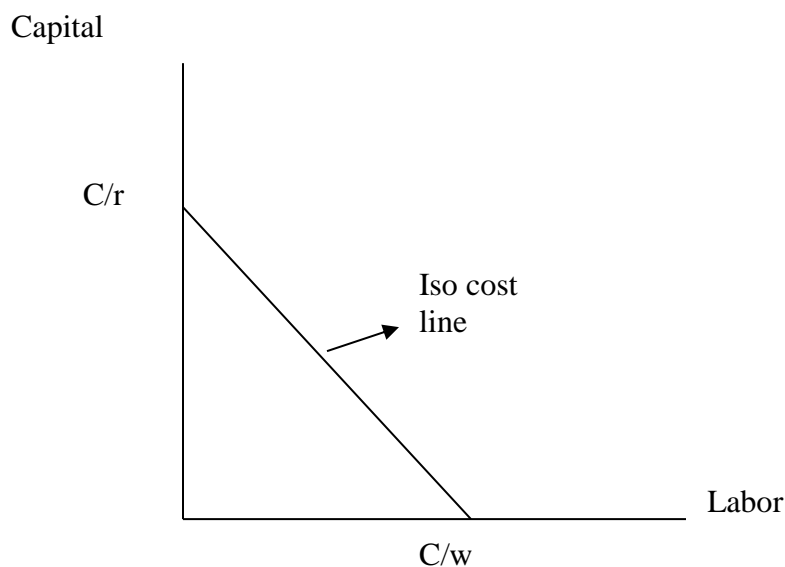


Fig: 3.18 the iso cost line: shows different combinations of labor and capital that the firm can buy given the cost out lay and prices of the inputs.

Now we are in a position to determine the firm's optimal in put combination. However, the problem of determining optimal input combination (economic efficiency) takes two forms. Some times, situations may happen when a firm has a constant cost outlay and seek to maximize its out put, given this constant and cost out lay and prices of in puts. Still, there are also situations when

the goal of the firm is to produce a predetermined (given) level of output with the least possible cost. Under we will discuss the two situations separately.

Case1: Maximization of output subject to cost constraint

Suppose a firm having a fixed cost out lay (money budget) which is shown by its iso-cost line. Here, the firm is in equilibrium when it produces the maximum possible out put, given the cost outlay and prices of input. The equilibrium point (economically efficient combination) is graphically defined by the tangency of the firm’s iso-cost line (showing the budget constraint) with the highest possible isoquant. At this point, the slope of the iso cost line ($\frac{w}{r}$) is equal to the slope

of the isoquant ($\frac{MP_L}{MP_K}$).

The condition of equilibrium under this case is, thus:

$$\frac{w}{r} = \frac{MP_L}{MP_K} \quad or \quad \frac{MP_L}{w} = \frac{MP_K}{r}$$

This is the first order (necessary) condition. The second order (sufficient) condition is that isoquant must be convex to the origin. See the following figure:

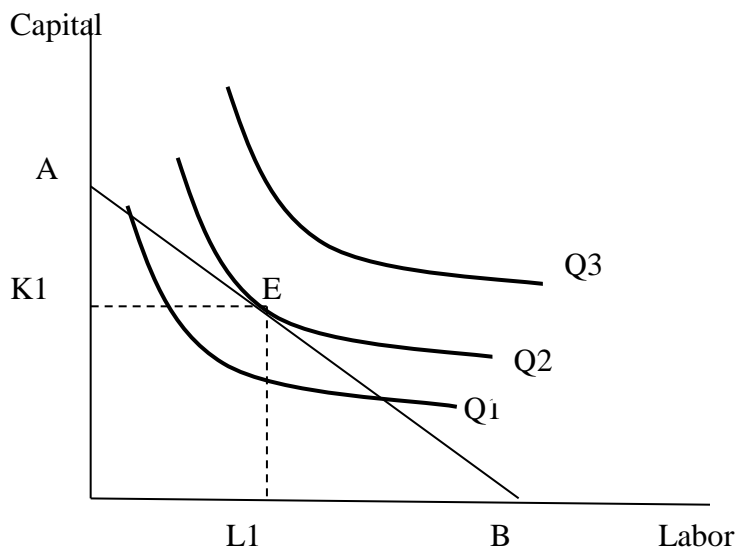


Fig: 3.19 the optimal combination of inputs (L_1 and K_1) is defined by the tangency of the iso-cost line (AB) and the highest possible isoquant (Q_2), at point E. At this point the slope of iso-cost line

$(\frac{w}{r})$ is equal to the slope of isoquant $X_2(\frac{MP_L}{MP_K})$. The second order condition is also satisfied by the convexity of the isoquant.

Dear learner, do you think that the point of tangency of the iso-cost line and the isoquant represents equilibrium point when the isoquant is concave?

If isoquant is concave to the origin the point of tangency of the iso cost line and the isoquant does not define the equilibrium combination of factor inputs. With a concave isoquant, we have a corner solution. Refer the figure below:

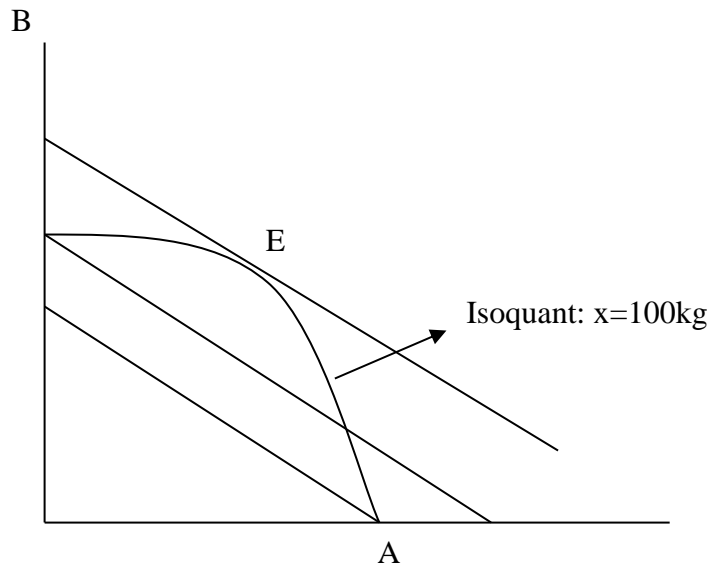


Fig: 3.20 Concave isoquant results in corner solution. The point of tangency between the isoquant and the iso-cost line does not satisfy the second order condition as the isoquant is concave. The same level of output($X=100Kg$) can be produced with a lower cost out lay at point A.

Mathematical derivation of the equilibrium condition

The problem can be stated as:

Maximize $X = f(L, K)$Objective function
 Subject to $\bar{C} = wL + rK$Constraint function
 or $\bar{C} = wL + rK - C = 0$

We use the lagrangian method to solve the problem.

The lagrangian equation is written as:

$$\phi = X - \lambda(\bar{C})$$

Then we find $\frac{\partial \phi}{\partial L}$, $\frac{\partial \phi}{\partial K}$, and $\frac{\partial \phi}{\partial \lambda}$ and set all of them equal to zero to solve for L and K .

That is,

$$\phi = X - \lambda(wL + rK - \bar{C})$$

$$\begin{aligned} \text{And, } \frac{\partial \phi}{\partial L} &= \frac{\partial X}{\partial L} - w\lambda = 0 \Rightarrow MP_L = w\lambda \Rightarrow \lambda = \frac{MP_L}{w} \\ \frac{\partial \phi}{\partial K} &= \frac{\partial X}{\partial K} - r\lambda = 0 \Rightarrow MP_K = r\lambda \Rightarrow \lambda = \frac{MP_K}{r} \\ \frac{\partial \phi}{\partial \lambda} &= -wL - rK + \bar{C} = 0 \Rightarrow wL + rK = \bar{C} \end{aligned}$$

Solving these equations simultaneously, we obtain the equilibrium condition

$$\frac{MP_L}{w} = \frac{MP_K}{r} \quad \text{or} \quad \frac{w}{r} = \frac{MP_L}{MP_K}$$

The second order condition (the convexity of isoquant) would be insured when:

$$\frac{\partial^2 X}{\partial L^2} < 0, \quad \frac{\partial^2 X}{\partial K^2} < 0 \quad \text{and} \quad \left(\frac{\partial^2 X}{\partial L^2} \right) \left(\frac{\partial^2 X}{\partial K^2} \right) < \left(\frac{\partial^2 X}{\partial L \partial K} \right)^2$$

Numerical Example

Suppose the production function of a firm is given as $X = 0.5L^{1/2}K^{1/2}$ prices of labor and capital are given as \$ 5 and \$ 10 respectively, and the firm has a constant cost out lay of \$ 600. Find the combination of labor and capital that maximizes the firm's out put and the maximum out put.

Solution

The condition of equilibrium is $\frac{MP_L}{w} = \frac{MP_K}{r}$ or $\frac{MP_L}{MP_K} = \frac{w}{r}$

$$MP_L = \frac{\partial X}{\partial L} = 0.25L^{-1/2}K^{1/2}$$

$$MP_K = \frac{\partial X}{\partial K} = 0.25L^{1/2}K^{-1/2}$$

Thus, the equilibrium exists when,

$$\frac{0.25L^{-1/2} K^{-1/2}}{0.25L^{-1/2} K^{-1/2}} = \frac{\$5}{\$10}$$

$$\frac{K}{L} = \frac{1}{2} \Rightarrow L = 2K \dots \dots \dots (1)$$

The constraint equation is:

$$wL + rK = C$$

$$5L + 10K = 600 \dots \dots \dots (2)$$

Solving equation (1) and (2) would give us the optimal combination of L and K.

$$L = 2K$$

$$5L + 10K = 600$$

$\Rightarrow L=60$ units and $K=30$ units.

Thus, the firm should use 60 units of labor and 30 units of capital to maximize its production (output). (Check the second order condition).

The maximum output can be found by substituting 60 and 30 for L and K in the production process.

Case -2: Minimization of cost for a given level of output

In this case, consider an entrepreneur (a firm) who wants to produce a given output (for example a bridge or a building or x tones of a commodity) with minimum cost outlay. That is, we have a single isoquant which denotes the desired level of output, but there are a set of isocost lines which denotes the different cost outlays. Higher isocost lines denote higher production costs. The production costs of a desired level of output will therefore be minimized when the isoquant line is tangent to the lowest possible isocost line (see fig) At the point of tangency, the slope of

the isoquant and isocost lines are identical. That is $\frac{w}{r} = \frac{MPL}{MPK}$

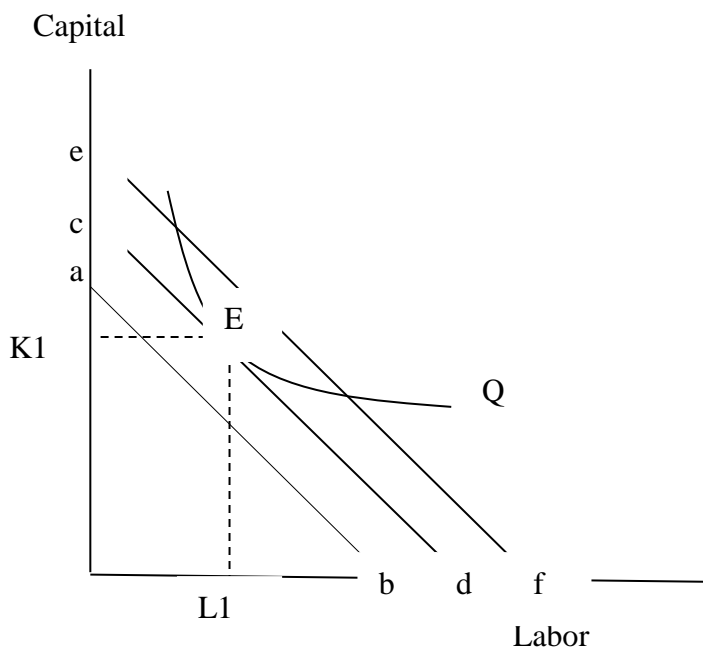


Fig: 3.21 the equilibrium combination of factors is K_1 and L_1 amounts of capital and labor respectively. Lower isocost lines such as 'ab' are economically desirable but unattainable given the desired level of output. So point E shows the least cost combination of labor and capital to produce X amount of output.

Now let us see the mathematical derivation of the equilibrium condition. As mentioned earlier, we minimize the cost of producing a given level of output.

Thus, the problem can be stated as:

$$\text{Minimize } C = f(q) = WL + rK \text{ ----- (Objective function)}$$

$$\text{Subject to } q = f(L, K) \text{ ----- (Constraint function)}$$

$$\text{Or } f(L, K) - q = 0$$

We use the LaGrange an method to obtain the equilibrium condition. Accordingly, the LaGrange an function will be:

$$\phi = C - \lambda(f(L, K) - q)$$

$$\phi = WL + rK - \lambda(f(L, K) - q)$$

The condition of equilibrium will be obtained by finding $\frac{\partial \phi}{\partial L}$, $\frac{\partial \phi}{\partial K}$ and $\frac{\partial \phi}{\partial \lambda}$ and then solving them simultaneously after equating each to zero.

That is

$$\frac{\partial \phi}{\partial L} = w - \lambda \frac{\partial f(L, K)}{\partial L} = 0$$

$$w - \lambda MPL = 0 \Rightarrow \lambda = \frac{w}{MPL}$$

$$\frac{\partial \phi}{\partial K} = r - \lambda \frac{\partial f(L, K)}{\partial K} = 0$$

$$r - \lambda MPK = 0 \Rightarrow \lambda = \frac{r}{MPK}$$

$$\frac{\partial \phi}{\partial \lambda} = f(L, K) - q = 0$$

Thus, the equilibrium condition is $\frac{w}{MPL} = \frac{r}{MPK}$

Rearranging the above condition, we obtain $\frac{MPL}{MPK} = \frac{w}{r}$

This condition $\left(\frac{MPL}{MPK} = \frac{W}{r}\right)$ is only a necessary condition .

The sufficient condition is that the isoquant must be convex to the origin. That is

$$\frac{\partial^2 q}{\partial L^2} < 0, \frac{\partial^2 q}{\partial k^2} < 0 \text{ and } \left(\frac{\partial^2 q}{\partial L^2}\right)\left(\frac{\partial^2 q}{\partial k^2}\right) > \left(\frac{\partial^2 X}{\partial L \partial K}\right)^2$$

Numerical example:

Suppose a certain contractor wants to maximize Π from building one bridge. The contractor uses both labor and capital, and efficient combinations of Labor and capital that are sufficient to make

a bridge is by the function $0.25 L^{\frac{1}{2}} K^{\frac{1}{2}}$. If the prices of labor (w) and capital (r) are \$ 5 and \$ 10 respectively.

Find the least cost combination of L and K, and the minimum cost.

Solution:

The contractor wants to build one bridge. Thus, the constraint equation can be written as $0.25 L^{\frac{1}{2}} k^{\frac{1}{2}} = 1$

$$MPL = 0.125 L^{-\frac{1}{2}} K^{\frac{1}{2}}$$

$$MPK = 0.125 L^{\frac{1}{2}} K^{-\frac{1}{2}}$$

The equilibrium condition is $\frac{MPL}{MPK} = \frac{W}{r}$

$$\frac{0.125 L^{-\frac{1}{2}} K^{\frac{1}{2}}}{0.125 L^{\frac{1}{2}} K^{-\frac{1}{2}}} = \frac{\$5}{\$10}$$

$$\frac{K}{L} = \frac{1}{2} \Rightarrow L = 2K$$

Substituting $L = 2K$ in the constraint equation we obtain

$$0.125 (2K)^{\frac{1}{2}} K^{\frac{1}{2}} = 1$$

$$0.125 \sqrt{2} \cdot K = 1$$

$$K = \frac{1}{0.125\sqrt{2}} = K = \frac{8}{\sqrt{2}}$$

$$L = 2K \Rightarrow \frac{16}{\sqrt{2}}$$

Therefore, efficient combination (least cost combination) of L and K are $\frac{16}{\sqrt{2}}$ and $\frac{8}{\sqrt{2}}$ respectively.

$$\text{The least cost is } C = 5 \left(\frac{16}{\sqrt{2}} \right) + 10 \left(\frac{8}{\sqrt{2}} \right) = \$ \frac{160}{\sqrt{2}}$$

Cost minimization with varying out put levels and the derivation of long run total cost curve

Dear learner, in the previous chapter we saw how a cost minimizing firm selects a combination of inputs to produce a given level of out put. Now we extend this analysis to see how the firm's costs depend on its out put level. To do best with, refer to the following figure

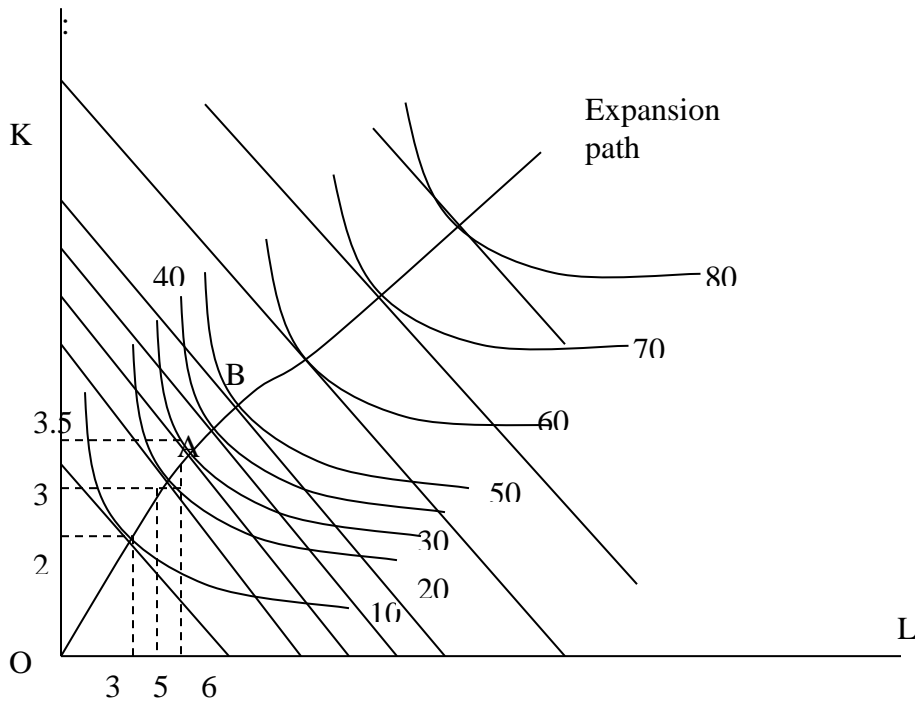


Fig 3.22

In the above figure, the firm is assumed to have increasing returns to scale up to point A (the distance between successive multiple levels of out put decreases), constant returns to scale between points A and B (the distance between successive multiple levels of out puts is constant) and decreasing returns to scale beyond point B (the distance between successive multiple levels of out put increases).

The curve passing through the points of tangency between the firm's iso cost lines and its isoquants is called *expansion path*. The expansion path denotes least cost combination of labor and capital required to produce different levels of out put.

To produce 10 units of out put, the firm uses 2K and 3L, to produce 20 units, the firm uses 3K and 5L, to produce 30 units it uses 3.5K and 6L, and so on. Hence, as out put expands at a constant amount, the units of labor and capital increases at a decreasing rate up to point A. That is, total cost of production increases at a decreasing rate up to point A. From point A to B, combination of labor and capital increase at a constant rate as out put increases at a constant rate. Hence, the long run total cost increases at a constant rate up to point B.

Beyond point B, to expand output at a constant rate, combination of labor and capital should be increased at an increasing rate. Assuming that the prices of inputs are constant, the long run total cost of production increases rapidly (at increasing rate) beyond point B. From the above discussion, we infer that the long run total cost curve assumes an inverse S-shape.

Summary

In economics, production means the act of creating those goods or services that have exchange values. The process of production requires inputs such as land, labor, capital and entrepreneurial ability. Fixed inputs are those inputs whose quantity can not readily be changed when market conditions indicate that an immediate change in output is required, Variable inputs are those inputs whose quantity can be changed almost instantaneously in response to derived changes in output.

A production function describes the maximum output that a firm can produce for each specified combinations of inputs. In the short run, one or more inputs to the production process are fixed. In the long run, all inputs are potentially variable.

In the short run the firm is said to be efficient when it operates over the range of employment of the variable inputs where marginal product of the variable input is positive but declining.

Analysis of long run production makes the use of isoquants (equal product lines). Isoquants show efficient combination of inputs that yield the firm equal level of output. Isoquants are always downward sloping. Higher isoquants show higher level of output. Isoquants can not intersect each other. They are convex to the origin, and they must be thin.

In the long run, efficient region of production is represented by the convex part of isoquants map.

UNIT FOUR

THEORY OF COSTS OF PRODUCTION

Introduction

In this unit, you will study the meaning and behaviors of costs of production, the relationship between production (output) and costs (i.e. cost function both in the short run and long run.)

Objectives

After successful completion of this unit, you will be able to:

- Explain different ways of measuring private costs, i.e. economic costs vs. accounting costs.
- Define the meaning and nature of cost functions both in the short run and long run.
- Explain the relationship between short run production function and short run cost function.
- State how learning and experience affects the costs of production.

4.1 Basic concepts

Dear learner! Do you know about costs?

(Use the space left below to write your response)

To produce goods and services, firms need factors of production or simply inputs. To acquire these inputs, they have to buy them from resource suppliers. Cost is, therefore, the monetary value of inputs used in production of an item.

We can identify two types of cost of production: *social cost* and *private cost*.

Social cost: is the cost of producing an item to the society. This cost is realized due to the fact that most resources used for production purpose are scarce and some production process, by their nature, emit dangerous chemicals, bad smell, etc to surrounding society.

For example, when a certain beer factory wants to produce beer in Ethiopia, the society as a whole also incurs a cost. Because, the next- best alternative of the raw material (such as barely) used for the production of beer is sacrificed. When the beer factories buy barley from the market, the amount of barely available for consumption by society may be reduced and the price may become dearer. Hence, the production of beer imposes an indirect cost on the society, moreover, by its nature; the production of beer emits bad chemicals to the environment, which pollutes waters, air, etc. To control the understandable consequences of the production process on the environment and their property, the society incurs cost.

Private cost: This refers to the cost of producing an item to the individual producer. It is the cost that the beer factory incurs to produce the beer, in our example:

Private cost of production can be measured in two ways:

i) Economic cost

In economics the cost of production to the individual producer includes the cost of all inputs used for the production of the item.

The producer may buy part of the inputs from the market. *For example*, he/ she hire workers, buy raw materials, the necessary machines, etc. the actual or out- of- pocket expenditures that the firm incurs to purchase these inputs from the market are called explicit costs.

But, the producer can also use his/ her own inputs which are not purchased from the market for the production purpose. For example, the producer may use his/ her own building as a production place, he/she may also manage his firm by himself instead of hiring another manager, etc. since these inputs are used for the purpose production, their value has to be estimated and included in the total cost of production. As to how to estimate the cost of these non- purchased inputs is concerned, we usually estimate their cost from what these inputs could earn in their best alternative use. For instance, if the firm uses his own building for production purpose, the cost of using this building for production is estimated by the rent income foregone. If the producer is a teacher with salary of 1000 birr per month and fruits his job to manage his factory, then the next best alternative of his labor is the salary that he sacrificed to be the manager of his factory. The estimated cost of there non- purchased inputs are called implicit costs.

Thus, in economics the cost of production includes the costs of all inputs used in the production process whether the inputs are purchased from the market or owned by the firm himself that is:

Economic cost: Explicit cost plus Implicit cost

ii) Accounting Cost

For accountant, the cost of production includes the cost of purchased inputs only. Accounting cost is the explicit cost of production only. More over, accountant's doesn't consider the cost of production from the opportunity cost of the resources point of view. To clarify the difference between accounting cost and economic cost on this regard, consider the following ***example***.

Suppose Bedele Brewery factory purchases 1000 quintals of barely for 200 birr per quintal in 1998 to use this barley for production purpose in the year 1999. However, suppose that the price of the barely has been increased to 300 birr per quintal in the year 1999.

-Now shall we use the actual price with which the barely was bought in 1998 or the current price (1999 price) to estimate the cost of barely in 1999?

In economics, the 1999 price should be taken because, though the barley was bought for 200 birr per quintal in 1998, the cost of using this barely for the production purpose in 1999 is the 300 birr per quintal, the amount of income that could be obtained if the barely were sold in the market.

But accountants use the 1998 price to estimate the cost of production in the year 1999.

4.2 Cost functions

Cost function shows the algebraically relation between the cost of production and various factors which determine it. Among others, the cost of production depends on the level of output produced, technology of production, prices of factors, etc. hence; cost function is a multivariable function.

Symbolically,

$$C = f(x, t, p_i)$$

Where c- is total cost of production

x - is the amount of output

T – is the available technology of production.

P_i – is the price of input

Graphically, cost functions can be illustrated by using a two- dimension diagrams. To do so, first we observe the relationship between the total cost of production and the level of output (the most factor determining the cost of production), by assuming that all other factors are constant. Then, the impact of change in “other factors” such as technology on the cost of production will be handled by shifting the total cost curves up ward or down- ward.

4.3 – Short run vs. long run costs

Dear learner! What do we mean by short run and long run in economics?

(Use the space left below to write your response)

Economics theory distinguishes between short run costs and long run costs. Short run costs are the costs over a period during which some factors of production (usually capital equipments and management) are fixed. The long- run costs are the cost over a period long enough to permit the change of all factor of production.

Short run costs of the traditional theory

In the traditional theory of the firm, total costs are split into two groups: total fixed costs and total variable costs:

$$TC = TFC + TVC$$

Where – TC is short run total cost

TFC is short run total fixed cost

TVC is short run total variable cost

By fixed costs, we mean a cost which doesn't vary with the level of out put. The fixed costs include:

- a. Salaries of administrative staff
- b. Expenses for building depreciation and repairs
- c. Expenses for land maintenance
- d. The rent of building used for production , etc

All the above costs are regarded as fixed costs because whether the firm produces much output or zero out put, these costs are unavoidable, and the firm can avoid fixed costs only if he / she shuts down the business stops operation.

Variable costs, on the other hand, include all costs which directly vary with the level of output.

The variable costs include:

- e. The cost of raw materials
- f. The cost of direct labor
- g. The running expenses of fixed capital such as fuel, electricity power, etc.

All these costs are regarded as variable costs because their amount depends on the level of output.

For example, if the firm produces zero output, the variable cost is zero.

Graphical presentation of short run costs.

Total fixed cost (TFC)

Graphically, TFC is denoted by a straight line parallel to the output axis. The point of intersection of the TFC line with the cost axis (vertical axis) shows the amount of the fixed. For example if the level of fixed cost is \$ 100, it can be shown as.

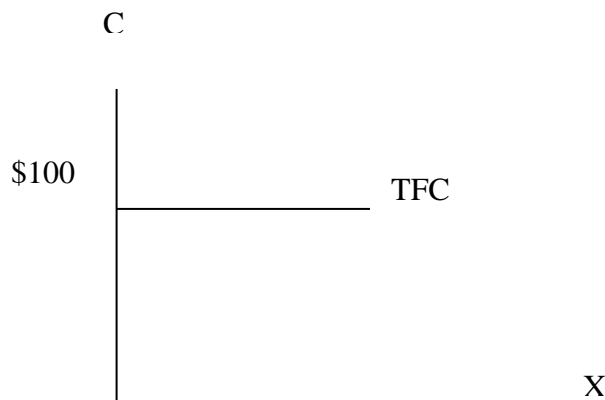


Fig 4.1

Total variable cost (TVC)

The total variable cost of a firm has an inverse s- shape. The shape indicates the law of variable proportions in production. According to this law, at the initial stage of production with a given plant, as more of the variable factor (s) is employed, its productivity increases. Hence, the TVC increases at a decreasing rate. This continues until the optimal combination of the fixed and variable factors is reached. Beyond this point, as increased quantities of the variable factors(s) are combined with the fixed factor (s) the productivity of the variable factor(s) declined, and the TVC

increases by an increasing rate. Thus, the TVC has an inverse s-shape due to the law of diminishing marginal returns.

Graphically, the TVC looks the following.

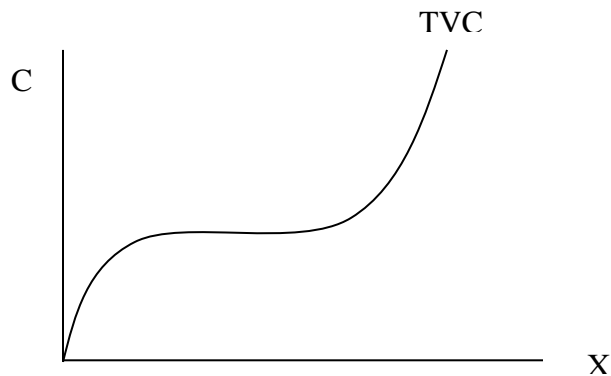


Fig 4.2

Total Cost (TC)

The total cost curve is obtained by vertically adding the TFC and the TVC i.e., by adding the TFC and the TVC at each level of output. The shape of the TC curve follows the shape of the TVC curve. i.e. the TC has also an inverse S-shape. But the TC curve doesn't start from the origin as that of the TVC curve. The TC curve starts from the point where the TFC curve intersects the cost axis.

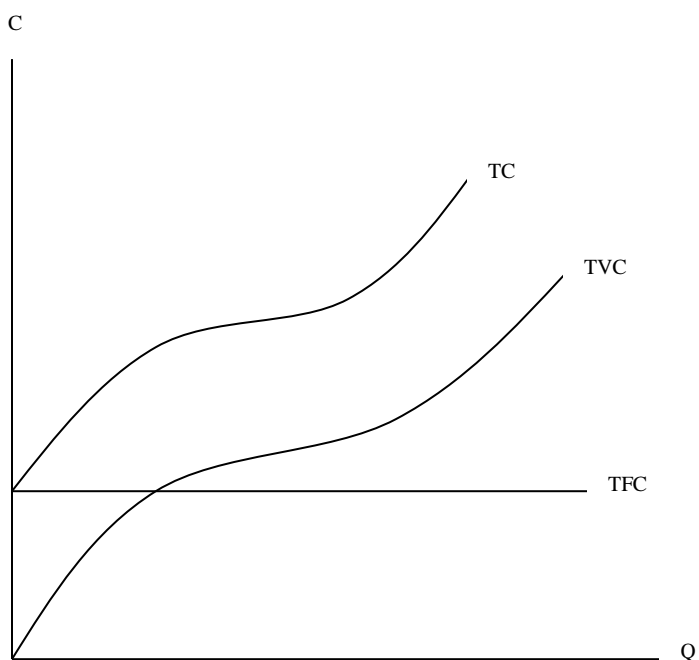


Fig 4.3 the TC and TVC curves has an inverse S- shape. The vertical distance between them (TFC) is constant.

Per unit costs (average costs)

From total costs we can derive per-unit costs. These are even more important in the short run analysis of the firm. Average fixed cost (AFC) - is found by dividing the TFC by the level of output.

Graphically, the AFC is a rectangular hyper parabola. The AFC curve is continuously decreasing curve, but decreases at a decreasing rate and can never be zero. Thus, AFC gets closer and closer to zero as the level of output increases, because a fixed amount of cost is being divided by increasing level of output.

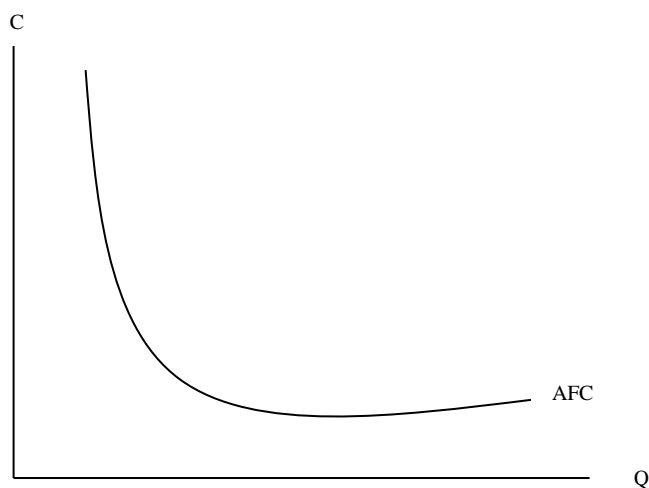


Fig 4.4 the average fixed cost curve is derived from the total fixed cost, and it represents the slope of straight lines drawn from the origin to a given point on the TFC curve.

Average variable cost (AVC)

The AVC is similarly obtained by dividing the TVC with the corresponding level of output.

$$AVC = \frac{TVC}{X}$$

Graphically, the AVC at each level of output is derived from the slope of a line drawn from the origin to the point on the TVC curve corresponding to the particular level of output.

The following graph clearly shows the process of deriving the AVC curve from the TVC curve.

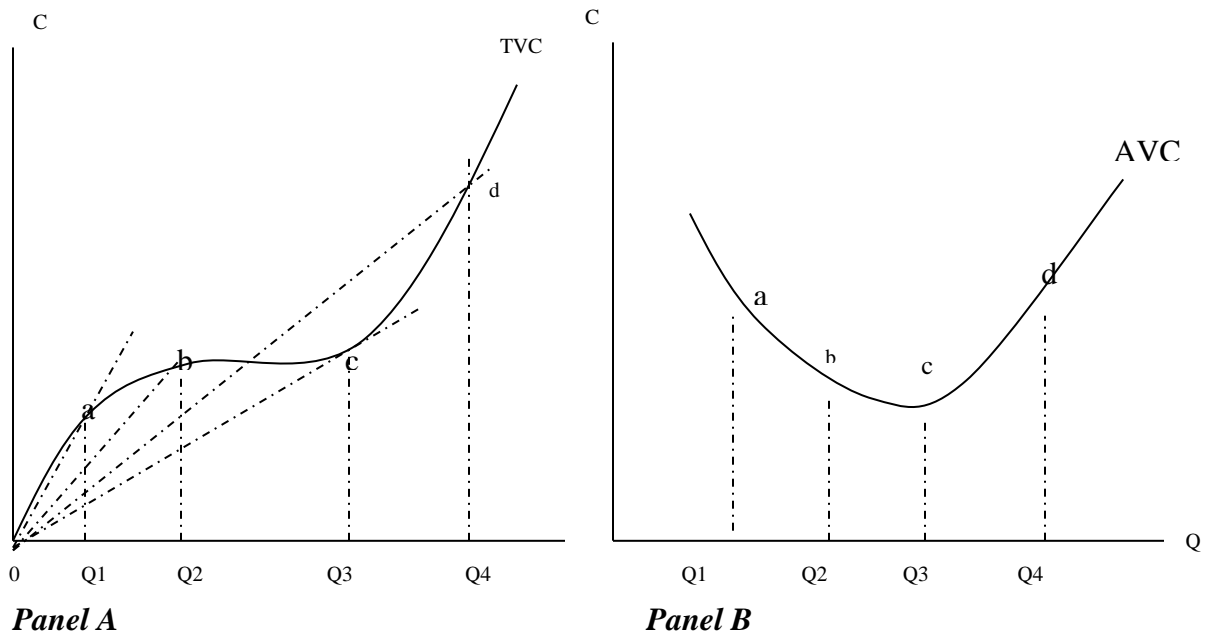


Fig 4.5 in the figure above, the AVC at Q_1 from panel A is given by the slope of the ray $0a$, the AVC at Q_2 is given by slope of the ray $0b$, and so on. The slope of the rays decreases until Q_3 and starts to rise beyond Q_3 .

It is clear from this figure that the slope of a ray through the origin declines continuously until the ray becomes tangent to the TVC curve at C. To the right of this point (Point c) the slope of the rays through the origin starts increasing. Thus, the short run AVC (SAVC now on) falls initially, reaches its minimum and then start to increase. Hence, the SAVC curve has a U-shape and the reason behind is the law of variable proportions. Had the TVC not been inverse S-shaped, the SAVC would never assume a U-shape.

Generally, at initial stage of production, the productivity of each additional unit a variable input increases, thus, the variable input requires to produce each successive units of output decreases at this stage, implying that the AVC (Variable Cost Incurred to produce a unit of output) decreases. This process continues until the point of optimal combination between the fixed input and the

variable input is reached. Beyond this point, the productivity of each additional unit of the variable combined with the existing fixed input decreases because the fixed input is over utilized. As the productivity of such variables decreases, more and more of the variables are required to produce successive units of the output, implying that the VC incurred to produce each successive unit (AVC) increases.

Average total cost (ATC) or simply, Average cost (AC)

ATC (or AC, now on) is obtained by dividing the TC by the corresponding level of output. It shows the amount of cost incurred to produce each unit of successive outputs.

$$AC = \frac{TC}{Q}$$

Or equivalently, $AC = \frac{TVC + TFC}{Q}$

$$= \frac{TVC}{Q} + \frac{TFC}{Q}$$

$$= AVC + AFC$$

Thus, AC can also be given as the vertical sum of AVC and AFC.

Graphically, AC curve can be obtained by vertically adding the AVC and AFC for each level of successive outputs. Alternatively, the AC curve can also be derived in the same way as the SAVC curve. The AC curve is U-shaped because of the law of variable proportions. Observe the figure that follows.

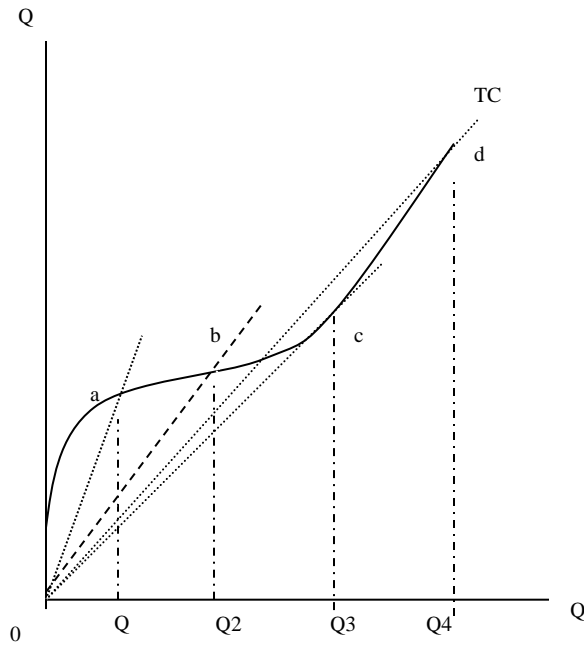
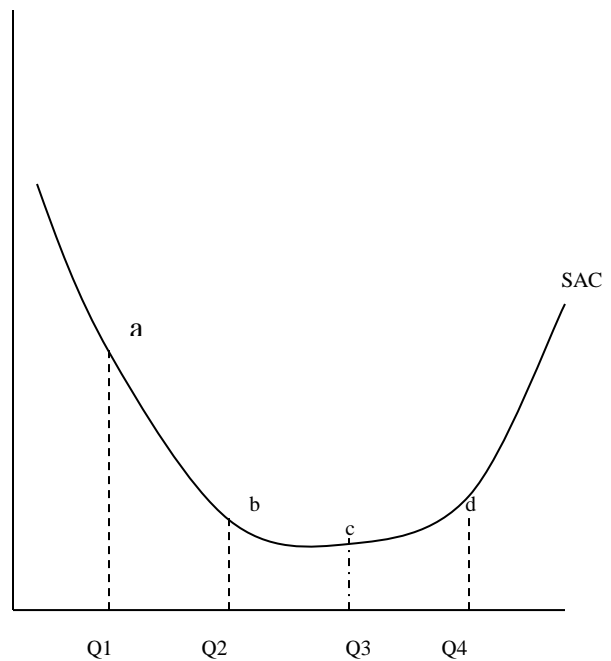


Fig 4.6 Panel-A



Panel-B

From this figure (Panel A), the AC at any level of output is the slope of the straight line from the origin to the point on the TC curve corresponding to that particular level of output. That is, for example, the AC of producing Q1 level of output is given by the slope of the line 0a, the AC of producing Q2 level of outputs is given by the slope of the line Ob and so on.

Marginal Cost (MC)

The marginal cost is defined as the additional cost that the firm incurs to produce one extra unit of the output. One thing to be noted here is that, the additional cost that the firm incurs to produce the 10th unit of output is not equal to the additional cost of producing the 1000th unit. They would be equal if the TC curve is straight line.

To sum up, the MC is the change in total cost which results from a unit change in output i.e. MC is the rate of change of TC with respect to output, Q or simply MC is the slope of TC function and given by:

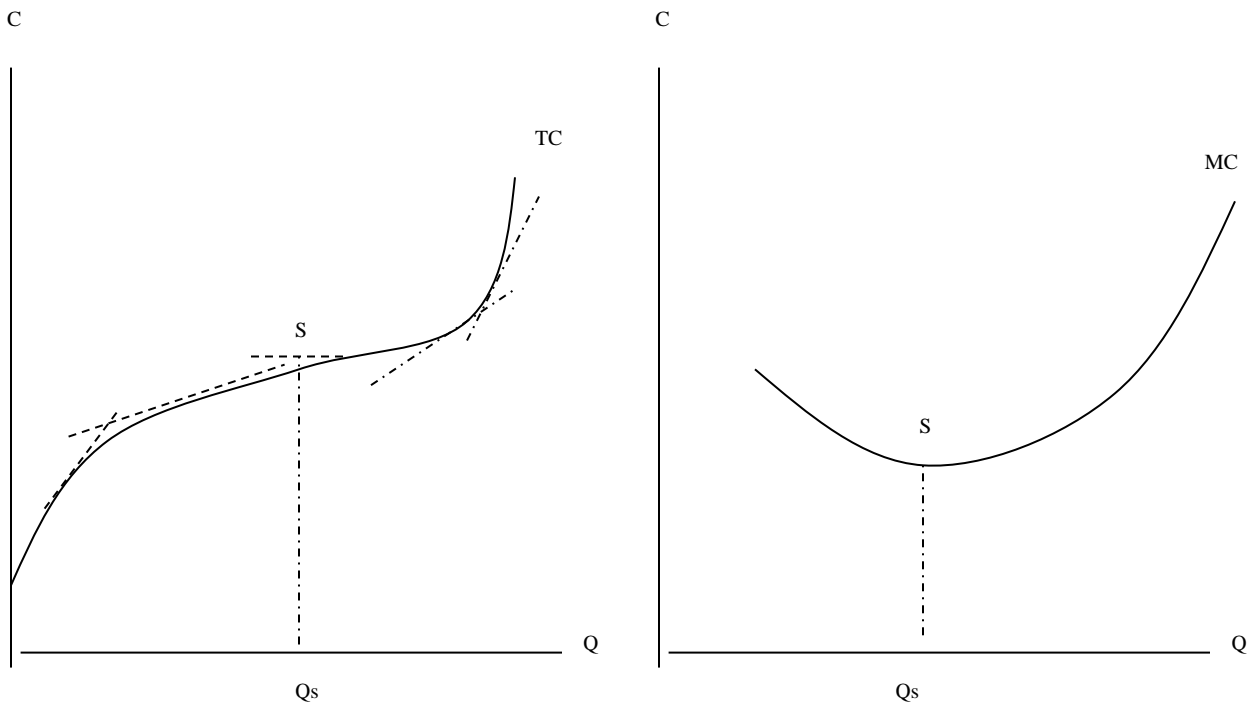
$$MC = \frac{dTC}{dQ}$$

In fact MC is also the rate of change of TVC with respect to the level of output.

$$MC = \frac{dTFC + dTVC}{dQ} = \frac{dTVC}{dQ}, \text{ since } \frac{dTFC}{dQ} = 0$$

Graphically, the MC the TC curve (or equivalently the slope of the TVC curve) obviously, the slope of curved lines at a given point is measured by constructing a tangent line to the curve at each point. So, the slope of the curve at a given point is equal to the slope of the tangent line at that specific point. Given the inverse S-shaped TC (or TVC) curve, the MC curve will be U-shaped. Thus given inverse S-shaped TC or TVC curve, the slope of the TC or TVC curve (i.e. MC) initially decreases, reaches its minimum and then starts to rise.

From this, we can logically infer that the reason for the U-shaped ness of MC is also the law of variable proportion. That is, had the TC or TVC curve not been inverse S-shaped, the MC curve have would never assumed the U-shape, and obviously, the TC or TVC is inverse S-shaped due to the law of variable proportions. Observe the figure that follows for more discussion. Fig 3.7



Panel-1

Panel-2

In Panel 2 the slope of the tangent lines to the TC curve (MC) decreases up to point **S** and then starts to rise.

In summary, AVC, ATC and MC curves are all U-shaped due to the law of variable proportions. The simplest total cost function which would incorporate the law of variable proportions is the cubic polynomial of the following form.

$$TC = b_0 + b_1Q - b_2Q^2 + b_3Q^3$$

Where Q- is the level of output and b_0 , b_1 , b_2 & b_3 – are none zero constants.

From this type of total cost function,

$$b_0 \text{ - represents the TFC, and } AFC = \frac{b_0}{Q}$$

$$b_1Q - b_2Q^2 + b_3Q^3 \text{ - represents TVC and}$$

$$AVC = \frac{b_1Q - b_2Q^2 + b_3Q^3}{Q} = b_1 - b_2Q + b_3Q^2$$

$$ATC = AFC + AVC$$

$$= \frac{b_0}{Q} + b_1 - b_2Q + b_3Q^2$$

4.4 The relationship between AVC, ATC and MC

Given $ATC = AVC + AFC$, AVC is part of the ATC. Both AVC and ATC are u – shaped, reflecting the law of variable proportions however, the minimum of ATC occurs to the right of the minimum point of the AVC (see the following figure) this is due to the fact that ATC includes AFC which continuously decreases as the level of output increases.

After the AVC has reached its lowest point and starts rising, its rise is over a certain range is more than off set by the fall in the AFC, so that the ATC continues to fall (over that range) despite the

increase in AVC. However, the rise in AVC eventually becomes greater than the fall in AFC so that the ATC starts increasing. The AVC approaches the ATC asymptotically as output increases.

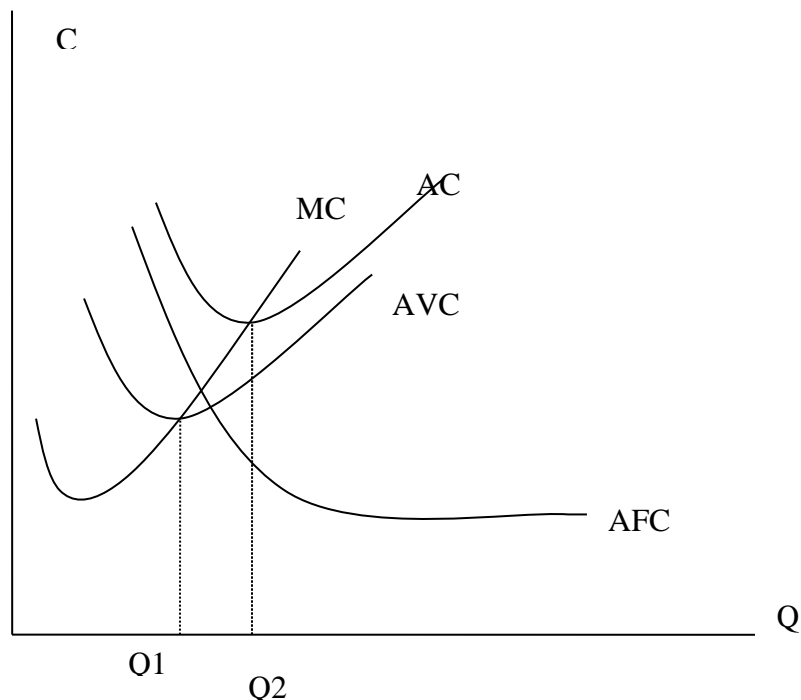


Fig 4.8

The AVC curve reaches its minimum point at Q_1 output and ATC reaches its minimum point at Q_2 . The vertical distance between ATC and AVC (AFC) decrease continuously as output increases. The MC curve passes through the minimum point of both ATC and AVC

Finally, the MC curve passes through the minimum point of both ATC and AVC curves.

This can be shown by using calculus.

Suppose the $TC = f(Q)$

$$MC = \frac{d(f(Q))}{dQ} = f'(Q)$$

$$AC = \frac{TC}{Q} = \frac{f(Q)}{Q}$$

$$AC = d \frac{d(f(Q))}{dQ} = \frac{(f(Q))Q - Q \cdot f'(Q)}{Q^2} \quad \text{Slope of}$$

But $f'(Q)$ is MC and Q^1 (or dQ/dQ) = 1

Thus, slope of

$$AC = \frac{MC \cdot Q - f(Q)}{Q^2} = \frac{\frac{MC}{Q} - \frac{f(Q)}{Q}}{Q}$$

$$\text{Slope AC} = \frac{1}{Q}(MC - AC), \text{ where } \frac{f(Q)}{Q} = AC$$

Now,

- i) when $MC < AC$, the slope of AC is negative, i.e. AC curve is decreasing (initial stage of production)
- ii) When $MC > AC$, the slope of AC is positive, i.e. the AC curve is increasing (after optimal combination of fixed and variable inputs.
- iii) When $MC = AC$, the slope of AC is zero, i.e. the AC curve is at its minimum point.

The relationship between AVC and MC can be shown in a similar fashion.

4.5 The relationship between short run per unit production and cost curves

Earlier in this chapter we have said that cost function is derived from production function. Now, let's see the important relation that per unit production curves (i.e. AP and MP of the variable input) and per unit cost curves (i.e. AVC and MC) have. The relationship is that the short run per unit costs are the mirror reflection (against the x-axis) of the short run production curves. That is the short run AVC is the mirror reflection of the short run AP of the variable input. When AP variable input increases, AVC decreases; when AP variable input reaches its maximum, the AVC reaches its maximum point, and finally when AP variable input starts to fall, the AVC curve starts to rise. The same relationship exists between the short run MP of variable input curve the MC curve. This can be shown algebraically by using a linear short run cost function.

Suppose the firm uses two inputs, labor L (which is variable) and capital (which is fixed input). And suppose that the prices of both factors are given and equal to w, and r respectively.

The total cost of production is then, $TC = rK + wL$

The first term (i.e. rK) is the fixed cost because both r and k are constant and the second term (i.e. wL) represents the variable cost.

Thus, $TVC = WL$

$$AVC = \frac{TVC}{Q} = \frac{WL}{Q} = W \cdot \frac{1}{\frac{Q}{L}} \text{ But, } \frac{Q}{L} \text{ represents APL}$$

$$\text{Therefore, } AVC = W \cdot \frac{1}{APL}$$

Hence, AVC and APL are inversely related. Similarly, MC and MPL ,

$$MC = \frac{dTC}{dQ} = \frac{dTVC}{dQ} \text{ (Remember that } MC = \frac{dTC}{dQ}$$

$$MC = \frac{d(W.L)}{dQ}$$

$$MC = W \cdot \frac{dL}{dQ} \dots\dots\dots \text{ (because } w \text{ is constant)}$$

$$MC = W \cdot \frac{dL}{dQ}$$

$$MC = W \cdot \frac{1}{\frac{dQ}{dL}}$$

$$MC = W \cdot \frac{1}{MPL}$$

$$\dots\dots\dots \text{ (Because } \frac{dQ}{dL} = MPL)$$

Hence, MC and MPL have also an inverse relation.

Graphically

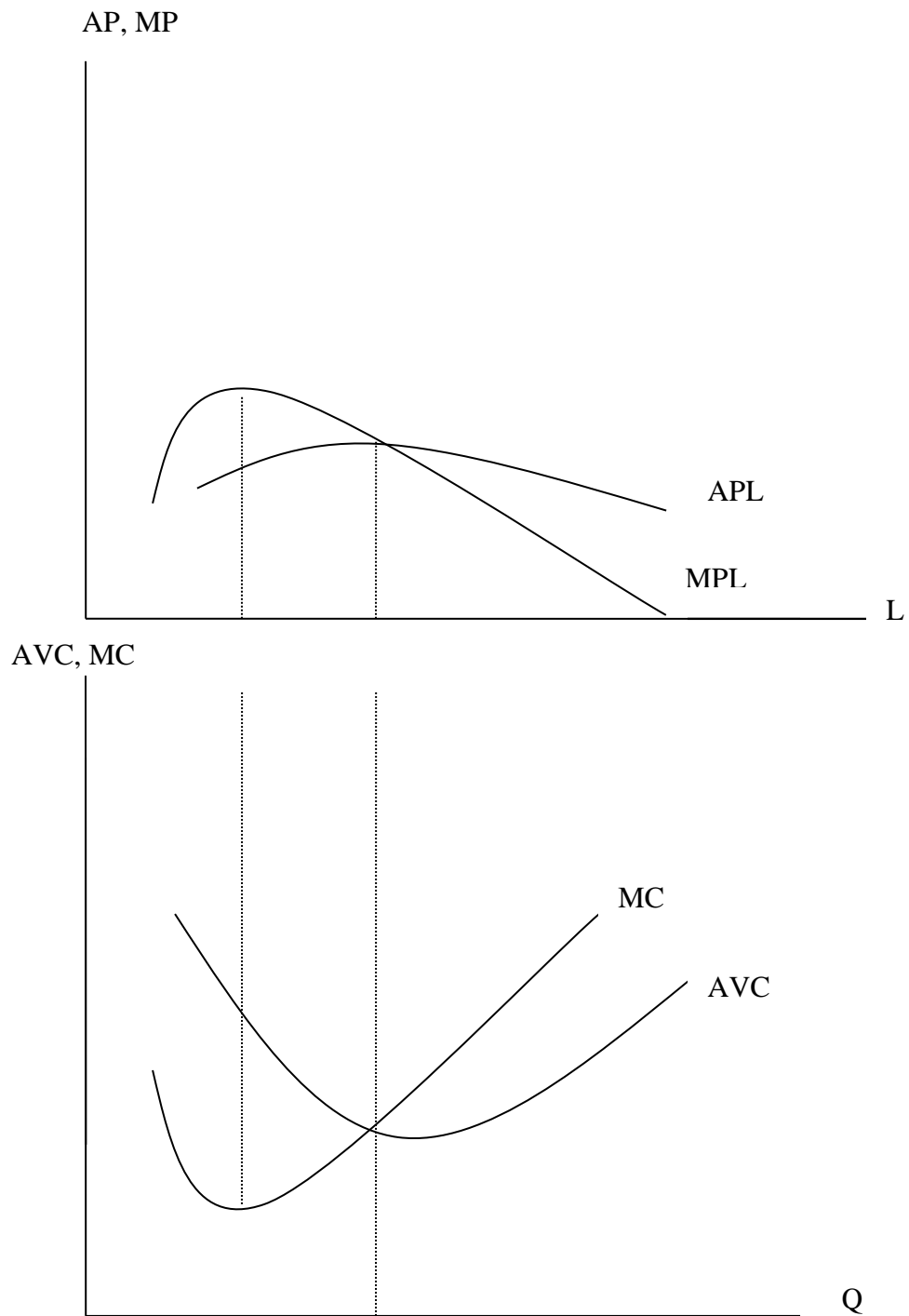


Fig 4.9 short run AVC and MC curves are the mirror reflection (along the horizontal axis) of short run AP_L and MP_L curves

4.5 Costs in the long run

Dear learner, in this section we will discuss the long run costs of a firm. The basic difference between long-run and short run costs is that in the short run, there are some fixed inputs which results in some amount of fixed costs. However, in the long run all factors are assumed to become variable. In the long run the firm can change the quantities of all inputs including the size of the plant. This implies that all costs are variable in the long-run in the sense that it is always possible to produce zero units of output at zero costs. That is, it is always possible to go out of business.

The long –run cost curve is a planning curve, in the sense that it is a guide to the entrepreneur in his decision to plan the future expansion of his plant.

Derivation of the long- run average cost curve

The long run average cost curve is derived from the short run average cost curves. Each point on the long run average cost (LAC, now on) corresponds to a point on the short run cost curve, which is tangent to the LAC at that point. Now let us examine in detail how the LAC is derived from the short run average cost (SAC) curves.

Assume that the available technology to the firm at a particular point of time includes three methods of production, each with a different plant size: a small plant, medium plant and large plant. The operation cost of the small plant is denoted by SAC1, the operating cost of the medium size plant is denoted by SAC2 and that of the large size plant is denoted by SAC3 in the following figure.

If the firm plans to produce x_1 units of output, it is well advised to choose the small size plant to minimize its cost. For example, if the firm choose to use the medium size plant to produce x_1 units of output, the per unit costs will be C_4 (a point corresponding to x_1 units of out put on the SAC2) but, the firm can produce x_1 units of output at a lower unit cost (c_1) if it uses the small size plant. Similarly, if it plans to produce x_2 units of output, it will choose the medium size plant. If the firm wishes to produce x_3 units, it will choose the large size plant.

If the firm starts with the small plant and its demand gradually increases, it will produce at lower costs (up to x_1 level of output). Beyond that level of output costs start increasing. If its demand reaches the level x_1 ” the firm can either continue to produce with the small plant or it can install the medium size plant. The decision, at this point, whether to install the medium size plant or not depends not on the costs but on the firm’s expectation about its future demand. If the firm expects

that the demand will expand further than x_1'' it will install a medium size plant because with this plant out puts larger than x_1'' are produced with a lower cost.

Similar considerations hold for the decision of the firm when it reaches the level x_2'' . If the firm expects its demand to stay constant at x_2'' level, the firm will not install the large plant, given that it involves a large investment which is profitable only if demand expands beyond x_2'' . If the firm expects that its demand will expand further, it will install the large size plant to reduce its cost. For example the level of output x_3 is produced at a cost c_3 with the large plant, while it costs c_2' if produced with the medium size plant ($c_2' > c_3$).

Now if we relax the assumption of the existence of only three plant sizes and assume that the available technology includes large number (infinite number) of plant sizes, each suitable for a certain level of output, the points of intersection of consecutive plants cost curves (which are the crucial points for the decision of whether to switch to a larger plant) are numerous and we obtain a continuous curve, which is the planning LAC curve of the firm.

The LAC curve is then the tangent to these SATC curves of various plant sizes and shows the minimum cost of producing each level of out put.

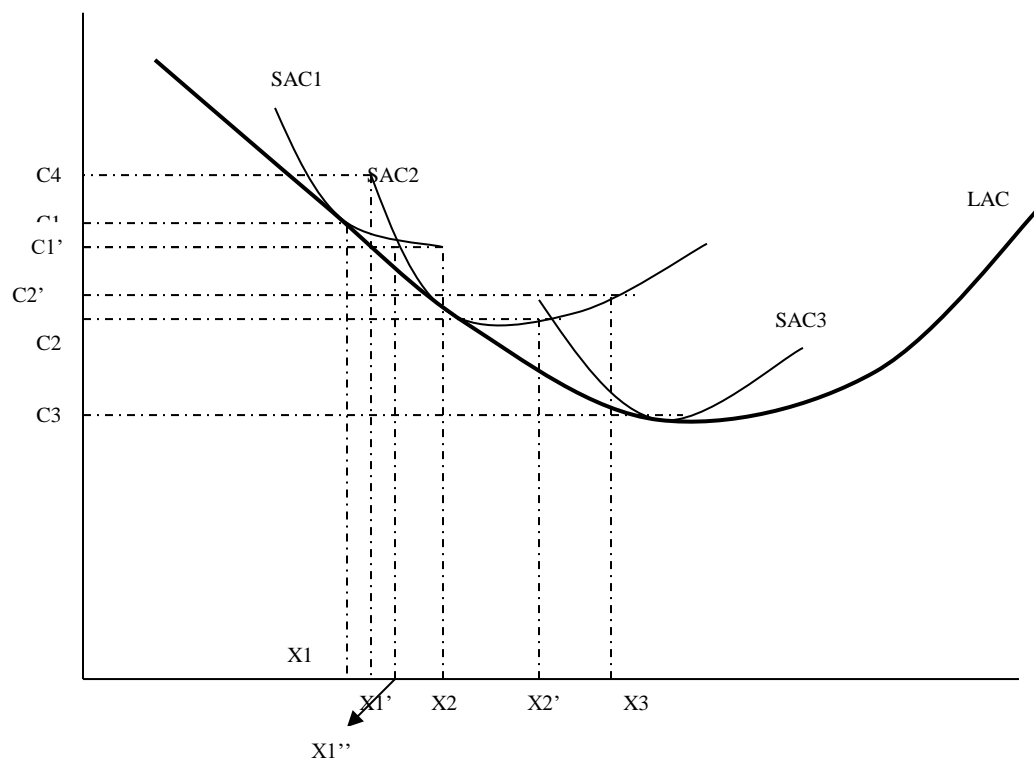


Fig: 4.10 the relation ship between LAC and short run average costs. The long run AVC curve is the lower envelope of the short run average costs of various plant sizes.

Assuming that there is infinite number of plant sizes, the LAC curve is a smooth curve tangent to each and every SAC curves corresponding to different plant sizes. See the following figure.

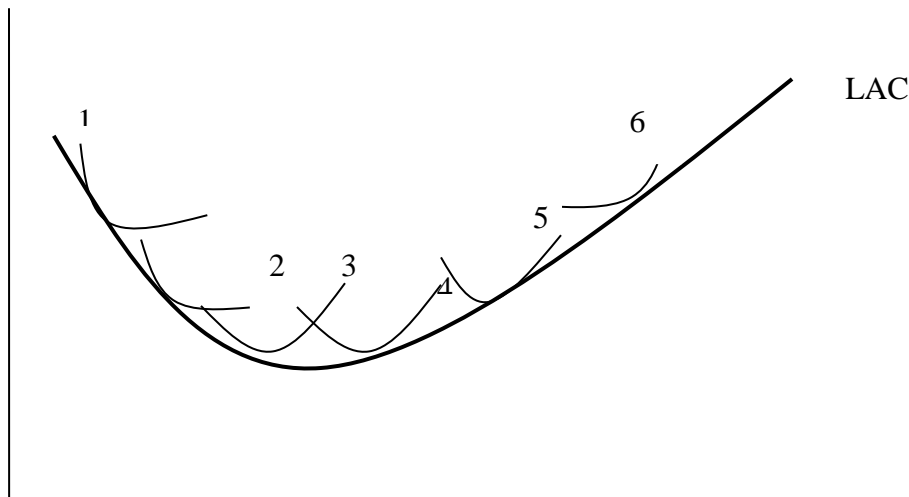


Fig: 4.11 the long run average cost curve, assuming that there are large number of plant sizes

In summary, the LAC curve shows the minimum per-unit cost of producing any level of output when the firm can build any desired scale of plant in the sense that the firm chooses the short –run plant which allows it to produce the anticipated (in the long run)out put at the least possible cost.

Why is the LAC U-shaped?

Dear learner, similar to the SAC curve, the LAC curve of a firm is also U-shaped, but the reason for the U-shapes ness of LAC curve is different from that of the SAC curve.

The LAC curve is U-shaped due to the laws of returns to scale(i.e increasing and decreasing returns to scale).that is, as out put expands from a very low levels increasing returns to scale prevails (i.e., out put rises proportionally more than inputs), and so the cost per-unit of out put falls(assuming that input prices remain constant).As out put continues expand, the forces of decreasing returns to scale eventually begin to over take the forces of increasing returns to scale and the LAC begins to rise.

In other words, the per unit costs of production decreases initially as the plant size increases, due to the economies of scale which larger plant size makes possible.

Economies of scale is the cost dimension of increasing returns to scale and thus, they are like the two sides of a coin. If a firm has increasing returns to scale in production (i.e., if it requires the firm less than double inputs to produce double output) the firm will have economies of scale in costs (it will require the firm less than double cost to produce double output). Thus, the reason for the decreasing part of the LAC curve is increasing returns to scale or economies of scale. Economies of scale may prevail for various reasons such as specialization of skills, lower prices for bulk-buying of raw materials, decentralization of management system and etc.

The traditional theory of the firm assumes that economies of scale exist only up to a certain size of plant, which is known as optimal plant size, because with this plant size all possible economies of scale are fully exploited. If the plant size increases further than this optimal size, diseconomies of scale will start to prevent, arising from managerial inefficiencies, the price advantage from bulk-buying may also stop beyond a certain limit etc. These diseconomies of scale will lead to an increasing LAC curve. Thus, the increasing portion of the LAC curve shows the existence of diseconomies of scale or decreasing returns to scale.

In general, the reason for the U-shapedness of the LAC curve are the existence of increasing returns to scale at the initial stage of expansion, decreasing returns to scale at a later stage of expansion.

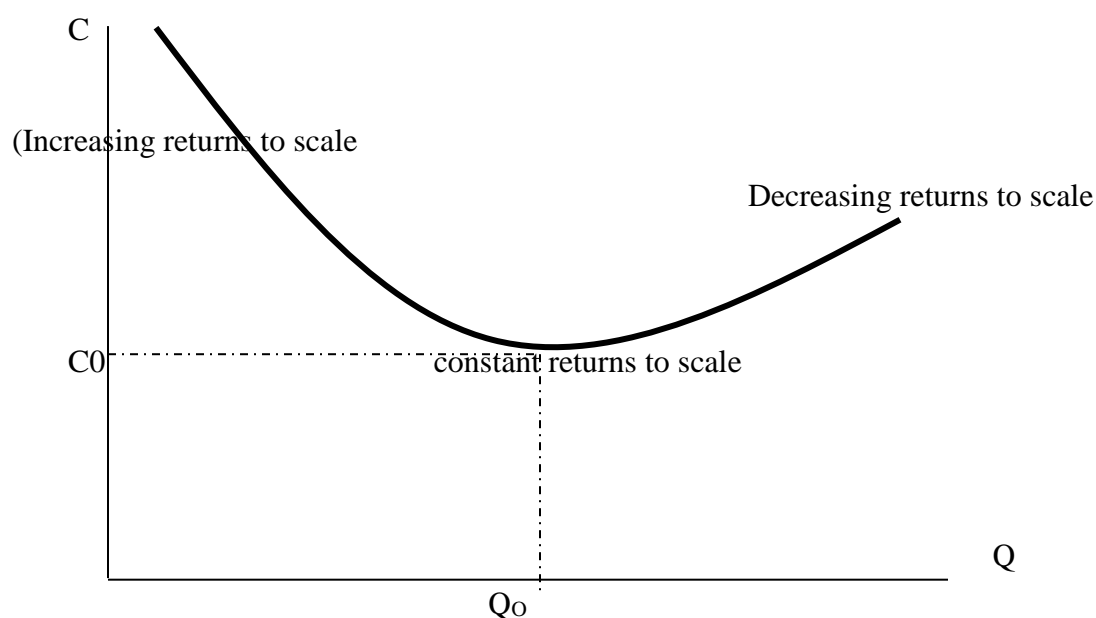


Fig 4.12 the LAC curve is U-shaped due to the combined effects of increasing, constant and decreasing returns.

The long-run marginal cost curve.

The long-run marginal cost curve (LMC) is derived from the short run MC curve but does not envelope them. The LMC is formed from points of intersection of the SMC curves with the vertical lines (to the x-axis) drawn from the points of tangency of corresponding SAC curves and the LAC curve.

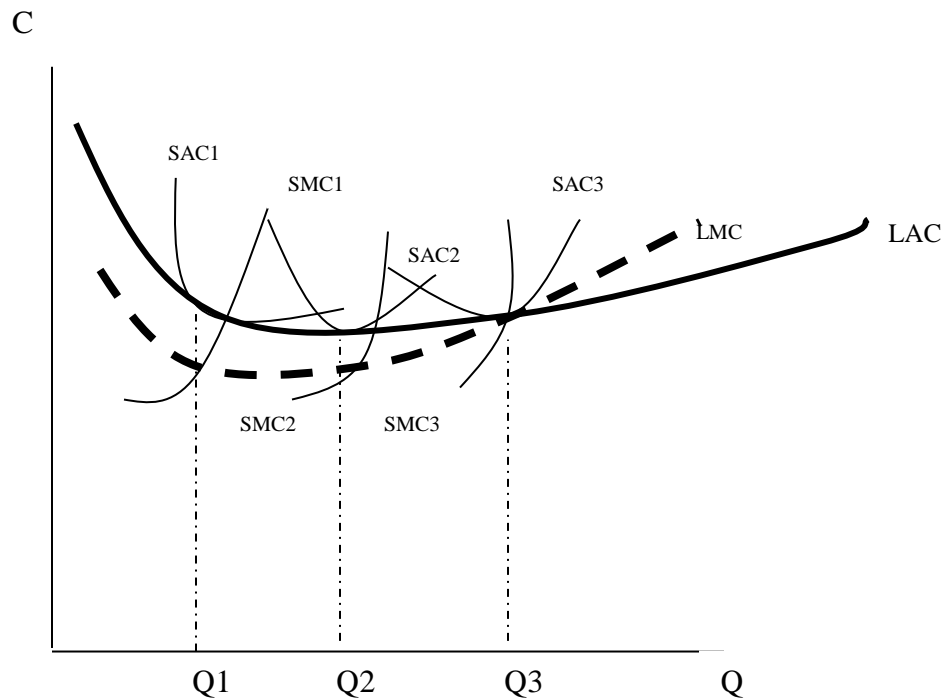


Fig4.13: Long run marginal cost curve; it is derived from the short run marginal cost curves by connecting the points of intersection of the vertical lines drawn from the point of tangency of SAC curves with the LAC curves with and the corresponding SMC curves.

Note that, the LMC curve passes through the minimum of the LAC curve.

Summary

Costs are the monetary value of inputs used for production purpose. Costs of production may involve explicit costs (costs of purchased inputs) and implicit cost (estimated costs of inputs owned by the producers itself).

Managers, investors and economists must take into account the opportunity cost associated with the use of a firm's resources the cost associated with the opportunities forgone when the firm uses its resources in its next best alternative.

Costs of production depend on several factors such as the quantity of production, the level of technology and input prices. Thus cost functions are a multivariate function.

In the short run, one or more of a firm's inputs are fixed. Thus, total cost of production can be divided to fixed cost and variable costs. Fixed costs are constant irrespective of the level output. A firm cannot avoid fixed cost even by producing zero level of output. Variable costs, on the other hand, vary with the level of output directly.

By dividing the total cost and total variable costs for the quantity of production, we obtain average cost (AC) respectively. In the short run, when not all inputs are variable, the AC and AVC curves assume a U- shape due to the law of variable proportions.

Marginal costs of production are additional costs incurred to produce one more unit of a commodity. The MC curve has a U- shape due to the law of variable proportions.

Short run marginal and average variable cost curves are a mirror reflection of the marginal product and average product of the variable input respectively.

In the long run, all inputs to the production process are variable. As a result, the choice of inputs depends both on the relative costs of a factor of production and on the extent to which the firm can substitute among inputs in its production process.

UNIT FIVE

PRICE AND OUT PUT DETERMINATION UNDER PERFECT COMPETITION

Introduction

In this chapter, we shall try to see how a given firm operating in a perfectly competitive market determines the profit maximizing level of output and price, and how equilibrium market price and level of output are determined in a perfectly competitive market. Our discussion starts with giving a brief description about perfect competition.

Objectives

After successful completion of this unit, you will be able to:

- Characterize a perfectly competitive market.
- Know how a perfectly competitive firm determines the profit maximizing output both in the long run and short run.
- Derive the short run supply schedule of an individual firm and industry.
- Explain when a perfectly competitive firm should decide to shut down.
- How a perfect competition results in efficient allocation of resources.

5.1 Perfect Competition

Definition and Assumptions

Perfect competition is a market structure characterized by a complete absence of rivalry among the individual firms. Thus, perfect competition in economic theory has a meaning diametrically opposite to the everyday use of this term.

Most of the time, we see business men using the word “Competition” as synonymous to “rivalry”. However, in theory, perfect competition implies no rivalry among firms

Assumptions

The model of perfect competition was constructed based on the following assumptions or imaginations.

1. *Large number of sellers and buyers.*

The perfect competitive market includes a large number of buyers and sellers. How large should the number of buyers and sellers is large to the extent that the market share of each firm (and buyer) is too small to have a perceptible effect on the price of the commodity. That is the

action of a single seller or buyer can not influence the market price of the commodity, since the firm or (the buyer) is too small in relation to the market.

2. *Products of the firms are homogeneous.*

This means the products supplied by all the firms in the market have uniform physical characteristics (are uniform in terms of quantity, quality etc) and the services associated with sales and delivery are identical. Thus buyers can not differentiate the product of one firm from the product of the other firm.

The assumptions of large number of sellers and of product homogeneity imply that the individual firm in pure competition is a price taker: its demand curve is infinitely elastic, indicating that the firm can sell any amount of out put at the prevailing market price. Since the share of the firm from the market supply is too small to affect the market price, the only thing that the firm can do is to sell any quantity demand at the ongoing market price. Thus, the demand curve that an individual firm faces is a horizontal line.

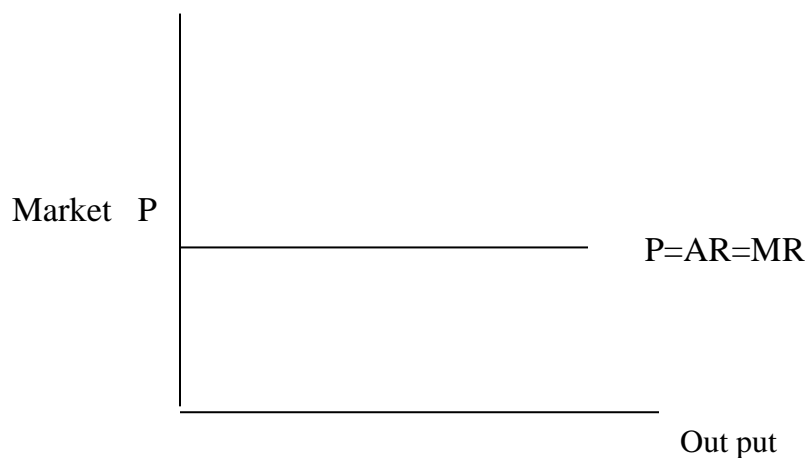


Fig 5.1 the demand curve indicates a single market price at which the firm can sell any amount of the commodity demanded. The demand curve also indicates the average revenue and marginal revenue of the firm.

3. *Free entry and exit of firms*

There is no barrier to entry and exit from the industry. Entry or exit may take time, but firms have freedom of movement in and out of the industry. If barriers exist the number of firms in the industry may be reduced so that one of them may acquire power to affect the market price.

4. The goal of all firms is profit maximization.

Of course, firms can have different objectives. Some firms may have the aim of making their product wise, others may want to maximize their sales even by cutting price, etc. But, in this model, it is assumed that the goal of all firms is to maximize their profit and no other goal is pursued.

5. No government regulation

By assumption, there is no government intervention in the market. That is there is no tax, subsidy etc. A market structure in which all the above assumptions are fulfilled is called pure competition. It is different from perfect competition which requires the fulfillment of the following additional assumptions.

6. Perfect mobility of factors of production

Factors of production (including workers) are free to move from one firm to another throughout the economy. Alternatively, there is also perfect competition in the market of factors of production.

7. Perfect knowledge

It is assumed that all sellers and buyers have a complete knowledge of the conditions of the prevailing and future market. That is all buyers and sellers have complete information about.

- ◆ The price of the product

- ◆ Quality of the product etc

Thus, a perfectly competitive market is a market which satisfies all the above conditions (assumptions). In reality, perfectly competitive markets are scarce if not none. But since the theory of perfectly competitive market helps as a bench mark to analyze the more realistic markets, it is very important to study it.

Given the above assumptions (based which the model of perfect competition was built), we will now examine how the firm operating in such a market determines the profit maximizing output both in the short run and in the long run. But to determine the profit maximizing output, first we have to see what the revenue and cost functions of the firms operating in perfectly competitive market looks like.

Costs under perfect competition

In the previous chapter, we have said that the per unit cost (AVC & AC) have U –shape due to the law of variable proportions (in the short run) and the law of returns to scale (in the long

run). There is no exception for firms operating under perfect competition i.e., their cost functions have the behavior mentioned in the last chapter.

Demand and revenue functions under perfect competition

Due to the existence of large number of sellers selling homogenous products, each seller is a price taker in perfectly competitive market. That is, a single seller cannot influence the market by supplying more or less of a commodity.

If, for example, the seller charges higher price than the market price to get larger revenue, no buyers will buy the product of this (the price raising) firm since the same product is being sold in the market at lower price by other sellers. Obviously, the firm will not also attempt to reduce the price. Thus firms operating in a perfectly competitive market are price takers and sell any quantity demanded at the ongoing market price.

Hence, the demand function that an individual seller faces is perfectly elastic (or horizontal line).

Graphically,

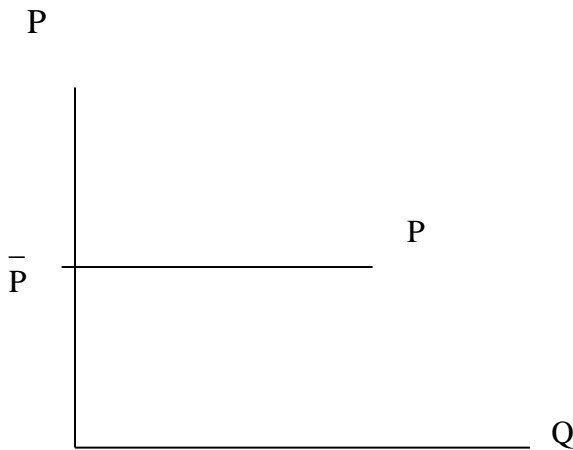


Fig 5.2 the demand curve that a perfectly competitive firm faces is horizontal line with intercept at the market price. This indicates that sellers sell any quantity demanded at the ongoing market price and buyers buy any amount they want at the ongoing market price.

From the buyers' side too, since there is large number of buyers in the market, a single buyer can not influence the market price.

Thus, in perfectly competitive market, both buyers and sellers are price takers. They take the price determined by the forces of market demand and market supply.

Given the horizontal demand function at the ongoing market price, the total revenue of a firm operating under perfect competition is given by the product of the market price and the quantity of sales, i.e.,

$$TR = P \cdot Q$$

Since the market price is constant at P^* , the total revenue function is linear and the amount of total revenue depends on the quantity of sales. To increase his total revenue, the firm should sell large quantity.

Graphically, the TR curve is as shown below.

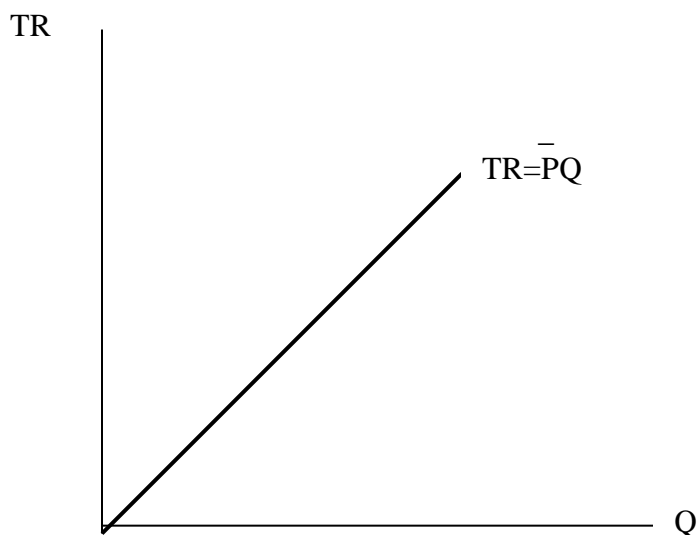


Fig 5.3 the total revenue of firm operating in a perfectly competitive market is linear (and increasing function) of the quantity of sales.

The marginal revenue (MR) and average revenue (AR) of a firm operating under perfect competition are equal to the market price. To see this, let's find the MR and AR functions from TR functions.

$$TR = PQ$$

By definition, MR is the change in total revenue that occurs when one more unit of the output is sold, i.e. $MR = \frac{dTR}{dQ} = P$. Hence $MR = P$

Average revenue is the TR divided by the quantity of sales. i.e. $AR = \frac{TR}{Q} = \frac{P \cdot Q}{Q} = P$ Hence,

$AR = P$.

Graphically, the demand curve represents the MR and AR of the firm

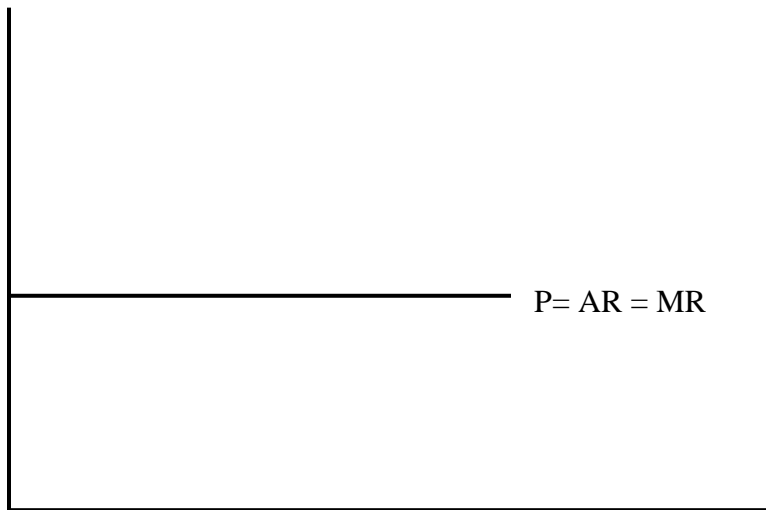


Fig: 5.4 the AR curve, MR curve and the demand curve of an individual firm operating under perfectly competitive market overlap.

5.3 Short run equilibrium of the firm

Dear learner! What do we mean when we say the firm is in equilibrium?

(Use the space left below to write your response)

If you don't know the answer, don't worry, just read what follows;

A firm is said to be in equilibrium when it maximizes its profit (Π). Profit is defined as the difference between total cost and total revenue of the firm:

$$\Pi = TR - TC$$

Under perfect competition, the firm is said to be in equilibrium when it produces that level of output which maximizes its profit, given the market price. Thus, determination of equilibrium of the firm operating in a perfectly competitive market means determination of the profit maximizing output since the firm is a price taker.

The level of output which maximizes the profit of the firm can be obtained in two ways:

- ➡ Total approach
- ➡ Marginal approach

Total approach

In this approach, the profit maximizing level of output is that level of output at which the vertical distance between the TR and TC curves is maximum. (Provided that the TR curve lies above the TC curve at this point).

Graphically

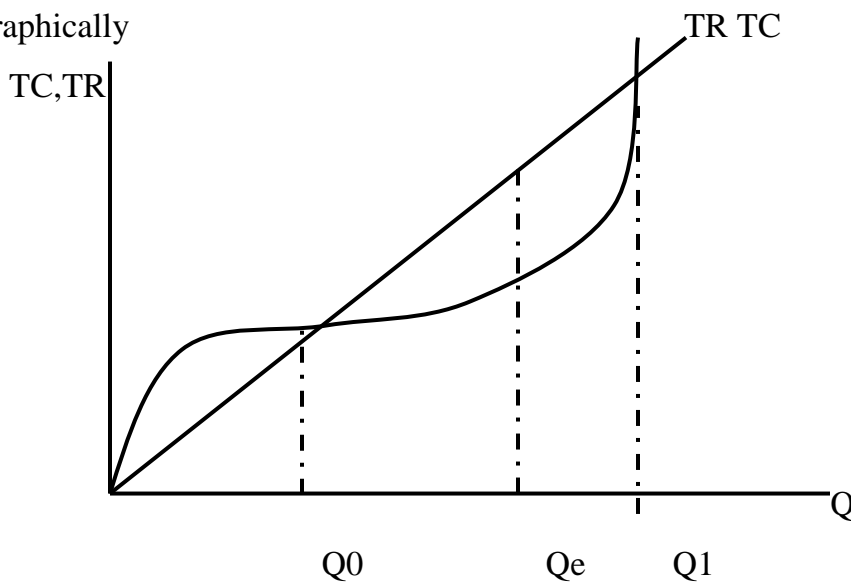


Fig:5.5 The profit maximizing output level is Q^e because it is at this output level that the vertical distance between the TR and TC curves (or profit) is maximum.

For all output levels below Q_0 and above Q_1 profit is negative because TC is above TR.

Marginal Approach

In this approach the profit maximizing level of output is that level of output at which:

$$MR=MC \text{ and}$$

MC is increasing

This approach is directly derived from the total approach. In figure 4.4, the vertical distance between the TR and TC curve is maximum where a straight line parallel to the TR curve is tangent to the TC curve. Or simply, the vertical distance between the TC and TR curves is maximum at output level where the slope of the two curves is equal. The slope of the TR curve constant and is equal to the MR or market price.

Similarly, the slope of the TC curve at a given level of output is equal to the slope of the tangent line to the TC curve at that level of output, which is equal to MC. Thus the distance between the TR and TC curves (Π) is maximum when MR equals MC.

Graphically, the marginal approach can be shown as follows.

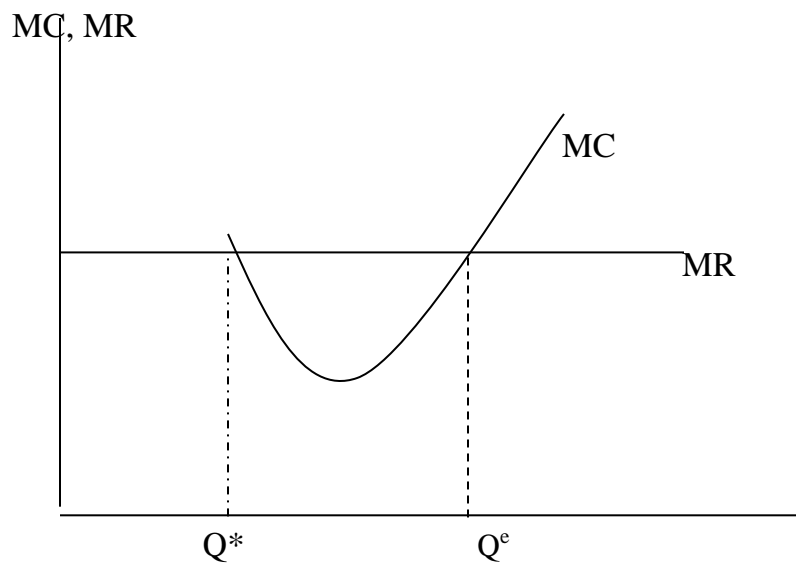


Fig 5.6: the profit maximizing output is Q^e , where $MC=MR$ and MC curve is increasing. At Q^* , $MC=MR$, but since MC is falling at this output level, it is not equilibrium output. For all output levels ranging from Q^* to Q^e the marginal cost of producing additional unit of output is less than the MR obtained from selling this output. Hence the firm should produce additional output until it reaches Q^e .

Mathematical derivation of the equilibrium condition

$$\text{Profit } (\Pi) = \text{TR} - \text{TC}$$

TC is a function of output, $\text{TC} = f(Q)$

TR is also a function of output, $\text{TR} = f(Q)$

Thus, profit is a function of output, $\Pi = f(Q)$

$$\Pi = TR - TC$$

To determine the profit maximizing output we find the first derivative of the Π function and equate the result to zero.

$$\begin{aligned} \frac{d\Pi}{dQ} &= \frac{dTR}{dQ} - \frac{dTC}{dQ} = 0 \\ &= MR - MC = 0 \end{aligned}$$

$$= MR = MC \text{ ----- (First order condition necessary condition)}$$

The equality of MC and MR is a necessary, but not sufficient condition. The sufficient condition for maximization of Π is that the second derivative of the Π function should be less than zero (or negative) i.e.

$$\frac{d^2\Pi}{dQ^2} < 0 \rightarrow \frac{d^2TR}{dQ^2} - \frac{d^2TC}{dQ^2} < 0$$

$$\frac{d^2TR}{dQ^2} = \frac{dMR}{dQ}, \text{ thus } \frac{d^2TR}{dQ^2} \text{ is the slope MR. Since MR is horizontal (or constant), the slope of}$$

MR is equal to zero.

$$\text{Like wise, } \frac{d^2TC}{dQ^2} \text{ is equal } \frac{dMC}{dQ} \text{ and thus, } \frac{d^2TC}{dQ^2} \text{ is the slope of MC, which is not constant}$$

$$\text{Thus, } \frac{d^2TR}{dQ^2} < \frac{d^2TC}{dQ^2} \text{ means}$$

- Slope of MR < Slope of MC
- 0 < Slope of MC or
- Slope of MC > 0 or
- MC is increasing..... Sufficient condition

Thus, the condition for profit maximization under perfect competition is

$$MR = MC \text{necessary condition and}$$

$$MC \text{ is increasing..... sufficient condition}$$

Conceptually, maximizing the difference between TR & TC means maximizing the area between the MR and MC curve, i.e., maximizing $\int (MR - MC)dQ$. And the area between the MC and MR would be maximal only when the firm produces Q^e level of output.

Dear learner, the fact that a firm is in the short run equilibrium does not necessarily mean that the firm gets positive profit. Whether the firm gets positive or zero or negative profit depends on the level of ATC at equilibrium thus;

- If the ATC is below the market price at equilibrium, the firm earns a positive profit equal to the area between the ATC curve and the price line up to the profit maximizing output (see fig5.7below)

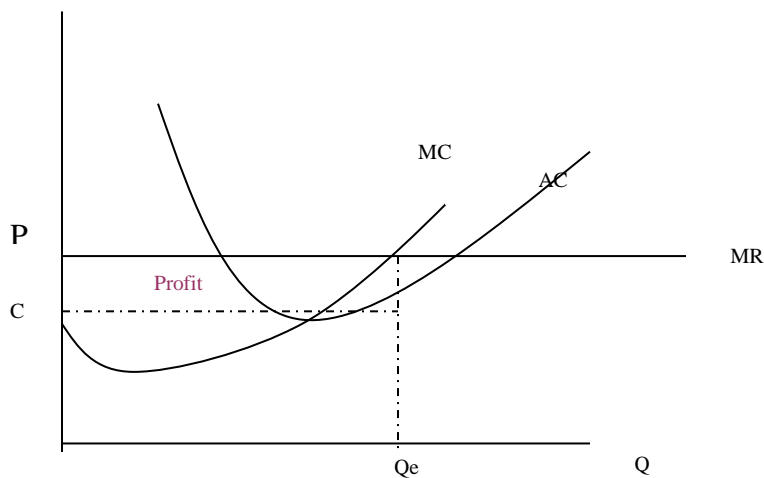


Fig 5.7 the firm earns a positive profit because price exceeds AC of production at equilibrium

- If the ATC is equal to the market price at equilibrium, the firm gets zero **II**.
- If the ATC is above the market price at equilibrium, the firm earns a negative profit (incurs a loss) equal to the area between the ATC curve and the price line.(see fig 5.8 below).

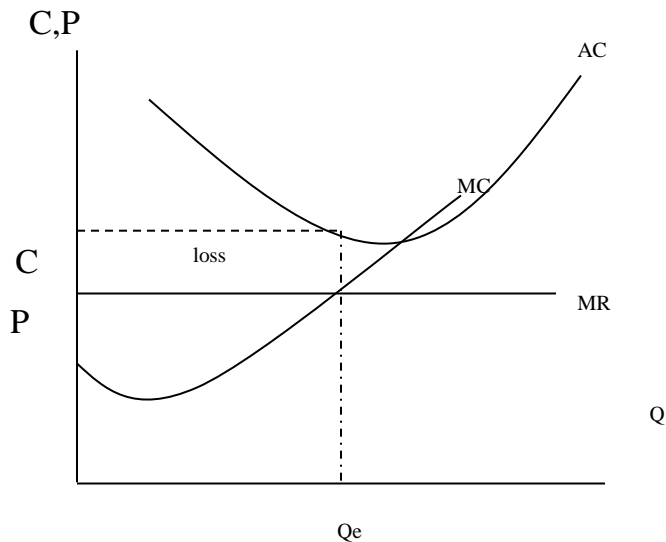


Fig 5.8 a firm incurs a loss because price is less than AC of production at equilibrium.

In this case, you may ask that “***why do the firm continue to produce if it had to incur a loss?***”

In fact, the firm will continue to produce irrespective of the existing loss as far as the price is sufficient to cover the average variable costs. In other words, the firm should continue producing as far as the TR sufficiently covers the total variable costs. This is so because if the firm stops production he will incur a loss which equals the total fixed cost. But, if it continues to produce the loss is less than the total fixed costs because the TR will cover some portion of the fixed costs in addition to the whole variable costs as far as it is greater than TVC.

However, if the market price falls below the AVC or alternatively, if the TR of the firm is not sufficient to cover at least the total variable cost, the firm should close (shut down) its factory (business). It will only lose the fixed costs; but if it continues operation while the TR is unable to cover even the variable costs, the loss is greater than the fixed costs since part of the variable cost is also not covered by the existing revenue.

To summarize, a firm may continue production even while incurring a loss (when $TC > TR$). This occurs as far as the TR is able to cover at least the TVC ($TR > TVC$). If the TR is less than the TVC, the firm is well advised to discontinue its operation so that the loss will be minimized. Hence, to continue its operation (or just to stay in the business) the firm should obtain the TR which can at least cover its variable costs. The following example will make the discussion clear.

Example:

Suppose a firm has a TFC of \$2,000, a TVC of \$ 5,000 and a TR of \$6,000 at equilibrium. Should the firm stop its operation? Why?

In fact the firm is incurring a loss of \$ 1,000 because TC ($2,000 + 5,000=7,000$) is greater than the total revenue. But the firm should continue production because the TR is greater than TVC. If the firm stops operation, it will lose the fixed cost (\$ 2,000). But if it continues production the loss is only \$ 1,000 (TR-TC). Thus, the firm requires a minimum TR of \$ 5,000 to continue operation. If

the TR is equal to \$ 5,000, the firm is indifferent in between choosing to continue or to discontinue its operations because in both cases the loss is equal to fixed costs. Thus the level output at which TR and TVCs are equal is called shut down out put level. In other words, shut down point is the point at which AVC equals the market price.

Equally important point is the point of break-even. Break-even point is the out put level at which market price is equal to the average cost of production so that the firm obtains only normal profit (zero profit).

Numerical example

Dear learner, now let us see how to determine the short run equilibrium of a firm operating in a perfectly competitive market by using a hypothetical example.

Suppose that the firm operates in a perfectly competitive market. The market price of his product is \$10. The firm estimates its cost of production with the following cost function:

$$TC=10q-4q^2+q^3$$

- A) What level of out put should the firm produce to maximize its profit?
- B) Determine the level of profit at equilibrium.
- C) What minimum price is required by the firm to stay in the market?

Solution

Given: p=\$10

$$TC= 10q - 4q^2+q^3$$

- A) The profit maximizing output is that level of output which satisfies the following condition

MC=MR &

MC is rising

Thus, we have to find MC& MR first

◆ MR in a perfectly competitive market is equal to the market price. Hence, MR=10

Alternatively, $MR = \frac{dTR}{dq}$ where TR= P.q = 10q

$$\text{Thus, } MR = \frac{d(10q)}{dq} = 10$$

$$MC = \frac{dTC}{dq} = \frac{d(10q - dq^2 + q^3)}{dq} = 10 - 8q + 3q^2$$

◆ To determine equilibrium output just equate MC & MR

And then solve for q.

$$10 - 8q + 3q^2 = 10$$

$$-8q + 3q^2 = 0$$

$$q(-8 + 3q) = 0$$

$$q = 0 \text{ or } q = 8/3$$

Now we have obtained two different output levels which satisfy the first order (necessary) condition of profit maximization

i.e. 0 & 8/3

◆ To determine which level of output maximizes profit we have to use the second order test at the two output levels i.e. we have to see which output level satisfies the second order condition of increasing MC.

◆ To see this first we determine the slope of MC

$$\text{Slope of MC} = \frac{dMC}{dq} = -8 + 6q$$

◆ At $q = 0$, slope of MC is $-8 + 6(0) = -8$ which implies that marginal cost is decreasing at $q = 0$. Thus, $q = 0$ is not equilibrium output because it doesn't satisfy the second order condition.

◆ At $q = 8/3$, slope of MC is $-8 + 6(8/3) = 8$, which is positive, implying that MC is increasing at $q = 8/3$

Thus, the equilibrium output level is $q = 8/3$

B) Above, we have said that the firm maximizes its profit by producing 8/3 units. To determine the firm's equilibrium profit we have calculate the total revenue that the firm obtains at this level of output and the total cost of producing the equilibrium level of output.

$$\begin{aligned} TR &= \text{Price} * \text{Equilibrium output} \\ &= \$10 * 8/3 = \$80/3 \end{aligned}$$

TC at $q = 8/3$ can be obtained by substituting 8/3 for q in the TC function, i.e.,

$$TC = 10(8/3) - 4(8/3)^2 + (8/3)^3 \approx 23.12$$

Thus the equilibrium (maximum) profit is

$$\Pi = TR - TC$$

$$= 26.67 - 23.12 = \$ 3.55$$

c) To stay in operation the firm needs the price which equals at least the minimum AVC. Thus to determine the minimum price required to stay in business, we have to determine the minimum AVC.

AVC is minimal when derivative of AVC is equal to zero

$$\text{That is: } \frac{dAVC}{dq} = 0$$

Given the TC function: $TC = 10q - 4q^2 + q^3$, there is no fixed cost i.e. TC is equal to the TVC.

$$\text{Hence, } TVC = 10q - 4q^2 + q^3$$

$$AVC = \frac{TVC}{q} = \frac{10q - 4q^2 + q^3}{q} = 10 - 4q + q^2$$

$$\frac{dAVC}{dq} = 0 \rightarrow \frac{d(10 - 4q + q^2)}{dq} = 0$$

$$= -4 + 2q = 0$$

$$\rightarrow q = 2 \text{ i.e. AVC is minimum when output is equal to 2 units.}$$

The minimum AVC is obtained by substituting 2 for q in the AVC function i.e., $\text{Min AVC} = 10 - 4(2) + 2^2 = 6$

Thus, to stay in the market the firm should get a minimum price of \$ 6.

Exercise:

Show that the break even price is also equal to \$ 6. What is the reason behind?

5.4 The short run supply curve of the firm and the industry

The short run supply curve of the firm

In the previous section, we have demonstrated how a competitive firm determines the level of output which maximizes its profit for a given market price. The profit maximizing level of output is defined by the point of equality of MC and market price (because market price is equal to MR in the perfectly competitive market). By repeating this analysis at different possible market prices, we observe how the equilibrium quantity supplied by the firm varies with the market price.

Now consider the figure 5.9 to understand how to derive the short run supply curve of a perfectly competitive firm.

Suppose that initially the market price and MR is \$6 and the demand curve is shown by line P1. Given the MC curve, the level of output which maximizes the firm's profit is defined by the point of intersection of the MC curve and the demand line (P1), which is equal to 50 units.

Now assume that the market price increases to \$7. This is shown by an upward shift of the demand curve (MR) to P2. Given the positive slope of MC, this higher demand (MR) curve cuts the MC curve at higher output level, 140. That is, when the market price increases from \$6 to \$7, the equilibrium quantity supplied by the firm increases from 50 units to 140 units. As the price increases further (say to \$8), the equilibrium output increases to 200 units.

This implies that the quantity supplied by the firm increases as the market price increases.

The firm, given its cost structure, will not supply any quantity (will shut down) if the price falls below \$6, because at a lower price than \$6, the firm can not cover its variable costs. Thus, supply is zero for all price levels below \$6 (minimum AVC)

If we plot the successive equilibrium points on a separate graph we observe that the supply curve of the individual firm overlaps with (is identical to) to part of its MC curve to the right of the shut down point.

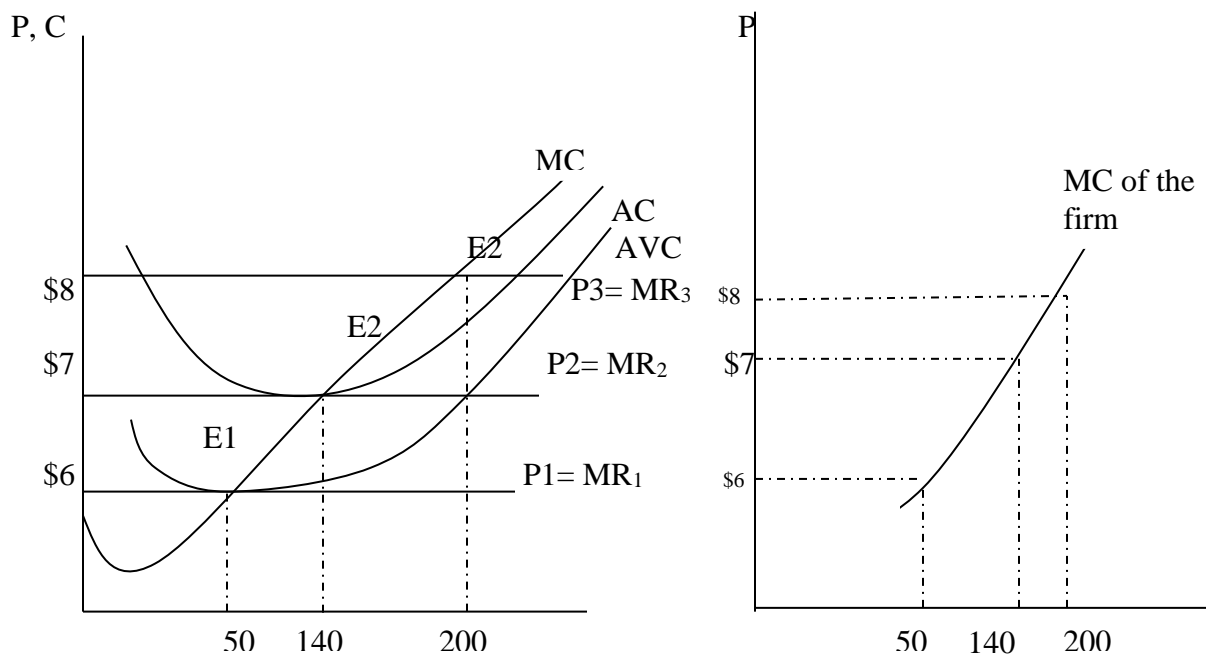


Fig.5.9 The short run supply curve of a perfectly competitive firm is obtained by connecting different equilibrium points E1, E2, E3 that occurs at successive price levels p1, p2 and p3 respectively. When the market price is \$6, the firm supplies 50 units to maximize its profit. As the price increases to \$7, the equilibrium quantity supplied increases to 140 units and so on.

Thus, the short run supply curve of a perfectly competitive firm is that part of MC curve which lies above the minimum average variable cost (Shut down point)

5.5 Short run supply curve of the industry

Dear learner, before we discuss the derivation of short run supply curve of the industry, let us see what the word '**industry**' refers to in this unit. The word 'industry' is defined as group of firms producing homogeneous products. Thus the industry supply is the total supply or market supply.

The industry –supply curve is the horizontal summation of the supply curves of the individual firms. That is, the total quantity supplied in the market at each price is the sum of the quantities supplied by all firms at that price. This is based on the assumption that the factor prices and the technology are given.

For detailed information as to how to derive the short run industry supply curve from the supply of individual firms, consider the following figure. S₁, S₂ and S₃ denote the supply curves of firms existing in a given industry. The industry supply curve is obtained by adding the quantities supplied by all the firms at each price. For example, at price which equals \$ 6, firm 1 supplies 50 units, firm 2 supplies 80 units & firm 3 supplies 120 units. The market supply at \$ 6 price is thus 250 units (50+80+120 units). The short run industry- supply is derived by repeating the above process at each price levels.

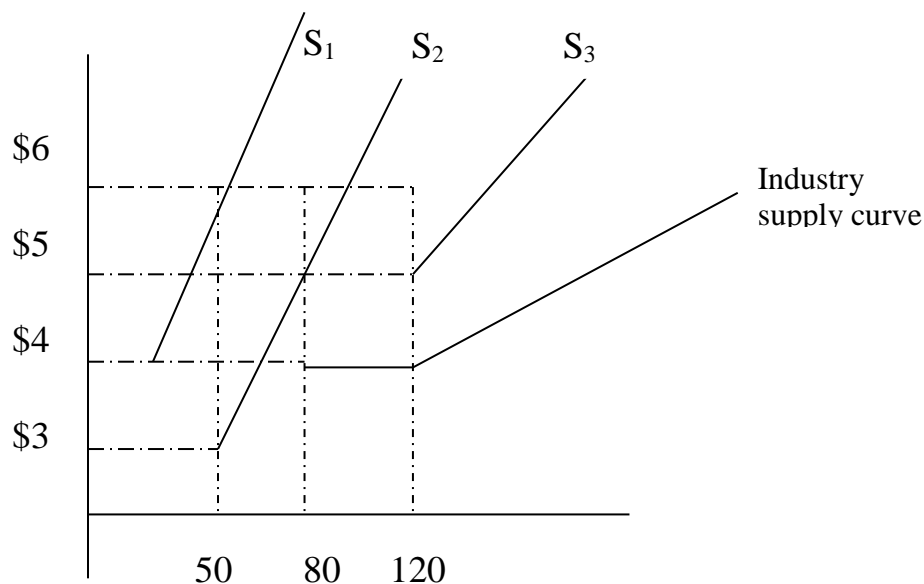


Fig. 5.10 the industry- supply curve is the horizontal summation (at each price) of the supply curves of all firms in the industry.

When the market price falls below \$ 4, only firm2 exists in the market. Thus, for prices below\$ 4, the industry supply curve is identical with the supply curve is identical with the supply curve of firm 2. Similarly, for price levels ranging from \$4 to \$5, only firm1 and firm2 are producing and searching in the market. Thus, the industry- supply curve for this range of price is the sum of the quantities supplied by firm 1 and firm 2, and so on.

5.6 Short run equilibrium of the industry

Dear learner, what is the necessary condition for a market to be equilibrium?

In our previous discussions, we have seen the short run equilibrium of the individual firm. At that time, we have said that the short run equilibrium of the firm is defined by the point of intersection of the horizontal MR curve (or the demand curve that the individual firm faces) and MC curve (or the supply curve of the firm). In other words, the short run equilibrium of an individual firm is defined by equality of MC of firm and the market price.

Short run equilibrium of the industry is defined by the intersection of the market demand and market supply. The intersection of market demand and market supply of a given commodity determines the equilibrium price and quantity of the commodity in the market.

While discussing the short run equilibrium of an individual firm we have said that the demand curve that an individual firm faces is horizontal line (perfectly elastic). This is due to the fact that; since there are large numbers of sellers in the market, an individual firm is too small to influence the market price. Rather, the firm sells any amount demanded at the prevailing market price.

Unlike the individual demand curve, the market demand curve (the total demand curve that the industry faces) is down-ward sloping, indicating that as the market price of the commodity increases, the total quantity demanded for the product decreases and vice versa. In fig.5.11 the industry is in equilibrium at price P_e , at which the quantity demanded and supplied is Q_e . At this equilibrium market price, individual firms can earn a positive profit, zero profit (normal profit) or even can incur a loss depending on their cost structures.

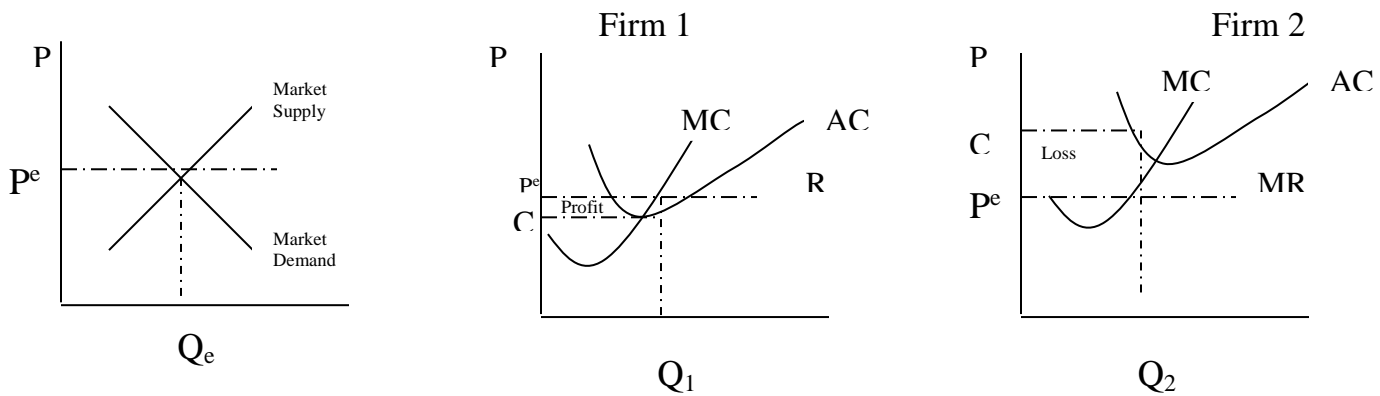


Fig5.11:- short run equilibrium of the industry. Short run equilibrium of the industry is defined by the intersection of the market demand and the industry supply Curve. At equilibrium price, P_e firm 1 gets a positive profit because the average cost of the firm at equilibrium is less than the market Price, p_e . On the other hand, firm 2 is incurring a loss as its average cost is higher than the market price.

5.7 The long-run Equilibrium

1-Equilibrium of an individual firm in the long run

In the long run, firms are in equilibrium when they have adjusted their plant size so as to produce at the minimum point of their long run AC curve, which is tangent (at this point) to the demand curve defined by the market price. That is, the firm is in the long run equilibrium when the market price is equal to the minimum long run AC. Thus since price is equal to long run AC (LAC now on) at the long run equilibrium, firms will be earning just normal profits (zero profits), which are included in the LAC. Firms get only normal profit in the long run due to two reasons.

First, if the firms existing in the market are making excess profits (the market price is greater than their LACs) new firms will be attracted to the industry seeking for this excess profit. The entry of new firms results in two consequences:

A. The entry of new firms will lead to a fall in market price of the commodity (which is shown by the down ward shift of the individual demand curve). This happens because entry of new firms will increase the market supply of the commodity (which is shown by the right ward shift of the industry supply), resulting in the lower market price. More over, if firms are getting excess profit, they have an incentive to expand their capacity of production, which increases the market supply and then reduces the market price.

B. More over, the entry of new firms results in an upward shift of the cost curves. This happens because, when new firms enter into the market the demand for factors of production increases which exerts an upward pressure on the prices of factors of production. An increase in the price of factors of production in turn shifts the cost curves up ward. These changes (decrease in the market price and upward shift of the cost curves) will continue until the LAC becomes tangent to the demand curve defined by the market price. At this time, entry of new firms will stop since there is no positive profit (since $P = LAC$) which attracts new firms in to the market.

Second, if the firms are incurring losses in the long run ($P < LAC$) they will leave the industry (shut down). This will result in higher market price (because market supply of the commodity decreases) and lower costs (because the market demand for inputs decreases as the number of firms in the market decreases). These changes will continue until the remaining firms in the industry cover their total costs inclusive of the normal rate of profit.

Thus, due to the above two reasons, firms can make only a normal profit in the long run.

The following figure shows how firms adjust to their long run equilibrium position excess profit (higher price than minimum lack) if the market price is p , the firm is making excess profit working with plant size whose cost is denoted by SAC, (short run average cost¹). It will therefore have an incentive to build new capacity or larger plant size and it moves along its LAC. At the same time new firms will be entering the industry attracted by the excess profits. As quantity supplied in the market increases (by the increased production of expanding old firms and by the newly established ones) the supply curve in the market will shift to the right and price will fall until it reaches the level of P_1 , at which the firms and the industry are in the long- run equilibrium.

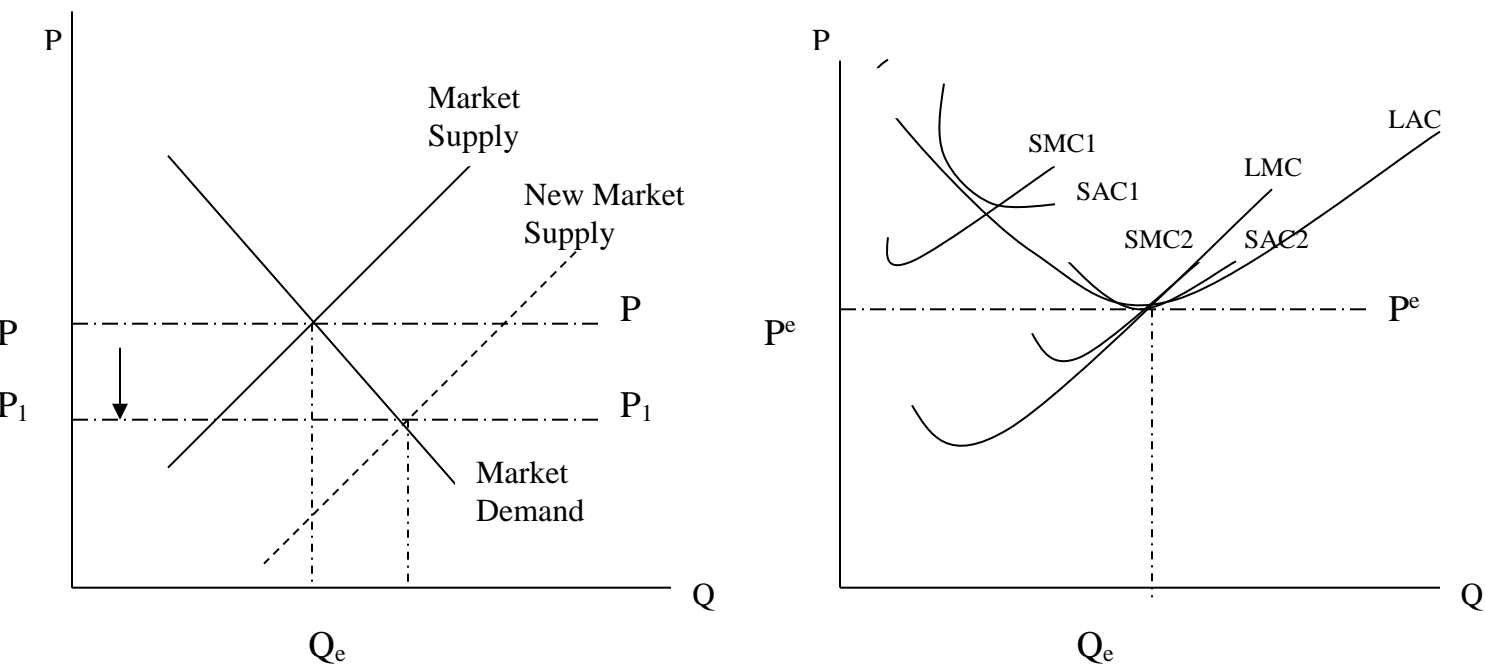


Fig5.12: Long run equilibrium of the firm. Entry of new firms reduces the market price from p to p_1 (in panel A) and the long run equilibrium is established at E (panel B).

The condition for the long run equilibrium of the firm is that the long run marginal cost (LMC) should be equal to the price and to the LAC i.e. $LMC = LAC = P$.

The firm adjusts its plant size to so as to produce that level of output at which the LAC is the minimum possible, given the technology and prices of inputs. At equilibrium the short – run marginal cost is equal to the long run marginal cost and the short –run average cost is equal to the long run average cost. Thus, given the above condition, we have,

$$SMC = LMC = SAC = LAC = P = MR$$

This implies that at the minimum point of the LAC the corresponding short run plant is worked at its optimal capacity so that the minimum of the LAC and SAC coincide.

Long run shut down decision

Dear learner, do you remember the short run shut down point of a firm? If you don't remember, please revise section-.for the time being the time being the following paragraph may remind you. (About the short-run shut down point)

In the short-run the firm should continue production as far as the market price is greater than the minimum AVC, If the market price falls below the minimum AVC, the firm is well advised to shut down because if it shut down it will lose only the fixed costs but if it continues production the loss is greater than the fixed cost.

The long-run shut down decision (point) is different from that of the short run. The firm shuts down if its revenue is less than its avoidable or a variable cost. In the long run all costs are variable because the firm can change the quantity of all inputs. Thus, in the long run the firm shuts down when its revenue falls below the long run total cost. In other words, in the long run shut down decision occurs if the market price falls below the minimum LAC of the firm.

The long-run supply curves the firm

Previously, we have noted that in the long run the firm shuts down if the market price is below the its minimum long run average cost. Thus, the firm will not supply for all price levels below the minimum LAC. On the other hand, the firm's long run equilibrium output is defined by the equality of the MR and its LMC. As a result, a firm's long-run supply curve is its LMC curve above the minimum of its long-run average cost curve.

Long run supply curve of the industry

The long run supply curve of the industry is the horizontal sum of the supply of individual firms just like the case of short run supply curve of the industry. Thus, the long run supply curve of the industry is upward sloping, provided that the firms are of different size. This is because, firms with relatively lower minimum LAC, are willing to enter the market than others. So that as the market price increased in the long run more firms will find it profitable to enter the market, resulting in upward sloping long-run supply curve of industry.

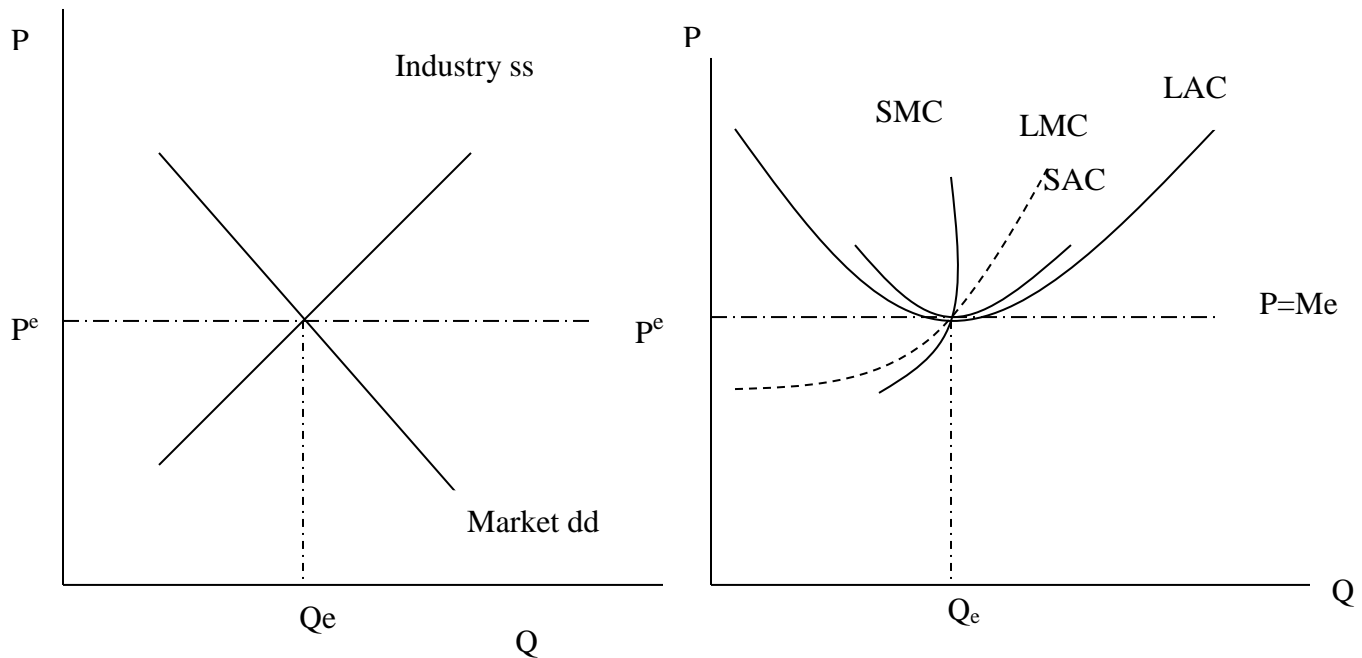
Long-run equilibrium of the industry

An industry is in the long-run equilibrium when the price is reached at which all firms are in equilibrium. That is, when all firms are producing at the minimum point of their LAC curve and

making just normal profits, the industry is said to be in the long-run equilibrium. Under these conditions there is no further entry or exit of firms in the industry (since all the firms are getting only normal profit), so that the industry supply remains stable.

The long-run equilibrium of the industry is shown by fig 5.13. At the market price, P , the firms produce at their minimum LAC, earning just normal profits. At this price all firms are in equilibrium because

$LMC=SMC=P=MR$ and they get only normal profit because $LAC=SAC=P$.



Industry equilibrium

Firm's equilibrium

Fig5.13: long-run equilibrium of the industry is defined by the price at which all individual firms are in equilibrium, marking just normal profit.

Dear learner, while the industry is in the short run equilibrium, we have seen that, individual firms can earn positive, normal or negative profits depending on the level of their AC s relative to the equilibrium market price. However, this is not the case in the long-run. That is, while the industry is in the long run .equilibrium all firms earn only normal profit.

5.8 Perfect competition and optimal resource allocation

In the perfect competition, the market mechanism leads to an optimal allocation of resources. The optimality is shown by the following conditions all of which prevail in the long run equilibrium of the industry;

- a) The output is produced at the minimum feasible cost. That is all firms produce at the minimum of their LAC.
- b) Consumers pay the minimum possible price which just covers the marginal cost of production, that is, price equals just opportunity cost so that the consumers are not exploited.
- c) Plants are used at full capacity in the long-run so that there is no waste of resources. That is, at the long run equilibrium the short run average cost is also minimum.
- d) Firms earn only normal profits.

These conditions justify the fact that perfect competition results in optimal resource allocation.

Summary

A perfectly competitive market is a market structure characterized by large number of buyers and sellers, homogenous product, free entry and exit, no government regulation and perfect knowledge of the market conditions,. Each buyer and seller in this market structure is a price taker due to the large number of buyers and sellers and homogeneity of the product.

Since an individual seller is a price taker, profit maximization in perfect competition involves determination that level of output which maximizes the firm's profit for the given market price. In the short run, the equilibrium (profit maximizing) level of output is that level of output which satisfies the condition: $MC = MR$ and MC is rising.

The fact that a perfectly competitive firm is in equilibrium doesn't necessarily mean that the firm enjoys a positive profit. Whether the firm gets positive or negative profit depends on the firm's average cost (AC) production with respect to the market price (P). If $AC > P$, $AC < P$ or $AC = P$ at equilibrium the firm will get negative, positive or zero profit respectively. When the market price of the commodity lower than the minimum AVC , the firm should shut down to minimize its losses. Thus, the firm would supply a commodity to the market only if price of the commodity is at least sufficient to cover the average variable cost. Hence, a firm's short run supply curve is that portion of MC curve which lies above the minimum AVC .

Short run equilibrium of the industry is defined by the intersection of market demand and industry supply curves.

In the long run, profit maximizing competitive firms chooses the output at which price is equal to long- run marginal cost and long –run average cost. In the long run, a firm can get only a normal profit.

Perfect competition results in efficient allocation of resources.

UNIT SIX

PRICE AND OUT PUT DETERMINATION UNDER MONOPOLY

Objectives

After successful completion of this chapter, you will be able to:

- Basic features of monopoly market and factors which give rise to monopoly.
- The nature of demands and revenue curves under monopoly.
- How to determine equilibrium price and out put under different conditions of monopoly such as Multiplan monopolists and price discriminating monopolists.
- Different types of price discrimination and conditions required to effectively exercise price discrimination.
- How monopoly results in welfare loss.

6.1 Introduction

Definition of monopoly

Dear learner, what do you know about monopoly?

(You can use the space left below to write your response)

In the last chapter we have seen perfectly competitive market structure in which there is large number of firms selling homogeneous products. Monopoly is quite opposite to perfectly competitive market. And it is defined as: ***a market situation in which a single seller sells a product or provides a service for which there is no close substitute.*** In monopoly there are no similar products whose prices or sales will influence the monopolist price or sales. In another words, cross elasticity between monopolist product and other commodities is zero or low. Since there is a single seller in monopoly market structure, the firm is at the same time the industry.

Common characteristics of monopoly

Monopoly markets share the following common characteristics.

1-Single seller and many buyers

There is a single seller who sells the product to many buyers.

2-Absence of close substitutes

A product produced by a monopolist has no close substitute so that consumers have no alternative choices to substitute one product for another.

3-Price maker

Dear learner, in perfectly competitive market, we have said that, both sellers and buyers are price takers. However, the monopolist is a price maker. Facing a downward sloped demand curve for its product, the monopolist can change its product price by changing the quantity of the product supplied. For example, the monopolist can increase the price of its product by decreasing the quantity of supply.

4-Barrier to entry

In monopoly, new competitors can not freely enter into the market due to some barriers which can be economical, technical, legal or other type of barriers.

6.2 Causes for the emergence of monopoly

Dear learner, think of any monopoly firm in our country and try to analyze the reason why the firm maintains its monopoly power. There are many factors that create monopoly and help the monopolists to maintain monopoly power. Some of the factors will be discussed below.

1. Ownership of strategic or key inputs.

A firm may own or control the entire supply of a raw material required for the production of a commodity. Such firms are not willing to sell the raw materials to another firm. For example, until the second world war, the aluminum Company of America (Alcoa) controlled practically the entire supply of Bauxite (the basic raw material necessary for the production of aluminum), giving it almost a complete monopoly in the production of aluminum in the United States. To come to our country, Ambo Mineral Water can be taken as an example. Ambo mineral water has monopolized the natural mineral water.

2. Exclusive knowledge of production technique.

Most of the beverage (soft drink) companies such as Coca Cola Company have maintained monopoly power over supply of their product partly due to exclusive knowledge of the ingredient chemicals required for the production of their product.

3. Patents and copyright

Patents and copyrights are government supported barriers to entry. Patents are granted by the government for 17 years as an incentive to investors. Authors of books, artistic works (such as cassette, video, etc) are the best examples of such monopoly. For example, no one, except Adama University, can copy and sell this course material as Adama University has an exclusive copy right over the material.

4. Government Franchise and License

Another cause for the emergence of monopoly is government franchise. Franchise is a promise by the government for a firm to prohibit the establishment of another firm (by another person) that produces the same product or offers the same service as the original one.

For example, when the first Bank in Ethiopia, Abyssinia Bank was established, Emperor Minilik has promised for the Egyptian firms (the owner of the Bank) that they will monopolize the Banking service in Ethiopia for 50 years. Postal service in Ethiopia, Ethiopian television, telecommunication service in Ethiopian etc. are other examples of monopoly.

5. Economies of scale may operate (i.e. the long run average cost may fall)

Another cause for the emergence of monopoly is economies of scale in production. A firm is said to have economies of scale if its long run average cost is declining. In such a situation, when the incumbent firm observes that new firms are entering into the market, it will produce large amount of output to minimize its unit cost of production and will charge a lower price than the new firms to deter entry. Such a monopoly is called natural monopoly.

Aside from the few cases of monopoly mentioned above, pure monopoly is rare and most governments discourage pure monopoly because monopoly is deemed to create inefficiency. For example, had it been the case that the telecommunication services are not monopolized in our country, their prices would have been lower. But through pure monopoly is rare, the pure monopoly model is useful for analyzing situations that approach pure monopoly and for other types of imperfectly competitive markets (i.e. monopolistic competition and oligopoly)

6.3 The demand and revenue curves of the monopoly firm

In the previous unit, we have seen that the perfectly competitive firm is a price taker and faces a demand curve that is horizontal or infinitely elastic at the price (determined by the intersection of the industry or market demand and supply) of the commodity. But, remember that the market

demand curve is down ward sloping. However, a monopolist firm is at the same time the industry and thus, it faces the negatively sloped market (industry) demand curve for the commodity. In other words, because a monopolist is the sole seller of a commodity, it faces a down ward sloping demand curve. This means, to sell more units of the commodity, the monopolist must lower the commodity price.

Conversely, if the monopolist decides to raise the price of the product, it will reduce the quantity of supply with out worrying about the competitors, who by charging lower prices would capture a large share of the market (customers) at the expense of him. So the monopolist can manipulate the price of its commodity by changing the quantity of supply. To sell more units of the commodity, the monopolist will charge lower price and vice versa. Hence, the demand curve facing the monopolist is negatively sloped, showing the inverse relationship between market price and quantity demanded.

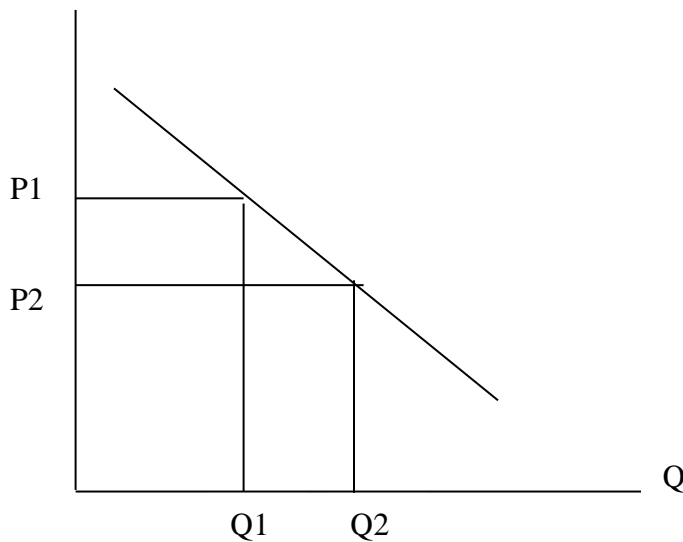


Fig.6.1 the demand curve facing the monopolist firm is down wards sloping. At price p1, the firm sells only Q1 outputs. To sell more units the firm should reduce the price.

Mathematically, assuming that the demand curve is linear, it can be written in the following form.

$$P = a - bQ$$

Where P – is the market price

Q – is the quantity of sales (quantity demanded)

a & b – are any positive constants

The total revenue of the monopolist can be obtained by multiply the market price with the quantity of sales

That is,

$$TR = P \cdot Q$$

Substituting $(a - bQ)$ for P

$$TR = (a - bQ) Q$$

$$TR = aQ - bQ^2$$

Hence the total revenue curve of the monopolist firm has an inverse U- shape. The total revenue of a monopolist firm first increases with the quantity of sales (over the elastic range of the demand curve), reaches its maximum (when the demand curve is unitary elastic), and finally decreases when quantity of sales increases (over the inelastic range of the demand curve) the following figure illustrates this fact.

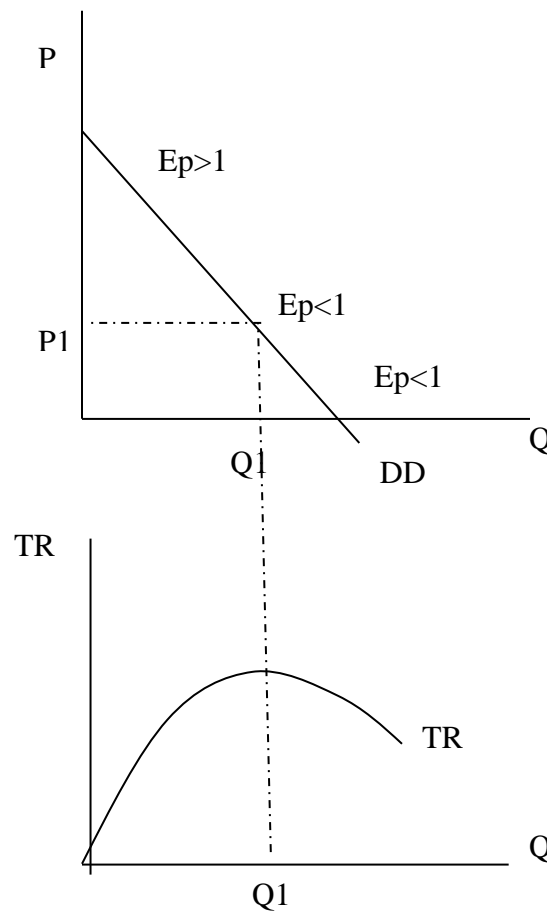


Fig: 6.2 the shape of total revenue curve and its relationship with the price elasticity of demand. When $E_p > 1$ TR and Q have positive relation, at a point where $E_p = 1$, TR curve reaches its maximum and when $E_p < 1$, TR and Q have negative relation.

The MR curve of monopolist firm is down ward sloping (decreases with quantity of sales). The fact that the monopolist must lower the price to increase its sales causes the MR to be less than price except for the first unit. This is so because when the firm reduces the commodity price to sell one more unit all units which would have been sold at the original higher price will now be sold at the new (lower) price. The following table may help you better understand this fact.

Price	Quantity	TR	AR	MR
\$ 6	0	0	6	-
5	1	5	5	5
4	2	8	4	3
3	3	9	3	1
2	4	8	2	-1
1	5	5	1	-3

The above table shows that as output increases the TR first increases, reaches its maximum (when the firm sells the third unit) and then starts to fall.

The MR is less P except for the first unit. For example, when the firm decreases the price from \$5 to \$4 marginal revenue decreases from \$5 to \$3. That is, at the second unit MR (\$3) is less than the P (\$4).

This is because, when the market price is \$5, the firm will sell one unit and will get a TR of \$5 and the MR of this first unit is \$5. When the price decreases to \$4, both the first and the second unit are sold at \$4 and the firm receives total revenue of \$8. Now, the MR that the firm obtains from the second unit is only \$3. Hence for a down ward sloping demand curves (in monopoly) the MR of the firm is less than the market price. Note that the AR of a monopolist is always identical to the P or demand curve.

In general, the MR curve of a monopolist firm is negatively sloped. The MR will be positive over the elastic range of the demand curve (because TR is increasing over this range), zero when the price elasticity of demand is unitary (because the TR is at its maximum level) and will have a negative sign over the inelastic range of the demand curve(because TR is decreasing).

The following figure illustrates the relationship between price elasticity of demand and MR

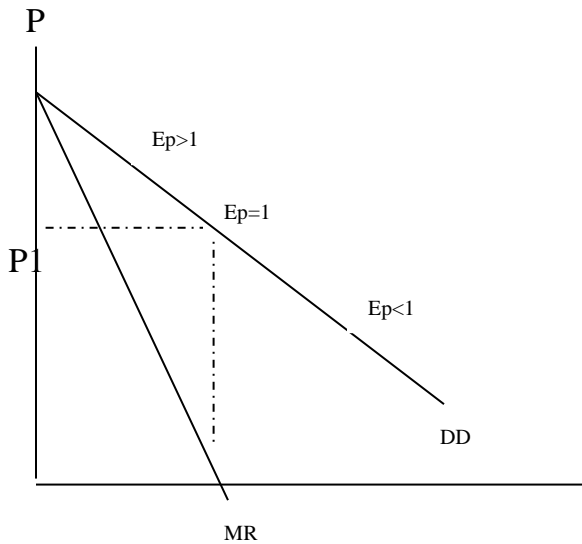


Fig: 6.3 the relationship between MR and P. The MR of a monopolist lies below the commodity price for each unit sold (except the first unit) and it is negative over the inelastic range of the demand curve.

Mathematically, it can be shown that MR is less (steeper) than the AR or demand curve.

Suppose a monopolist's demand curve is given by

$$P = a - bQ$$

Where a & b - are any positive constants

P & Q - are price and quantity.

$$\begin{aligned} TR &= P \cdot Q = (a - bQ) \cdot Q \\ &= aQ - bQ^2 \end{aligned}$$

By definition MR is change in TR that happens due to a one unit change in quantity of sales.

Symbolically,

$$MR = \frac{dTR}{dQ} = \frac{d(aQ - bQ^2)}{dQ} = a - 2bQ$$

Thus, $MR = (a - 2bQ)$ has a slope which equals twice the slope of demand (average revenue) curves. This implies that MR is less than AR or demand or price.

Dear learner, we have seen that a monopolist firm faces a down ward sloping demand curve. Exception to the law of demand under monopoly is that the firm can increase the quantity of sales only through promotional activities (without price cut).

Profit maximization in the short run

Dear learner, do you remember how a perfectly competitive firm maximizes its profit? In this section, we examine the determination of equilibrium price and out put by a monopolist in the short run. We will also show that a monopolist, like a perfectly competitive firm, can incur losses in the short run. Finally, we demonstrate that, unlike the case of the perfectly competitive firm, the monopolist's short run supply curve can not be derived from its short run marginal cost curve (the supply curve of the monopolist is indeterminate).

To start with, it was discussed in the last chapter that in a perfectly competitive market price is given and profit maximization involves only looking for the profit maximizing unit of output, given the market price. But, under monopoly, the firm is a price maker and has a power to alter the level of output. Thus, profit maximization under monopoly involves determination of the price and output combination that yields the firm the maximum possible profit.

Price and out put combination that maximizes the monopolist profit can be determined in the similar fashion as that of the perfectly competitive firm. That is, price- output combination that yields the monopolist the maximum profit can be determined in two ways:

1. Total approach
2. Marginal approach

Now let us see the two approaches one by one.

1. Total approach

In this approach the profit maximizing unit of output is defined as that level of output where the positive difference between TR and TC is maximal or the negative difference between TR and TC is minima. The equilibrium price can be determined by dividing the TR corresponding to the

equilibrium output level to the equilibrium output. The following figure tells more about this approach.

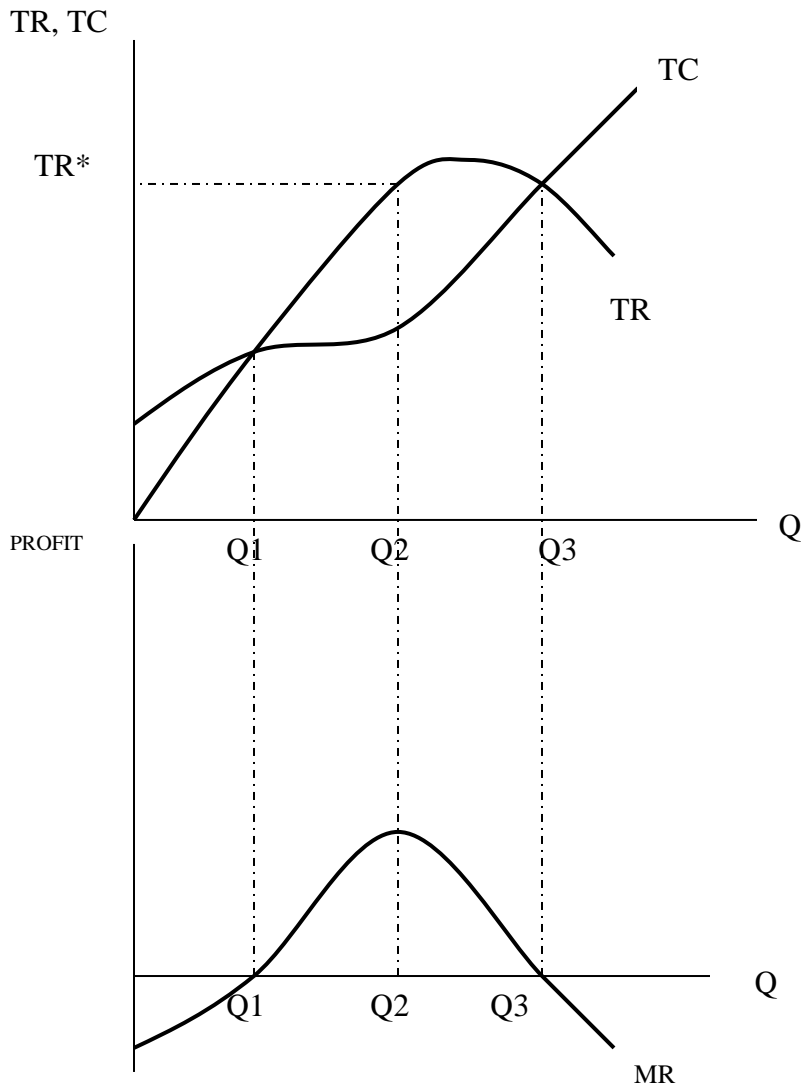


Fig 6.3 Short –run equilibrium of the monopolist Total approach: The TR of the monopolist has an inverse U shape because the monopolist must lower the commodity price to sell additional units. The STC has the usual shape. The total profit is maximized at Q2, where the positive difference between the TR and STC is the greatest. Profit is negative for output levels below Q1 and above Q. In this approach the profit maximizing price is given by the ratio of TR to Q2.*

2. Marginal approach

In this approach the profit maximizing level of output is that level of output at which marginal cost curve cuts the marginal revenue curve from below. The equilibrium (profit maximum) price is the price corresponding to the equilibrium quantity from the demand curve.

Consider the following figure:

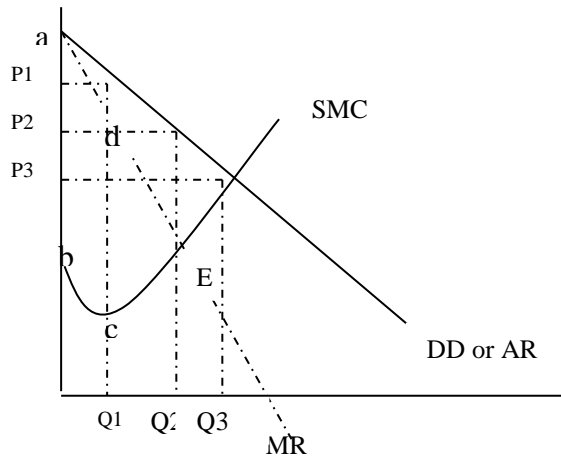


Fig. 6.4 Short-run equilibrium of the monopolist: marginal approach. Equilibrium output is Q_2 , where MC and MR curves intersect each other and MC curve is up ward sloping. Equilibrium price is the price corresponding to the equilibrium quantity, Q_2 (i.e. p_2).

Note that, a monopolist charges a price which exceeds the MC of production, unlike the case of the perfectly competitive firm. Now, how can we be sure that Q_2 is the profit maximizing unit of output? To answer this question, note that in the total approach the level of profit at a given level of output is the vertical distance between the TR and TC (i.e., $\Pi = TR - TC$.)

In the marginal approach, however, the level of profit at a given level of output is not the distance between the MR and MC curves. Rather it is the area between marginal revenue and marginal cost curves starting from the origin up to the given level of output. Symbolically, the level of profit say at Q_2 level of output is:

$$\Pi \text{ at } Q_2 = TR \text{ at } Q_2 - TC \text{ at } Q_2 \text{ ----- Total approach}$$

$$\Pi \text{ at } Q_2 = \int_0^{Q_2} (MR - MC) dQ \text{----- Marginal approach}$$

Given the level of profit as the area between the MR and MC, let's come back to our question above.

In the above figure, we have said that the equilibrium price is Q_2 and the level of profit is the area between that part of MR and MC curves between the origin and Q_2 (area abE).

Now we are going to prove whether this level of output is actually the profit maximizing level of output. To prove this, suppose initially that the monopolist produces a smaller quantity Q_1 and receives the higher price, P_1 . The level of profit at this level of output the area between that part of MR and MC curves ranging from the origin up to Q_1 (i.e. area abcd). Hence the firm loses the level of profit given by the area cde by producing Q_1 level of output instead of Q_2 . Thus, any level of output below Q_2 can not yield the firm the maximum profit. Similarly, it can be shown in the same way that any level of output above Q_2 can not maximize the firm's profit.

In other words, for any level of output below Q_2 , MR is greater than the MC, implying that each additional unit of output yields larger additional (marginal) revenue to the firm than the additional cost of producing it. Hence the firm should produce additional units until Q_2 . On the other hand, for all levels of output above Q_2 , the MC of producing additional unit of output is greater than the MR obtained from it. Hence, the firm should not expand its output above Q_2 . This argument can prove the fact that Q_2 is the profit maximizing level of output.

Mathematically, the profit maximizing condition of $MR = MC$ and MC is increasing can be shown as follows.

$$\Pi = TR - TC$$

$$\Pi \text{ is maximized when } \frac{d\Pi}{dQ} = 0$$

$$\text{That is, } \frac{D\Pi}{dQ} = \frac{dTR}{dQ} - \frac{dTC}{dQ} = 0$$

→ $MR - MC = 0$

→ $MR = MC$ first order condition

The second order condition of profit maximization is

That: $\frac{d^2\pi}{dQ^2} < 0$

That is, $\frac{d^2\pi}{dQ^2} = \frac{d^2TR}{dQ^2} - \frac{d^2TC}{dQ^2} < 0$

$\frac{dMR}{dQ} - \frac{dMC}{dQ} < 0$ (Because $\frac{d^2TR}{dQ^2} = \frac{d\left(\frac{dTR}{dQ}\right)}{dQ} = \frac{dMR}{dQ}$ and the same for MC)

Slope of MR- slope of MC < 0

Slope of MC > slope of MR ----- the second order condition

Numerical example

Suppose the monopolist faces a market demand function given by $P=40-Q$. The firm has a fixed cost of \$ 50 and its variable cost is given as $TVC=Q^2$ determine:

- a) the profit maximizing unit of output and price
- b) the maximum profit

Solution

Given: $p=40-Q$

$TFC=50$

$TVC= Q^2$

- a) equilibrium condition is $MR=MC$, and slope of $MC >$ slope of MR .

$TR=P.Q = (40-Q) Q =40Q- Q^2$

$TC=TFC+TVC =50 + Q^2$

Now, $MR = \frac{dTR}{dQ} = \frac{d(400 - Q^2)}{dQ} = 40 - 2Q$

$$MC = \frac{dTC}{dQ} = \frac{d(50 + Q^2)}{dQ} = 2Q$$

$$MR=MC \quad 40-2Q=2Q$$

$$40=4Q$$

$$Q=10$$

Second order condition: slope of $MR = \frac{dMR}{dQ} = -2$

Slope of $MC = \frac{dMC}{dQ} = 2$

$$\frac{dMC}{dQ} > \frac{dMR}{dQ}$$

the second order condition is met

Thus, the profit maximizing level of output is 10 and the profit maximizing price is obtained by substituting the profit maximizing quantity (10) in the demand function.

$$\text{Thus, } P = 40 - Q$$

$$P = 40 - 10 = 30$$

b) The maximum profit is the level of profit obtained from selling 10 units at \$ 30 each.

$$\Pi = TR - TC$$

$$\text{But } TR = P \cdot Q$$

$$= \$ 30 * 10 = \$ 300$$

$$TC = 50 + Q^2 = 50 + 10^2 = \$ 150$$

The maximum Π is thus \$ 300 - \$ 150 = \$ 150.

Exercise:

Suppose the monopolist faces the market demand function given by $Q = \frac{144}{P^2}$. The AVC of the

firm is given as $AVC = Q^{1/2}$ and the firm has a fixed cost of \$ 5

- a) determine equilibrium P&Q
- b) determine the maximum profit

6.5 Mark up pricing

Although prices can be determined by equating MC and MR, most managers have only limited knowledge of the AR and MR functions that their firm faces. Mark- up pricing helps us to translate the equilibrium condition $MR = MC$ into a convenient form that can easily be applied in practice. Accordingly,

$$P = \frac{MC}{1 + \frac{1}{ed}}$$

This formula is derived from the equilibrium condition $MR = MC$ as follows.

The TR of the monopolist is:

$$TR = P \cdot Q$$

Marginal revenue is:

$$MR = \frac{dTR}{dQ} = \frac{d(P \cdot Q)}{dQ} = \frac{dP \cdot Q}{dQ} + \frac{dQ}{dQ} \cdot P \quad \dots\dots (1) \quad (\text{Product rules of differentiation.})$$

On the other hand the price elasticity of demand, ep is

$$ep = \frac{dQ}{dP} \cdot \frac{P}{Q} \quad \text{and by rearranging (Multiplying both sides by } \frac{Q}{P} \text{)}$$

$$\frac{dQ}{dP} = ep \cdot \frac{P}{Q} \quad \text{----- (2)}$$

Taking the reciprocal of both sides we obtain :

$$\frac{dP}{dQ} = \frac{P}{eP \cdot Q} \quad \text{----- (3)}$$

Substituting equation (3) in equation (1) we get:

$$MR = \frac{P}{eP \cdot Q} \cdot Q + P$$

$$MR = P \left(\frac{1}{ep} + 1 \right) \quad \text{----- (4)}$$

By substituting equation (4) in the equilibrium condition $MR = MC$ we get the following

$$P\left(1 + \frac{1}{ep}\right) = MC$$

$$P = \frac{MC}{1 + \frac{1}{ep}}, \text{ where } Ep = \frac{dQ}{dP} \cdot \frac{p}{Q}$$

Here, $|Ep|$ should be greater than one. Other wise the price would be negative. In other words, if price elasticity of demand is inelastic (or $|Ep| < 1$), it implies that MR is negative, which requires the MC to be negative for equilibrium to occur. But, MC can never be negative. ***Hence, a monopolist operates only over the elastic range of its demand curve.***

Numerical example:

Suppose the monopolist's total cost function is given as $TC = 10 + 1.5Q$. The firm estimates the price elasticity of demand to be -4, determine the profit maximizing price.

Solution:

Given: $TC = 10 + 1.5 Q$

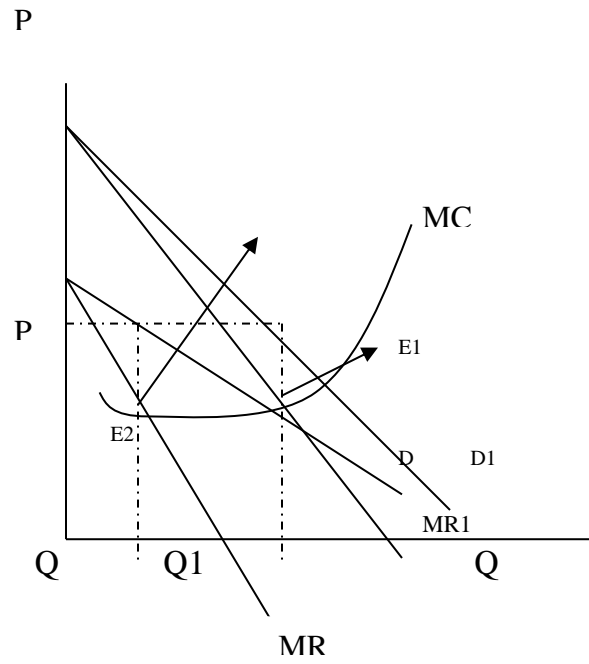
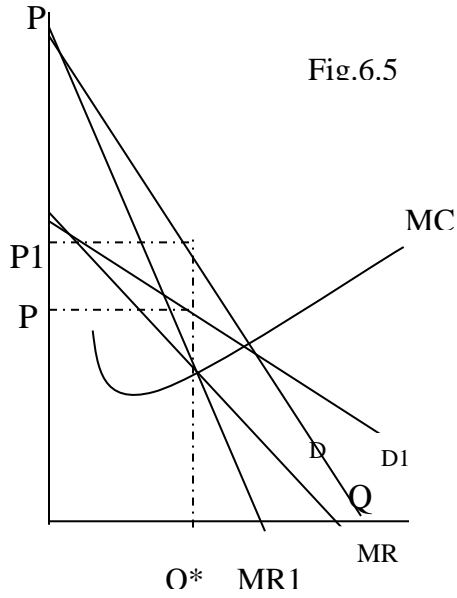
$$Ed = -4$$

Required: $P=?$

$$P = \frac{MC}{1 + \frac{1}{ed}} ; MC = \frac{dTC}{dQ} = \$1.5. \text{ Thus, } P = \frac{1.5}{1 - \frac{1}{4}} = \frac{1.5}{3/4} = \$2$$

6.6 Absence of unique supply Curve under Monopoly

Under Perfect competition, you remember that firms have unique supply curve. That is there is unique supply price for each unit of output supplied. In monopoly supply price is not unique. A given quantity could be supplied at different prices and different quantities can be sold at the same price, depending on market demand and marginal revenue. Hence there is no one to one correspondence between P and Q under monopoly. Consider the following figures.



Panel-1

In this panel, the same quantity Q^ is sold at different prices depending on the market demand. If the market demand is $D1$ and the MR curve is $MR1$, equilibrium occurs when $MR1$ cuts MC curve and the equilibrium price and quantity are $P1$ and Q^* . If the market demand for the monopolist product decreases to D , the monopolist can still sell the same quantity Q^* by lowering the price. So, there is no unique (or one to one correspondence) between P & Q , as the same Q^* is matched with two different price, P & $P1$*

Panel-2

In this panel, initially equilibrium is $E1$ (where $MR1=MC$) and equilibrium P & $Q1$. when the demand for monopolist product decreases to D the new equilibrium becomes $E2$ where the new $MR=MC$. At the new equilibrium, price is the same, but the monopolist sell only Q amount of output i.e. the monopolist sells lower quantity at the original price when the dd decreases.

Therefore, there is no unique supply curve under monopoly.

6.7 Long – run Equilibrium under Monopoly

The monopolist’s long run condition is different from the perfectly competitive firms’ long run situation in respect of the entry of new firms into an industry. In perfectly competitive market there is free entry in the long run. Nevertheless, entrance is barred by several factors in monopoly. More over, we have seen that a perfectly competitive firm can earn only normal profit in the long run. The monopolist firm can, however, get a positive profit even in the long run because there are entry barriers that discourage new firms to enter the industry, attracted by the positive profit.

Let us now examine the long run equilibrium situation for single plant monopolist. If the monopolist incur loss in the short run ($SAC > P$) and if there is no plant size that will result in super normal profit in the long run given the market size, the monopolist must stop operation (shut down). If the monopolist makes ($P > SAC$) in the short run in a given plant, the monopolist not only continue its operation but also looks for different plant size to expand, so that could maximize profit in the long run. But at what output level the monopolist maximizes its profit? A monopolist maximizes its long run profit when it produces and sells that output level where $LMC = MR$, slope of LMC being greater than the slope of MR at the point of intersection, and the optimal plant size is the one whose SAC curve is tangent to the LAC at the point corresponding to long run equilibrium output.

Let's illustrate the equilibrium situation graphically.

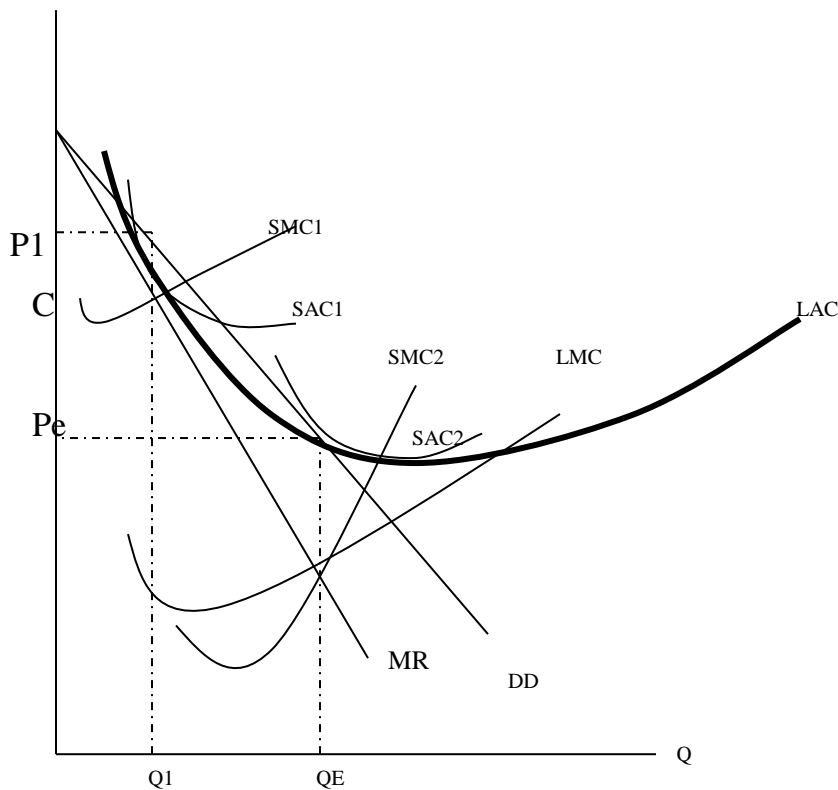


Fig 6.6 Suppose initially the monopolist builds the plant size having the costs SAC_1 and SMC_1 the equivalence of SMC_1 and MR leads into producing and marketing output levels Q_1 and P_1 , making a unit profit of $P_1 - C$, since the monopolist is making a positive profit, it decide to continue its operation and looks for a more profitable plant size in the long run. This long run plant is attained when $LMC = MR$, and the corresponding output level and price are Q_e and P_e respectively.

Finally, it should be noted that there is no certainty in the long run that the monopolist will reach the optimal plant size (minimum LAC), as in perfectly competitive case. The monopolist may reach optimal plant size or even may exceed the optimal size if the market demand allows him (or if there is enough demand which absorb that level of output).

6.8 Monopoly Power

Pure monopoly is rare. It is common to see market in which several firms compete with one another. Although many firms compete with one another some firms may have greater monopoly power than the others. Hence they can affect the market price more than other firms. You may think that firms which share the larger part of the market supply have greater monopoly power. But this can not be necessarily true. What matters is the consumers' preference for the firm's product. If most consumers prefer the product of the firm to other substitutes, the firm has greater monopoly power than other firms in the market and the firm can slightly increase the price of his commodity being confident that he will not lose its customers.

Now let us come to discuss measurement of monopoly power. The important distinction between a perfectly competitive firm and a monopolist is that: for the competitive firm price equals marginal cost; and for the firm with some monopoly power price exceeds marginal cost. Therefore, a natural way to measure a monopoly power is to examine the extent to which the profit maximizing price exceeds marginal cost. In particular, we can use the mark up ratio of price minus marginal cost to price that we introduced earlier. This measure of monopoly power, introduced by an Economist Abba Lerner in 1934, is called Lerner index of monopoly power. Lerner index (L) is the difference between price and marginal cost, divided by price.

Symbolically, $L = \frac{P - MC}{P}$ L - is learner index of monopoly power

L always has a value between zero and one. For perfectly competitive firm, $P = MC$, so that $L = 0$ i.e. there is no monopoly power in perfect competition. Hence, the larger L is the greater the degree of monopoly power.

The learner index of monopoly power (L) can also be expressed in terms the price elasticity of demand for the firm's product as follows.

$L = \frac{-1}{ep}$, where $ep = \frac{dQ}{dP} \cdot \frac{P}{Q}$ and $|ep| > 1$, since the monopolist operates only over the elastic range of its demand curve.

Proof: From the mark-up pricing, we know that

$$P = \frac{MC}{1 + \frac{1}{ep}}$$

Re arranging this we get

$$1 + \frac{1}{ep} = \frac{MC}{P}$$

$$\frac{1}{ep} = \frac{MC}{P} - 1$$

$$\frac{1}{ep} = \frac{MC - P}{P}$$

$$-\frac{1}{ep} = \frac{P - MC}{P}$$

But, $\frac{P - MC}{P}$ is the learner index of monopoly power (L),

$$\text{Thus, } L = \frac{-1}{ep}$$

6.9 The multi- plant monopolist

We have seen that a monopolist maximizes its profit by producing that level of output where MR equals MC. For many firms, however, production takes place in two or more different plants whose operating costs can differ. To minimize transport cost, to approach the consumers or for different reasons a monopolist may establish more than one plant in different areas. The operating costs of these plants can also vary due to many reasons such as variation in prices of raw materials, wage

of labors etc. Now let's examine how a monopolist facing such cases maximizes its profit by taking the following a two- plant monopoly firm as an example. Data regarding cost and revenue is given in a table below.

Out put and sales	Price	Marginal revenue	Marginal cost Plant -1	Marginal cost Plant-2	Multi plant Marginal cost
1	5.0	-	1.92	2.04	1.92
2	4.5	4	2.00	2.14	2.00
3	4.1	3.30	2.08	2.24	2.04
4	3.8	2.9	2.16	2.34	2.08
5	3.55	2.55	2.24	2.44	2.14
6	3.35	2.35	2.32	2.54	2.16
7	3.2	2.30	2.40	2.64	2.24
8	3.08	2.24	2.48	2.74	2.24
9	2.98	2.18	2.56	2.84	2.32
10	2.98	2.08	2.64	2.94	2.34

Given this information, how can the monopolist decide the total production and how much of that output each plant should produce?

The logic used in choosing output levels is very similar to that of the single-plant firm. We can find the answer intuitively in two steps.

Step 1 - What ever the total output, it should be divided between the two plants so that marginal cost is the same in each plant. Other wise, the firm could reduce its cost by reallocating production. For example, if marginal cost at Plant-1 were higher than at Plant-2, the firm could produce the same output at a lower total cost by producing less output at plant -1 and more output at plant-2. Thus, for equilibrium to occur marginal cost at firm-1 (MC_1) must equal marginal cost at firm- 2 (MC_2) i.e. $MC_1 = MC_2$

Step-2 We know that the total output must be such that marginal revenue equals the multi plant marginal cost. Now it is essential to know first how the multi -plant marginal cost is derived from each plant marginal costs. If the firm wants to produce the first unit, it should produce it in plant 1 because, the MC is lower in plant 1 than in plant 2 (i.e. $1.92 < 2.04$). Hence, MC of producing the first unit for the multi –plant monopolist is 1.92. If output is to be two units or if the firm wants to add one more units, the second unit should also be produced in plant 1 because the MC of the second unit in plant 1 is less than MC of producing one unit in plant 2 (i.e. $2.00 < 2.04$). Hence, multi-plant marginal cost for the second unit is \$2. If three units are to be produced, plant 2 will enter into production since the MC of producing one unit in plant 2 (2.04) is less than marginal cost of producing the third unit in plant 1, & 2.08.

Hence, multi-plant MC for the third unit is 2.04, the derivation of multi-plant marginal cost continues in the same manner.

Once, multi-plant marginal cost is derived, the only thing left to obtain equilibrium total output is equating the multi plant MC with the marginal revenue. So in the above table, equilibrium output is 8 units where MC of multi-plant = Marginal revenue (i.e. $2.24 = 2.24$).

Now the remaining issue will be how to allocate the total production between plants 1 and 2. The multi plant monopolist allocates production in a way that each plants MC equals common value of multi plant MC and marginal revenue. The common value of multi plant MC and marginal revenue is 2.24. Thus it follows that the allocation of production is in a way that MC of plant-1 = 2.24 and MC of plant-2 = 2.24

i.e. Plant 1 produces 5 Units (because at 5 units $MC_1 = 2.24$)

Plant 2 produces 3 units (because at 3 units $MC_2 = 2.24$)

In short, the condition of equilibrium in multi- plant monopolist is: $MR = MC$ of multi plant monopolist and to allocate the total out put among each plant, the condition must satisfy:

$MC_1 = MR = MC$ of multi plant monopolist

$MC_2 = MR = MC$ of multi plant monopolist

$MR = MC_1 = MC_2$

Graphically, the above table (problem) can be represented as follows

P, MC, MR

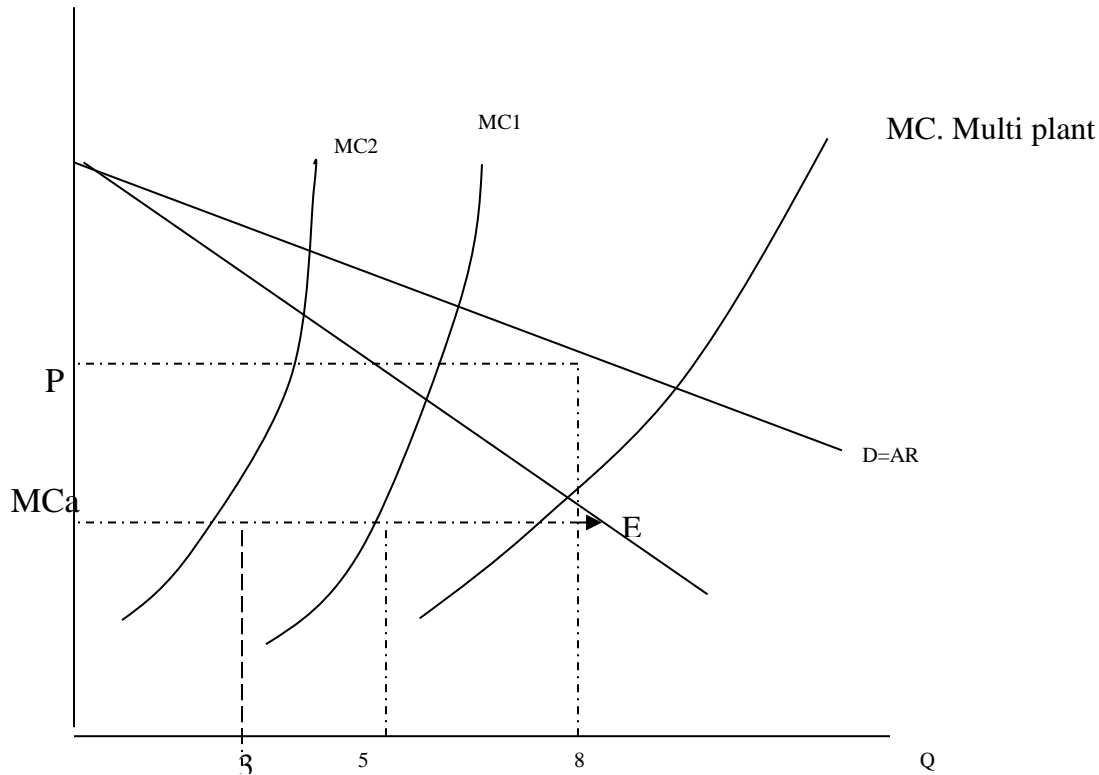


Fig 5.7 Multi-plant monopolist equilibrium. MC_1 and MC_2 denote the MCs of production in plants 1&2 respectively. MC_m denotes the marginal cost of multi-plant firm which is derived from MC_1 and MC_2 . Note that, MC_m is obtained from MC_1 and MC_2 by adding the levels of out put produced in the two plants at equal marginal costs. E.g. when marginal cost is MC_a , the firm produces 3 units in plant 1 and 5 units in plant 2 and the monopolist marginal cost of producing the 8th unit is MC_a . The Multiplant monopolist's equilibrium is defined by point E and the two firms 1and2 produce 5 and 3 units respectively.

Now let us derive this rule algebraically,

Let Q_1 and C_1 be the output and production cost of plant1 and $C_1 = f(Q_1)$

- Q_2 and C_2 be the output and production cost of plant-2 and $C_2 = f(Q_2)$ and

- $Q_T = Q_1 + Q_2$ is the total output of the firm.

All output, whether they are produced in plant 1 or in plant2 will be sold at uniform market price,

Say P then the total profit of the monopolist is

$$\pi = P \cdot Q_T - C_1 - C_2$$

$$\pi = PQ_1 + PQ_2 - C_1 - C_2,$$

To maximize profit $\frac{d\Pi}{dQ_1}$ and $\frac{d\Pi}{dQ_2}$ must be equal to zero

$\pi = TR - C_1 - C_2$, because $PQ_1 + PQ_2 = TR$ the condition of equilibrium is:

$$\frac{d\Pi}{dQ_1} = \frac{dTR}{dQ_1} - \frac{dC_1}{dQ_1} - \frac{dC_2}{dQ_2} = 0 \rightarrow MR_1 - MC_1 = 0$$

$$\frac{d\Pi}{dQ_2} = \frac{dTR}{dQ_2} - \frac{dC_1}{dQ_2} - \frac{dC_2}{dQ_2} = 0 \rightarrow MR_2 - MC_2 = 0$$

Note that:

$$\frac{dC_2}{dQ_1} = 0 \text{ because } C_2 = f(Q_2) \text{ and}$$

$$\frac{dC_1}{dQ_2} = 0 \text{ because } C_1 = f(Q_1)$$

The equilibrium condition is, thus

$$MR_1 = MC_1$$

$$MR_2 = MC_2$$

But $MR_1 = MR_2$ because all outputs whether they are produced in plant1 or plant 2 are sold at the same market price.

$$\text{Let } MR_1 = MR_2 = MR$$

Then the above equilibrium condition can be written as:

$$MR = MC_1 \text{ and}$$

$$MR = MC_2$$

Equivalently, it can be written as $MR = MC_1 = MC_2$

Now let us see one numerical example

Suppose Ethiopian Electric Light and Power Corporation (EELPC) is a multi plant monopolist having two plants, Tekeze plant (plant1) and Fincha plant (Plant2). The operating costs of the two plants are given as follows:

Tekeze Plant: $TC_1 = 10 Q_1^2$ and where Q_1 - Amount of electric power produced in Tekeze

Fincha plant: $TC_2 = 20 Q_2^2$ Q_2 – amount of electric power produced in Fincha

EELPC estimates the demand for electric power by the following function

$P = 700 - 5Q$ where P - is price (total in million birr) per Giga watt and

Q - is the total amount of Giga watt sold and $Q = Q_1 + Q_2$

Note that a Giga watt of electric power, whether it comes from Fincha or Tekeze plant worth equal price

a) What level of output (electric power) should EELPC produce and what price per Kilowatt should it charge to maximize its profit?

b) How much of the total output should be produced in each plant?

c) Suppose that recently the Tekeze plant is suffering from siltation problem (which leads to additional cost of cleaning the dum), but Fincha plant is not. How should EELPC adjust Q_1 , Q_2 and Q_T and P to maximize its profit?

Solution

a) The equilibrium condition is:

$$MR = MC_1$$

$$MR = MC_2$$

$$TR = P \cdot Q$$

$$= (700 - 5Q) Q = 700Q - 5Q^2$$

$$MR = \frac{dTR}{dQ} = 700 - 10Q, \text{ where } Q = Q_1 + Q_2$$

$$\text{Thus, } MR = 700 - 10Q_1 - 10Q_2$$

$$MC_1 = \frac{dTC_1}{dQ_1} = 20Q_1$$

$$MC_2 = \frac{dTC_2}{dQ_2} = 40Q_2$$

Now the equilibrium occurs when:

$$700 - 10Q_1 - 10Q_2 = 20Q_1 \text{ and}$$

$$700 - 10Q_1 - 10Q_2 = 40Q_2$$

Re-arranging the above equations we get the following simultaneous equation.

$$30Q_1 + 10Q_2 = 700$$

$$10Q_1 + 50Q_2 = 700$$

Solving the above equations simultaneously, we get

$$Q_1 = 20 \text{ giga watts}$$

$$Q_2 = 10 \text{ giga watts}$$

$$\begin{aligned} \text{The profit maximizing level of output is, thus, } Q_1 + Q_2 \\ = 30 \text{ giga watt} \end{aligned}$$

To determine the equilibrium price we substitute the total output (30) in the demand function:

$$\begin{aligned} \text{Accordingly, } P &= 700 - B(30) \\ &= 550 \text{ mill birr} \end{aligned}$$

b) The Tekeze plant should produce 20 giga watts and the Fincha plant should produce 10 giga watts

c) To answer this question let us graphically present the problem. In the following figure suppose MC_1 , MC_2 and MC_m denote the initial marginal cost of Tekeze, Fincha and the multi-plant (EELPC), and MC_1' and MC_m' denote the new marginal costs. Note that the silt problem will increase the MC_1 to MC_1' and as a result MC_m will increase to MC_m' .

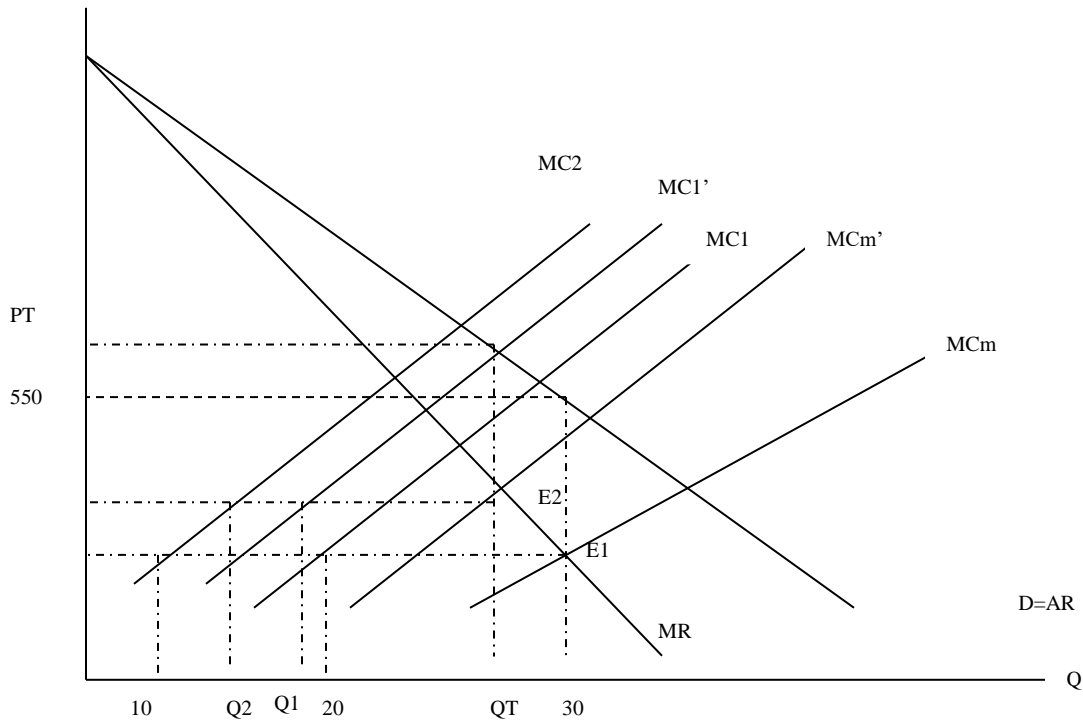


Fig.6.5

Initially, the total output was 30 giga watts. out of which, 20 gw is produced in Tekeze and 10gw in Fincha.

Due to siltation problem MC1 shifts up ward to MC1' and MCm shifts up ward to MCm'. The new equilibrium takes place at E₂ i.e. the total output decreases from 30 to Q_T, output of Tekeze plant decreases from 20 to Q₁.

And that of Fincha increases from 10 to Q₂. Hence the firm will re allocate some of its output from Tekeze to Fincha and will decrease the total output from Tekeze to Fincha

And will decrease the total output. As to equilibrium price is concerned, it increases from 550 to P_T.

6.10 Price Discrimination

Price discrimination refers to the charging of different prices for the same good. But not all price differences are price discrimination. If the costs of offering a certain uniform commodity (service) to different group of customers are different (say due to difference in transport costs), price of the commodity may differ for each group owing to this cost difference. But this can not be considered

as price discrimination. A firm is said to be price discriminating if it is charging different prices for the same commodity with out any justification of cost differences.

By practicing price discrimination, the monopolist can increase its total revenue and profits.

Necessary conditions for price discrimination

For a firm to effectively practice price discrimination the following necessary conditions should be fulfilled.

1-There should be effective separation of markets for different classes of consumers, so that buyers of low price market can not resale the commodity in high price market.

A market is said to be effectively separated if one of the following points is met:

- Geographical variation with high transport cost so that the inter market price margin is unable to cover the transport expense.

E.g. Domestic Vs international markets.

- Exclusive use of the commodity. For some services resale is inherently difficult. For example you can not resale Doctor's services, Entertainment shows.

- Lack of distribution channels

2. The second necessary condition to successfully practice price discrimination is that the price elasticity of demand should be different in each sub market.

For example, a movie theatre knows that college students and old people differ in their willingness to pay for a ticket and can exercise discrimination by charging the college students a higher price.

This condition can be justified by using the markup formula. Suppose the firm has a marginal cost of MC and the price elasticity's of demand for its product into different markets are ed_1 and ed_2

Then the price in each market is

$$P_1 = \frac{MC}{1 + \frac{1}{ed_1}}, \text{ and } P_2 = \frac{MC}{1 + \frac{1}{ed_2}}$$

If $ed_1 = ed_2$, P_1 will be automatically equal to P_2 .

Hence, $ed_1 = ed_2$ for the prices to differ.

3- Lastly, the market should be imperfectly competitive. In other words, the seller of the product should have some monopoly power (it should not be price taker) to practice price discrimination.

6.11 Degrees (types) of price discrimination

The degree of price discrimination refers to the extent to which a seller can divide the market and can take advantage of it in extracting the consumer Surplus. In economics literature, there are three degrees of price discrimination. These are discussed one by one here under.

1-First degree price discrimination (Perfect price discrimination)

This is a price discrimination in which the monopolist attempts to entirely take away the consumers surplus. Ideally, a firm would like to charge each customer the maximum price that the customer is willing to pay for each unit bought. We call this maximum price the consumer's **reservation price** and obviously, the consumers' reservation prices are different due to the differences in their economic status or the value they attach to a commodity. The practice of charging each customer his/her reservation price is called first degree price discrimination. Note that the consumer's willingness to pay reservation price for a given commodity varies with the quantities of the commodity the consumers own. The law of diminishing marginal utility implies that a consumer's willingness to pay for successive units of a commodity declines because the marginal utilities of these successive units decline. Hence, in the first degree price discrimination prices differ across customers, and a given customer may pay more for the initial units than for others (successive units).

First degree price discrimination is the limiting case of price discrimination, the monopolist, in this case, individually negotiate with each buyer and sell each unit of the output at the corresponding price given on the demand curve of the consumer, then receiving the entire of consumer's surplus.

For example, a doctor who knows his patients' paying capacity charges high price for the richest patients' and low price for the poor patients for identical services.

This is practiced to increase revenue. If the doctor fixes the price at the richest patients' level, no poor will afford to pay and the doctor will not get revenue from the poor. On the other hand, the doctor would not fix the price at the poorest patients' level for all patients because he knows that the rich can pay more and he will exploit the rich. Lawyers also practice the same discrimination for identical legal service.

Perfect price discrimination is efficient as it maximizes the total welfare, where welfare is defined as the sum of consumer surplus and producer surplus. That is, there is no welfare loss associated with first degree price discrimination equilibrium. The problem with perfect price discrimination is that it hurts consumers because the monopolist will take the entire of the consumer surplus. The other problem with perfect discrimination is that it involves high transaction costs; it is too difficult and costly to gather information about each customer's price sensitively.

2-Second degree price discrimination (block pricing)

Many firms are unable to determine which customers have the highest reservation prices. Such firms may know, however, that most customers are willing to pay more for the first unit than for successive units. This is due to the fact the typical customer's demand curve is downward sloping. Such a firm can price discriminate by letting the price each customer pays vary with the number of units the customer buys. The act of charging different prices for different quantities of purchases is called second degree price discrimination or some times called quantity discrimination. In second degree price discrimination the price varies only with quantity: all customers pay the same price for a given quantity.

In second degree price discrimination, the monopolist attempts to take the major part of the consumer surplus instead of the entire of it.

Block pricing can feasibly be implemented where:

- the number of consumers is large and price rationing can be effective e.g. electricity and telephone services.
- the demand curves of all customers are identical and
- a single rate is applicable for a large number of buyers.

Graphically, block pricing can be explained as follows:

A monopolist that practices second degree price discrimination charges the price OP_1 , for the first OQ units, OP_2 for the next Q_1 Q_2 units and OP_3 for Q_2 Q_3 units. By doing so, the monopolist will increase its total revenue by extracting the major part the consumer surplus.

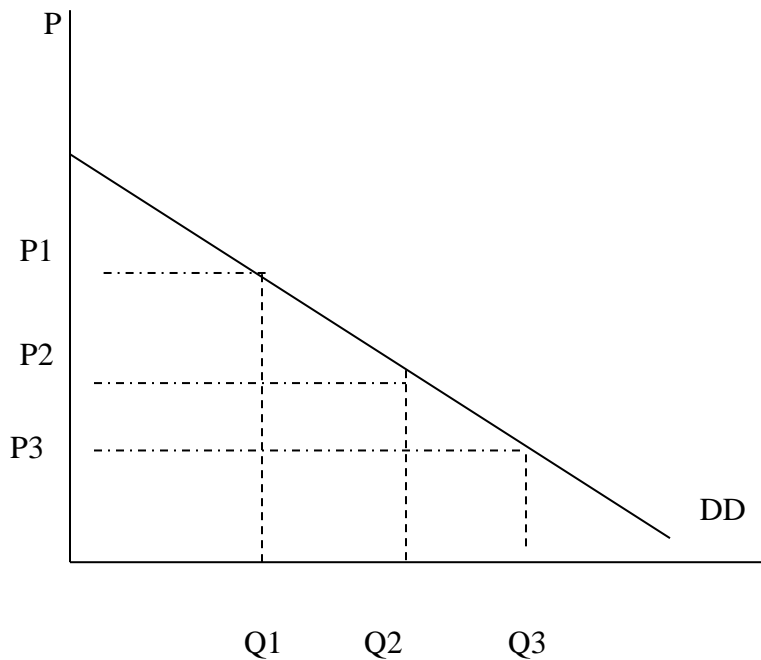


Fig.6.7 Second price degree price discrimination. The monopolist receives a price OP_1 , for each unit sold to a given customer for the first OQ_1 units, OP_2 for the next $Q_1 Q_2$ units and OP_3 for the next $Q_2 Q_3$ units. By so doing, the monopolist will receive total revenue of $OP_1 A B C D E$. If the monopolist charges a uniform price of OP_3 , its total revenue will only be $OP_3 E Q_3$. Hence, block pricing will enable him receive large total revenue than uniform pricing.

Note that not all quantity discounts are a form of price discrimination. Some times selling in large quantities may reduce the unit costs of sales and as a result a firm may charge a relatively lower per unit price for large sales than small sales. Such an action can not be regarded as price discrimination.

3-Third degree price discrimination (multi-market price discrimination)

Typically, a firm does not know the reservation price for each of its customers. But, the firm may know which groups of customers are likely to have higher reservation prices than others. In such a situation the firm may divide potential customers in to two or more groups and set a different price for each group. Such an action of charging different prices in different markets is called ***third degree price discrimination***. All units of the good sold to customer with in a group (in one market) are sold at a single price, but prices will differ among the different groups or markets.

For simplicity, let us assume that there are only two markets. To maximize profits, the monopolist must produce the level of output (defined by $MC=MR$) and sell that output in the two markets in such a way that the marginal revenue of the last unit sold in each market is the same. This will require the monopolist to sell the commodity at higher price in the market with the less elastic demand.

For example, suppose that a monopolist has 100 units of a commodity to be sold in one or both of two sub markets. How should the monopolist allocate the 100 units between the two markets to maximize its profit? Suppose, initially, that the monopolist simply sold 50 units in each market and also assume that the marginal revenue of the last unit sold in market 1 is 5 and the marginal revenue of the last unit sold in market 2 is 3.

In this case, the monopolist can increase its total revenue by decreasing the number of units sold in market 2 and increasing the number of units sold in market 1. Hence, if one less unit is sold in market 2, total revenue falls by \$3. But by selling this unit in market 1 total revenue increases by \$5. So, by reallocating its sales from market 2 to market 1 the monopolist can increase its total revenue by \$2 (\$5-3\$). Obviously, reallocation of sales will increase the firm's total revenue until the marginal revenue of the last unit sold in each market gets equal.

Thus we can conclude that to maximize the total revenue received from the sale of a given quantity of a commodity, the monopolist should allocate the total quantity in each sub market in such a way that the marginal revenue of the last unit sold in each sub market is the same. Symbolically, the equilibrium condition for a third degree price discriminating monopolist is: $MC=MR_1=MR_2$.

Now let us derive this equilibrium condition algebraically.

Assume that the firm sells its product in two markets and the demand functions of the segmented markets are given as:

$$P_1 = f_1(Q_1) \text{ and } P_2 = f_2(Q_2)$$

And suppose that cost function of the firm is

$$TC = f(Q), \text{ where } Q=Q_1+Q_2$$

Q1 is the amount of the commodity sold in market 1

Q2 is the amount sold in market 2.

The total profits of the monopolist (Π) are equal to the total revenue it receives from selling the commodity in the two markets (i.e., TR1 + TR2) minus the total cost of producing the total output (TC), that is,

$$\Pi = TR_1 + TR_2 - TC \text{----- (1)}$$

But $TR_1 = P_1 \cdot Q_1$ and $TR_2 = P_2 \cdot Q_2$

For Π to be maximized, the first derivatives of the Π function with respect to Q_1 and Q_2 should be equal to zero. That is,

$$\frac{\partial \Pi}{\partial Q_1} = \frac{\partial \Pi}{\partial Q_2} = 0 \text{----- (2)}$$

$$\text{But, } \frac{\partial \Pi}{\partial Q_1} = 0 \rightarrow \frac{\partial TR_1}{\partial Q_1} + \frac{\partial TR_2}{\partial Q_2} - \frac{\partial TC}{\partial Q_1} = 0 \text{----- (3)}$$

$$MR_1 - MC_1 = 0 \text{ or } MR_1 = MC_1$$

$$\text{and } \frac{\partial \Pi}{\partial Q_2} = 0 \rightarrow \frac{\partial TR_1}{\partial Q_2} + \frac{\partial TR_2}{\partial Q_2} - \frac{\partial TC}{\partial Q_2} = 0 \text{----- (4)}$$

$$\rightarrow MR_2 - MC_2 = 0 \text{ or } MR_2 = MC_2$$

Note that:

$$\frac{\partial TR_2}{\partial Q_1} = 0 \text{ in equation (3) because } TR_2 \text{ is only a function of } Q_2.$$

$$\text{and } \frac{\partial TR_1}{\partial Q_2} = 0 \text{ in equation (4) because } TR_1 \text{ is only a function of } Q_1.$$

Now the equilibrium condition equation (2) & (4) can be summarized as.

$$MR_1 = MC_1$$

$$MR_2 = MC_2 \text{----- (5)}$$

But, note that the monopolist produce its commodity in one plant, and the fact that the commodity is sold in market 1 or market 2 has no impact on the marginal cost of the commodity (i.e., $MC_1=MC_2$)

Suppose—that $MC_1 = MC_2 = MC$

The equilibrium condition equation (5) can be reduced

in to $MR_1 = MC$

$MR_2 = MC$

Or $MR_1 = MR_2 = MC$ ----- (6)

Thus, the equilibrium condition of a third degree price discriminating monopolist is $MR = MR_2 = MC$.

Numerical example

Suppose a monopolist sells its commodity in U.S.A and Ethiopian markets. The demand function for the monopolist’s product in U.S.A market is given as $P_u = 100 - Q_u$ and in Ethiopian market the demand function is $P_e = 80 - 2Q_e$, where P_u and Q_u denote price and quantity demanded in U.S A and P_e and Q_e denote price and quantity demanded in Ethiopia.

The monopolist has 55 units of the commodity.

- a) How many units should be sold in Ethiopia and U.S.A?
- b) In which country should the firm charge larger price? Why?

Solutions

a) the monopolist allocates its product in such a way that $MR_u = MR_e$

$$MR_u = \frac{\partial TR_u}{\partial Q_u}, \text{ and, } MR_e = \frac{\partial TR_e}{\partial Q_e}$$

But, $TR_u = P_u \cdot Q_u$

$$= (100 - Q_u) Q_u = 100 Q_u - Q_u^2 \text{ and}$$

$TR_e = P_e \cdot Q_e$

$$= (80 - 2Q_e) Q_e = 80Q_e - 2Q_e^2$$

$$MR_u = \frac{d(100Q_u - Q_u^2)}{dQ_u} = 100 - 2Q_u \quad \text{and}$$

$$MR_e = \frac{d(100Q_e - Q_e^2)}{dQ_e} = 80 - 4Q_e$$

The Firm maximizes its revenue when the condition $MR_u = MR_e$ is fulfilled. That is, it maximizes its revenue when:

$$100 - 2Q_u = 80 - 4Q_e \quad \text{or}$$

$$2Q_u - 4Q_e = 20 \quad \text{----- (1)}$$

$$\text{More over, we know that } Q_u + Q_e = 55 \quad \text{----- (2)}$$

Solving equations (1) and (2) simultaneously, we obtain $Q_u = 40$ units and $Q_e = 15$ units.

Thus, the monopolist should sell 40 units in U.S.A and 15 units in Ethiopia to maximize its TR, the demand functions in each country.

$$P_u = 100 - Q_u$$

Substituting 15 for Q_e , we get $P_e = \$ 50$

Hence, the firm should charge higher price in U.S.A. the reason is that the price elasticity of demand for the firm's commodity is lower in U.S.A than Ethiopia.

That is,

Price elasticity of demand in U.S.A (E_u) is

$$E_u = \frac{dQ_u}{dP_u} \cdot \frac{P_u}{Q_u}$$

$$= -1 \cdot \frac{60}{40} = -3/2 = 1.5$$

and price elasticity of demand in Ethiopia (i.e.) is

$$E_e = \frac{dQ_e}{dP_e} \cdot \frac{P_e}{Q_e}$$

$$= -1 \cdot \frac{50}{2} = -50/2 = 1.67$$

$E_u < E_e$ which implies that a one percent increase in price of the commodity reduces the amount of sales by a lower percent (1.5%) in U.S.A than in Ethiopia (1.67%). In other words, U.S.A. citizens are less sensitive to a price change than Ethiopians so that the firm can charge higher price in U.S.A.

The fact that the firm should charge a higher price in the market having lower price elasticity of demand can be shown algebraically as follows:

You know that the marginal revenues in two markets (market 1 and 2) having price elasticities of demand, E_{d1} and E_{d2} respectively are given as:

$$MR_1 = P_1 \left(1 + \frac{1}{E_{d1}} \right) \text{ Where } MR_1 \text{ and } P_1 \text{ are marginal revenue and price in market 1}$$

$$MR_2 = P_2 \left(1 + \frac{1}{E_{d2}} \right) \text{ Where } MR_2 \text{ and } P_2 \text{ are marginal revenue and price in market 2}$$

For optimal allocation of the commodity between the two markets, $MR_1 = MR_2$ i.e.,

$$P_1 \left(1 + \frac{1}{E_{d1}} \right) = P_2 \left(1 + \frac{1}{E_{d2}} \right)$$

or

$$\frac{P_1}{P_2} = \frac{1 + \frac{1}{E_{d2}}}{1 + \frac{1}{E_{d1}}}$$

If $|E_{d2}| > |E_{d1}|$, the ratio $\frac{1 + \frac{1}{E_{d2}}}{1 + \frac{1}{E_{d1}}}$ will be greater than one (i.e. $1 + \frac{1}{E_{d2}} > 1 + \frac{1}{E_{d1}}$), which implies that

P_1/P_2 will be greater than one (i.e., $P_1/P_2 > 1$)

Therefore, if $|E_{d2}| > |E_{d1}|$, $P_1 > P_2$

Hence, the larger the price elasticity of demand, the lower the price to be charged.

Numerical example

Suppose Ethiopian Air lines (EAL) flies only one route: from Addis Ababa to Dubai. EAL knows that two different types of people fly to Dubai. Type **A** consists of rich merchants flying to Dubai for business purposes with demand for flight of

$Q_A = 260 - 0.4P_A$. Type **B** consists of poor ladies flying to Dubai in search of jobs (such as house maid) whose total demand is $Q_B = 240 - 0.6P_B$.

Assume that EAL has a running cost of \$30,000 plus \$100 per passenger and it has decided to charge different prices for the two groups of passengers.

- a. How many tickets should EAL sell to each group?
- b. How much price should EAL charge each group?
- c. Suppose now that EAL is prohibited by the Ethiopian government to exercise such discrimination. How many tickets should the EAL sell to maximize its profit and at what price?

Solution

Given

$$TC = 30,000 + 100Q$$

Where $Q = Q_A + Q_B$

$Q_A = 260 - 0.4P_A$ Merchants demand function:

$$P_A = 650 - 2.5Q_A$$

$Q_B = 240 - 0.6 P_B$ Ladies demand function

$$P_B = 400 - \frac{5}{3}Q_B$$

a) The equilibrium condition is that

$$MC = MRA = MRB$$

$$\text{But } MC = \frac{dTC}{dQ} = 100$$

$$MRA = \frac{dTRA}{dQ_A}, \text{ and } TRA = Q_A.P_A = 650Q_A - 2.5 Q_A^2$$

$$\text{Thus, } MR_A = 650 - 5Q_A$$

$$\text{Like wise } MR_B = 400 - \frac{10}{3}QB$$

The equilibrium condition is thus presented as:

$$100 = 650 - 5QA$$

$$100 = 400 - \frac{10}{3}QB$$

Solving the above equations simultaneously, we get

$$QA = 110 \text{ and}$$

$$QB = 90$$

Therefore, EAL should sell 110 tickets of **A** type and 90 tickets of **B** type passengers.

b) Substituting the above quantities in their respective demand functions, we get

$$\begin{aligned} PA &= 650 - 2.5 QA \\ &= 650 - 2.5 (110) \\ &= \$ 375 \end{aligned}$$

$$\begin{aligned} PB &= 400 - \frac{5}{3}QB \\ &= 400 - \frac{5}{3}(90) = \$ 250 \end{aligned}$$

Hence, the EAL should charge \$ 375 for the **A** type passengers and \$ 250 for the **B** type passengers.

c) If EAL decides to charge a uniform price, the equilibrium price will be obtained first by deriving the market demand function and then by using the usual method ($MC = MR$)

$$\text{Market demand (Q)} = QA + QB$$

$$Q = 260 - 0.4 PA + 240 - 0.6 PB$$

$$\text{Since prices are uniform } PB = PA = P$$

Thus the market demand equation becomes $= 500 - P$ or

$$P = 500 - Q$$

$$TR = P.Q = 500 Q - Q^2$$

$$MR = 500 - 2Q$$

Given $MC = 100$, Equilibrium occurs when $MC = MR$, i.e.

$$100 = 500 - 2Q$$

$$Q = 200, \text{ and } P = 500 - Q = \$300$$

That is, EAL should sell 200 tickets at a price of \$ 300 each to maximize its profit.

6.12 Social costs of monopoly: the dead weight loss

In a competitive market, price equals marginal cost of production. Monopoly power, on the other hand, implies that price exceeds marginal cost. Because monopoly power results in higher prices and lower quantities produced, we would expect it to make consumers worse off and the firm better off. But suppose we value the welfare of consumers the same as that of producers. In aggregate, does monopoly power make consumers and producers better off or worse off?

To answer this question, suppose an industry operating under perfectly competitive situation is suddenly monopolized. We can answer the questions by comparing the consumer and producer surplus that results when a competitive industry produces a good with the surplus that results when a monopolist supplies the entire market.

Referring to the following figure, suppose DD represents the market demand curve, MR represents the corresponding marginal revenue.

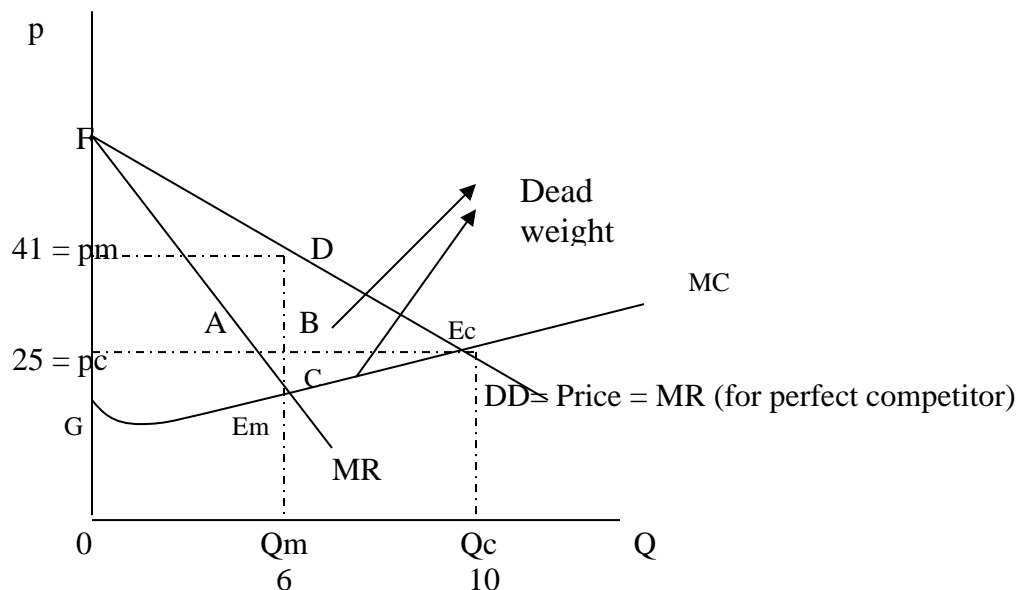


Fig.6.8

Here, we use consumers' and producers' surplus as a measure of welfare of each. Consumer surplus is the area between the demand curve and equilibrium price and producer surplus is the area between the equilibrium price and marginal cost curve.

- A perfect competitor's equilibrium occurs when MC equal price or marginal revenue at E_c and the equilibrium price and quantity are P_c & Q_c respectively. Here the consumer's

surplus is the area above the dropped line $P_c E_c$ and below the demand curve i.e. area of $\Delta P_c F E_c$. On the other hand the producer surplus is the area below the dropped line $P_c E_c$ and above the MC curve.

- A monopolist equilibrium occurs when $MC = MR$ i.e. at E_m and the equilibrium price and quantity become P_m and Q_m respectively. Hence, in monopoly lower quantity is sold at higher price. The new consumers' welfare is the area above the dropped line $P_m D$ and below the demand curve (i.e. area of $\Delta P_m F D$) where as the producers surplus becomes the area below the dropped line $P_m D$ and above MC curve to the left of Q_m (i.e. the area $G P_m D E_m$)
- Thus monopoly power reduces the consumers' surplus by the amount which equals area $A+B$. But increases the producers' surplus by the area $A-C$. The net welfare effect (loss) is obtained by deducting the welfare loss of consumers from the welfare gain of producers i.e.,

$$\begin{aligned} \text{Net welfare} &= \text{Welfare gain by producers} - \text{Welfare loss by consumers} \\ &= A-C - (A+B) \\ &= A-C - A-B \\ &= -C -B \text{ or } -(C+B) \end{aligned}$$

Thus monopoly results in a welfare loss which is given by the area $(C+B)$

This area is called dead weight loss. It is gained neither by producers nor by consumers.

The other disadvantage (Social cost) of monopoly is that it discourages innovations. Monopolist may feel secure and have no incentive to innovate new product (technology) since there are no competitors.

Summary

Monopoly is a market structure in which a single seller sells a product for which there is no close substitute. Ownership of strategic inputs, economies of scale in production, patents and copy right, and exclusive knowledge of production are among the factors that cause emergence of monopoly. A monopolist firm is a price setter. It can increase the price of its commodity by decreasing the quantity of supply and vice versa. That is he firm can sell more units in the market by reducing the price of the commodity. Thus, a monopolist faces a negatively sloped demand curve.

In monopoly, since the firm is a price setter, profit maximization involves determining the price output combination that yields the monopolist the maximum profit. Unlike the case of perfect competitive firm, a monopolist has no unique supply curve. The short run profit maximization condition of the monopolist is identical with that of a perfect competitive firm.

In the long run a monopolist maximizes its profit by equating marginal revenue (MR) with long run Marginal cost (MC). The optimal plant size is the one whose short run average cost curve is tangent to the long run AC curve at equilibrium output. Unlike perfectly competitive firm, a monopolist can get positive profits in the long run due to entry barriers.

Sometimes, monopoly firms undertake production in two or more different plants. The operating costs of this plant may also differ. In a situation the profit maximizing condition is that the marginal costs of the different plants should be equal to the multi-plant marginal cost.

To increase profits, monopolists discriminate prices. Price discrimination requires different conditions to be fulfilled. For a monopolist to effectively exercise price discrimination, there must be effective separation (segmentation) of the different markets, price elasticities of demand in different markets should not be equal, and the monopolist should possess a strong monopoly power. The equilibrium condition of a price discriminating monopolist is that the marginal revenues of each market should be equal to marginal costs of production.

Monopoly market results in inefficiency (dead weight losses).

1. Introduction to Macroeconomics

Points to be remembered:

Economy: A system of providing living to people.

Microeconomics: Study of the behavior of individual, small, isolated and disaggregated units.

Macroeconomics: Study of groups and broad aggregates of the economy.

Firm: An individual producing unit.

Industry: A group of firms producing identical or closely related goods.

The term microeconomics and macroeconomics were first given by **Ragner Frisch** in 1933.

Prof. J.M. Keynes is known as father of modern macroeconomics.

Macroeconomics became popular after *great depression of 1929- 33*.

Prof. J.M. Keynes wrote the book *General Theory of Employment, Interest and Money* in 1936.

Meaning of Macroeconomics:

The term macro has been derived from Greek word '*makros*' which means *large*. It is the study of aggregates or groups or the entire economy such as gross domestic product, total employment, aggregate demand, aggregate supply, total savings, general price level, etc.

Scope of Macroeconomics:

Macroeconomics has a wider scope than microeconomics. The study of macroeconomics extends to the following areas:

- Theory of National Income;
- Theory of Employment;
- Theory of Money;
- Theory of General Price Level;
- Theory of International Trade; and
- Theory of Economic Growth.

Goals and Importance of Macroeconomics:

- To Achieve Higher Level of Gross Domestic Product;
- To Achieve Higher Level of Employment;
- Stability of Prices;
- Formulation of Economic Policies; and
- Achievement of Economic Development.

Tools of Macroeconomics:

Fiscal Policy: relates to the management of government revenue, expenditure and debt to achieve favorable effects and avoid unfavorable effects on income, output and employment.

Monetary Policy: relates to the management of money supply and credit to step up business activities, promote economic growth, stabilize the price level, achievement of full employment and equilibrium in balance of payments.

Income Policy: through this policy direct control is exercised over prices and wages.

Major Issues and Concerns of Macroeconomics:

- Employment and Unemployment;
- Determination of National Income (or GNP);
- General Price Level and Inflation;
- Business Cycle;
- Stagflation;
- Economic Growth;
- Balance of Payments and Exchange Rate.

Post- Keynesian Developments in Macroeconomics:

- Monetarism;
- Supply- side Economics;
- Rational Expectations Theory.

Monetarism:

Monetarists led by American economist Milton Friedman criticised Keynes' macroeconomics and developed a new idea that monetary policy is the prime engine in causing fluctuations in economic activity by bringing about change in aggregate demand. He stressed that even the

Great Depression of 1930s was primarily caused by tight monetary policy adopted at that time.

There are two differences (issues) between the monetarists and Keynesians. First issue or difference relates to the relationship between money supply and inflation. The second relates to the role of government in the economy.

Monetarists believe that inflation is always and everywhere a monetary phenomenon. According to them, inflation is caused by rapid expansion of money supply in the economy and suggest a constant growth rate of money supply to control inflation.

As Keynesians emphasize that active role should be played by the government to control business cycles and achieve economic stability. Like classical economists, monetarists also believe that free-market economy is inherently stable and if the economy departs from the state of full-employment, full-employment equilibrium is restored through automatic adjustments in it.

Supply-Side Economics:

The failure of Keynesians to deal with stagflation (high inflation with high rate of unemployment) led supply-side economics. Supply-side economists pointed out that it was supply-shocks, delivered among others by reduction in oil supplies and increase in oil prices that caused the problem of stagflation. As a result of contraction in supply due to the adverse supply shocks, given the aggregate demand curve, price level and inflation rate could rise on the one hand and aggregate output could fall giving rise to more unemployment on the other.

Supply-side economists suggest that for the expansion in aggregate supply and thereby increase in employment opportunities, incentives to work, save and invest more were required to be promoted.

New Classical Macroeconomics or Rational Expectation Theory:

New classical macroeconomics also opposed to Keynesian macroeconomic theory and policy which focused on aggregate demand for goods and services. According to the new classical macroeconomic theory, consumers, workers and producers behave rationally to promote their interests and welfare. On the basis of their rational expectations, based on all the available information, they make quick adjustments in their behaviour. Therefore, according to the new classical macroeconomists or the proponent of rational expectation theory, involuntary unemployment cannot prevail. A significant difference between the Keynesian theory and rational expectation theory may be noted here. In the Keynesian theory deficit in government budget leads to increase in aggregate demand and will therefore promote private investment. On the other hand, according to rational expectations theory, budget deficit will cause rate of interest to rise which will discourage private investment.

Like Friedman and other monetarists, supporters of rational expectations theory are opposed to the active role by the government.

2. National Income Accounting

2.1. Concepts and definitions

National income accounting is a bookkeeping system that a government uses to measure the level of the country's economic activity in a given time period. It is also a set of rules and definitions for measuring economic activity in the aggregate economy (the economy as a whole). Measures of national income and output are used by macroeconomists to estimate the total market value of final goods and services produced in an economy in a given year. The most important measures are GDP and GNP. The measurement of GDP (and GNP) and its components are known as national income accounting.

Gross Domestic Product (GDP)

GDP is the total market value of all final goods and services produced within a territory of a country in a given year irrespective of its ownership. Thus, GDP counts income according to where it is earned rather than who owns the factors of production.

Four things to note about this definition of GDP:

➤ **Market value**

- ✓ GDP is a market value—goods and services are valued at their market prices.
- ✓ To add apples, oranges and computers, we add the market values so we have a total value of output in dollars/ birr. Thus by converting everything to its birr value we have a uniform measure.

➤ **Final goods and services**

- ✓ A **final good** (or service), is an item bought by its final user during a specified time period.
- ✓ Excluding intermediate goods and services avoids double counting. So a car will be counted in the GDP but the steel used to make the car is not counted separately. Why? Because the value of the car already reflects the value of the steel, rubber, plastic, etc. that goes into it.

➤ **Produced within a country**

- ✓ GDP measures production within a nation's borders – domestic production – no matter who owns the resources. The Ethiopian GDP counts only those goods and services

produced in the physical borders of Ethiopia. So Toyota made in Addis Ababa are counted in Ethiopian GDP, *Toyota is a* Japanese company. However, products made by Ethiopians or Ethiopian Companies in other parts of the world do not count in Ethiopian GDP. Thus, Where the good or service is produced is important.

➤ **In a given time period**

- ✓ GDP measures production during a specific time period, normally a year or a quarter of a year.
- ✓ It will be useful here to mention what happens to the value of an item that is produced in a given year, but does not sell in the year in which it is produced. At the end of the year it is included in inventories for that year and is thereby included in that year's GDP.

Gross National Product (GNP)

GNP is the total market value of all final goods and services produced by the factors of production of a country in a given year irrespective of its territorial location.

- ✓ It is important to differentiate Gross Domestic Product (GDP) from Gross National Product (GNP). GDP includes only goods and services produced within the geographic boundaries of a country, regardless of the producer's nationality. GNP doesn't include goods and services produced by foreign producers, but do include goods and services produced by Ethiopian firms operating including in foreign countries. The difference between GDP and GNP corresponds to *net factor income from abroad*. When GDP exceeds GNP residents of a given country are earning less abroad than foreigners are earning in that country. In Ethiopia, GDP has exceeded GNP since 1981 (based on the data available in World Development Indicators CD-ROM, 2000) but the gap is well below 1% (0.75% to be exact) during 1981 – 1998.
- ✓ **In simple words, GDP is territorial while GNP is national.**
- ✓ Net factor income from abroad= (factor payments from abroad) – (factor payments to abroad).
$$\text{GNP} = \text{GDP} + \text{Net factor income from abroad}$$

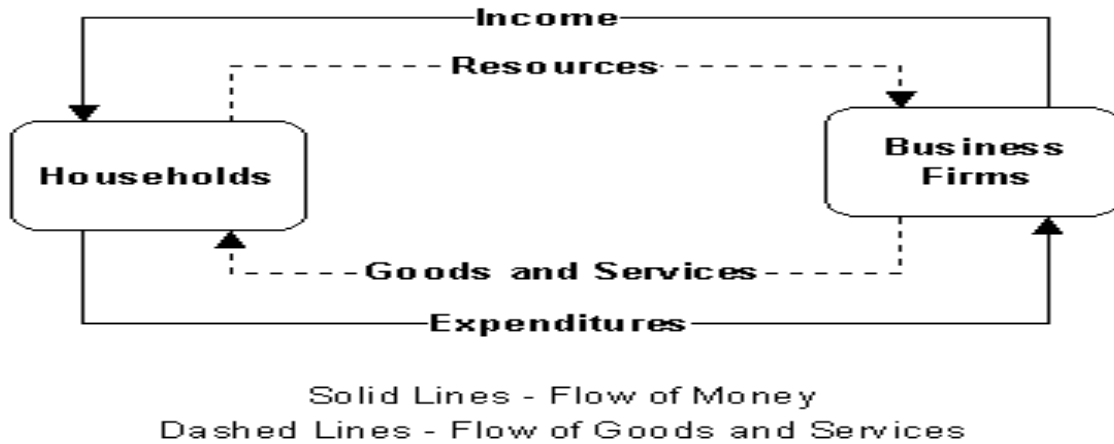
Importance of National Income Estimates:

- Indicator of Economic Progress;
- Measure of Economic Growth;

- Comparison with other Countries;
- Significance in Business Policy Making;
- Significance for Trade Unions;
- Signification for Economic Analysis, etc.

Circular Flow of Income:

Flow of Income in a Two- sector Economy:



Some Concepts of National Income and Related Aggregates:

The important concepts related to national income are:

- Gross Domestic Product (GDP);
- Gross Nation Product (GNP);
- Net Domestic Product (NDP);
- Net Nation Product (NNP);
- Personal Income;
- Personal Disposable Income;

Gross Domestic Product (GDP): is the market value of all final goods and services produced within domestic territory of the country during a year.

Features of GDP:

- It includes only final goods and services produced in the domestic territory of a country;
- It includes consumption of fixed capital (depreciation);
- It is estimated at the prevailing prices.

Gross Nation Product (GNP): is Gross Domestic Product plus net factor income from abroad. **GNP** is the money value of all final goods and services produced in the domestic territory of a country during a year plus Net factor income from abroad i.e.

$$\text{GNP} = \text{GDP} + \text{NFIA}$$

Where,

GNP = Gross National Product

GDP= Gross Domestic Product

NFIA= Net factor income from abroad

Net Factor Income from Abroad (NFIA): is the difference between the income received from abroad for rendering factor services by the normal residents of the country to the rest of the world and income paid for the factor services rendered by nonresidents in the domestic territory of a country.

Net Domestic Product (NDP): is the net market value of all the final goods and services produced in domestic territory of a country during a year. Net market value of the goods is equal to the market value of goods minus depreciation.

$$\text{NDP} = \text{GDP} - \text{D (or CCA)}$$

Where,

NDP= Net Domestic Product

GDP= Gross Domestic Product

D = Depreciation

CCA= Capital Consumption Allowances

Net National Product (NNP): is the net market value of all the final goods and services produced by the normal residents of a country during a year.

$$\text{NNP} = \text{GNP} - \text{D}$$

Where,

NNP = Net National Product

GNP = Gross National Product

D = Depreciation

Personal Income: Personal income is sum of all kinds of income received by the individuals from all sources.

Personal Disposable Income: It refers to that part of personal income which is actually available to households for consumption and saving. In other words, the income that individual households have to spend or save after payment of personal taxes is called disposable income.

Personal Disposable Income = Personal Income – personal tax

2.2. Approaches to national income accounting

National income worked as important indicators to measure the level of economic growth and welfare of the country. National income data facilitate the formulation of plans and fixing the targets of development. National incomes also narrate the picture of saving, investment, distribution of national income, consumption and employment level of the country. For this overall these aspects we need national income and associated estimates.

There are three approaches of the measurement of the national income. They are as follows:

1. Expenditure approach

2. Income approach or Factor income in production process

3. Value Added approach or Product approach

2.2.1 The Expenditure Approach

The **expenditure approach** determines aggregate demand, or Gross Domestic Expenditure, by summing consumption, investment, government expenditure and net exports:

i. Consumption (C): The value of all goods and services bought by households. It includes:

- **durable goods** that last a long time e.g.: cars, home appliances
- **nondurable goods** that last a short time e.g.: food, clothing
- **Services** work done for consumers e.g.: dry cleaning, air travel

ii. Investment (I): in the context of macroeconomics, means **capital** investment. Capital means goods which are used to produce other goods or services. This category includes:

- **Business Fixed Investment (Nonresidential):** Spending on plant and equipment that firm will use to produce other goods & services.

- **Residential Fixed Investment:** spending on housing units
- **Inventory Investment:** change in the value of all firms' inventories

What is important to remember is that **capital investment** is totally different from **financial investment**. Financial investment means buying a financial asset. But when you buy a financial asset, like a share, no good is produced. Capital investment means the purchase of actual goods or services from producers of those goods. When Ethiopian Airlines buys a 777 aircraft from Boeing that is capital investment -- a purchase of a good that Boeing had to produce.

- **Depreciation:** the part of capital stock used up in current year production process

iii. Government Spending (G): Federal, state, and local government purchases of goods and services are counted here. This does not count transfer payments, such as welfare, since this just transfers income and does not represent production of goods and services.

iv. Net Exports (X - M). Net exports are exports minus imports.

Putting all of the expenditures together we have the identity

$GDP = C + I + G + X - M$ Where, C = Personal consumption expenditures

I = Gross private domestic investment

G = Government expenditures

X = Gross exports of goods and services

M = Gross imports of goods and services

GDP = Gross Domestic Product

$GNP = GDP + \text{Net Factor Income from abroad (NFIA)}$

Precautions of Expenditure Approach

- A. Expenditure on second hand goods should not included
- B. Expenditure on bonds and shares should not be included
- C. All government expenditure on transfer payments, such as Unemployment benefits, old age Pensions, scholarships etc. should not be included.
- D. Expenditure on all intermediate goods and services should not be included.

Formula: National Income Accounting:

D = Depreciation

IBT = Indirect business taxes

NDP = Net Domestic Product

NI = National Income

PI = Personal Income

DI = Disposable income

Note: (X - M) is often written as "NX," which stands for "Net Exports"

$$GDP = C + I + G + (X - M)$$

$$GNP = C + I + G + (X - M) + NFIA$$

$$GNI = C + I + G + (X - M) + NFI - IBT$$

$$NI = C + I + G + (X - M) + NFI - IBT - D$$

$$NDP = GDP - \text{depreciation}$$

$$NI = NDP - IBT + \text{net foreign factor income}$$

Example

Personal Consumption	3,657	Rental Income	17
Depreciation	400	Corporate Profits	341
Wages	3,254	Exports	673
Indirect Business Taxes	500	Net Foreign Factor Income	-20
Interest	530	Proprietor's Income	403
Domestic Investment	741	Imports	704
Government Expenditures	1,098		

Let me demonstrate calculating the GDP using the Expenditures Approach with the above hypothetical data: $Y = C + I + G + NX$

$$Y = 3,657 + 741 + 1,098 + (673 - 704)$$

$$Y = 3,657 + 741 + 1,098 - 31$$

$$Y = 5,465$$

2.2.2. Income approach

The second approach to measuring GDP is to look at income instead of expenditures. If someone bought it, then someone is being paid to make it. According to the income approach, national income is estimated by adding incomes earned by all factors of production for their factor services during a year. The factor services include land, labour, capital and enterprises. These factor services received the income against their services.

The factors income distributed as follow:

Factors of Production	Factors Income
1. Land	1. Rent
2. Labour	2. Wages or Salaries
3. Capital	3. Interest
4. Enterprises (entrepreneur)	4. Profit

- ✓ **Employee compensation** consists of payments for labor such as salaries and wages.
- ✓ **Rents** are payments for use of land and buildings.
- ✓ **Interest** includes payments

for loans by households to firms.

- ✓ **Profits** are payments to the owners of firms, such as Corporate profits
- ✓ **Proprietors income** (the profits of partnerships and solely owned businesses, like a family restaurant)
- ✓ **Depreciation** (the amount of capital that has worn out during the year)

Precautions while estimating National Income through Income Method

- A. All government expenditure on transfer payments, such as Unemployment benefits, old age Pensions, scholarships etc. should not be included. This is because they are received without rendering any productive services.
- B. Incomes earned through illegal activities like smuggling, tax evasion etc. should not be included.
- C. Windfall gains like lottery income should not be included.
- D. Money received for sale and purchase of second hand goods and bonds and share should not be included in factor income.
- E. Imputed rent of owner occupied houses should be properly counted and included.

Example

I will demonstrate calculating the GDP using the Income Approach with the hypothetical data given earlier:

$$\text{National Income} = \text{Compensation to Employees (Wages)} + \text{Rents} + \text{Interest} + \text{Proprietor's Income} + \text{Corporate Profits}$$

$$\text{National Income} = 3,254 + 17 + 530 + 403 + 341 = 4,545$$

$$Y (\text{GDP}) = \text{National Income} + \text{Indirect Business Taxes} + \text{Depreciation} - \text{NFFI}$$

$$Y = 4,545 + 500 + 400 - (-20) = 5,465$$

This is the same value for the GDP received when calculating it using the Expenditures Approach.

2.2.3. Value added approach

The economy consists of either public or private enterprises. The enterprises contribute to the production with the help of given factors of production. In the economy the produced goods are useful for producers and individuals or we can say that they are either consumer goods or the producer goods. In the same way there are various people who are not directly producing. But without providing their services the economy cannot run. For example the services of doctor, teachers, Police, bank services etc. are necessary for economic development. Those people who are providing these services are paid in terms of the wages and salaries. Therefore, Value added approach or product approach of the estimation of the national income states that if we add all the money value of the final goods and services in domestic territory of the in an accounting year time gives us **Gross Domestic Product** at market prices.

Precautions while estimating National Income through Value Added Method

- A. Net Increase in stocks should be included.
- B. Own account production of fixed assets by all the producer enterprises should be included.
- C. Non- marketed goods and services for self-consumption should not be included.
- D. Imputed rent of owner occupied houses should be properly counted and included.
- E. Sales and purchase of second hand goods should not be included.

F. Trading of stocks and bonds should not count in the estimation of national income because it does not represent the production of new assets.

It is the sum of value-added at all stages in the production of final goods. Example:

- A farmer grows a bushel of wheat and sells it to a miller for \$1.00.
- The miller turns the wheat into flour and sells it to a baker for \$3.00.
- The baker uses the flour to make a loaf of bread and sells it to an engineer for \$6.00.
- The engineer eats the bread.

Compute & compare value added at each stage of production and GDP

$$\begin{aligned} \text{GDP} &= \text{value of final goods produced (\$ 6.00)} \\ &= \text{sum of value added at all stages of production (\$1 + \$2 + \$3 = \$6)}. \end{aligned}$$

The national accounts identity is the basic principle underlying the measurement of GDP since it implies that the total value of output produced could be calculated in any of three separate ways, each of which would necessarily give the same answer. Thus, the three approaches must yield the same results because the total expenditures on goods and services must by definition be equal to the value of the goods and services produced which must be equal to the total income paid to the factors that produced these goods and services.

Note: When using the actual data for a large economy like the United States, the Expenditures Approach and the Income Approach do not yield exactly the same value. However, it turns out to be close

Measuring Changes in GDP over Time

The GDP is often used to measure the growth in an economy over time. If the GDP is rising, we assume the economy is growing; if the GDP is falling, the economy is shrinking and probably is

in the middle of an economic downturn. Since the GDP measures the **value** of the final goods and services **produced** in the domestic economy in a given year, however, the GDP can rise from one year to the next for one of three reasons: either because the economy has **produced** more from one year to the next, because of the **value** of the product has gone up from year to year, or both. Since the value is measured in dollar prices, the GDP would go up from one year to the next, even if you produced exactly the same amount of output in both years but the prices of the products were to rise. It is therefore important to distinguish between what is called the Nominal GDP and what is called the Real GDP.

Nominal GDP vs. Real GDP

In measuring GDP, we use prices to measure the value of goods and services produced. Using the current prices to value current production is known as **nominal GDP**. The problem with nominal GDP is that a change in nominal GDP can be due to either (1) a change in the production of goods and services, or (2) a change in the prices of those goods and services, or (3) both. So an increase in prices will cause nominal GDP to rise, even if production has not changed at all. This gives a misleading picture of how well our economy is doing. It also makes it difficult to compare production from year to year, since prices change every year. To address the price problem, we also construct a measure of GDP that takes price changes into account. **Real GDP** values goods and services in any given year by using the prices of a set **base period**. By holding prices constant, real GDP measures only the changes in production from year to year. Changes in real GDP are used to measure economic growth. We choose a base year for measuring real GDP.

Let's continue our example above. The table below has the prices and production of apples, computers and pizza for three years:

Table 2.2: Economy producing three goods in three different years

Year	Apples		computers		pizza	
	price	quantity	price	quantity	price	quantity
1997	\$0.45	475	\$1100	70	\$7	380
1998	\$0.48	510	\$1050	85	\$8	390
1999	\$0.50	500	\$1000	100	\$9	400

First, let's calculate nominal GDP for each year. This means we take the price for that year times the quantity for that year.

$$\text{nominal GDP 1997} = (.45)(475) + (1100)(70) + (7)(380) = \$ 79,873.75$$

$$\text{nominal GDP 1998} = (.48)(510) + (1050)(85) + (8)(390) = \$ 92,614.80$$

$$\text{nominal GDP 1999} = (.50)(500) + (1000)(100) + (9)(400) = \$103,850$$

From 1997 to 1999 nominal GDP has increased by

$$\frac{103,580 - 79,873.75}{79873.75} \times 100 = 30.02\%$$

But this increase includes changes in BOTH price and production

Now let's calculate real GDP, using 1998 as the base year. This means for each year we will value output using 1998 prices.

$$\text{real GDP 1997} = (.48)(475) + (1050)(70) + (8)(380) = \$ 76,786$$

$$\text{real GDP 1998} = (.48)(510) + (1050)(85) + (8)(390) = \$ 92,614.80$$

$$\text{real GDP 1999} = (.48)(500) + (1050)(100) + (8)(400) = \$108,440$$

From 1997 to 1999 real GDP has increased by

$$\frac{108,440 - 76,786}{76,786} \times 100 = 41.22\%$$

But this increase includes ONLY changes in production, because prices are held constant at their 1998 levels. Note how the real GDP increase is greater. This is because in calculating nominal GDP, computer prices are falling over time even though computer production is increasing.

Note that real GDP = nominal GDP in the base year, since both measures use the same prices and same production.

Price Indices and GDP Growth

Inflation is an upward movement in the average level of prices and deflation is a downward movement in the average level of prices. The price level is measured by a price index (the average level of prices in one period relative to their average level in an earlier period). The two most common price indices are called the Consumer Price Index (CPI) and the GDP Deflator. We will discuss the CPI in a future lesson.

GDP Price Index

The GDP Deflator includes all of the items (C, I, G, and NX) included in the GDP. When comparing the value of the GDP from year to year, we use the GDP Deflator to make a valid comparison, i.e. one that takes into account the changes in prices that have occurred in the economy between the two years.

For example, let's assume we have a very simple economy that only produces three products: pineapples, computer, and beach umbrellas. The prices and outputs of these items in the current and base years are as follows:

	Current Period	Base Period

ITEM	Output	Price	Expenditures	Price	Expenditures
Pineapples	4,240	\$1.30	\$5,512	\$1.00	\$4,240
computer	5,000	\$10.00	\$50,000	\$8.00	\$40,000
Umbrellas	1,060	\$100.00	\$106,000	\$100.00	\$106,000

The Nominal GDP will be the value of the current year's output using the current year's prices:

$$\begin{aligned} \text{Nominal GDP in the Current Period} &= (4,240 \times \$1.30) + (5,000 \times \$10.00) + (1,060 \times \$100.00) \\ &= \$156,512 \end{aligned}$$

The Real GDP will be the value of the current year's output using the base year's prices:

$$\begin{aligned} \text{Real GDP in the Current Period} &= (4,240 \times \$1.00) + (5,000 \times \$8.00) + (1,060 \times \$100.00) \\ &= \$150,240 \end{aligned}$$

$$\text{GDP Deflator} = (\text{Nominal GDP} / \text{Real GDP}) \times 100$$

$$(\$156,512 / \$150,240) \times 100 = 104.175$$

Notice that the GDP Deflator is equal to the Nominal GDP divided by the Real GDP and multiplied by 100.

This deflator tells us that there has been 4.175% inflation over the period from the base year to the current year. Note that if the prices had been the same in the base year and in the current year, the Nominal and the Real GDP would have been the same in the current year and the deflator would have equaled 100. A deflator of 100 indicates no inflation between the two periods. A deflator greater than 100 indicates inflation and a deflator less than 100 indicates deflation (a decline in the average price level from one year to the next).

2.3. Measurement problems and Limitations of GDP as a Measure of Welfare

Gross Domestic Product (GDP) is essentially an indicator of aggregate economic activity. In addition to that it is also frequently used to describe social welfare. The idea behind this is that

GDP tends to correlate with consumption, which in turn is commonly used as a proxy for welfare. In other words, the more people consume, the happier they are supposed to be.

Now, this line of argument seems a little too simplistic. Assuming causality based on a simple correlation between GDP and welfare may lead to false conclusions which can be highly problematic especially for policy makers. Hence it is important to look at the limitations of GDP as a welfare indicator and to consider possible alternative approaches.

Limitations of GDP

There are several limitations of GDP as a welfare indicator. Most of them can be traced back to the fact that in essence GDP is not supposed to measure well-being. As a result the concept does not account for various important factors that influence social welfare. To keep things simple the most relevant limitations are listed below:

- **GDP does not incorporate any measures of welfare:** This is probably the most obvious issue. As mentioned before, GDP only describes the value of all finished goods produced within an economy over a set period of time. There are multiple ways to calculate and measure GDP, but neither of them includes any indicator of welfare or well-being. Even though this does not necessarily mean GDP cannot be a good indicator of welfare, the fact that it is used as a "proxy of a proxy" should be kept in mind as it significantly affects its validity.
- **GDP only includes market transactions:** As a result, it does not account for domestic or voluntary work, even though these activities have a considerable positive impact on social welfare, as they complement the market economy and thus improve the standard of living. On the other hand GDP does not include black market transactions or other illegal activities that may have a substantial negative impact on overall social well-being.
- **GDP does not describe wealth distribution:** If there is a high degree of wealth inequality, the majority of people do not really benefit from an increased economic output because they cannot afford to buy most of the goods and services. Thus to accurately describe social welfare it is essential to consider wealth distribution.

- **GDP does not describe what is being produced:** Since GDP measures the value of all finished goods and services within an economy; it also includes products that may have negative effects on social welfare. Think of a country with an extremely strong armaments industry that represents most of its GDP. If the arms are sold and used within the country itself, overall social welfare will most likely decrease. Of course this also holds true for other goods and services that may have adverse effects on society.
- **GDP ignores externalities:** Economic growth usually goes hand in hand with increased exploitation of both renewable and non-renewable resources. Due to this overuse, more and more negative externalities arise (e.g. pollution, overfishing) and social welfare will decrease as a result. This effect is not included in GDP at all.

If we look at these aspects, the major issue with GDP as a welfare indicator becomes quite obvious. It suggests that a higher GDP always increases social well-being. However at one point the positive effects resulting from the increase in consumption opportunities may be outweighed by the negative effects associated with the limitations mentioned above. Hence although GDP may on certain occasions be a good proxy for social welfare, it results in a biased description that may lead to unfavorable conclusions.

3. Economic Performance and Business Cycle

3.1. Definitions and Concepts of Business Cycle

The **business cycle** or **economic cycle** refers to the fluctuations of economic activity about its long term growth trend. *Inflation, growth, and unemployment are related through the business cycle.* These fluctuations are often measured using the real gross domestic product. Shortly, the business cycle is the periodic but irregular up-and-down movement in the economy. That is, output is not always at its trend level, that is, the level corresponding to full employment of the factors of production. Rather output fluctuates around the trend level. During expansion (or recovery) the employment of factors of production increased, and that is a source of increased production. Conversely, during a recession unemployment increases and less output is produced than can in fact be produced with the existing resources and technology. Deviations of output from trend are referred to as the **output gap**.

The output gap measures the gap between actual output and the output the economy could produce at full employment given the existing resources. Full employment output is also called **potential output**.

$$\text{Output gap} \equiv \text{actual output} - \text{potential output}$$

3.2. Phases of business cycle

Four Phases of Business Cycle

Business Cycle (or Trade Cycle) is divided into the following four phases:-

Prosperity Phase: Expansion or Boom or Upswing of economy

Recession Phase: from prosperity to depression (upper turning point)

Depression Phase: Contraction or Downswing of economy

Recovery Phase: from depression to prosperity (lower turning Point)

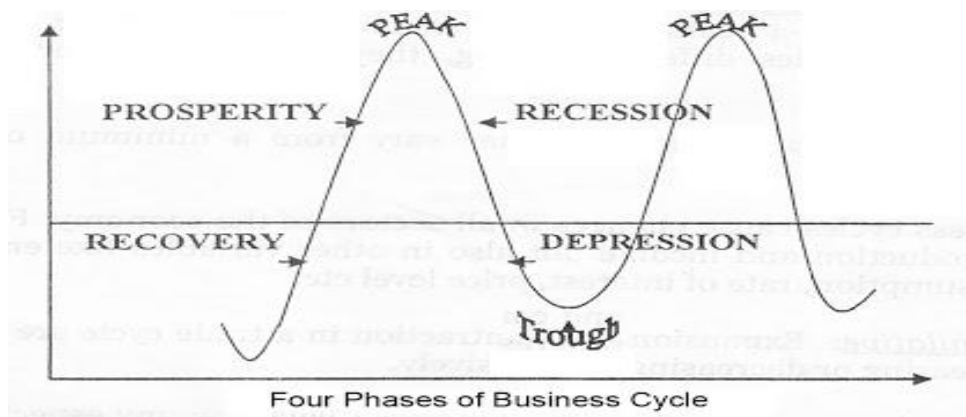
There are two turning points: The business cycle is characterized by two turning points namely **peaks** and **troughs**.

- The **peak** is the highest level of real GDP in the cycle. Each peak indicates an economy operating at close to full capacity, so that national product and national income corresponds to a very high degree of utilization of factors of production.

- A **trough** is the lowest level of real GDP observed over the business cycle. A trough is reached when the economy begins to pull out of recession. During this time there is an excessive amount of unemployment and idle productive capacity. Businesses are more likely to fail because of low demand for their products.

✚ Diagram of Four Phases of Business Cycle

The four phases of business cycles are shown in the following diagram:-



The business cycle starts from a trough (lower point) and passes through a recovery phase followed by a period of expansion (upper turning point) and prosperity. After the peak point is reached there is a declining phase of recession followed by a depression. Again the business cycle continues similarly with ups and downs.

Explanation of Four Phases of Business Cycle

✚ The four phases of a business cycle are briefly explained as follows:-

1. Prosperity Phase

When there is an expansion of output, income, employment, prices and profits, there is also a rise in the standard of living. This period is termed as prosperity phase.

The **features of prosperity** are:-

1. High level of output and trade
2. High level of effective demand
3. High level of income and employment

4. Rising interest rates
5. Inflation
6. Large expansion of bank credit
7. Overall business optimism.
8. A high level of MEC (Marginal efficiency of capital) and investment

Due to full employment of resources, the level of production is maximum and there is a rise in **GNP** (Gross National Product). Due to a high level of economic activity, it causes a rise in prices and profits. There is an upswing in the economic activity and economy reaches its **Peak**. This is also called as a **Boom Period**.

2. Recession Phase

The turning point from prosperity to depression is termed as recession phase. During a recession period, the economic activities slow down. When demand starts falling, the overproduction and future investment plans are also given up. There is a steady decline in the output, income, employment, prices and profits. The businessmen lose confidence and become pessimistic (negative). It reduces investment. The banks and the people try to get greater liquidity, so credit also contracts. Expansion of business stops, stock market falls. Orders are cancelled and people start losing their jobs. The increase in unemployment causes a sharp decline in income and aggregate demand. Generally, recession lasts for a short period.

3. Depression Phase

When there is a continuous decrease of output, income, employment, prices and profits, there is a fall in the standard of living and depression sets in.

The **features of depression** are:-

1. Fall in volume of output and trade
2. Fall in income and rise in unemployment
3. Decline in consumption and demand
4. Fall in interest rate
5. Deflation
6. Contraction of bank credit

7. Overall business pessimism
8. Fall in MEC (Marginal efficiency of capital) and investment

In depression, there is under-utilization of resources and fall in GNP (Gross National Product). The aggregate economic activity is at the lowest, causing a decline in prices and profits until the economy reaches its **Trough** (low point).

4. Recovery Phase

The turning point from depression to expansion is termed as Recovery or **Revival** Phase. During the period of revival or recovery, there are expansions and rise in economic activities. When demand starts rising, production increases and this causes an increase in investment. There is a steady rise in output, income, employment, prices and profits. The businessmen gain confidence and become optimistic (positive). This increases investments. The stimulation of investment brings about the revival or recovery of the economy. The banks expand credit, business expansion takes place and stock markets are activated. There is an increase in employment, production, income and aggregate demand, prices and profits start rising, and business expands. Revival slowly emerges into prosperity, and the business cycle is repeated.

Thus we see that, during the expansionary or prosperity phase, there is inflation and during the contraction or depression phase, there is a deflation.

3.3 Causes and Effects of business cycle

Interest rates: Changes in the interest rate affect consumer spending and economic growth. For example, if interest rates are cut, this reduces borrowing costs; this leads to higher spending and economic growth. However, if the Central Bank increases interest rates to reduce inflation, this will tend to reduce consumer spending and investment, leading to an economic downturn and recession. See: Interest rate cycle. High-interest rates in 1991-92 were a major factor in the recession of that year. The cut in interest rates post 1992 helped the economy to recover.

Consumer and business confidence: People are easily influenced by external events. If there is a succession of bad economic news, this tends to discourage people from spending and investing, making a small downturn in to a bigger recession. But, when the economy recovers this can cause a positive movement effect. Economic growth encourages consumers to borrow and banks to lend.

This causes higher economic growth. Confidence is an important factor in causing the business cycle.

Monetary effect: business cycle is caused by the expansion and contraction of bank credit. The business cycle will be in the upward phase when the banking system creates more money. Results in expansion. Soon the bank starts restricting credits as they exhaust their reserves, results in contraction.

Purchasing power: another cause of business cycle is under consumption or over consumption. During the period of boom, the income of wealthier people will increase and they will start investing it in to the production, as a result supply increase.

Over investment: according to Frederick A. Hayek, if money supply increases the investors will invest more money in the capital goods which results in expansion but soon it turns into contraction because of over investment.

Human psychology: psychological cause refers to men's attitude of mind towards actual economic conditions. When businessman expects better times, they expand investment, this gives rise to expansion. A small shock or small failure will reverse the process, expansion comes to an end.

Cyclical changes in weather: A cyclical change in the weather is also a cause of the business cycle. If the weather is favorable then agricultural production will increase and business cycle also will see the expansion.

3.4 Measures to control business cycle

The following are the main measures which can be suggested for the effective control of business cycle fluctuation.

1. Monetary Policy
2. Fiscal Policy
3. State Control of Private Investment
4. International Measures to Control of Business Cycle
5. Reorganization of Economic System

1. Monetary Policy A Control of Business Cycle

Monetary policy as measure to control business cycle fluctuation refers to all those measures which are taken with a view to control money and credit supply in the country. When we are in the state of full employment and we are facing inflation, a deflationary policy may be adopted. The central bank can reduced the quantity of money in circulation. The bank can adopt different measures for this purpose, like increase in the bank rate, selling of securities in the market, increasing the reserve ratio of the member banks etc. On the other hand, in case of deflation the central bank can adopt inflationary monetary policy by lowering the bank rates or purchase of securities. Monetary policy has achieved a very limited success in the past, because central bank has not full power over the supply of money and credit in the country. Moreover, the quantity of money has failed during the world depression of 1930s.

2. Fiscal Policy Measure to Control of Business Cycle Fluctuation

Fiscal policy as measure to control business cycle fluctuation nowadays is considered to be a powerful anti-cycle weapon in the hands of the government. Fiscal policy involves the process of shaping the public finance (income and expenditure) with a view of reduce fluctuations in the business cycle and attainment of full employment without inflation. In case of inflation the governments reduces the public work programs, imposing heavy taxes on business profits to discourage private investment, reduces purchasers power, taking loans from the people. All these fiscal measures greatly help in reducing the inflationary trend in the economy.

If the economy facing depression, the government increases it expenditure on public works programs like construction of new canals, new roads, buildings etc. Increase in government expenditure, income, employment, profit and consumption of the people. In order to encourage private investment the government reduces taxes on profit. All these fiscal measures to control business cycle sets in upswing in the economy.

3. State Control of Private Investment

Some economists have suggested that if a government takes control of private investment is a tool to control of business cycle fluctuations can be controlled within the limits. The other economists, who disagree with the above view state that if a government takes control of private investment, private investment will be discouraged. Low investment will reduce employment and income.

4. International Measures Control of Business Cycle

Today, every country has trade relations with the rest of the world. If there is inflation or deflation in one country, it can be easily carried to other countries. The example of great depression can be

given. Business cycle is an international phenomenon and it should be tackled on international level. Different measures to control business cycle fluctuations have been suggested by some well-known economists these are:

- Control of International Production
- International Bill Stock Control
- International Investment Control

5. Reorganization of Economic System

Some economists suggest that there should be complete reorganization of the whole economic system to control of business cycle fluctuation. The capitalistic system of production should be replaced by the socialistic system of production. In socialistic economy, there are few chances of cyclic fluctuations. In 1930, when all capitalist countries of the world were suffering from depression, it was only socialist countries which were free from such crisis.

4. MACRO ECONOMIC POLICY

Macroeconomic Policies are tools used by the Government to manage and influence the performance and behavior of the economy. The set of government rules and regulations to control or stimulate the aggregate indicators of an economy frames the macroeconomic policy. Aggregate indicators involve national income, money supply, inflation, unemployment rate, growth rate, interest rate and many more. In short, policies framed to meet the macro goals.

These are important because they affect the economy in which businesses operate. Macroeconomic policy is usually implemented through two sets of tools: **fiscal and monetary policy**. Both forms of policy are used to stabilize the economy, which can mean boosting the economy to the level of GDP consistent with full employment. Macroeconomic policy focuses on limiting the effects of the business cycle to achieve the economic goals of price stability, full employment, and growth.

The **Key objectives** of Macroeconomic policies are:

- Full employment of resources (Full and Stable Employment)
- Price Stability (little or no inflation putting upward pressure on price)
- Economic Growth (National Income)
- Balance of Payments Stability (Payment Surplus/deficit)
- Appropriate distribution of Income and Wealth

To achieve these objectives can be very difficult because conflicts between macroeconomic objectives exist. **For example**, the achievement of full employment may lead to excessive inflation because of the increase of level of aggregate demand within an economy. Macroeconomic policies can influence the economy and businesses through fiscal and monetary policies.

Fiscal policy is the macroeconomic policy where the government makes changes in government spending or tax to stimulate growth. Monetary policy deals with changes in money supply or changes with the parameters that affects the supply of money in the economy. Contract laws, debt management policy, income policy are some of the other macroeconomic policies designed to modify macroeconomic indicators of the economy.

Economists and politicians all have their favourite approaches. Many favour a combination of taxing and public spending, while others advocate regulating the supply of money. In practice, economic policy involves a mixture of the two, but during the past several decades each method has at one time or another been dominant.

6.1 Fiscal Policy

Fiscal policy is the use of government's revenue and expenditure as instruments to influence the economy. Examples of such tools are [expenditure](#), [taxes](#), [debt](#). For example, if the economy is producing less than potential output, government spending can be used to employ idle resources and boost output. Government spending does not have to make up for the entire output gap. There is a [multiplier effect](#) that boosts the impact of government spending. For instance, when the government pays for a bridge, the project not only adds the value of the bridge to output, but also allows the bridge workers to increase their consumption and investment, which helps to close the output gap. When fiscal policy is used to try to increase output and reduce unemployment, it is called expansionary; when fiscal policy is used to try to lower inflation, it is referred to as contractionary.

The government can **manipulate budgets** to influence the level of **aggregate demand** and activity in the economy and this refers to fiscal policies. It covers:

- Government Spending
- Taxation
- Government Borrowing

The role of the Chancellor of the Exchequer in the UK is to balance the budget in two circumstances:

1. **Deficit**- Additional spending financed through borrowings
2. **Surplus**- Spare funds use to pay off public debts

To see how fiscal policy works, consider a period of high unemployment and business stagnation. America has suffered through numerous such periods such as in the early 1980s and 1991 to late 1993. The national government attempts to revive industry and create jobs by injecting billions of dollars into the economy. It does so by cutting taxes, thereby leaving individuals and businesses

with more to spend; by purchasing goods and services (such as building bridges and buying airplanes); and by making direct payments to individuals (social security or unemployment insurance, for example). In theory at least, the net effect is to raise **aggregate demand**, the total goods and services citizens and businesses can afford to buy. A rise in demand causes industries to manufacture more products, hire additional labor, and invest in new buildings and machinery, all of which helps business and trade.

Government spending, moreover, has a **multiplier effect**. The billions of dollars allocated to public projects go into the pockets of carpenters, steelworkers, truck drivers, and thousands of other labourers, who spend the money on food, clothing, housing, medical care, automobiles, and recreation. Workers in these industries, in turn, spend their wages on additional goods and services. Gradually the government's dollars trickle through the economy.

In times of prosperity, on the other hand, demand may exceed supply. The excess causes prices to increase and, unless stopped, leads to **inflation**, a condition in which the value of money decreases as prices rise. When this happens, the government reverses mechanisms by cutting spending, raising taxes, or both. The result is less money in the hands of consumers and business, and less money means lower aggregate demand, which causes prices to level off.

Fiscal policy thus strives to smooth out the business cycle by manipulating the federal budget to maintain just enough demand to keep people working but not so much as to fuel inflation. By adjusting spending and taxation, the government can in principle maintain high levels of employment and stable prices.

6.2 Monetary Policy

Monetary policy attempts to control the amount of money in circulation or the cost and availability of credit. Monetary policy falls within the area of the [Federal Reserve System](#), the **nation's central bank**.

Central banks implement monetary policy by controlling the money supply through several mechanisms. Typically, central banks take action by issuing money to buy bonds (or other assets), which boosts the supply of money and lowers interest rates, or, in the case of contractionary monetary policy, banks sell bonds and take money out of circulation. Usually policy is not implemented by directly targeting the supply of money.

Central banks continuously shift the money supply to maintain a targeted fixed interest rate. Some of them allow the interest rate to fluctuate and focus on [targeting inflation rates](#) instead. Central banks generally try to achieve high output without letting loose monetary policy that creates large amounts of inflation.

Monetary policy is used to influence the rate of interest and the money supply in an economy.

Monetary policies impact:

- Cost of finance
- Level of consumer demand
- Level of exchange rates
- Level of inflation

The Government is careful when changing interest rates as the effects are uncertain. For example if interest rates were raised, this would discourage expenditure due to the rise in cost of credit. However effects will vary because they might affect some sectors of business greater than others because credit-based purchases are typically consumer durable goods and houses, this may result in instability in some sectors.