Mathematical Economics

UTKAL UNIVERSITY

NAAC ACCREDITED A+
DDCE
Education for All

DIRECTORATE OF DISTANCE AND CONTINUING EDUCATION
Utkal University, Bhubaneswar - 7, Odisha

E-mail: helpline@ddceutkal.ac.in
Website: www.ddceutakal.ac.in

Mathematical Economics

Suman Kalyan Chakraborty

This SIM has been prepared exclusively for Directorate of Distance and Continuing Education (DDCE), Utkal University. It conforms to the syllabi and contents as approved by Utkal University.

From the Director's Desk

The Directorate of Distance and Continuing Education, originally established as the University Evening College way back in 1962, has travelled a long way in the last 57 years. **'EDUCATION FOR ALL'** is our motto. Increasingly, the Open and Distance Learning institutions are aspiring to provide education for anyone, anytime and anywhere. DDCE, Utkal University has been constantly striving to rise up to the challenges of Open Distance Learning system. Nearly one lakh students have passed through the portals of this great temple of learning. We may not have numerous great tales of outstanding academic achievements but we have great tales of success in life, of recovering lost opportunities, tremendous satisfaction in life, turning points in career and those who feel that without us they would not be where they are today. There are also flashes when our students figure in best ten in their honours subjects. Our students must be free from despair and negative attitude. They must be enthusiastic, full of energy and confident of their future. To meet the needs of quality enhancement and to address the quality concerns of our stakeholders over the years, we are switching over to self-instructional material printed courseware. We are sure that students would go beyond the courseware provided by us. We are aware that most of you are working and have also family responsibility. Please remember that only a busy person has time for everything and a lazy person has none. We are sure that you will be able to chalk out a well planned programme to study the courseware. By choosing to pursue a course in distance mode, you have made a commitment for self-improvement and acquiring higher educational qualification. You should rise up to your commitment. Every student must go beyond the standard books and self-instructional course material. You should read number of books and use ICT learning resources like the internet, television and radio programmes, etc. As only limited number of classes will be held, a student should come to the personal contact programme well prepared. The PCP should be used for clarification of doubt and counseling. This can only happen if you read the course material before PCP. You can always mail your feedback on the courseware to us. It is very important that one should discuss the contents of the course materials with other fellow learners.

We wish you happy reading.

DIRECTOR

SYLLABUS

Unit – I

1. Theory of Consumer Behaviour: Marshallian Demand Analysis, Indifference Curve Analysis, Slutsky Equation, Comparative Static Theorems, Theory of Revealed Preference.

Unit – II

1. Theory of the Firm: Production and Cost Function, Homogeneous Production Functions, Euler's Theorem, Cobb-Douglas Production Function, CES Production Function, Characteristics of Production Possibility Sets, Duality Relationship between Production and Cost Functions, Comparative Static Result, Joint Production.

2. Market Equilibrium: Factor Market Equilibrium, Stability of Equilibrium.

Unit – III

1. Linear Programming: Basic Theorems, Theory of the Simplex Method (Non-degeneracy Excluded), Duality Theorems, Complimentary Slackness Theorem.

2. Non-linear Programming: Kuhn-Tucker Optimality Criteria, Kuhn-Tucker Results and Saddle Points with Special Reference to Concave Programming.

Unit – IV

1. Theory of Games: Two-person Zero-sum Game, Saddle Point, Pure and Mixed Strategies, Equivalence of Matrix Game and Linear Programming, Rectangular Game and its Solution.

2. The Leontief System: Static Input-Output Analysis, Hawkins-Simon Theorem, A Linear Programming Interpretation, Theorem on Non-substitution.

Unit – V

1. Theories of Economic Growth: Harrod-Domar Model, Neo-classical Model of Solow, Neo-Keynesian Model of Passineti, Two-sector Model of Uzawa, Optional Economic Growth – Ramsey Problem.

2. Multisector Growth Models: Von-Neumann Growth Model, Concept of Efficiency and Optimisation for Von-Neumann Model, Turnpike Theorems – Samuelson, Turnpike Result.

CONTENTS

UNIT – I

UNIT – II

UNIT – III

UNIT – I

Chapter	
1	**THEORY OF THE CONSUMER BEHAVIOUR**

Objectives

The objectives of this lesson are to learn:

- Marshallian demand analysis
- Indifference curve analysis
- Slutsky equation
- Comparative Static Theorem
- Theory of Revealed Preference

Structure:

1.1 THEORY OF CONSUMER BEHAVIOUR: MARSHALLIAN DEMAND ANALYSIS

Cardinal Utility Analysis

From time to time, different theories have been advanced to explain consumer's demand for a good and to derive a valid demand theorem. Cardinal utility analysis is the oldest theory of demand which provides an explanation of consumer's demand for a product and derives the law of demand which establishes an inverse relationship between price and quantity demanded of a product.

Introduction

The price of a product depends upon the demand for and the supply of it. In this part of the book, we are concerned with the theory of consumer's behaviour, which explains his demand for a good and the factors determining it. Individual's demand for a product depends upon price of the product, income of the individual and the prices of related goods.

It can be put in the following functional form:

$D_x = f(P_x, I, P_y, P_z, T, etc.)$

where D_x stands for the demand of good X, P_x for price of good X, I for individual's income, P_y and P_z for the prices of related goods and T for tastes and preferences of the individual. But among these determinants of demand, economists single out price of the good in question as the most important factor governing the demand for it. Indeed, the function of a theory of consumer's behaviour is to establish a relationship between quantity demanded of a good and its own price, and to provide an explanation for it.

Recently, cardinal utility approach to the theory of demand has been subjected to severe criticisms, and as a result, some alternative theories, namely, Indifference Curve Analysis, Samuelson's Revealed Preference Theory, and Hicks' Logical Weak Ordering Theory have been propounded.

Assumptions of Cardinal Utility Analysis

Cardinal utility analysis of demand is based upon certain important assumptions. Before explaining how cardinal utility analysis explains consumer's equilibrium in regard to the demand for a good, it is essential to describe the basic assumptions on which the whole utility analysis rests. As we shall see later, cardinal utility analysis has been criticised because of its unrealistic assumptions.

The basic assumptions or premises of cardinal utility analysis are as follows:

The Cardinal Measurability of Utility

The exponents of cardinal utility analysis regard utility to be a cardinal concept. In other words, they hold that utility is a measurable and quantifiable entity. According to them, a person can express utility or satisfaction he derives from the goods in the quantitative cardinal terms. Thus, a person can say that he derives utility equal to 10 units from the consumption of a unit of good A, and 20 units from the consumption of a unit of good B.

Moreover, the cardinal measurement of utility implies that a person can compare utilities derived from goods in respect of size, i.e., how much one level of utility is greater than another. A person can say that the utility he gets from the consumption of one unit of good B is double the utility he obtains from the consumption of one unit of good A.

According to Marshall, marginal utility is actually measurable in terms of money. Money represents the general purchasing power and it can therefore be regarded as a command over alternative utility-yielding goods. Marshall argues that the amount of money which a person is prepared to pay for a unit of a good rather than go without it is a measure of the utility he derives from that good.

Thus, according to him, money is the measuring rod of utility. Some economists belonging to the cardinalist school measure utility in imaginary units called "utils". They assume that a consumer is capable of saying that one apple provides him utility equal to 4 utils. Further, on this ground, he can say that he gets twice as much utility from an apple as compared to an orange.

The Hypothesis of Independent Utilities

The second important tenet of the cardinal utility analysis is the hypothesis of independent utilities. On this hypothesis, the utility which a consumer derives from a good is the function of the quantity of that good and of that good only. In other words, the utility which a consumer obtains from

a good does not depend upon the quantity consumed of other goods; it depends upon the quantity purchased of that good alone.

On this assumption, then the total utility which a person gets from the whole collection of goods purchased by him is simply the total sum of the separate utilities of the goods. Thus, the cardinalist school regards utility as 'additive', i.e., separate utilities of different goods can be added to obtain the total sum of the utilities of all goods purchased.

Constancy of the Marginal Utility of Money

Another important assumption of the cardinal utility analysis is the constancy of the marginal utility of money. Thus, while the cardinal utility analysis assumes that marginal utilities of commodities diminish as more of them are purchased or consumed, but the marginal utility of money remains constant throughout when the individual is spending money on a good and due to which the amount of money with him varies. Daniel Bernoulli first of all introduced this assumption but later Marshall adopted this in his famous book "Principles of Economics".

As stated above, Marshall measured marginal utilities in terms of money. But measurement of marginal utility of goods in terms of money is only possible if the marginal utility of money itself remains constant. It should be noted that the assumption of constant marginal utility of money is very crucial to the Marshallian analysis, because otherwise Marshall could not measure the marginal utilities of goods in terms of money. If money which is the unit of measurement itself varies as one is measuring with it, it cannot then yield correct measurement of the marginal utility of goods.

When price of a good falls and as a result the real income of the consumer rises, marginal utility of money to him will fall but Marshall ignored this and assumed that marginal utility of money did not change as a result of the change in price. Likewise, when price of a good rises, the real income of the consumer will fall and his marginal utility of money will rise. But Marshall ignored this and assumed that marginal utility of money remains the same. Marshall defended this assumption on the ground that "his (the individual consumer's) expenditure on any one thing is only a small part of his whole expenditure."

Introspective Method

Another important assumption of the cardinal utility analysis is the use of introspective method in judging the behaviour of marginal utility. "Introspection is the ability of the observer to reconstruct events which go on in the mind of another person with the help of self-observation. This form of comprehension may be just guesswork or intuition or the result of long lasting experience."

Thus, the economists construct with the help of their own experience the trend of feeling which goes on in other men's mind. From his own response to certain forces and by experience and observation, one gains understanding of the way other people's minds would work in similar situations. To sum up, in introspective method, we attribute to another person what we know of our own mind, i.e., by looking into ourselves, we see inside the heads of other individuals.

So, the law of diminishing marginal utility is based upon introspection. We know from our own mind that as we have more of a thing, the less utility we derive from an additional unit of it. We conclude from it that other individuals' mind will work in a similar fashion, i.e., their marginal utility for a good will diminish as they have more units of it.

With the above basic premises, the founders of cardinal utility analysis have developed two laws which occupy an important place in economic theory and have several applications and uses.

These two laws are:

1. Law of Diminishing Marginal Utility and

2. Law of Equi-marginal Utility.

It is with the help of these two laws about consumer's behaviour that the exponents of cardinal utility analysis have derived the law of demand. We explain below these two laws in detail and how law of demand is derived from them.

Law of Diminishing Marginal Utility

An important tenet of cardinal utility analysis relates to the behaviour of marginal utility. This familiar behaviour of marginal utility has been stated in the Law of Diminishing Marginal Utility according to which marginal utility of a good diminishes as an individual consumes more units of a good. In other words, as a consumer takes more units of a good, the extra utility or satisfaction that he derives from an extra unit of the good goes on falling.

It should be carefully noted that it is the marginal utility and not the total utility that declines with the increase in the consumption of a good. The law of diminishing marginal utility means that the total utility increases at a decreasing rate.

Marshall who has been a famous exponent of the cardinal utility analysis has stated the law of diminishing marginal utility as follows:

"The additional benefit which a person derives from a given increase of his stock of a thing diminishes with every increase in the stock that he already has."

This law is based upon two important facts. First, while the total wants of a man are virtually unlimited, each single want is satiable. Therefore, as an individual consumes more and more units of a good, the intensity of his want for the good goes on falling and a point is reached where the individual no longer wants any more units of the good, i.e., when saturation point is reached, marginal utility of a good becomes zero. Zero marginal utility of a good implies that the individual has all that he wants of the good in question.

The second fact on which the law of diminishing marginal utility is based is that the different goods are not perfect substitutes for each other in the satisfaction of various wants. When an individual consumes more and more units of a good, the intensity of his particular want for the good diminishes but if the units of that good could be devoted to the satisfaction of other wants and yielded as much satisfaction as they did initially in the satisfaction of the first want, marginal utility of the good would not have diminished.

It is obvious from above that the law of diminishing marginal utility describes a familiar and fundamental tendency of human nature. This law has been arrived at by introspection and by observing how consumers behave.

Illustration of the Law of Diminishing Marginal Utility

Consider Table 1.1 where we have presented the total and marginal utilities derived by a person from cups of tea consumed per day. When one cup of tea is taken per day, the total utility derived by the person is 12 utils. And because this is the first cup, its marginal utility is also 12 utils with the consumption of 2nd cup per day, the total utility rises to 22 utils but marginal utility falls to 10. It will be seen from the table that as the consumption of tea increases to six cups per day, marginal utility from the additional cup goes on diminishing (i.e., the total utility goes on increasing at a diminishing rate).

However, when the cups of tea consumed per day increases to seven, then instead of giving positive marginal utility, the seventh cup gives negative marginal utility equal to –2 utils. This is because too many cups of tea consumed per day (say more than six for a particular individual) may cause acidity and gas trouble. Thus, the extra cups of tea beyond six to the individual in question gives him disutility rather than positive satisfaction.

Table 1.1: Diminishing Marginal Utility

Cup of Tea Consumed/Day (Q)	Total Utility (Utils) TU	Marginal Utility (Utils) ΔTU/ΔQ
1	12	12
2	22	10
3	30	8
4	36	6
5	40	4
6	41	1
7	39	-2
8	34	-5

Fig. 1.1 illustrates the total utility and the marginal utility curves. The total utility curve drawn in Fig. 1.1 is based upon three assumptions. Firstly, as the quantity consumed per period by a consumer increases, his total utility increases but at a decreasing rate. This implies that as the consumption per period of a commodity by the consumer increases, marginal utility diminishes as shown in the lower panel of Fig. 1.1. Secondly, as will be observed from the figure when the rate of consumption of a commodity per period increases to Q_4, the total utility of the consumer reaches its maximum level. Therefore, the quantity Q_4 of the commodity is called satiation quantity or satiety point. Thirdly, the increase in the quantity consumed of the good per period by the consumer beyond the satiation point has an adverse effect on his total utility, i.e., his total utility declines if more than Q_4 quantity of the good is consumed.

This means beyond Q_4, marginal utility of the commodity for the consumer becomes negative and will be seen from the lower panel of Fig. 1.1 beyond the satiation point Q_4 marginal utility curve MU goes below the X-axis indicating it becomes negative beyond quantity Q_4 per period of the commodity consumed.

It is important to understand how we have drawn the marginal utility curve. As stated above, marginal utility is the increase in total utility of the consumer caused by the consumption of an additional unit of the commodity per period. We can directly find out the marginal utility of the successive units of the commodity consumed by measuring the additional utility which a consumer obtains from successive units of the commodity and plotting them against their respective quantities.

However, in terms of calculus, marginal utility of a commodity X is the slope of the total utility function $U = f(Q_x)$. Thus, we can derive the marginal utility curve by measuring the slope at various points of the total utility curve TU in the upper panel of Fig. 1.1 by drawing tangents at them. For instance, at the quantity Q_1, marginal utility (i.e., $dU/dQ = MU_1$) is found out by drawing tangent at point A and measuring its slope which is then plotted against quantity in the lower panel of Fig. 1.1. In the lower panel, we measure marginal utility of the commodity on the Y-axis. Likewise, at quantity Q_2, marginal utility of the commodity has been obtained by measuring slope of the total utility curve TU at point B and plotting it in the lower panel against the quantity Q_2.

It will be seen from the figure that at Q_4 of the commodity consumed, the total utility reaches at the maximum level T. Therefore, at quantity Q_4, the slope of the total utility curve is zero at this point. Beyond the quantity Q_4, the total utility declines and marginal utility becomes negative. Thus, quantity Q_4 of the commodity represents the satiation quantity.
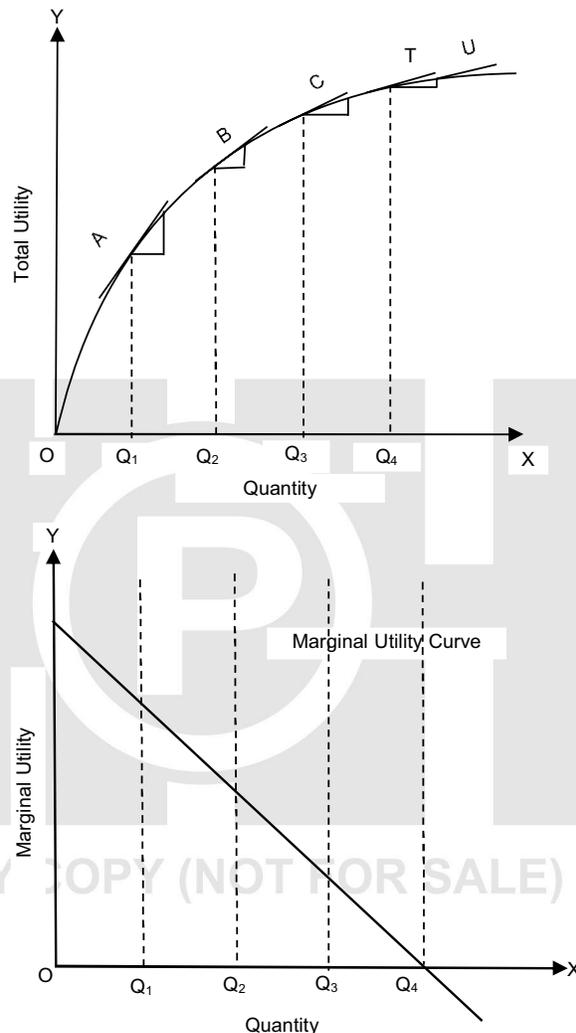
Fig. 1.1: Total Utility and Marginal Utility

Another important relationship between total utility and marginal utility is worth noting. At any quantity of a commodity consumed, the total utility is the sum of the marginal utilities. For example, if marginal utility of the first, second and third units of the commodity consumed are 15, 12, and 8 units, then the total utility obtained from these three units of consumption of the commodity must equal 35 units $(15 + 12 + 8 = 35)$.

Similarly, in terms of graphs of total utility and marginal utility depicted in Fig. 1.1, the total utility of the quantity Q_4 of the commodity consumed is the sum of the marginal utilities of the units of commodity up to point Q_4, i.e., the entire area under the marginal utility curve MU in lower panel up to the point Q_4 is the sum of marginal utilities which must be equal to the total utility Q_4T in the upper panel.

Marginal Utility and Consumer's Tastes and Preferences

The utility people derive from consuming a particular commodity depends on their tastes and preferences. Some consumers like oranges, others prefer apples and still others prefer bananas for consumption. Therefore, the utility which different individuals get from these various fruits depends on their tastes and preferences.

An individual would have different marginal utility curves for different commodities depending on his tastes and preferences. Thus, utility which people derive from various goods reflect their tastes and preferences for them. However, it is worth noting that we cannot compare utility across consumers. Each consumer has a unique subjective utility scale. In the context of cardinal utility analysis, a change in consumer's tastes and preferences means a shift in his one or more marginal utility curves.

However, it may be noted that a consumer's tastes and preferences do not frequently change, as these are determined by his habits. Of course, tastes and preferences can change occasionally. Therefore, in economic theory, we generally assume that tastes or preferences are given and relatively stable.

Significance of Diminishing Marginal Utility

The significance of the diminishing marginal utility of a good for the theory of demand is that it helps us to show that the quantity demanded of a good increase as its price falls and *vice versa.* Thus, it is because of the diminishing marginal utility that the demand curve slopes downward. If properly understood, the law of diminishing marginal utility applies to all objects of desire including money.

But it is worth mentioning that marginal utility of money is generally never zero or negative. Money represents purchasing power over all other goods, i.e., a man can satisfy all his material wants if he possesses enough money. Since man's total wants are practically unlimited, therefore, the marginal utility of money to him never falls to zero.

The marginal utility analysis has a good number of uses and applications in both economic theory and policy. The concept of marginal utility is of crucial significance in explaining determination of the prices of commodities. The discovery of the concept of marginal utility has helped us to explain the paradox of value which troubled Adam Smith in "The Wealth of Nations."

Adam Smith was greatly surprised to know why water which is so very essential and useful to life has such a low price (indeed no price), while diamonds which are quite unnecessary have such a high price. He could not resolve this water-diamond paradox. But modern economists can solve it with the aid of the concept of marginal utility.

According to the modern economists, the total utility of a commodity does not determine the price of a commodity and it is the marginal utility which is crucially important determinant of price. Now, the water is available in abundant quantities so that its relative marginal utility is very low or even zero. Therefore, its price is low or zero. On the other hand, the diamonds are scarce and therefore their relative marginal utility is quite high and this is the reason why their prices are high.

Prof. Samuelson explains this paradox of value in the following words:

"The more there is of a commodity, the less the relative desirability of its last little unit becomes, even though its total usefulness grows as we get more of the commodity." So, it is obvious why a large amount of water has a low price or why air is actually a free good despite its vast usefulness. The many later units pull down the market value of all units.

Besides, the Marshallian concept of consumer's surplus is based on the principle of diminishing marginal utility.

Consumer's Equilibrium: Principle of Equi-marginal Utility

Principle of equi-marginal utility occupies an important place in cardinal utility analysis. It is through this principle that consumer's equilibrium is explained. A consumer has a given income which he has to spend on various goods he wants. Now, the question is how he would allocate his given money income among various goods, that is to say, what would be his equilibrium position in respect of the purchases of the various goods. It may be mentioned here that consumer is assumed to be 'rational', i.e., he carefully calculates utilities and substitutes one good for another so as to maximise his utility or satisfaction.

Suppose there are only two goods X and Y on which a consumer has to spend a given income. The consumer's behaviour will be governed by two factors – first, the marginal utilities of the goods and second, the prices of two goods. Suppose the prices of the goods are given for the consumer.

The law of equi-marginal utility states that the consumer will distribute his money income between the goods in such a way that the utility derived from the last rupee spent on each good is equal. In other words, consumer is in equilibrium position when marginal utility of money expenditure on each good is the same. Now, the marginal utility of money expenditure on a good is equal to the marginal utility of a good divided by the price of the good. In symbols,

$MU_m = MU_x/P_x$

where MU_m is marginal utility of money expenditure, MU_x is the marginal utility of X and P_x is the price of X. The law of equi-marginal utility can therefore be stated thus: the consumer will spend his money income on different goods in such a way that marginal utility of money expenditure on each good is equal, i.e., consumer is in equilibrium in respect of the purchases of two goods X and Y when

$MU_x/P_x = MU_y/P_y$

Now, if MU_x/P_x and MU_y/P_y are not equal and MU_x/P_x is greater than MU_y/P_y, then the consumer will substitute good X for good Y. As a result of this substitution, the marginal utility of good X will fall and marginal utility of good Y will rise. The consumer will continue substituting good X for good Y until MU_x/P_x becomes equal to MU_y/P_y. When MU_x/P_x becomes equal to MU_y/P_y, the consumer will be in equilibrium.

But the equality of MU_x/P_x with MU_y/P_y can be achieved not only at one level but at different levels of expenditure. The question is how far does a consumer go in purchasing the goods he wants. This is determined by the size of his money income. With a given income and money expenditure a rupee has a certain utility for him. This utility is the marginal utility of money to him.

Since the law of diminishing marginal utility applies to money income also, the greater the size of his money income the smaller the marginal utility of money to him. Now, the consumer will go on purchasing goods until the marginal utility of money expenditure on each good becomes equal to the marginal utility of money to him.

Thus, the consumer will be in equilibrium when the following equation holds good:

$MU_x/P_x = MU_y/P_y = MU_m$

where MU_m is marginal utility of money expenditure, i.e., the utility of the last rupee spent on each good).

If there are more than two goods on which the consumer is spending his income, the above equation must hold good for all of them. Thus,

$MU_x/P_x = MU_y/Py = \ldots\ldots = MU_m$

Let us illustrate the law of equi-marginal utility with the aid of an arithmetical table 1.2 given below.

Table 1.2: Marginal Utility of Goods X and Y

Units	MU_x (Utils)	MU_y (Utils)
1	20	24
2	18	21
3	16	18
4	14	15
5	12	9
6	10	3

Let the prices of goods X and Y be ₹ 2 and ₹ 3 respectively. Reconstructing the above table by dividing marginal utilities (MU) of X by ₹ 2 and marginal utilities (MU) of Y by ₹ 3, we get the Table 1.3.

Table 1.3: Marginal Utility of Money Expenditure

Units	MU_x/P_x	MU_y/P_y
1	10	8
2	9	7
3	8	6
4	7	5
5	6	3
6	5	1

Suppose a consumer has money income of ₹ 24 to spend on the two goods. It is worth noting that in order to maximise his utility, the consumer will not equate marginal utilities of the goods because prices of the two goods are different. He will equate the marginal utility of the last rupee (i.e., marginal utility of money expenditure) spent on these two goods.

In other words, he will equate MU_x/P_x with MU_y/P_y while spending his given money income on the two goods. By looking at the Table 1.3, it will become clear that MU_x/P_x is equal to 5 utils when the consumer purchases 6 units of good X and MU_y/P_y is equal to 5 utils when he buys 4 units of good Y. Therefore, consumer will be in equilibrium when he is buying 6 units of good X and 4 units of good Y and will be spending (₹ 2 × 6 + ₹ 3 × 4) = ₹ 24 on them that are equal to consumer's given income. Thus, in the equilibrium position where the consumer maximises his utility,

$MU_x/P_x = MU_y/P_y = MU_m$

$10/2 = 15/3 = 5$

Thus, marginal utility of the last rupee spent on each of the two goods he purchases is the same, i.e., 5 utils.

Consumers' equilibrium is graphically portrayed in Fig. 1.2. Since marginal utility curves of goods slope downward, curves depicting and MU_x/P_x and MU_y/P_y also slope downward. Thus, when the consumer is buying OH of X and OK of Y, then

$MU_x/P_x = MU_y/P_y = MU_m$
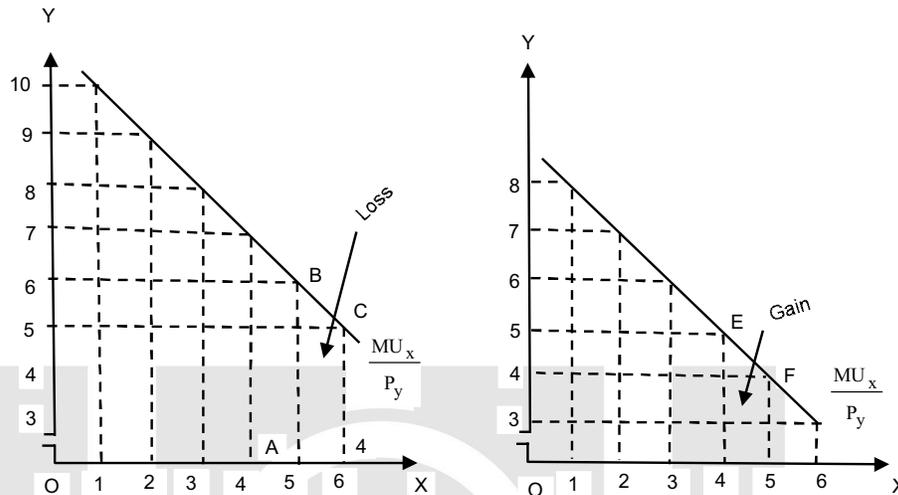
Fig. 1.2: Equi-marginal Utility Principle and Consumer's Equilibrium

Therefore, the consumer is in equilibrium when he is buying 6 units of X and 4 units of Y. No other allocation of money expenditure will yield him greater utility than when he is buying 6 units of commodity X and 4 units of commodity Y. Suppose the consumer buys one unit less of good X and one unit more of good Y.

This will lead to the decrease in his total utility. It will be observed from Fig. 1.2(a) that the consumption of 5 units instead of 6 units of commodity X means a loss in satisfaction equal to the shaded area ABCH, and from Fig. 1.2(b), it will be seen that consumption of 5 units of commodity Y instead of 4 units will mean gain in utility equal to the shaded area KEFL. It will be noticed that with this rearrangement of purchases of the two goods, the loss in utility ABCH exceeds gain in utility KEFL.

Thus, his total satisfaction will fall as a result of this rearrangement of purchases. Therefore, when the consumer is making purchases by spending his given income in such a way that $MU_x/P_x = MU_y/P_y$, he will not like to make any further changes in the basket of goods and will therefore be in equilibrium situation by maximising his utility.

Limitations of the Law of Equi-marginal Utility

Like other laws of economics, law of equi-marginal utility is also subject to various limitations. This law, like other laws of economics, brings out an important tendency among the people. It is not necessary that all people exactly follow this law in the allocation of their money income and therefore all may not obtain maximum satisfaction.

This is due to the following reasons:

1. For applying this law of equi-marginal utility in the real life, consumer must weigh in his mind the marginal utilities of different commodities. For this, he has to calculate and compare the marginal utilities obtained from different commodities.

But it has been pointed out that the ordinary consumers are not so rational and calculating. Consumers are generally governed by habits and customs. Because of their habits and customs, they spend particular amounts of money on different commodities, regardless of whether the particular allocation maximises their satisfaction or not.

2. For applying this law in actual life and equate the marginal utility of the last rupee spent on different commodities, the consumers must be able to measure the marginal utilities of different commodities in cardinal terms. However, this is easier said than done. It has been said that it is not possible for the consumer to measure utility cardinally.

   Being a state of psychological feeling and also there being no objective units with which to measure utility, it is cardinally immeasurable. It is because of the immeasurability of utility in cardinal terms that the consumer's behaviour has been explained with the help of ordinal utility by J.R. Hicks and R.G.D. Allen.

3. Another limitation of the law of equi-marginal utility is found in case of indivisibility of certain goods. Goods are often available in large indivisible units. Because the goods are indivisible, it is not possible to equate the marginal utility of money spent on them. For instance, in allocating money between the purchase of car and foodgrains, marginal utilities of the last rupee spent on them cannot be equated.

An ordinary car costs about ₹ 300,000 and is indivisible, whereas foodgrains are divisible and money spent on them can be easily varied. Therefore, the marginal utility of rupee obtained from cars cannot be equalised with that obtained from foodgrains. Thus, indivisibility of certain goods is a great obstacle in the way of equalisation of marginal utility of a rupee from different commodities.

An illustration: The following table gives an individual's utility schedules for goods X_1 and X_2. If the prices of X_1 and X_2 are ₹ 2.00 each and that the individual has ₹ 20.00 of Income, which she spends on X_1 and X_2, what is the individual's equilibrium purchase of X_1 and X_2?

Q.	1	2	3	4	5	6	7	8	9	10	11
MU_1	16	14	11	10	9	8	7	6	5	3	1
MU_2	15	13	12	8	6	5	4	3	2	1	0

Solution: The individual's equilibrium purchase is given by the conditions $MU_1|P_1 = MU_2|P_2$ and the budget constraint must be fully satisfied from the above table an derive the following:

Q.	1	2	3	4	5	6	7	8	9	10	11
$MU_1|P_1$	8	7	5.5	5	4.5	4	3.5	3	2.5	1.5	0.5
$MU_2|P_2$	7.5	6.5	6	4	3	2.5	2	1.5	1	0.5	0

At $X_1 = 6$

$MU_1|P_1$ is 4

At $X_2 = 4$ units

$MU_2|P_2$ is 4

Hence, $MU_1|P_1 = MU_2|P_2 = 4$

The amount spent is $P_1X_1 + P_2X_2$ which is ₹ 20.00 ($2 \times 6 + 2 \times 4 = 12 + 8$). Many income is also ₹ 20.00. Hence, the budget constraint is satisfied. The equilibrium purchase is $X_1 = 6$ units and $X_2 = 4$ units.

Since MU_1 falls from 16 to 1 as X_1 increase from 1 to 11 and MU_2 defines from 15 to zero as X_2 increase from 1 to 11, the second order condition is fulfilled.

When there are more than an combination of two goods (x_1, x_2) at which the equi-marginal principle holds, one has to take recourse to the budget constraint to obtain the equilibrium combination and all other combinations violating the budget constraint have been rejected.

It should be noticed that when the consumers n goods, the law of equi-marginal utility would then read as:

$MU_1|P_1 = MU_2|P_2 = MU_3|P_3 = ……. = MU_n|P_n = \lambda$(the marginal utility of money).

With the second-order conditions (the law of eventual diministring marginal utility must hold for each of the n goods.)

Further Illustration: What is the maximum total utility, which the consumer derives from consuming 6 units of X_1 and 4 units of X_2?

Remember total utility equal sum of marginal utilities.

From goods X_1 the total utility derived is $16 + 14 + 11 + 10 + 9 + 8$ which equal 68 units of utility. From good X_2 it is $15 + 13 + 12 + 8$, which equal 48. Hence, the maximum total utility derived from consuming 6 units of X_1 and 4 units of X_2 is $(68 + 48)$, which equal 116 units of utility. At the given prices of X_1 and X_2 any other combination of these goods would generate less than 116 units of total utility.

Derivation of Demand Curve and the Law of Demand

We now turn to explain how the demand curve and law of demand is derived in the marginal utility analysis. As stated above, the demand curve or law of demand shows the relationship between price of a good and its quantity demanded. Marshall derived the demand curves for goods from their utility functions.

It should be further noted that in his utility analysis of demand, Marshall assumed the utility functions of different goods to be independent of each other. In other words, Marshallian technique of deriving demand curves for goods from their utility functions rests on the hypothesis of additive utility functions, i.e., utility function of each good consumed by a consumer does not depend on the quantity consumed of any other good.

As has already been noted, in case of independent utilities or additive utility functions, the relations of substitution and complementarity between goods are ruled out. Further, in deriving demand curve or law of demand, Marshall assumes the marginal utility of money expenditure (MU_m) in general to remain constant.

We now proceed to derive demand curve from the law of equi-marginal utility. Consider the case of a consumer who has a certain given income to spend on a number of goods. According to the law of equi-marginal utility, the consumer is in equilibrium in regard to his purchases of various goods when marginal utilities of the goods are proportional to their prices.

Thus, the consumer is in equilibrium when he is buying the quantities of the two goods in such a way that satisfies the following proportionality rule:

$MU_x/P_x = MU_y/P_y = MU_m$

where MU_m stands for marginal utility of money income in general.

With a certain given income for money expenditure, the consumer would have a certain marginal utility of money (MU_m) in general. In order to attain the equilibrium position, according to the above proportionality rule, the consumer will equalise his marginal utility of money (expenditure) with the ratio of the marginal utility and the price of each commodity he buys.

It follows therefore that a rational consumer will equalise the marginal utility of money (MU_m) with MU_x/P_x of good X, with MU_m/P_y of good Y and so on. Given *Ceteris Paribus* assumption, suppose price of good X falls. With the fall in the price of good X, the price of good Y, consumer's income and tastes remaining unchanged, the equality of the MU_x/P_x with MU_y/P_y and MU_m in general would be disturbed.

With the lower price than before, MU_x/P_x will be greater than MU_y/P_y or MU_m (It is assumed of course that the marginal utility of money does not change as a result of the change in the price of one good). Then, in order to restore the equality, marginal utility of X or MU_x must be reduced. And the marginal utility of X or MU_x can be reduced only by the consumer buying more of the good X.

It is thus clear from the proportionality rule that as the price of a good falls, its quantity demanded will rise, other things remaining the same. This will make the demand curve for a good downward sloping. How the quantity purchased of a good increases with the fall in its price and also how the demand curve is derived in the cardinal utility analysis is illustrated in Fig. 1.3.
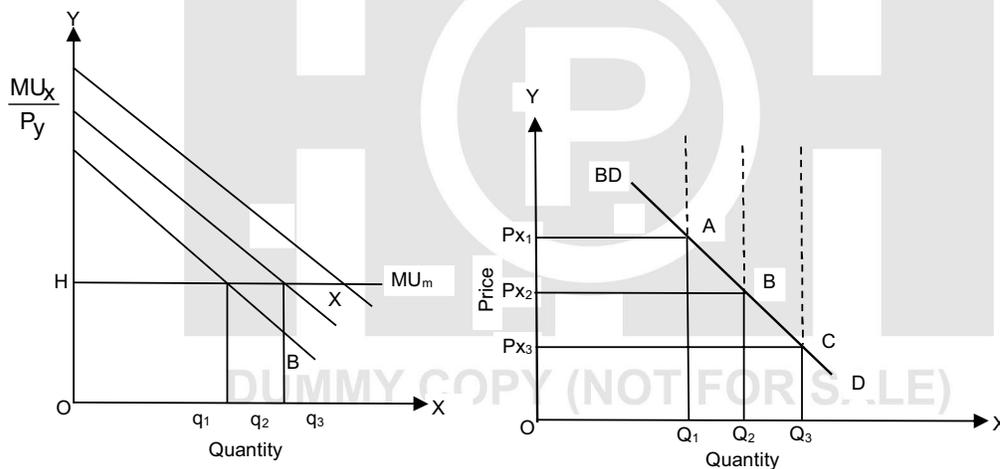
Fig. 1.3: Derivation of Demand Curve

In the upper portion of Fig. 1.3, on the Y-axis MU_x/P_x is shown, and on the X-axis, the quantity demanded of good X is shown. Given a certain income of the consumer, marginal utility of money in general for him is equal to OH. The consumer is buying O_{q1} of good X when price is P_{x1} since at the quantity O_{q1} of X, marginal utility of money OH is equal to MU_x/P_{x1}.

Now, when price of good X falls to P_{x2}. The curve will shift upward to the new position MU_x/P_{x2}. In order to equate marginal utility of money (OH) with the new MU_x/P_{x2}, the consumer increases the quantity demanded to Oq_2. Thus, with the fall in price of good X to P_{x2}, the consumer buys more of it.

It should be noted that no account is taken of the increase in real income of the consumer as a result of fall in price of good X. This is because if change in real income is taken into account, then marginal utility of money will also change and this would have an effect on the purchases of goods. Marginal utility of money can remain constant in two cases. First, when the elasticity of marginal

utility curve (price elasticity of demand) is unity so that even with increase in the purchase of a commodity following the fall in price, the money expenditure made on it remains the same.

Second, marginal utility of money will remain approximately constant for small changes in price of unimportant goods, i.e., goods which account for negligible part of consumer's budget. In case of these unimportant goods, increase in real income following the fall in price is negligible and therefore can be ignored.

At the bottom of Fig. 1.3, the demand curve for X is derived. In this lower panel, price is measured on the Y-axis. As in the upper panel, the X-axis represents quantity. When the price of good X is P_{x1}, the relevant curve of MU/P is MU_x/P_{x1} which is shown in the upper panel. With MU_x/P_{x1}, he buys Oq_1 of good X. Now, in the lower panel, this quantity Oq_1 is directly shown to be demanded at the price P_{x2}.

When price of X falls to P_{x2}, the curve of MU/P shifts upward to the new position MU_x/P_{x2}. With MU_x/P_{x2}, the consumer buys Oq_2 of X. This quantity Oq_2 is directly shown to be demanded at price P_{x2} lower panel. Similarly, by varying price further, we can know the quantity demanded at other prices. Thus, by joining points A, B and C, we obtain the demand curve DD. The demand curve DD slopes downward which shows that as price of a good falls, its quantity purchased rises.

Critical Evaluation of Marshall's Cardinal Utility Analysis

Cardinal utility analysis of demand which we have studied above has been criticised on various grounds.

The following shortcomings and drawbacks of cardinal utility analysis have been pointed out:

1. Cardinal measurability of utility is unrealistic: Cardinal utility analysis of demand is based on the assumption that utility can be measured in absolute, objective and quantitative terms. In other words, it is assumed in this analysis that utility is cardinally measurable. According to this, how much utility a consumer obtains from goods can be expressed or stated in cardinal numbers such as 1, 2, 3, 4 and so forth. But in actual practice utility, cannot be measured in such quantitative or cardinal terms.

Since utility is a psychic feeling and a subjective thing, it cannot be measured in quantitative terms. In real life, consumers are only able to compare the satisfactions derived from various goods or various combinations of the goods. In other words, in the real life, consumer can state only whether a good or a combination of goods gives him more or less, or equal satisfaction as compared to another. Thus, economists like J.R. Hicks are of the opinion that the assumption of cardinal measurability of utility is unrealistic and therefore it should be given up.

2. Hypothesis of independent utilities is wrong: Utility analysis also assumes that utilities derived from various goods are independent. This means that the utility which a consumer derives from a good is the function of the quantity of that good and of that good alone. In other words, the assumption of independent utilities implies that the utility which a consumer obtains from a good does not depend upon the quantity consumed of other goods; it depends upon the quantity purchased of that good alone.

On this assumption, the total utility which a person gets from the whole collection of goods purchased by him is simply the total sum of the separate utilities of various goods. In other words, utility functions are additive.

Neo-classical economists such as Jevons, Menger, Walras and Marshall considered that utility functions were additive. But in the real life, this is not so. In actual life, the utility or satisfaction

derived from a good depends upon the availability of some other goods which may be either substitutes for or complementary with each other. For example, the utility derived from a pen depends upon whether ink is available or not.

On the contrary, if you have only tea, then the utility derived from it would be greater but if along with tea you also have the coffee, then the utility of tea to you would be comparatively less. Whereas pen and ink are complements with each other, tea and coffee are substitutes for each other.

It is thus clear that various goods are related to each other in the sense that some are complements with each other and some are substitutes for each other. As a result of this, the utilities derived from various goods are interdependent, i.e., they depend upon each other. Therefore, the utility obtained from a good is not the function of its quantity alone but also depends upon the existence or consumption of other related goods (complements or substitutes).

It is thus evident that the assumption of the independence of utilities by Marshall and other supporters of marginal utility analysis is a great defect and shortcoming of their analysis. As we shall see below, the hypothesis of independent utilities along with the assumption of constant marginal utility of money reduces the validity of Marshallian demand theorem to the one-commodity model only.

3. Assumption of constant marginal utility of money is not valid: An important assumption of cardinal utility analysis is that when a consumer spends varying amount on a good or various goods or when the price of a good changes, marginal utility of money remains unchanged. But in actual practice, this is not correct. As a consumer spends his money income on the goods, money income left with him declines.

With the decline in money income of the consumer as a result of increase in his expenditure on goods, the marginal utility of money to him rises. Further, when price of a commodity changes, the real income of the consumer also changes. With this change in real income, marginal utility of money will change and this would have an effect on the demand for the good in question, even though the total money income available with the consumer remains the same.

But utility analysis ignores all this and does not take cognizance of the changes in real income and its effect on demand for goods following the change in price of a good. As we shall see below, it is because of the assumption of constant marginal utility of money that Marshall ignored the income effect of the price change which prevented Marshall from understanding the composite character of the price effect, i.e., price effect is the sum of substitution effect and income effect).

Moreover, as we shall see later, the assumption of constant marginal utility of money together with the hypothesis of independent utilities renders the Marshall's demand theorem to be valid in case of one commodity. Further, it is because of the constant marginal utility of money and therefore the neglect of the income effect by Marshall that he could not explain Giffen Paradox.

According to Marshall, utility from a good can be measured in terms of money, i.e., how much money a consumer is prepared to sacrifice for a good). But, to be able to measure utility in terms of money marginal utility of money itself should remain constant. Therefore, assumption of constant marginal utility of money is very crucial to Marshallian demand analysis. On the basis of constant marginal utility of money, Marshall could assert that "utility is not only measurable in principle" but also "measurable in fact".

But, as we shall see below, in case a consumer has to spread his money income on a number of goods, there is a necessity for revision of marginal utility of money with every change in price of a

good. In other words, in a multi-commodity model, marginal utility of money does not remain invariant or constant.

Now, when it is realised that marginal utility of money does not remain constant, then Marshall's belief that utility is 'measurable in fact' in terms of money does not hold good. However, if in marginal utility analysis, utility is conceived only to be 'measurable in principle' and not in fact, then it practically gives up cardinal measurement of utility and comes near to the ordinal measurement of utility.

4. Marshallian demand theorem cannot genuinely be derived except in a one-commodity case: J.R. Hicks and Tapas Majumdar have criticised Marshallian utility analysis on the ground that "Marshallian demand theorem cannot genuinely be derived from the marginal utility hypothesis except in a one-commodity model without contradicting the assumption of constant marginal utility of money. In other words, Marshall's demand theorem and constant marginal utility of money are incompatible except in a one-commodity case. As a result, Marshall's demand theorem cannot be derived in the case when a consumer spends his money on more than one good.

In order to know the truth of this assertion, consider a consumer who has a given amount of money income to spend on some goods with given prices? According to utility analysis, the consumer will be in equilibrium when he is spending money on goods in such a way that the marginal utility of each good is proportional to its price. Let us assume that, in his equilibrium position, consumer is buying q_1 quantity of a good X at a price p_1. Marginal utility of good X, in his equilibrium position, will be equal to its price p_1 multiplied by the marginal utility of money (which, in Marshallian utility analysis, serves as the unit of measurement).

Thus, in the equilibrium position, the following equation will be fulfilled:

$MU_x/ = MU_m \times p_1$

Since the consumer is buying q_1 quantity of good X at price p_1, he will be spending p_1q_1 amount of money on it. Now, suppose that the price of good X rises from p_1 to p_2. With this rise in price of X, all other things remaining the same, the consumer will at once find himself in disequilibrium state, for the marginal of good X will now be less than the higher price multiplied by the marginal utility of money (Mu_m) which is assumed to remain unchanged and constant. Thus, now there will be

$MU_x < MU_y \, p_2$

In order to restore his equilibrium, the consumer will buy less of good X so that the marginal utility of good X (MU_x) would rise and become equal to the product of p_2 and MU_m. Suppose in this new equilibrium position, he is buying q_2 of good X which will be less than q_1. With this, he will now be spending p_2q_2 amount of money on good X. Now, the important thing to see is that whether his new expenditure p_2q_2 on good X is equal to, smaller or greater than p_1q_1.

This depends upon the elasticity of marginal utility curve, i.e., price elasticity of demand. If the elasticity of marginal utility curve of good X is unity, then the new expenditure on good X (i.e., p_2q_2) after the rise in its price from p_1 to p_2 will be equal to the initial expenditure p_1q_1. When the monetary expenditure made on the good remains constant as a result of change in price, then the Marshallian theory is valid.

But constant monetary expenditure following a price change is only a rare phenomenon. However, the Marshallian demand theory breaks down when the new expenditure p_2q_2 after the rise in price, instead of being equal is smaller or greater than the initial expenditure p_2q_2.

If elasticity of marginal utility curve is greater than one, i.e., price demand for the good is elastic), then the new expenditure p_2q_2, after the rise in price from p_1 to p_2, will be less than the initial expenditure p. On the other hand, if the elasticity of marginal utility curve is less than unity, then the new expenditure p_2q_2 after the rise in price will be greater than the initial expenditure p_1q_1.

Now, if the new expenditure p_2q_2 on good X is less than the initial expenditure p_1q_1 then more money will be left with the consumer to spend on goods other than X. And if the new expenditure p_2q_2 on good X is greater than the initial expenditure p_1q_1 on it, then less money would be left with him to spend on goods other than X.

In order that the consumer spends the entire amount of money available with him, then in case of new expenditure p_2q_2 on good X being smaller or greater than initial expenditure p_1q_1 on it, the expenditure or goods other than X and therefore consumer's demand for them will change.

But in Marshallian theoretical framework, this further adjustment in consumer's expenditure on goods other than X can occur only if the unit of utility measurement, i.e., the marginal utility of money revised or changed. But Marshall assumes marginal utility of money to remain constant.

Thus, we see that marginal utility of money cannot be assumed to remain constant when the consumer has to spend his money income on a number of goods. In case of more than one good, Marshallian demand theorem cannot be genuinely derived while keeping the marginal utility of money constant.

If, in Marshallian demand analysis, this difficulty is avoided by giving up the assumption of constant marginal utility of money, then money can no longer provide the measuring rod, and we can no longer express the marginal utility of a commodity in units of money. If we cannot express marginal utility in terms of common numeraire (which money is defined to be), then the cardinality of utility would be devoid of any operational significance.

Only in case there is one good on which the consumer has to spend his money, Marshallian demand theorem can be validity derived. To conclude, in the words of Majumdar, "Except in a strictly one-commodity world, therefore, the assumption of a constant marginal utility of money would be incompatible with the Marshallian demand theorem."

Without the assumption of an invariant unit of measurement, the assertion of measurability would be entirely meaningless. The necessity and the possibility of revision of the unit of utility measurement, following every change in price, had been assumed away in Marshallian theory under the cover of 'other things remaining the same' clause.

5. Cardinal utility analysis does not split up the price affect into substitution and income effects: The third shortcoming of the cardinal utility analysis is that it does not distinguish between the income effect and the substitution effect of the price change.

We know that when the price of a good falls, the consumer becomes better off than before, i.e., a fall in price of a good brings about an increase in the real income of the consumer. In other words, if with the fall in price the consumer purchases the same quantity of the good as before, then he would be left with some income.

With this income, he would be in a position to purchase more of this good as well as other goods. This is the income effect of the fall in price on the quantity demanded of a good. Besides, when the price of a good falls, it becomes relatively cheaper than other goods, and as a result, the consumer is induced to substitute that good for others. This results in increase in quantity demanded of that good. This is the substitution effect of the price change on the quantity demanded of the good.

With the fall in price of a good, the quantity demanded of it rises because of income effect and substitution effect. But cardinal utility analysis does not make clear the distinction between the income and the substitution effects of the price change. In fact, Marshall and other exponents of marginal utility analysis ignored income effect of the price change by assuming the constancy of marginal utility of money. Thus, according to Tapas Majumdar, "the assumption of constant marginal utility of money obscured Marshall's insight into the truly composite character of the unduly simplified price-demand relationship".

They explained the changes in demand as a result of change in the price of a good on the basis of substitution effect on it. Thus, marginal utility analysis does not tell us about how much quantity demanded increases due to income effect and how much due to substitution effect as a result of the fall in price of a good. J.R. Hicks rightly remarks, "that differentiate between income effect and substitution effect of a price change is accordingly left by the cardinal theory as an empty box which is crying out to be filled." In the same way, Tapas Majumdar says, "The efficiency and precision with which the Hicks-Allen approach can distinguish between the income and substitution effects of a price change really leaves the cardinal argument in a very poor state indeed."

6. Marshall could not explain Giffen Paradox: By not visualising the price effect as a combination of substitution and income effects and ignoring the income effect of the price change, Marshall could not explain the Giffen Paradox. He treated it merely as an exception to his law of demand. In contrast to it, indifference curve analysis has been able to explain satisfactorily the Giffen good case.

According to indifference curve analysis, in case of a Giffen Paradox or the Giffen good, negative income effect of the price change is more powerful than substitution effect so that when the price of a Giffen good falls, the negative income effect outweighs the substitution effect with the result that, quantity demanded falls.

Thus, in case of a Giffen good, quantity demanded varies directly with the price and the Marshall's law of demand does not hold good. It is because of the constant marginal utility of money and therefore the neglect of the income effect of price change that Marshall could not explain why the quantity demanded of the Giffen good falls when its price falls and rises when its price rises. This is a serious lacuna in Marshallian's utility analysis of demand.

7. Marginal utility analysis assumes too much and explains too little: Marginal utility analysis is also criticised on the ground that it takes more assumptions and also more severe ones than those of ordinal utility analysis of indifference curve technique. Marginal utility analysis assumes, among others, that utility is cardinally measurable and also that marginal utility of money remains constant. Hicks-Allen's indifference curve analysis does not take these assumptions and even then it is not only able to deduce all the theorems which cardinal utility analysis can but also deduces a more general theorem of demand.

In other words, indifference curve analysis explains not only that much as cardinal utility analysis does but even goes further and that too with fewer and less severe assumptions. Taking less severe assumption of ordinal utility and without assuming constant marginal utility of money, analysis is able to arrive at the condition of consumer's equilibrium, namely, equality o Marginal Rate of Substitution (MRS) with the price ratio between the goods, which is similar to the proportionality rule of Marshall. Further, since indifference curve analysis does not assume constant marginal utility of money, it is able to derive a valid demand theorem in a more than one commodity case.

In other words, indifference curve analysis clearly explains why in case of Giffen goods quality demanded increases with the rise in price and decreases with the fall in price. Indifference curve analysis explains even the case of ordinary inferior goods (other than Giffen goods) in a more analytical manner.

It may be noted that even if the valid demand is derived for the Marshallian hypothesis, it would still be rejected because "better hypothesis" of indifference preference analysis was available which can enunciate more general demand theorem (covering the case of Giffen goods) with fewer, less severe and more realistic assumptions.

Because of the above drawbacks, cardinal utility analysis has been given up in modern economic theory and demand is analysed with new approaches to demand theory.

1.2   INDIFFERENCE CURVE ANALYSIS

Assumptions

1. Rationality: The consumer is assumed to be rational if he aims at the maximisation of his utility, given his income and market prices. It is assumed that he has full knowledge (certainty) of all relevant information.

2. Utility is ordinal: It is taken as axiomatically true that the consumer can rank his preferences (order the various 'baskets of goods') according to the satisfaction of each basket. He need not know precisely the amount of satisfaction. It suffices that he expresses his preference for the various bundles of commodities. It is not necessary to assume that utility is cardinally measurable. Only ordinal measurement is required.

3. Diminishing marginal rate of substitution: Preferences are ranked in terms of indifference curves, which are assumed to be convex to the origin. This implies that the slope of the indifference curves increases. The slope of the indifference curve is called the marginal rate of substitution of the commodities. The indifference curve theory is based, thus, on the axiom of diminishing marginal rate of substitution.

4. The total utility of the consumer depends on the quantities of the commodities consumed.

$U = f(q_1, q_2, \ldots, q_x, q_y, \ldots, q_n)$

5. Consistency and transitivity of choice: It is assumed that the consumer is consistent in his choice, i.e., if in one period he chooses bundle A over B, then he will not choose B over A in another period if both bundles are available to him.

The consistency assumption may be symbolically written as follows:

If A > B, then B > A

Similarly, it is assumed that consumer's choices are characterised by transitivity: if bundle A is preferred to B, and bundle B is preferred to C, then bundle A is preferred to C.

Symbolically, we may write the transitivity assumption as follows:

If A > B, and B > C, then A > C

(a) Equilibrium of the consumer: To define the equilibrium of the consumer i.e., his choice of the bundle that maximises his utility), we must introduce the concept of indifference

curves and of their slope (the marginal rate of substitution), and the concept of the budget line. These are the basic tools of the indifference curve approach.

(b) Indifference curves: An indifference curve is the locus of points – particular combinations or bundles of goods – which yield the same utility (level of satisfaction) to the consumer, so that he is indifferent as to the particular combination he consumes.

An indifference map shows all the indifference curves which rank the preferences of the consumer. Combinations of goods situated on an indifference curve yield the same utility. Combinations of goods lying on a higher indifference curve yield higher level of satisfaction and are preferred. Combinations of goods on a lower indifference curve yield a lower utility.

An indifference curve is shown in Fig. 1.4 and a partial indifference map is depicted in Fig. 1.5. It is assumed that the commodities y and x can substitute one another to a certain extent but are not perfect substitutes.
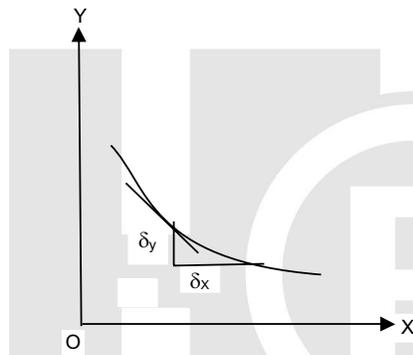
Fig. 1.4: Indifference Curve   **Fig. 1.5: Indifference Map**
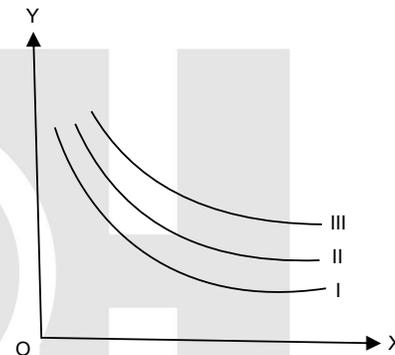
The negative of the slope of an indifference curve at any one point is called the marginal rate of substitution of the two commodities, x and y, and is given by the slope of the tangent at that point.

[Slope of indifference curve] $= -dy/dx = MRS_{x,y}$

The 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 x so that the consumer maintains the same level of satisfaction. With this definition, the proponents of the indifference curve approach thought that they could avoid the non-operational concept of marginal utility.

In fact, what they avoid is the assumption of diminishing individual marginal utilities and the need for their measurement. The concept of marginal utility is implicit in the definition of the MRS, since it can be proved that the marginal rate of substitution (the slope of the indifference curve) is equal to the ratio of the marginal utilities of the commodities involved in the utility function.

$MRS_{x,y} = MU_x/MU_y$ or $MRS_{y,x} = MU_y/MU_x$

Furthermore, the indifference-curves theorists substitute the assumption of diminishing marginal utility with another which may also be questioned, namely the assumption that the indifference curves are convex to the origin, which implies diminishing MRS of the commodities.

Properties of the Indifference Curves

(a) An indifference curve has a negative slope, which denotes that if the quantity of one commodity (y) decreases, the quantity of the other (x) must increase, so that the consumer stays at the same level of satisfaction.

(b) The farther away from the origin an indifference curve lies, the higher the level of utility. It denotes bundles of goods on a higher indifference curve are preferred by the rational consumer.

(c) Indifference curves do not intersect. If they intersect then the point of their intersection would imply two different levels of satisfaction, which is impossible.

Proof: The slope of a curve at any one point is measured by the slope of the tangent at that point. The equation of a tangent is given by the total derivative or total differential, which shows the total change of the function as all its determinants change.

The total utility function in the case of two commodities x and y is

$$U = f(x, y)$$

The equation of an indifference curve is

$$U = f(x, y) = k$$

where k is a constant. The total differential of the utility function is

$$dU = dU/dy \cdot dy + dU/dx \cdot dx = (MU_y)dy + (MU_x)dx$$

It shows the total change in utility as the quantities of both commodities change. The total change in U caused by changes in y and x is (approximately) equal to the change in y multiplied by its marginal utility, plus the change in x multiplied by its marginal utility.

Along any particular indifference curve, the total differential is by definition equal to zero. Thus, for any indifference curve,

$$dU = (MU_y)dy + (MU_x)dx = 0$$

Rearranging, we get

either $-dy/dx = MU_x/MU_y = MRS_{x, y}$ or $-dx/dy = MU_y/MU_x = MRS_{y, x}$

The 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 the marginal rate of substitution of the commodities is diminishing. This axiom is derived from introspection, like the 'law of diminishing marginal utility' of the cardinalist school.

The axiom of decreasing marginal rate of substitution expresses the observed behavioural rule that the number of units of x the consumer is willing to sacrifice in order to obtain an additional unit of y increases as the quantity of y decreases. It becomes increasingly difficult to substitute x for y as we move along the indifference curve. In Fig. 1.8, the fifth unit of y can be substituted for x by the consumer giving up x_1x_2 of x; but to substitute the second unit of y and still retain the same satisfaction, the consumer must give up a much greater quantity of x, namely $x_3 x_4$.

The budget constraint of the consumer

The consumer has a given income which sets limits to his maximising behaviour. Income acts as a constraint in the attempt for maximising utility. The income constraint, in the case of two commodities, may be written as:

$Y = P_x q_x + P_y q_y$ ...(1.1)

We may present the income constraint graphically by the budget line, whose equation is derived from expression 1.1, by solving for q_y

$q_y = 1/P_y\ Y - P_x/P_y\ q_x$

Assigning successive values to q_x (given the income, Y and the commodity prices, P_x, P_y), we may find the corresponding values of q_y. Thus, if $qx = 0$ i.e., if the consumer spends all his income on y), then the consumer can buy Y/P_y units of y. Similarly, if $q_y = 0$ i.e., if the consumer spends all his income on x), then the consumer can buy Y/P_x units of x.

This assumption implies that the commodities can substitute one another, but are not perfect substitutes. If the commodities are perfect substitutes the indifference curve becomes a straight line with negative slope (Fig. 1.6). If the commodities are complements the indifference curve takes the shape of a right angle (Fig. 1.7).
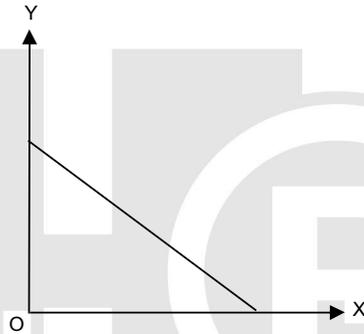
Fig. 1.6: Perfect Substitute   **Fig. 1.7: Complementary Goods**
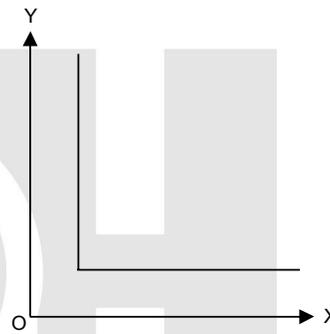
In the first case the equilibrium of the consumer may be a corner solution, that is, a situation in which the consumer spends all his income on one commodity. This is sometimes called 'monomania'. Situations of 'monomania' are not observed in the real world and are usually ruled out from the analysis of the behaviour of the consumer. In the case of complementary goods, indifference curve analysis breaks down, since there is no possibility of substitution between the commodities.
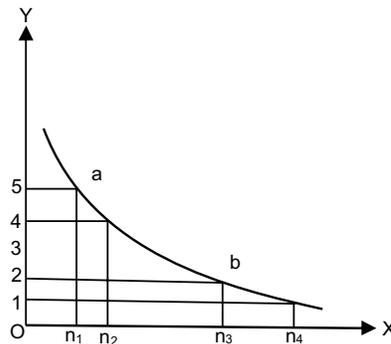
Fig. 1.8: Slope of Budget Line   **Fig. 1.9: Budget Line**
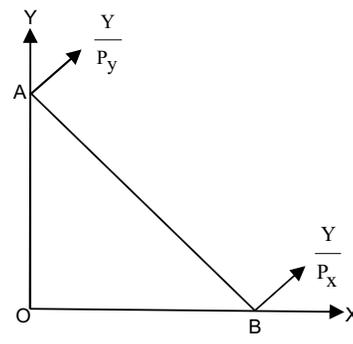
In Fig. 1.9, these results are shown by points A and B. If we join these points with a line we obtain the budget line, whose slope is the ratio of the prices of the two commodities. Geometrically, the slope of the budget line is

$OA/OB = Y/P_y/Y/P_x = P_x/P_y$

Mathematically, the slope of the budget line is the derivative

$\partial q_y/\partial q_x = P_x/P_y$

Derivation of the Equilibrium of the Consumer

The consumer is in equilibrium when he maximises his utility, given his income and the market prices. Two conditions must be fulfilled for the consumer to be in equilibrium.

The first condition is that the marginal rate of substitution be equal to the ratio of commodity prices

$MRS_{x, y} = MU_x/MU_y = P_x/P_y$

This is a necessary but not sufficient condition for equilibrium. The second condition is that the indifference curves be convex to the origin. This condition is fulfilled by the axiom of diminishing $MRS_{x, y}$, which states that the slope of the indifference curve decreases (in absolute terms) as we move along the curve from the left downwards to the right.

Graphical Representation of the Equilibrium of the Consumer

Given the indifference map of the consumer and his budget line, the equilibrium is defined by the point of tangency of the budget line with the highest possible indifference curve (point e in Fig. 1.10).
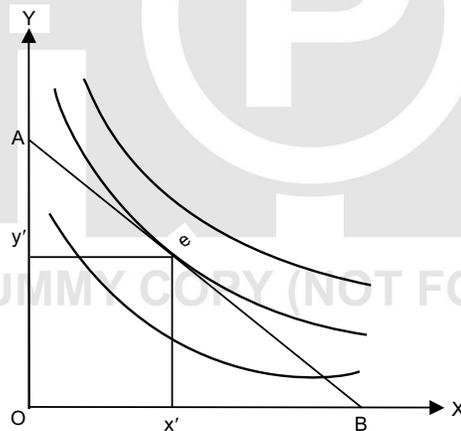
Fig. 1.10: Consumer Equilibrium

At the point of tangency, the slopes of the budget line (P_x/P_y) and of the indifference curve $(MRS_{x, y} = MU_x/MU_y)$ are equal:

$MU_x = MU_y = P_x/P_y$

Thus the first-order condition is denoted graphically by the point of tangency of the two relevant curves. The second-order condition is implied by the convex shape of the indifference curves. The consumer maximises his utility by buying x and y of the two commodities.

Notes

Mathematical Derivation of the Equilibrium

Given the market prices and his income, the consumer aims at the maximisation of his utility. Assume that there are n commodities available to the consumer, with given market prices P_1, P_2, .., P_n. The consumer has a money income (V), which he spends on the available commodities.

Formally, the problem may be stated as follows:

- **Maximise** $U = f(q_1, q_2, \ldots, q_n)$

- **Subject** to $\overset{n}{\underset{i=1}{\mathring{a}}}\ q_i P_i = q_1 P_1, q_2 P_2, \ldots, q_n P_n = Y$

We use the 'Lagrangian multiplier' method for the solution of this constrained maximum. The steps involved in this method may be outlined as follows:

- Rewrite the constraint in the form

    $(q_1 P_1 + q_2 P_2 + \ldots + q_n P_n - Y) = 0$

- Multiply the constraint by a constant A, which is the Lagrangian multiplier

    $\lambda(q_1 P_1 + q_2 P_2 + \ldots + q_n P_n - Y) = 0$

- Subtract the above constraint from the utility function and obtain the 'composite function'

    $\phi = U - \lambda(q_1 P_1 + q_2 P_2 + \ldots + q_n P_n - Y) = 0$

It can be shown that maximization of the 'composite' function implies maximisation of the utility function.

The first condition for the maximisation of a function is that its partial derivatives be equal to zero. Differentiating ϕ with respect to $q_1, \ldots, q_n$ and λ, and equating to zero, we find

$$\delta\phi = \frac{\delta U}{\delta q_1} - \lambda(P_1) = 0$$

$$\frac{\delta\phi}{\delta q_1} = \frac{\delta U}{\delta q_2} - \lambda(P_2) = 0$$

$$. \qquad . \qquad .$$

$$. \qquad . \qquad .$$

$$. \qquad . \qquad .$$

$$\frac{\delta\phi}{\delta q_n} = \frac{\delta U}{\delta q_n} - \lambda(P_n) = 0$$

$$\frac{\delta\phi}{\delta\lambda} = -(q_1 P_1 + q_2 P_2 + \ldots + q_n P_n - Y) = 0$$

From the equations, we get

$$\frac{\delta U}{\delta q_1} = \lambda P_1$$

$$\frac{\delta U}{\delta q_2} = \lambda P_2$$

$$. \qquad .$$

$$\frac{\delta U}{\delta q_n} = \lambda P_n$$

But,

$$\frac{\delta U}{\delta q_1} = MU_1, \ \frac{\delta U}{\delta q_2} = MU_2, \ \ldots \ \frac{\delta U}{\delta q_n} = MU_n$$

Substituting, and solving for λ we get

$$\lambda = \frac{MU_1}{P_1} = \frac{MU_2}{P_2} = \ldots = \frac{MU_n}{P_n}$$

Alternatively, we may divide the preceding equation corresponding to commodity x, by the equation which refers to commodity y, and get

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

We observe that the equilibrium conditions are identical in the cardinalist approach and in the indifference curve approach. In both theories, we have

$$\frac{MU_1}{P_1} = \frac{MU_2}{P_2} = \ldots = \frac{MU_x}{P_x} = \frac{MU_y}{P_y} = \ldots = \frac{MU_n}{P_n}$$

Thus, although in the indifference curve approach, cardinality of utility is not required, the MRS requires knowledge of the ratio of the marginal utilities, given that the first-order condition for any two commodities may be written as

$$MU_x/MU_y = P_x/P_y = MRS_{x,y}$$

Hence, the concept of marginal utility is implicit in the definition of the slope of the indifference curves, although its measurement is not required by this approach. What is needed is a diminishing marginal rate of substitution, which of course does not require diminishing marginal utilities of the commodities involved in the utility function.

Derivation of the Demand Curve Using the Indifference Curve Approach

Graphical Derivation of the Demand Curve

As the price of a commodity, for example of x, falls, the budget line of the consumer shifts to the right, from its initial position (AB) to a new position (AB) due to the increase in the purchasing power of the given money income of the consumer. With more purchasing power in his possession, the consumer can buy more of x (and more of y). The new budget line is tangent to a higher indifference curve (e.g., curve II). The new equilibrium occurs to the right of the original equilibrium (for normal goods) showing that as price falls, more of the commodity will be bought.

If we allow the price of x to fall continuously and we join the points of tangency of successive budget lines and higher indifference curves, we form the so-called price-consumption line (Fig. 1.11), from which we derive the demand curve for commodity x. At point e_1, the consumer buys quantity x_1 at price y_1. At point e_2 the price, y_2, is lower than y_1, and the quantity demanded has increased to x_2, and so on. We may plot the price-quantity pairs defined by the points of equilibrium (on the price-consumption line) to obtain a demand curve, as shown in Fig. 1.12.
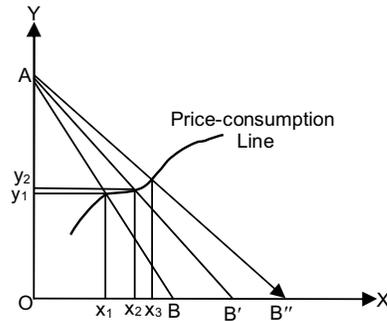
Fig. 1.11: Price-consumption Line
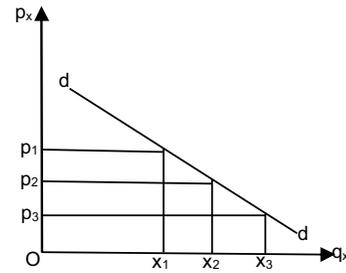
Fig. 1.12: Demand Curve

The demand curve for normal commodities will always have a negative slope, denoting the 'law of demand' (the quantity bought increases as the price falls). In the indifference curve approach, the 'law of demand' is derived from what is known as Slutsky's theorem, which states that the substitution effect of a price change is always negative (relative to the price if the price increases, the quantity demanded decreases and *vice versa*). The formal proof of Slutsky's theorem involves sophisticated mathematics. However, we may show graphically the implications of this theorem.

We saw that a fall in the price of x from P_1 to P_2 resulted in an increase in the quantity demanded from x_1 to x_2. This is the total price effect which may be split into two separate effects, a substitution effect and an income effect. The substitution effect is the increase in the quantity bought as the price of the commodity falls, after 'adjusting' income so as to keep the real purchasing power of the consumer the same as before.

This adjustment in income is called compensating variation and is shown graphically by a parallel shift of the new budget line until it becomes tangent to the initial indifference curve (Fig. 1.13). The purpose of the compensating variation is to allow the consumer to remain on the same level of satisfaction as before the price change. The compensated budget line will be tangent to the original indifference curve (I) at a point (e'_1) to the right of the original tangency (e_1), because this line is parallel to the new budget line which is less steep than the original one when the price of x falls. The movement from point e_1 to e'_1 shows the substitution effect of the price change. The consumer buys more of x now that it is cheaper, substituting y for x.
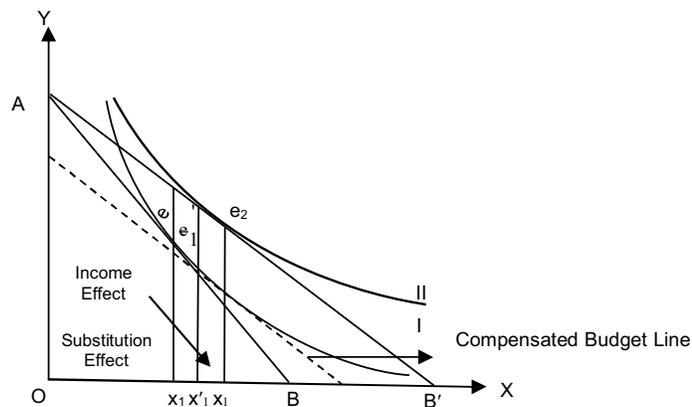
Fig. 1.13: Compensating Variation (Compensated Budget Line)

However, the compensating variation is a device which enables the isolation of the substitution effect, but does not show the new equilibrium of the consumer. This is defined by point e_2 on the higher indifference curve II. The consumer has in fact a higher purchasing power, and, if the commodity is normal, he will spend some of his increased real income on x, thus moving from x'_1 to x_2. This is the income effect of the price change.

The income effect of a price change is negative for normal goods and it reinforces the negative substitution effect (Fig. 1.14). If, however, the commodity is inferior, the income effect of the price change will be positive: as the purchasing power increases, less of x will be bought. Still for most of the inferior goods, the negative substitution effect will more than offset the positive income effect, so that the total price effect will be negative. Thus, the negative substitution effect is in most cases adequate for establishing the law of demand.

(It is when the income effect is positive and very strong that the 'law of demand' does not hold. This is the case of the Giffen goods, which are inferior and their demand curve has a positive slope. Giffen goods are very rare in practice.)

It should be noted that although Slutsky's theorem can be proved mathematically, its proof is based on the axiomatic assumption of the convexity of the indifference curves.

Mathematical derivation of the demand curve. The demand curve may be derived from the equilibrium condition

$$\frac{MU_x}{P_x} = \frac{MU_y}{P_y} = \ldots = \frac{MU_n}{P}$$

and the budget constraint

$$Y = \sum_{i=1}^{n} P_l q_l$$

For example, assume that there are only two commodities and that the total utility function is multiplicative of the form

$$U = \frac{1}{4} q_x q_y$$

The marginal utility of x and y are

$$MU_x = \frac{dU}{dq_x} = \frac{1}{4} q_y$$

and $MU_y = \dfrac{\delta U}{\delta q_y} = \dfrac{1}{4} q_x$

Substituting the marginal utilities with equilibrium condition, we get

$$\frac{(1/4)q_y}{P_x} = \frac{(1/4)q_x}{P_y}$$

or $q_y P_y = q_x P_x$

Note: That the equality of expenditures of the two commodities is not a general rule; the expenditures depend on the specific form of the utility function.

We may derive the demand for commodity x by substituting $q_y P_y$ with budget constraint.

$$q_y P_y + q_x P_x = Y$$

$$\Rightarrow 2q_x P_x = Y$$

$$\Rightarrow q_x = \frac{1}{2P_x} Y$$

Thus, the demand for x is negatively related to its own price P_x and positively to income Y.

Similarly, the demand for Y is obtained by substituting $q_x P_x$ in the budget constraint

$$q_y = 1/2P_y \, Y$$

In our particular example, the demand curves are symmetric due to the particular multiplicative form of the consumer's utility function which we assumed.

Criticism of the Indifference Curve Approach

The indifference-curves analysis has been a major advance in the field of consumer's demand. The assumptions of this theory are less stringent than for the cardinal utility approach. Only ordinality of preferences is required, and the assumption of constant utility of money has been dropped.

The methodology of indifference curves has provided a framework for the measurement of the 'consumer's surplus', which is important in welfare economics and in designing government policy.

Perhaps, the most important theoretical contribution of this approach is the establishment of a better criterion for the classification of goods into substitutes and complements. Earlier theorists were using the total effect of a price change for this purpose, without compensating for the change in real income. The classification was based on the sign of the cross-elasticity of demand.

$$e_{yx} = \partial q_y / \partial P_x . \, P_x / q_y$$

Where the total change in the quantity of y was considered as a result of a change in the price of x. A positive sign of the cross-elasticity implies that x and y are substitutes; a negative sign implies that the commodities are complements. This approach may easily lead to absurd classifications if the change in the price of x is substantial.

For example, if the price of beef is halved, it is almost certain that both the consumption of beef and of pork will be increased, due to the increase of the real income of the consumer. This would imply a negative cross-elasticity for pork, and hence pork would be classified as a complementary commodity to beef.

Hicks suggested measuring the cross-elasticity after compensating for changes in real income. Thus, according to Hicks, goods x and y are substitutes if, after compensating for the change in real income (arising from the change in the price of x) a decrease in the price of x leads to a decrease in the quantity demanded of y.

Although this criterion is theoretically more correct than the usual approach based on the total change in the quantity of y as a result of a change in the price of x, in practice, its application is impossible because it requires knowledge of the individual preference functions, which cannot be statistically estimated. On the other hand, the usual approach of the total price effect is feasible because it requires knowledge of the market demand functions which can be empirically estimated.

Although the advantages of the indifference curve approach are important, the theory has indeed its own severe limitations. The main weakness of this theory is its axiomatic assumption of the existence and the convexity of the indifference curves. The theory does not establish either the

existence or the shape of the indifference curves. It assumes that they exist and have the required shape of convexity.

Furthermore, it is questionable whether the consumer is able to order his preferences as precisely and rationally as the theory implies. Also, the preferences of the consumers change continuously under the influence of various factors, so that any ordering of these preferences, even if possible, should be considered as valid for the very short run. Finally, this theory has retained most of the weaknesses of the cardinalist school with the strong assumption of rationality and the concept of the marginal utility implicit in the definition of the marginal rate of substitution.

Another defect of the indifference curve approach is that it does not analyse the effects of advertising, of past behaviour (habit persistence), of stocks and of the interdependence of the preferences of the consumers, which lead to behaviour that would be considered as irrational, and hence is ruled out by the theory. Furthermore speculative demand and random behaviour are ruled out these factors are very important for the pricing and output decisions of the firm.

1.3   SLUTSKY EQUATION

The Slutsky Substitution Effect

The concept of substitution effect was put forward by J.R. Hicks. There is another important version of substitution effect put forward by E. Slutsky. The treatment of the substitution effect in these two versions has a significant difference.

Since Slutsky substitution effect has an important empirical and practical use, we explain below Slutsky's version of substitution effect in some detail.

In Slutsky's version of substitution effect when the price of good changes and consumer's real income or purchasing power increases, the income of the consumer is changed by the amount equal to the change in its purchasing power which occurs as a result of the price change. His purchasing power changes by the amount equal to the change in the price multiplied by the number of units of the good which the individual used to buy at the old price.

In other words, in Slutsky's approach, income is reduced or increased (as the case may be), by the amount which leaves the consumer to be just able to purchase the same combination of goods, if he so desires, which he was having at the old price.

That is, the income is changed by the difference between the cost of the amount of good X purchased at the old price and the cost of purchasing the same quantity of X at the new price. Income is then said to be changed by the cost difference. Thus, in Slutsky substitution effect, income is reduced or increased not by compensating variation as in case of the Hicksian substitution effect but by the cost difference.

Slutsky Substitution Effect for a Fall in Price

Slutsky substitution effect is illustrated in Fig. 1.14. With a given money income and the given prices of two goods as represented by the price line PL, the consumer is in equilibrium at Q on the indifference curve IC_1 buying OM of X and ON of Y.

Now, suppose that price of X falls, price of Y and money income of the consumer remaining unchanged. As a result of this fall in price of X, the price line will shift to PL′ and the real income or the purchasing power of the consumer will increase.

Now, in order to find out the Slutsky substitution effect, consumer's money income must be reduced by the cost difference or, in other words, by the amount which will leave him to be just able to purchase the old combination Q, if he so desires.
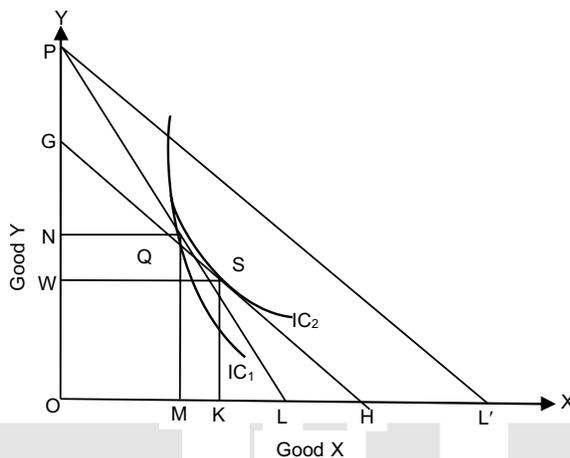
Fig. 1.14: Slutsky Substitution Effect
(for a fall/decline in Price)

A price line GH parallel to PL′ has been drawn which passes through the point Q. It means that income equal to PG in terms of Y or L′H in terms of X has been taken away from the consumer, and as a result, he can buy the combination Q, if he so desires, since Q also lies on the price line GH.

Actually now, he will not buy the combination Q since X has now become relatively cheaper and Y has become relatively dearer than before. The change in relative prices will induce the consumer to rearrange his purchases of X and Y. He will substitute X for Y.

But in this Slutsky substitution effect, he will not move along the same indifference curve IC_1, since the price line GH, on which the consumer has to remain due to the new price-income circumstances is nowhere tangent to the indifference curve IC_1.

The price line GH is tangent to the indifference curve IC_2 at point S. Therefore, the consumer will now be in equilibrium at a point S on a higher indifference curve IC_2. This movement from Q to S represents Slutsky substitution effect according to which the consumer moves not on the same indifference curve, but from one indifference curve to another.

A noteworthy point is that movement from Q to S as a result of Slutsky substitution effect is due to the change in relative prices alone, since the effect due to the gain in the purchasing power has been eliminated by making a reduction in money income equal to the cost difference.

At S, the consumer is buying OK of X and OW of Y; MK of X has been substituted for NW of Y. Therefore, Slutsky substitution effect on X is the increase in its quantity purchased by MK and Slutsky substitution effect on Y is the decrease in its quantity purchased by NW.

Slutsky Substitution Effect for a Rise in Price

We have graphically explained above Slutsky substitution effect for a fall in price of good X. It will be instructive to explain it also for a rise in price of X. This is demonstrated in Fig. 1.15. Initially, the consumer is in equilibrium at point Q on the indifference curve IC_1, prices of the two goods and

his money income being given. Now, suppose that price of good X rises, price of Y remaining unchanged.

As a result of the rise in price of X, budget line will shift downward to PL″ and consumer's real income or purchasing power of his given money income will fall. Further, with this price change, good X has become relatively dearer and good Y relatively cheaper than before.

In order to find out Slutsky substitution effect in this present case, consumer's money income must be increased by the 'cost difference' created by the price change to compensate him for the rise in price of X. In other words, his money income must be increased to the extent which is just large enough to permit him to purchase the old combination Q, if he so desires, which he was buying before.

For this, a budget line GH has been drawn which passes through point Q. It will be evident from the figure that, PG (in terms of Y) or L″H (in terms of X) represents 'cost difference' in this case. With budget line GH, he can buy if he so desires the combination Q, which he was buying at the previous price of X.
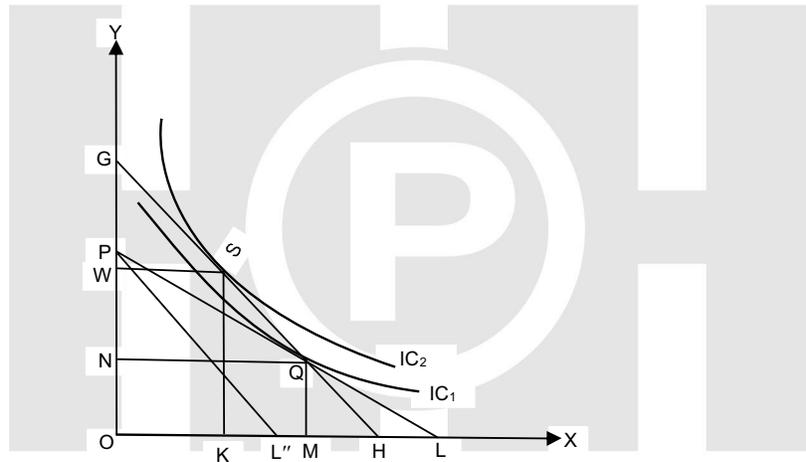
Fig. 1.15: Slutsky Substitution Effect (for a rise in price)

But actually, he will not buy combination Q, since on budget line GH, X is relatively dearer than before, he will therefore replace some X by Y (i.e., he will substitute of Y for X). As shown in Fig. 1.15 with budget line GH, he is in equilibrium position at S on a higher indifference curve of IC$_2$ and is buying OK of X and OW of Y.

MK of X has been replaced by AW of Y. Movement from point Q to S is the result of Slutsky substitution effect; the effect due to the fall in purchasing power has been cancelled by giving him money equal to PG of Y or L″H of X. In this present case of stipulated rise in price of X, Slutsky substitution effect on X is the fall in its quantity brought by MK and Slutsky substitution effect on Y is the increase in its quantity brought by NW.

From the above analysis, it is clear that whereas Hicks-Allen substitution effect takes place on the same indifference curve, Slutsky substitution effect involves the movement from one indifference curve to another curve, a higher one.

The difference between the two versions of the substitution effect arises solely due to the magnitude of money income by which income is reduced or increased to compensate for the change in income. The Hicksian approach just restores to the consumer his initial level of satisfaction,

whereas the Slutsky approach "overcompensates" the consumer by putting him on a higher indifference curve.

Merits and Demerits of Hicksian and Slutsky Methods

Prof. J.R. Hicks points out that the method of adjusting the level of money income by the compensating variation has the merit that on this interpretation, the substitution effect measures the effect of change in relative price. With real income constant, the income effect measures the effect of the change in real income. Thus, the analysis which is based upon the compensating variation is a resolution of the price change into two fundamental economic 'directions'. We shall not encounter a more fundamental distinction upon any other route.

But Slutsky method has a distinct advantage in that it is easier to find out the amount of income equal to the 'cost difference' by which income of the consumer is to be adjusted. On the other hand, it is not so easy to know the compensating variation in income.

Thus, the cost-difference method has the advantage of being dependent on observable market data, while for knowing the amount of compensating variation in income, knowledge of indifference curves i.e., tastes and preferences of the consumer between various combinations of goods is required.

It follows from what has been said above that both the cost difference and compensating variation methods have their own merits. While the law of demand can be easily and adequately established by the method of cost difference, method of compensating variation is very useful for the analysis of consumer's surplus and welfare economics.

With the help of the cost difference, the income effect can be easily separated from the substitution effect but the substitution effect so found out involves some gain in real income (since it causes movement from a lower indifference curve to a higher indifference curve). It is because of this that, on cost difference method, substitution effect is not a theoretically distinct concept.

Numerical Examples

Let us explain the concept of cost difference and Slutsky substitution effect with a numerical example stated below:

When the price of petrol is ₹ 20.00 per litre, Amit consumes 1,000 litres per year. The price of petrol rises to ₹ 25.00 per litre. Calculate the cost difference equal to which the Government should give him extra money income per year to compensate him for the rise in price of petrol. Will Amit be better off or worse off after the price rise plus the cash compensation equal to the cost difference than he was before? What will happen to petrol consumption?

As explained above, the cost difference is equal to AP.Q where AP stands for the change in price of a good and Q stands for the quantity of commodity he was consuming prior to the change in price. Thus, in our above example,

ΔP = ₹ 25 – 20 = ₹ 5

Q = 1,000 litres per year

Cost difference = $\Delta P.Q$ = ₹ 5 × 1,000 = ₹ 5,000

Now, with higher price of petrol of ₹ 25.00 per litre and cash compensation of ₹ 5,000 equal to the cost difference, he can buy, if he so desires, the original quantity of 1,000 litres of petrol per year. However, he may not buy this original quantity of petrol in the new price-income situation if his

satisfaction is maximum at some other point. Consider Fig. 1.16 where we measure petrol on the X-axis and money income representing other goods on the f-axis.

Suppose, BL₁ is the initial budget line when price of petrol is ₹ 20.00 per litre and consumer is in equilibrium at point Q on the indifference curve IC1 where he is consuming 1,000 litres of petrol per year. Now, with the rise in price of petrol to ₹ 25.00 per litre, suppose the budget line shifts to BL₂.

Now, if to compensate for the rise in price, his, money income is raised by ₹ 5,000, i.e., equal to the cost difference, then the budget line shifts in a parallel manner to the left so that it reaches the position GH which passes through the original point of consumption Q.

As shown in Fig. 1.16 will reveal that the consumer with higher price of petrol and having received monetary compensation equal to the cost difference of ₹ 5,000 will not be in equilibrium at the original point Q and instead he will maximise his satisfaction in the new situation at point S on a higher indifference curve IC₂ where his consumption of petrol has decreased to ON litres i.e., the decrease in consumption of petrol by MN is the Slutsky substitution effect). Since, with the rise in price and simultaneous increase in his income equal to the cost difference has enabled him to attain a higher indifference curve, he has become better off than before the rise in price.
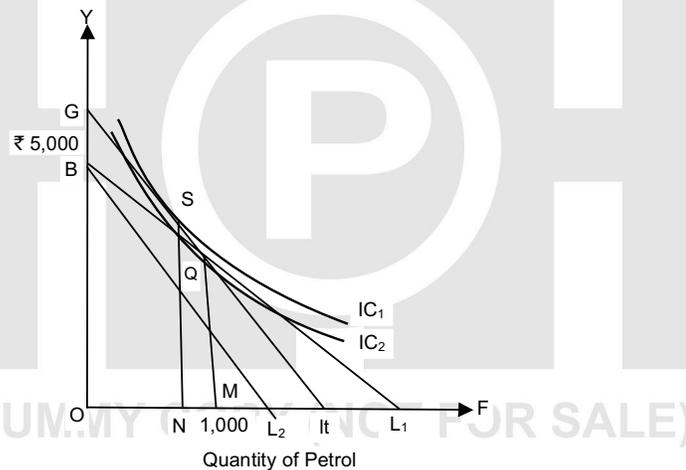
Fig. 1.16: Cash Compensation for a Rise in Price of Petrol

Price Effect Broken Up into Income and Substitution Effects: Slutsky Method

In our discussion of substitution effect, we explained that Slutsky presented a slightly different version of the substitution and income effects of a price change from the Hicksian one. In fact, it was Slutsky who first of all divided the price effect into income and substitution effects. His way of breaking up the price effect is shown in Fig. 1.17. With a certain price-income situation, the consumer is in equilibrium at Q on indifference curve IC₁.
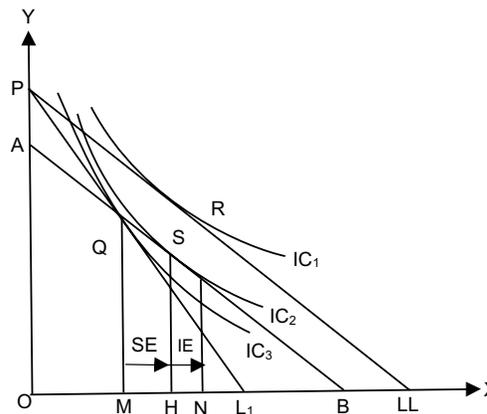
**Fig. 1.17: Price Effect is Decomposed into Substitution and Income Effects
with Slutsky's Cost-difference Method**

With a fall in price of X, other things remaining the same, budget line shifts to PL_2. With budget line PL_2, the consumer would now be in equilibrium at R on the indifference curve IC_3. This movement from Q to R represents the price effect.

As a result of this, he buys MN quantity of good X more than before. Now, in order to find out the substitution effect, his money income be reduced by such an amount that he can buy, if he so desires, the old combination Q.

Thus, a line AB, which is parallel to PL_2, has been so drawn that it passes through point Q. Thus, PA in terms of good Y represents the cost difference. With budget line AB, the consumer can have combination Q if he so desires, but actually he will not buy combination Q because X is now relatively cheaper than before. It will pay him to substitute X for Y.

With budget line AB, he is in equilibrium at S on indifference curve IC_2. The movement from Q to S represents Slutsky substitution effect which induces the consumer to buy MH quantity more of good X. If now the money taken away from him is restored to him, he will move from S on indifference curve IC_2 to R on indifference curve IC_3.

This movement from S to R represents income effect. Thus, movement from Q to R as a result of price effect can be divided into two steps. First, movement from Q to S as a result of substitution effect and secondly, movement from S to R as a result of income effect.

It may be pointed out here again that, unlike the Hicksian method, Slutsky substitution effect causes movement from a lower indifference curve to a higher one. While separately discussing substitution effect above, we pointed out the merits and demerits of the Hicksian and Slutskian methods of breaking up the price effect.

Slutsky Equation

We have graphically shown above how the effect of change in price of a good can be broken up into its two component parts, namely, substitution effect and income effect. The decomposition of price effect into its two components can be derived and expressed mathematically.

Suppose price of good X falls, its substitution effect on quantity demanded of the good arises due to substitution of the relatively cheaper good X for the now relatively dearer good Y and as a result in the Hicksian method, the consumer moves along the same indifference curve so that his level of utility remains constant.

The overall effect of change in its own price on the quantity demanded can be expressed as dq_x/dp_x and the substitution effect can be expressed $\partial p_x/\partial p_x|_{u=\bar{u}}$. The term $\partial q_x/\partial p_x|_{u=\bar{u}}$ shows change in quantity demanded resulting from a relative change in price of X while utility or satisfaction of the consumer remains constant.

However, expressing income effect of the price change mathematically is rather a difficult thing. Suppose a unit change in income (∂I) causes a (∂q_x) change in quantity demanded of the good. This can be written as $\partial q_x/\partial I$. But how much income changes due to a change in price of the good is determined by how much quantity of the good (q_x) the consumer was purchasing on the one hand and change in price of the good (∂p_x) that has taken place on the other. The change in income due to a change in price can be measured by $q_x(\partial p_x)$. How much this change in income will affect the quantity demanded of the good X is determined by $\partial q_x/\partial I$ which shows the effect of a unit change in income on the quantity demanded of the good X.

Thus, the overall effect of change in price of the good X on its quantity demanded can be expressed by the following equation which is generally called Slutsky equation because it was Russian economist E. Slutsky who first of all divided the price effect into substitution effect and income effect.

$$\partial q_x/\partial p_x = \partial q_x/\partial p_x|_{u=\bar{u}} + q_x.\partial p_x.\partial q_x/\partial I$$

The first term on the right hand side of the equation represents the substitution effect obtained after income of the consumer has been adjusted to keep his level of utility constant. The second term on the right hand side of the equation shows the income effect of the fall in price of the good. The term $q_x.\partial P_x$ measures the increase in income or purchasing power caused by the fall in price and $\partial q_x/\partial I$ measures the change in quantity demanded resulting from a unit increase in income (I). Therefore, income Affect of the price change is given by $q_x.\partial p_x.\partial q_x/\partial I$.

Since the fall in price increases income or purchasing power of the consumer which in case of normal goods leads to the increase in quantity demanded of the good, sign of the income effect has been taken to be positive.

Further, a point needs to be clarified. In the above analysis of Slutsky equation, we have considered the substitution effect when with a change in price, the consumer is so compensated as to keep his real income or purchasing power constant. In obtaining Slutsky substitution effect, income of the consumer is adjusted to keep his purchasing power (i.e., real income) constant so that he could buy the original combination of goods if he so desires. On the other hand, in the Hicksian substitution effect, with a change in price of a good, money income with the consumer is so adjusted that his satisfaction remains constant.

In fact, Hicks interprets real income in terms of satisfaction obtained by a consumer. This difference was later emphasised by J.R. Hicks, but since it was Slutsky who first of all split up the price effect into substitution effect and income effect, the above equation is popularly known as Slutsky equation. It is proper to call it Slutsky-Hicks equation.

$$\partial q_x/\partial p_x = \partial q_x/\partial p_x|_{u=\bar{u}} + q_x.\partial p_x.\partial q_x/\partial I \qquad\qquad ...(1.2)$$

An important result follows from the Slutsky equation. If the commodity is a normal good, then $\partial q_x/\partial I$ is positive by definition. It follows that a fall in price will lead to the increase in income causing increase in quantity demanded of the good and therefore the expression for income effect of the price change $q_x.\partial p_x.(\partial q_x/\partial I)$ is taken to be positive in the Slutsky equation (1.2) above. Besides, since the substitution effect is always negative, a fall in the relative price of a good will cause the

increase in its quantity demanded. Therefore, Slutsky equation tells us that when commodity X is normal, the price effect dq_x/dp_x is necessarily negative implying that fall in price will cause quantity demanded of the good to increase. Thus, in case of normal goods, both the substitution effect and income effect work in the same direction and reinforce each other. Thus, in case of normal goods a fall in price of a commodity leads to the increase in quantity demanded due to both the substitution effect and income effect.

On the other hand, if price of the commodity rises, then due to the negative substitution effect, the consumer will buy less of the good, his purchasing power remaining the same. Therefore, in case of rise in price of a good, the first term in the right side of Slutsky equation, namely, $\partial q_x/\partial p_x|_{u=\bar{u}}$ will have a negative sign. Further, rise in price of a good causes income of the consumer to fall, and income effect will lead to the decrease in quantity demand of good and therefore, the second term $(q_x . \partial p_x . \partial q_x/\partial I)$ on the right hand side of the equation will have a negative sign in case of normal goods. Thus, in case of rise in price of a good, both the substitution effect and also income effect (if it is a normal good) will work in the same direction to reduce the quantity demanded of the good whose price rises.

The second important conclusion which follows from Slutsky equation is that as the quantity of commodity (q_x) consumed becomes smaller and smaller, the income effect of the price change will become smaller and smaller. Thus, if the quantity consumed of a commodity is very small, then the income effect is not very significant.

First consider the following optimization problem and its comparative statics:

Maximize $U(x_1, x_2)$

With respect to x_1 and x_2

Subject to the constraint that

$p_1 x_1 + p_2 x_2 = y$

The first order conditions for the maximizing values of x_1, and x_2 are:

$\delta u | \delta x_1 - \lambda P_1 = 0$

$\delta u | \delta x_2 - \lambda P_2 = 0$

when λ_2 Lagrangian multiplier.

Comparative Statics

The differentiation of the first order equations with respect to p_1 yields

$u_{1,1} \delta x_1 | \delta p_1 + u_{1,2} \delta x_2 | \delta p_1 - p_1 \delta\lambda | \delta p_1 = \lambda$

$u_{2,1} \delta x_1 | \delta p_1 + u_{2,2} \delta p_2 | \delta p_1 - p_2 \delta\lambda | \delta p_1 = 0$

where, $u_{i,j}$ is $\delta^2 u | \delta x_j \delta x_i$

The differentiation of the budget constraint with respect to p_1 yields an equation that can be put into the form:

$- p_1(\delta x_1 | \delta p_1) - p_2(\delta x_2 | \delta p_1) = x_1$

These equations in matrix form are:

$$| \ u_{1,1} u_{1,2} - p_1 \ | \ | \ \delta x_1 | \ \delta p_1 \ | \qquad | \ \lambda \ |$$

$$| \ u_{2,1} u_{2,2} - p_1 \ | \ | \ \delta x_2 \ | \ \delta p_1 \ | \ = \ | \ 0 \ |$$

$$| - p_1 - p_2 \ 0 \ | \ | \ \delta \lambda \ | \ \delta p_1 \ | \qquad | \ x_1 \ |$$

Let the 3×3 matrix on the left, which happens to be called a bordered Hessian matrix, be denoted an H. The column vector of the x_1, x_2 and λ values can be represented as partial derivatives with respect to p_1 and are denoted by

$\delta x \mid \delta p_1$. The matrix equation then become

$H(\delta x \mid \delta p_1) = (\lambda, 0, x_1)^T$

Where the column vector on the right is represented in the form of a transpose of a row vector.

Thus, the solution is $\delta x \mid \delta p_1 = H^{-1}(\lambda, 0, x_1)^T$

Which can be, decomposed into two terms, i.e.,

$\delta x \mid \delta p_1 = H^{-1}(\lambda, u, 0)^T + H^{-1}(u, 0, x_1)^T$

There two terms represent the substitution effect and the income effect, respectively. This assertion is to be proven.

The Income Effect

Where the first order conditions are differentiated with respect to income y the result is:

$$| \ u_{1,1} u_{1,2} - p_1 \ | \ | \ \delta x_1 \mid \delta y \ | \ = \ | \ 0 \ |$$

$$| \ u_2, u_{2,2} - p_1 \ | \ | \ \delta x_1 \mid \delta y_1 \ | \ = \ | \ 0 \ |$$

$$| -p_1 - p_2 \ 0 \ | \ | \ \delta x \mid \delta y \ | \qquad \cdot \quad | -1 \ |$$

Thus,

$\delta x \mid \delta y = H^{-1} (0, 0, -1)^T$

The second term in the equation for the effect of a change in p_1,

$H^{-1}(u, 0, x_1)^T$

can be respected as

$-x_1 H^{-1} (0, 0, -1)^T$

and hence as $-x_1(\delta x \mid \delta y)$, then this term is the income effect. To establish the substitution effect another optimization problem has to be considered.

Minimizing the Cost of Achieving a given Utility Level

The optimization problem is:

Minimize $C = p_1 x_1 + p_2 x_2$

with respect to x_1 and x_2 subject to the constraint $u(x_1, x_2) = u_0$

The first order conditions for this optimization problem are:

$p_1 - (1/\mu)\delta u/\delta x_1 = 0$

$p_2 - (1/\mu)\delta u/\delta x_2 = 0$

Where the Lagrangian multiplier has been expressed as $(1 \mid \mu)$ in order that the first conditions can be written as

$$\delta u/\delta x_1 - \mu p_1 = 0$$

$$\delta u/\delta x_2 - \mu p_2 = 0$$

The differentiation of there equations with respect to p_1 yields essentially the same for first two equation as for the first optimization problem. It the previous case the constraint was differentiated with respect to p_1. In this case the result is

$$\delta u/\delta x \cdot (\delta x_1/\delta p_1) + \delta u/\delta x_2/\delta p_1 = 0$$

but, because of the first order conditions that

$$\delta u/\delta x_1 = \mu p_1 \text{ and } \delta u/x_2 = \mu p_2$$

the condition can be expressed as

$$-p_1\, \delta x_1/\delta p_1 - p_2\, \delta x_2/\delta p_1 = 0$$

The matrix form of the equation is thus,

$$H(\delta x/\delta p_1) = (\mu, u, 0)^T$$

and other solution is

$$\delta x/\delta p_1 = H^{-1}(\mu, u, 0)^T$$

This is the same as the other term in the comparative static analysts. For the first optimization problem with $\lambda = \mu$.

The same results apply for changes in p_2.

Thus, we have Slutsky's equation:

$$(\delta x/\delta p_i)y = (\delta x/\delta p_i)u - x_i(\delta x/\delta y)p$$

Where, $\delta x/\delta p_i$ and $\delta x/\delta y$ represents the impact of a change in the price p_i and many income y on the vector of quantities demanded and the Lagrangian multiplier. The notation ( )y, ( )u, and ( )p indicates that the derivatives inside the parentheses are with respectively, money income y held constant utility (real income) u held constant and all preious P held constant.

Addendum

In the preceding only a change in p_1 was considered. There are analogous for the impact of a change as p_2. Rather then present those equations separately it is more interesting to present the comparative static analysis of price and money income changes.

The full set of equations which derive from the first order conditions is:

$$\mid u_1,\ _1u_1,\ 0_2\ -p_1 \mid$$

$$\mid u_2,\ _1u_2,\ 0_2\ -p_1 \mid\ \mid \delta x/\delta p_1,\ \delta x/\delta p_2,\ \delta x/\delta y\mid = \begin{vmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ x_1 & x_2 & -1 \end{vmatrix}$$

$$\mid -p_1\, -p_2\quad 0 \mid$$

Where X is the column vector $(x_1, x_2, \lambda)^T$.

Thus, the solution is

$$| \delta x/\delta p_1 \; \delta x/\delta p_2 \; \delta x/\delta y| = H^{-1} \begin{vmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ x_1 & x_2 & -1 \end{vmatrix}$$

Where, H is the bordered Hessian matrix.

1.4 COMPARATIVE STATIC THEOREMS

The Rybczynski Theorem

The Rybczynski Theorem (RT) says that if the endowment of some resource increases, the industry that uses that resource most intensively will increase its output while the other industry will decrease its output. The relative factor intensity is measured by the ratio of factor used in each industry.

The theorem suggests that unbalanced growth in factor supplies tends, at constant commodity prices, to lead to strong asymmetric changes in output level of two types of industries – capital-intensive and labour-intensive.

If the member of factors and commodities are evenly matched and two commodities (such as food and cloth) are not jointly produced, then this asymmetry entails that growth in one factor, such as labour, acts as a force to cause an actual fall in the production of other commodity.

Let us suppose that cloth is capital-intensive and food is labour-intensive. Now, if capital stock grows, the output of cloth will increase. However, the production of more cloth will lead to an increase in the demand for labour.

If more labour is not used with capital, then the productivity of labour is bound to fall. As a result, there will be shortage of labour in the labour-intensive sector (food). And the end result is a fall in the output of food. Thus, capital accumulation will lead to a fall in output of the labour-intensive industry. The RT can be presented in two types of models – a linear model and a non-linear model. In Fig. 1.18, we show the RT in a linear model.

Here, we have the following two linear constraints.
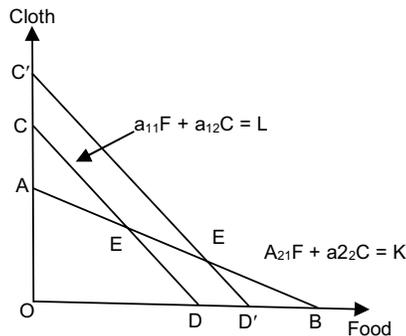
Fig. 1.18: RT in a Linear Model

$a_{11}F + a_{12}C = L$ (Labour)

$a_{21}F + a_{22}C = K$ (Capital)

Notes

where a_{11}, a_{12}, etc. are input coefficients of the Leontief type. Optimum production of the two goods and the maximum income occur at point E. The slope of the labour constraint is a_{11}/a_{12} and the slope of the capital constraint is $-a_{21}/a_{22}$

Since a_{11}/a_{12} . a_{21}/a_{22}, $a_{11}/a_{21} > a_{12}/a_{22}$

This means that, industry 1, food, is labour-intensive and industry 2, cloth, is capital-intensive.

If the amount of labour is increased, the labour constraint shifts to the right, and a new output mix is produced at point E′. At point E′, output of food, the labour-intensive good, is increases, while the output of the capital-intensive good (cloth) falls.

Thus, the RT simply suggests that if product prices are held constant, the an increase in the supply of one factor, say labour, will lead to an increase in output of the labour-intensive good and a fall in output of the capital-intensive good.

Algebraic Proof

The RT can also be proved by using a system of simultaneous equations.

In this case, the solution values of the model are determined by capital and labour constraints:

$$a_{11}F + a_{12}C = L \text{ (Labour)} \qquad \qquad \text{...(1.3)}$$

$$a_{12}F + a_{22}C = K \text{ (Capital)} \qquad \qquad \text{...(1.4)}$$

Solving by Cramer's rule, we get,

$$F = \frac{a_{22}L - a_{12}K}{a_{11}a_{22} - a_{22}a_{21}} \qquad \qquad \text{...(1.5)}$$

$$C = \frac{-a_{21}L + a_{11}K}{a_{11}a_{22} - a_{12}a_{21}}$$

The factor intensities determine the sign of the denominator. If, for example, the food (F) industry is relatively labour-intensive, then $\frac{K_1}{L_1} > \frac{K_2}{L_2}$.

This is equivalent to $\frac{a_{11}}{a_{21}} > \frac{a_{12}}{a_{22}}$ or $a_{11}a_{22} - a_{12}a_{21} > 0$.

If the denominator is positive, then by simple differentiation,

$$\frac{\partial F}{\partial L} = \frac{a_{22}}{(+)} \quad \frac{\partial F}{\partial K} = -\frac{a_{22}}{(+)}$$

$$\frac{\partial C}{\partial L} = -\frac{a_{21}}{(+)} \quad \frac{\partial C}{\partial K} = -\frac{a_{11}}{(+)} \qquad \qquad \text{...(1.6)}$$

These are Rybczynski results. Moreover, with any change in some factor endowment, the output that is intensive in that factor will change in greater absolute proportion. than the parameter change.

For example, $F = \alpha Lt\beta K$.

where, $\alpha = \dfrac{a_{22}}{a_{11}a_{22} - a_{12}a_{21}}$ and $\beta = \dfrac{-a_{12}}{a_{11}a_{22} - a_{12}a_{21}}$

Thus, $\dfrac{\Delta F/F}{\Delta L/L} = \dfrac{L}{F} \cdot \dfrac{\partial F}{\partial L} = \dfrac{F}{L} \; \alpha = \dfrac{aF}{aL + bK}$

Since, $\beta < O,\ \alpha L > \alpha L + \beta K \Rightarrow \dfrac{L}{F} \cdot \dfrac{\partial F}{\partial L} > 1$

An identical result follows for the change in C with respect to a change in capital. The output levels of each industry responds elastically to change in factor endowments in which they are intensive.

The RT in Non-linear Models

The RT can be interpreted as the comparative statics of the two-factor variable proportion model with respect to changes in endowments. One basic hypothesis of Samuelson's factor price equalisation theorem which includes the assumption that one industry is always more labour-intensive than the other is that a change in the factor endowment of either labour or capital (or both) will have no effect on factor prices.

According to factor price equalisation theorem, factor prices are functions of output prices only. This follows from the fact that input coefficients are functions of the factor price ratio $a_{ij} = a^* (w/r)$ and $w = w^* (p_1\ p_2)$.

Thus, the first important result is that

$\partial w/\partial L = \partial w/\partial K = \partial r/\partial L = \partial r/ok$

Here, we hold output prices constant and change only the factor endowments. We may now consider the effects of changing the endowment of capital, say, on output levels. Since the price of each commodity is equal to unit cost under competitive conditions we have

$$a_{L_1}^* w + a_{k_1}^* r\ =\ P_1 \qquad\qquad …(1.7)$$

$$a_{L_2}^* w + a_{K_2}^* r = P_2$$

and we have two linear constraints as expressed by (1.3) which are rewritten here:

$$a_{L_1}^* F + a_{L_2}^* C = L \qquad\qquad …(1.3(a))$$

$$a_{K_1}^* F + a_{K_2}^* C = K \qquad\qquad …((1.3(b))$$

Here, we have assumed linear homogeneity of the production function to express the *s as function of the factor price ratio (w/r) only. If now either labour or capital changes, then a_{ij}*s remains constant, since $[\partial_{aij}/\partial(w/r)].\ [\partial(w/r)/dL] = 0$ and since the $[(\partial w/r)/\partial L] = 0$ from (1.4).

Let us now differentiate equations (1.3) partially w.r.t. labour (we drop the asterisks for simplicity).

$$a_{L1}\dfrac{\partial F}{\partial L}\ + a_{L2}\dfrac{\partial C}{\partial L}\ \equiv 1$$

$$a_{K1}\dfrac{\partial F}{\partial L}\ + a_{K2}\dfrac{\partial C}{\partial L}\ \equiv 0$$

Now, by using Cramer's rule, we get

$$\dfrac{\partial F}{\partial L} \equiv \dfrac{a_{K2}}{A} = \dfrac{a_{K_2}}{a_{L_1} a_{K_1} - a_{L_1} a_{K_1}} \qquad\qquad …(1.3(c))$$

$$\dfrac{\partial C}{\partial L} \equiv -\dfrac{a_{K_1}}{A} = -\dfrac{a_{K_1}}{a_{L_1} a_{K_2} - a_{L_2} a_{K_1}} \qquad\qquad …(1.3(d))$$

Under the assumption that the industry 1 is labour-intensive, $A > 0$ and then $\frac{\partial F}{\partial L} > 0, \frac{\partial C}{\partial L} < 0$.

Similarly, differentiating (1.3) w.e.f. capital yields

$$\frac{\partial F}{\partial K} \equiv -\frac{a_{22}}{A} = \frac{-a_{L_2}}{a_{L_1}a_{K_2} - a_{L_2}a_{K_1}} \qquad \ldots(1.3(e))$$

$$\frac{\partial C}{\partial K} \equiv -\frac{a_{L_1}}{A} = \frac{-a_{L1}}{a_{L_1}a_{K_1} - a_{L_1}a_{K_1}} \qquad \ldots(1.3(f))$$

Again, assuming industry 1 to be labour-intensive, (1.3(c)) says that $\frac{\partial F}{\partial K} < 0$ and (1.3(d)) hows that $\frac{\delta C}{\delta K} < 0$.

These results known as Rybczynski theorem states that an increase, say, in the endowment of labour (holding output prices constant) will increase the output of the labour-intensive (food) industry and decrease that of the capital-intensive (cloth) industry.

The converse is also true. An increase in the endowment of capital, *ceteris paribus*, will increase the output of the capital-intensive cloth industry and decrease that of the labour-intensive food industry. However, all these repercussions will leave factor prices unchanged.

Duality of the Two Theorems

There is a reciprocity relationship between the SST and RT. This states that, in any general equilibrium model, the effect of an increase in commodity price (say P_i) on the return to a factor (say w_i) is the same as the effect of an increase in the corresponding factor endowment (v_i) on the output of commodity. However, in each case, some other set of variables are held constant. Thus,

$$\partial w_i/\partial p_j = \partial x_j/\partial v_i$$

with all other commodity prices and all factor endowments held constant in the left hand derivative and all other endowments and all commodity prices held constant in the right hand derivative. This relationship reveals the dual nature of SST and RT theorems.

If an increase in the price of cloth lowers agricultural wages, then an increase in the endowment of rural labour (a constant wage) would lower the output of cloth. In each case, it is the presumed capital- intensity of cloth which is operative.

In the 2×2 case, both the SST and RT, thus, reflect the magnification effects that stem directly from the assumed lack of joint production. Putting a hat (a) over a variable to indicate relative changes, if cloth is capital-intensive and food is labour-intensive and if the relative price of cloth rises, then

$$\hat{r} > \hat{p}_c > \hat{p}_F > \hat{w} \qquad \ldots(1.8)$$

In addition, if the economy grows, but with capital growing at a faster rate than labour

$$\hat{x}_c > \hat{K} > \hat{L} > \hat{x}_F \qquad \ldots(1.9)$$

Inequality (1.8) shows that commodity price changes are trapped between factor price changes (since both the factors are required to produce each good), while inequality (1.9) shows that in order to absorb endowment changes, the composition of outputs (each of which uses both the factors) must

change more drastically. SST stressed the first inequality in (1.8) while RT focused on the last inequality in (1.9), assuming L equals zero.

The Findlay-Grubert Theorem

Findlay and Grubert (1959) have analysed the effects of neutral and factor-saving technological progress on the level of production in a simple two-good two-factor model by assuming constant commodity prices. The geometrical techniques they used are brilliant. However, they do not explicitly state their definition of technological progress but they do implicitly assume the Hicks' definition.

Consider a standard neo-classical production function, $Q = F(K, L, t)$ where Q, K, L are respectively output, capital and labour, and t is a shift parameter. Technological progress takes place where $\partial F/\partial t > 0$. The problem is to compare a point in the initial isoquant to a point on the post-invention or shifted isoquant. This requires a definition and each definition transforms the general function (F) into an essentially specific form with well-defined properties. According to Hicks, technological progress (process innovation) is neutral if it increases the marginal products of both K and L in the same proportion. Technological progress is K-saving (or L-using) if it increases the marginal product of labour more than that of capital and K-using (or L-saving) if it increases the marginal product of capital more than that of labour. The dual of the Hicks' definition is that a shift is neutral, K-saving or K-using according to whether, at a given wage-rental ratio, the shift respectively leaves unchanged, decreases or increases the initial capital-labour ratio. These shifts are illustrated in Fig. (a), (b), (c) where the initial equilibrium is given by point E. Technological progress shifts the isoquants inward indicating the less quantities of factors of production are required to produce the same level of output (Q) and the post-invention equilibrium is given by point E′.

Considering the Hicks-neutral technological progress, the Findlay-Grubert theorem may be stated as:

"For any given fixed quantities of capital and labour, if the Hicks-neutral technological progress takes place in one of the two sectors, irrespective of the relative factors intensities, then at constant commodity prices, the level of output of that sector must increase while the level of output of the other sector must fall."
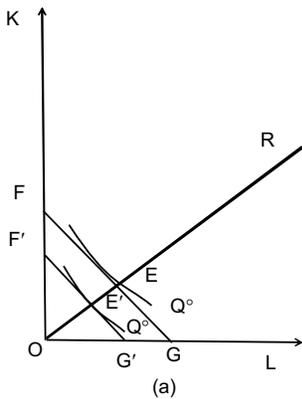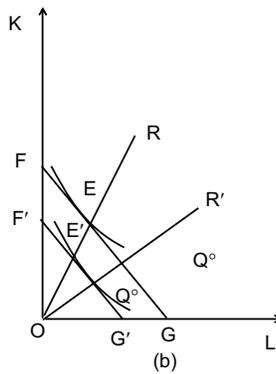
(a)	(b)	(c)
Hicks-neutral	**Hicks-saving**	**Hicks K-using**
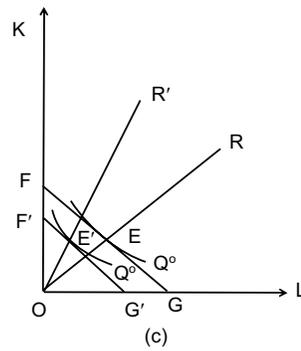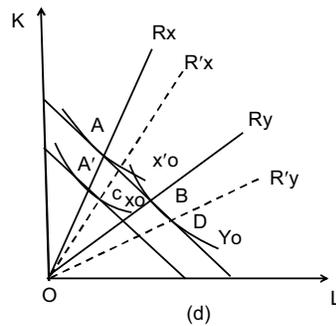
Because of neutral technological progress in the X sector, equilibrium moves from A and B points to A′ and D.
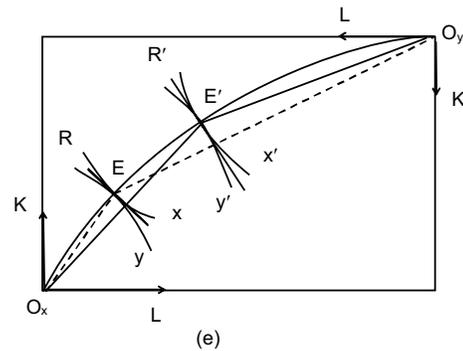
Since w/r decreases, capital intensity goes down in both X and Y sectors, general equilibrium moves from E to E′.

Fig. 1.19: The Findlay-Grubert Theorem

1.5   THEORY OF REVEALED PREFERENCE

The Concept of Revealed Preference

Prof. Samuelson has invented an alternative approach to the theory of consumer behaviour which, in principle, does not require the consumer to supply any information about himself.

If his tastes do not change, this theory, known as the Revealed Preference Theory (RPT), permits us to find out all we need to know just by observing his market behaviour, by seeing what he buys at different prices, assuming that his acquisitions and buying experiences do not change his preference patterns or his purchase desires.

Given enough such information, it is even theoretically possible to reconstruct the consumer's indifference map.

Samuelson's RPT is based on a rather simple idea. A consumer will decide to buy some particular combination of items either because he likes it more than the other combinations that are available to him or because it happens to be cheap. Let us suppose, we observe that of two collections of goods offered for sale, the consumer chooses to buy A, but not B.

We are then not in a position to conclude that he prefers A to B, for it is also possible that he buys A, because A is the cheaper collection, and he actually would have been happier if he got B. But price information may be able to remove this uncertainty.

If their price tags tell us that A is not cheaper than B (or, B is no-more-expensive than A), then there is only one plausible explanation of the consumer's choice—he bought A because he liked it better.

More generally, if a consumer buys some collection of goods, A, rather than any of the alternative collections B, C and D and if it turns out that none of the latter collections is more expensive than A, then we say that A has been revealed preferred to the combinations B, C and D or that B, C and D have been revealed inferior to A.
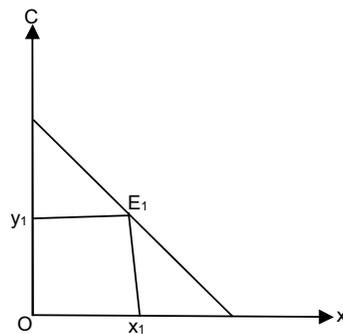
Fig. 1.20: Revealed Preference

Therefore, if the consumer buys the combination $E_1(x_1, y_1)$ of the goods X and Y and does not buy the combination $E_2(x_2, y_2)$ at the prices $(p'_x, p'_y,)$ of the goods, then we would be able to say that he prefers combination E_1 to combination E_2, if we obtain.

$$p'_x x_1 + p'_y y_1 \geq p'_x x_2 + p'_y y_2$$

The complete set of combinations of the goods X and Y to which a particular combination is revealed preferred can be found with the aid of the consumer's price line. Let us suppose that the consumer's budget line is L_1M_1 in Fig. 1.21 and he is observed to purchase the combination $E_1(x_1, y_1)$ that lies on this line.

Now, since the costs of all the combinations that lie on the budget line are the same as that of E_1 and since the costs of all the combinations that lie below and to the left of the budget line are lower than that of E_1, we may say that E_1 is revealed preferred to all the combinations lying on or below the consumer's budget line.

Again, since the costs of the combinations that lie above and to the right of the budget line are higher than that of E_1, we cannot say that the consumer prefers E_1 to these combinations when he is observed to buy E_1, because here E_1 is the cheaper combination.

We have to note the difference between "preference" and "revealed preference". Combination A is "preferred" to B implies that the consumer ranks A ahead of B.

But A is "revealed preferred to B" means A is chosen when B is affordable (no-more-expensive). In our model of consumer behaviour, we generally assume that people are choosing the best combination they can afford that the choices they make are preferred to the choices that they could have made. That is, if (x_1, y_1) is directly revealed preferred to (x_2, y_2), then (x_1, y_1) is, in fact, preferred to (x_2, y_2).

Let us now state the RP principle more formally:

Let us suppose, the consumer is buying the combination (x_1, y_1) at the price set (p'_x, p'_y) Let us also suppose that another combination is (x_2, y_2), such that $p'_x x_1 + p'_y y_1 \geq p'_x x_2 + p'_y y_2$. Now, if the consumer buys the most preferred combination subject to his budget constraint, then we will say the combination (x_1, y_1) is strictly preferred to combination (x_2, y_2).

Assumptions

With the help of the simple principle of RP, we may build up a powerful theory of consumer demand. The assumptions that we shall make here are:

(a) The consumer buys and uses only two goods (X and Y). The quantities x and y of these goods are continuous variables.

(b) Both these goods are of MIB (more-is-better) type. This assumption is also known as the assumption of monotonicity. This assumption implies that the ICs of the consumer are negatively sloped.

(c) The consumer's preferences are strictly convex. This assumption implies that the ICs of the consumer would be convex to the origin, which again implies that there would be obtained only one point (the point of tangency) on the budget line of the consumer that would be chosen by him over all other affordable combinations.

This assumption is very important. On the basis of this assumption, we shall obtain a one-to-one relation between the consumer's price-income situation or budget line and his equilibrium choice—for any particular budget line of the consumer, there would be obtained one and only one equilibrium combination of goods and for any combination to be an equilibrium one, there would be obtained one and only one budget line.

(d) The fourth assumption of the RP theory is known as the weak axiom of RP (WARP). Here, we assume that if the consumer chooses the combination $E_1(x_1, y_1)$ over another affordable combination $E_2(x_2, y_2)$ in a particular price-income situation, then under no circumstances would he choose E_2 over E_1 if E_1 is affordable.

In other words, if a combination E_1 is revealed preferred to E_2, then, under no circumstances, E_2 can be revealed preferred to E_1.

(e) The fifth assumption of the RP theory is known as the strong axiom of RP (SARP). According to this assumption, if the consumer, under different price-income situations, reveals the combination E_1 as preferred to E_2, E_2 to E_3,…, E_{k-1} to E_k, then E_1 would be revealed preferred to E_k and E_k would never (under no price-income situation) be revealed preferred to E_1.

Revealed Preference—Direct and Indirect

If RP is confined to only two combinations of goods, E_1 and E_2, and if, in a particular price-income situation, $E_1(x_1, y_1)$ is revealed preferred to combination $E_2(x_2, y_2)$, then it is said that E_1 is directly revealed preferred to E_2.

But if preferences are considered for more than two combinations and if preferences are established by way of transitivity of RP, then it is a case of indirectly revealed preference. For example, if E_1 is revealed preferred to E_2,…, E_{k-1} to E_k, then by SARP, we say E_1 is indirectly revealed preferred to E_k.

Violation of the WARP

Let us consider Fig. 1.21. Here, let us suppose that, under the price-income situation represented by the budget line L_1M_1, the consumer purchases the combination $E_1(x_1, y_1)$ and he reveals combination $E_1(x_1, y_1)$ as preferred to $E_2(x_2, y_2)$.

For here he chooses E_1 over the affordable combination E_2. Again, let us suppose that when the budget line of the consumer changes from L_1M_1 to L_2M_2, the consumer buys the combination $E_2(x_2, y_2)$, although he could have obtained the affordable combination $E_1(x_1, y_1)$, i.e., under L_2M_2, E_2 is revealed preferred to E_1.

What we have seen here is that under the budget line, L_1M_1, the combination E_1 is revealed
preferred to E_2 and under a different budget line L_2M_2, E_2 is revealed preferred to E_1. Obviously, the
consumer here violates the WARP.

The reason for this violation may be that the consumer here does not attempt to obtain the most
preferred combination subject to his budget constraint; or, it may be that his taste or some other
element in his economic environment has changed which should have remained unchanged by our
assumptions.

Now, whatever may be the reason for the violation of WARP, this violation is not consistent
with the model of consumer behaviour that we are discussing.

The model assumes that the consumer wants to maximise his level of satisfaction and, that is
why, when he chooses a particular combination, say, E_1 subject to his budget, that must be the most
'preferred' to all other affordable combinations, and none of these 'other' combinations can be
'preferred' to E_1 under a different budget. WARP puts emphasis on this simple but important point.
We may give the formal statement of WARP in the following way.
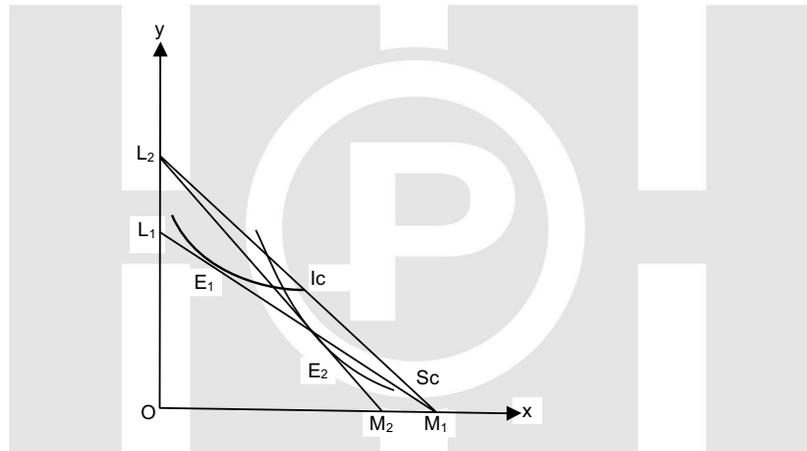
Fig. 1.21: WARP Model Violated
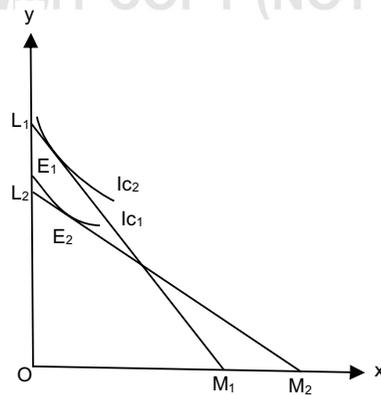
Fig. 1.22: WARP Model not Violated

If a particular combination $E_1(x_1, y_1)$ is directly revealed by the consumer as preferred to a
different combination $E_2(x_2, y_2)$, then E_2 would never be revealed by the consumer as preferred to E_1.

In other words, if the consumer is observed to purchase $E_1(x_1, y_1)$ at the price set $(p_x^{(1)}, p_y^{(1)})$ and $E_2(x_2, y_2)$ at the price set $(p_x^{(1)}, p_y^{(2)})$, then if (1.10) below holds, then (1.11)) must never hold:

$$P_x^{(1)}x_1 + P_y^{(1)}y_1 \geq P_x^{(1)}x_2 + P_y^{(1)}y_2 \qquad \qquad \qquad \ldots(1.10)$$

$$P_x^{(1)}x_2 + P_y^{(2)}y_2 \geq P_x^{(1)}x_1 + P_y^{(2)}y_1 \qquad \qquad \qquad \ldots(1.11)$$

As we have seen, WARP has been violated in Fig. 1.21, when the consumer buys combination E_1 on L_1M_1 and E_2 on L_2M_2. Here, the preference ordering of the consumer breaks down. It may be verified in Fig. 1.21 that the IC tangent to L_1M_1 at E_1 and the IC tangent to L_2M_2 at E_2 cannot be non-intersecting in this case.

In Fig. 1.22, on the other hand, let us suppose, the consumer buys the combination E_1 on L_1M_1 and the combination E_2 on L_2M_2. Here, when he buys E_1, he chooses E_1 over the affordable combination E_2, i.e., E_1 is revealed preferred to E_2. But when he buys E_2, he chooses E_2 over an unaffordable E_1, i.e., E_2 is not revealed preferred to E_1.

Therefore, here, WARP is not violated, and so, here the preference ordering of the consumer does not break down. It may be seen in Fig. 1.22 that the IC tangent to L_1M_1 at E_1 and the IC tangent to L_2M_2 at E_2 would be non-intersecting.

Significance of the SARP

Let us now discuss the significance of the strong axiom of revealed preference (SARP). According to this axiom, if the consumer reveals a combination $E_1(x_1, y_1)$ as preferred to another combination $E_2(x_2, y_2)$ and if $E_2(x_2, y_2)$ is revealed preferred to $E_3(x_3, y_3)$, then E would always be revealed preferred to E_3.

This may be called the transitivity of revealed preferences. Now, if the consumer is a utility-maximising one, then the transitivity of revealed preferences would lead to transitivity of preferences—if E_1 is preferred to E_2 and E_2 to E_3, then E_1 would be preferred to E_3.

But this is necessary to ensure that the ICs are non-intersecting and the non-interesting ICs are necessary for arriving at the utility-maximising solution. It is evident that if any of the WARP and SARP is violated, then utility maximisation cannot be achieved by the consumer.

Revealed Preference Theory and the Slutsky Theorem

Let us now see how the RPT can be used to prove the Slutsky Theorem which states that if the Income Effect (IE) for a commodity is ignored, then its demand curve must have a negative slope. To explain this, we shall take the help of Fig. 1.23.
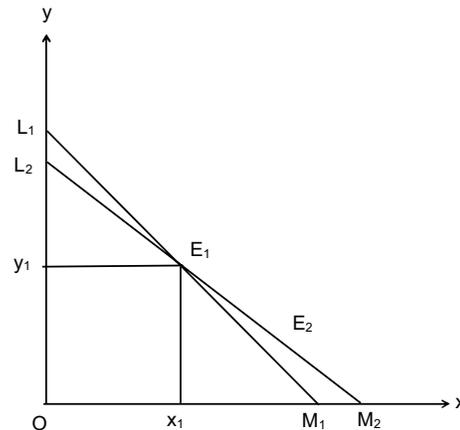
Fig. 1.23: Deduction of Slutsky Theorem from RPT

In this figure, let $E_1(x_1, y_1)$ represent the combination of goods that the consumer initially purchases when his budget line is L_1M_1. We want to show here that a *ceteris paribus* fall in the price of good X from L_1M_1 will increase the purchase of the good if we ignore the income effect, i.e., if we consider only the Substitution Effect (SE).

Let us suppose that the imaginary budget line for Slutsky SE is L_2M_2. This line will be flatter than L_1M_1, since the price of X has fallen, *ceteris paribus*, and this line (L_2M_2) will pass through the combination E, so that, as per the Slutsky condition, the consumer might be able to buy the initial combination, if he liked, under the changed circumstances.

Let us now see, because of the SE, the point the consumer may select on the imaginary budget line L_2M_2 (if it is to be different from E), would be a point like E_2 to the right of the point E_1. To prove that this must be so, we have to note that selection of any point on L_2M_2 such as E_3 which lies to the left of E_1, is ruled out by WARP.

This is because, initially, E_1 has been revealed preferred to E_3, since E_3 lies below L_1M_1. But if E_3 were chosen when the price line was L_2M_2, it (E_3) is revealed preferred to E_1 since E_1 is no-more-expensive than E_3 (for they both lie on the same budget line L_2M_2). In that case, we obtain that E_1 is revealed preferred to E_3, and *vice versa*, which violates WARP.

Thus, no point on L_2M_2 which, like E_3, lies to the left of E_1, can be chosen. On the other hand, if the consumer chooses a point like E_2 on L_2M_2 to the right of E_1, then there is no harm to the weak axiom, because when he purchases E_2, E_2 is revealed preferred to the no-more-expensive combination E_1 but, initially, when he purchased E_1 (on L_1M_1) and not a point like E_2, he did this, because E_1 was cheaper than these points.

From the analysis, it is clear that the SE of a fall in the price of X will generally increase the demand for the relatively cheaper commodity X at a point like E_2 to the right of E_1. Thus, the Slutsky theorem is deduced from the revealed preference approach.

We have seen that if the price of X falls, *ceteris paribus*, and if the income effect of this price fall is ignored, then the SE would increase the demand for X, i.e., the demand curve for X would be negatively sloped, and the law of demand is obtained.

From Revealed Preference to Preference

The principle of Revealed Preference (RP) is rather simple, but at the same time, it is very powerful. Supported by the assumptions we have made, the RPT enables us to obtain the consumer's preference pattern or Indifference Curves (ICs) from his revealed preferences.

No introspective data are required from the consumer to achieve this task. If we know the price-income situation of the consumer as represented by his budget line and his point of revealed preference on the line, we would be able to derive his IC that passes through this point. The process of obtaining the IC is described below.
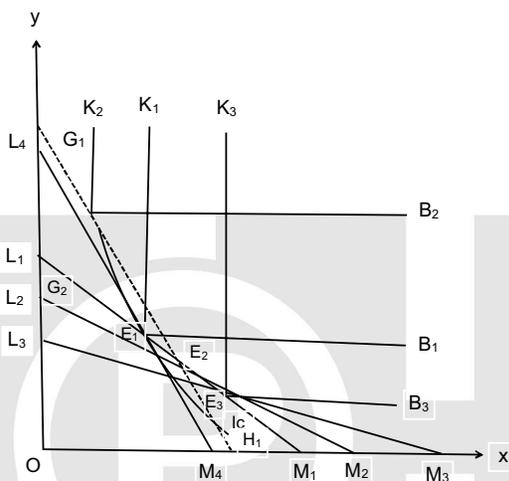
Fig. 1.24: Derivation of IC from the Revealed Preference System

Let us suppose that the budget line of the consumer is L_1M_1 in Fig. 1.24 and the combination of the goods that the consumer is observed to purchase, is $E_1(x_1, y_1)$. As we know, the consumer here prefers the point E, directly to all other points on the budget line or in the area OL_1M_1. For, in spite of all these points being within his budget, he purchases E_1. "All these points" are considered to be "worse" than E_1.

On the other hand, the costs of all the combinations lying to the right of the budget line L_1M_1 are more than that of the point E_1, or, E_1 is cheaper than these points. Apparently, the consumer chooses E_1 over these points because they are more expensive, and we cannot say anything about the 'revealed' preference of E_1 to any of these points.

That is why, the area in the commodity space to the right of L_1M_1 is known as the area of ignorance. We shall see, however, with the help of the assumptions of the RPT, that some of the points in the area of ignorance are directly or indirectly preferred or inferior to E_1 and some of the points are indifferent with E_1.

These latter points that are indifferent with E_1 give us the Indifference Curve (IC) passing through E_1. Let us now see how we can derive this curve.

At the very first, let us consider the area $K_1E_1B_1$. The commodity combinations (except E_1) belonging to this area are directly preferable to the consumer to E_1, since all these combinations have more of either one or both of the goods than the point E_1. These combinations may be called the "better" combinations.

So far, we have obtained that the consumer directly prefers E_1 to the points to the left of the budget line L_1M_1, i.e., those lying in the area OL_1M_1 and he directly prefers the points lying in the area $K_1E_1B_1$ to E_1. Therefore, his IC through the point E_1 if it is obtained, would be spread in the space between these two areas, and it would touch the line L_1M_1 and the area $K_1E_1B_1$ at the point E_1.

Let us now consider the points in the area of ignorance lying above the line L_1M_1 and outside the area $K_1E_1B_1$. At first, we shall try to identify the points that the consumer prefers less to E_1—these points may be called the "worse" points. In order to do this, let us consider any point E_2 lying on L_1M_1 to the right of E_1.

Let us suppose that the consumer is observed to purchase E_2 when his budget line is L_2M_2. He, therefore, reveals the point E_2 as preferred to the points to the left of the budget line L_2M_2. Since E_1 has already been revealed preferred to E_2, the consumer prefers E_1 to all these points lying in the area OL_2M_2.

Since a portion of this area, *viz.*, $OL_2E_2M_1$ belongs to the area OL_1M_1, here the net increase in the area of "worse" points is obtained to be $E_2M_1M_2$. The consumer prefers E_1 indirectly to the points of this area through the combination E_2—he prefers E_1 to E_2 and E_2 to these points.

We may again increase the area of "worse" points to the right of E_1 by considering any other point E_3 lying on the line L_1M_1 to the right of E_2. Let us suppose that the consumer purchases E_3 when the budget line is L_3M_3, i.e., he reveals the point E_3 as preferred to the points lying in the area OL_3M_3.

So far, we have seen how we may lessen the area of ignorance by considering the points on the budget line L_1M_1 to the right of E_1. We may also do this job by considering points on L_1M_1 to the left of E_1. Let us suppose, E_4 is any point on L_1M_1 to the left of E_1 and the consumer is observed to purchase E_4 when his budget line is L_4M_4.

The point E_4, therefore, is revealed preferred to the points lying in the area OL_4M_4. But the point E_1 has already been revealed preferred to point E_4 and so the consumer prefers E_1 to these points. Here, if we leave out the common portion of areas OL_1M_1 and OL_4M_4, we obtain that the consumer indirectly prefers E_1 to the points of $E_4L_1L_4$.

Therefore, now, we have been able to reduce the area of ignorance by $E_4L_1L_4$. We may, in this way, go on reducing the area of ignorance by considering more points on L_1M_1 lying to the left of the point E_1.

So far, we have reduced the area of ignorance by increasing the area of "worse" combinations. We may now see how we may increase the area of "better" combinations outside the area $K_1E_1B_1$ and thereby reduce further the area of ignorance. Let us suppose that the consumer is observed to purchase the point E_5 when his budget line is $G_1E_1H_1$.

Here, the consumer will prefer all the points in the area $K_2E_5B_2$ to the point E_5, since these points have more of either one or both the goods. Also, it is now revealed that the consumer prefers E_5 to E_1, for he chooses E_5 over the affordable E_1. Therefore, what we obtain here is that the points lying in the area $K_2E_5B_2$ are "better" than the point E_1.

Here, if we leave out the portion of $K_2E_5B_2$ which is in common with $K_1E_1B_1$, we find that there has been a net increase in the area of "better" points and net decrease in the area of ignorance—this net increase is represented by the area lying in between the lines K_2E_5, K_1T and E_5T.

Here, we would find that the area of the "better" points gets an increase by the area in between the lines RB_1, RE_6 and E_6B_3 plus the area E_1E_6R. Therefore, these areas are also added to the area of "better" combinations and the area of ignorance is reduced accordingly.

In Fig. 1.25, we have seen that on the basis of the idea of revealed preference and with the help of the assumptions made, we may go on increasing the area of the combinations that are "worse" than a particular combination E_1 from below and we may also go on increasing the area of the combinations that are "better" than E_1 from above.

In the limit, the area between these two areas would get reduced to a border line curve of indifference. By applying the advanced methods of calculus and also intuitively, we may obtain that this indifference curve of the consumer would pass through the point E_1 would lie in between the two paths like $K_2E_5E_1E_6B_3$ and L_4E_4E, E_2SM_3 and would be convex to the origin.

We have seen how we may obtain a consumer's IC through any particular combination E_1. Applying the same process, we may obtain his IC through any other point in the commodity space, i.e., we would obtain his indifference map.

Let us now see with the help of Fig. 1.25, how we may conclude intuitively that the borderline between the areas of "better" and "worse" combinations than any point E_1 is an IC through that point.
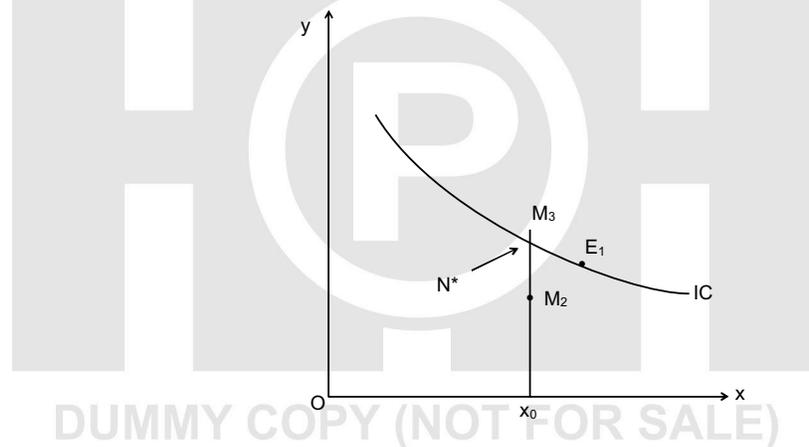
Fig. 1.25: Intuitive Deduction of an IC from the Areas of
Better and Worse Combinations

In Fig. 1.25, we have represented that the areas of better and worse combinations than E_1 have been made to advance towards each other and in the limit the gap between them looks like an IC, and actually, it would be an IC, passing through E_1. We can understand this in the following way.

Let us move vertically from one point to another in the commodity space of Fig. 1.25 starting from any point like $N(x^0, y_1)$ of the area of worse combinations. As we move upwards vertically, the quantity of good X remains the same at x_0 and that of good Y increases, and ultimately, very near the border of the "worse" area, we shall arrive at a point like $N_2(x_0, y_2)$.

Let us suppose, if we still move upwards slightly beyond N_2, we shall arrive at a point $N_3(x_0, y_3)$ in the area of "better" combinations. Now, we may easily understand intuitively that there lies a point $N^*(x_0, y^*)$, $y_2 < y^* < y_3$, in the infinitesimally small vertical gap between the points N_2 and N_3 which is neither worse nor better than E_1 but which is indifferent with E_1.

Therefore, if we join the points E_1 and the points like N^* by a curve, we would obtain the required IC through E_1.

Indifference Curve, Revealed Preference and Cost of Living Index

Let us first consider two price index formulas. One is Laspeyre's formula and the other is Paasche's formula. Laspeyre's price index number is the ratio of two aggregates—aggregate of current year prices at base year quantities and that of base year prices at base year quantities. Let us suppose that an individual purchases two goods.

The base year and current year prices of the goods are p_{01}, p_{02} and pt_1, pt_2. Also, the base year and current year quantities of the goods purchased by the consumer are q_{01}, q_{02} and q_{t1}, q_{t2}. Then Laspeyre's price index would be

$$L = \frac{P_{t1} \cdot q_1}{P_{01} \cdot q_{01}} + \frac{p_{t2} \cdot q_{02}}{P_{02} \cdot q_{02}} \qquad \qquad \ldots(1.12(a))$$

Here, the base year quantities of the goods have been taken as weights of their prices. L gives us the price index in the current year if the base year price index is 1. For example, if $L = 1.5$, then we obtain that current year price index is 1.5 when the base year price index is 1, i.e., the prices in the current year are 50 per cent more than those in the base year.

Laspeyer's price index may be interpreted in another way. The numerator of the right hand side of (1.12(a)) gives us the cost of the base year basket of goods (q_{01}, q_{02}) at the current year prices (p_{t1}, p_{t2}), and the denominator gives us the cost of buying the same basket of goods at the base year prices (p_{01}, p_{02}).

Looking in this way, $L = 1.5$ gives us that the cost of buying the base year basket of goods has increased by 50 per cent in the current year over the base year, i.e., the Laspeyre's price index number L may also be considered as the Laspeyre's cost of living index number.

Let us now come to Paasche's price index number which is the ratio of the aggregate of current year prices at current year quantities and that of the base year prices at current year quantities. Therefore, we obtain Paasche's price index number as

$$P = \frac{P_{t1} \cdot q_{t1}}{P_{01} \cdot q_{t1}} + \frac{p_{t2} \cdot q_{t2}}{P_{02} \cdot q_{t2}} \qquad \qquad \ldots(1.12(b))$$

Here, the current year quantities of the goods have been taken as the weights of their prices. Just like the Laspeyre's price index number, Paasche's price index number may also be considered as the Paasche's cost of living index number. It gives us the percentage increase in the cost of buying the current year basket of goods in the current year over the base year.

Let us now come to the total expenditure of the consumer in the base year and in the current year. In the base year, his total expenditure is, say, E_0, and he purchases the quantities q_{01} and q_{02} at prices p_{01} and p_{02}. Therefore, his budget line in the base year is

$$E_0 = p_{01}q_{01} + p_{02}q_{02} \qquad \qquad \ldots(1.12(c))$$

Similarly, in the current year, his total expenditure is, say, E_t, and he purchases the quantities q_{t1} and q_{t2} at the prices p_{t1} and p_{t2}. Therefore, his budget line in the current year is

$$E_t = p_{t1}q_{t1} + p_{t2}q_{t2} \qquad \qquad \ldots(1.12(d))$$

Since it is assumed that expenditure equals income, E_t/E_0 gives us the index of change in the consumer's income in the current year over the base year, i.e., the index of money income change is

$$E = \frac{E_t}{E_0} = \frac{P_{t1}q_{t1}}{P_{01}q_{01}} + \frac{P_{t2}q_{t2}}{P_{02}q_{02}} \qquad \qquad \ldots(1.12(e))$$

Let us now suppose that

$$p_{t1}q_{01} + p_{t2}q_{02} < E_t \qquad \qquad ...(1.12(f))$$

This means that the cost of the base year basket at current year prices is less than the current year expenditure. In other words, in the current year, the consumer might purchase the base year basket, if he so desired, but he chose not to buy this basket. This means that he prefers the current year basket to the base year basket, i.e., he is better off in the current year than in the base year.

Dividing both sides of inequality (1.12(f)) by E_0, we get

$$\frac{q_{01}p_{t1}}{q_{01}p_{01}} + \frac{q_{02}p_{t2}}{q_{02}p_{02}} < E_t/E_0 = E$$

or, $L < E$

i.e., $E > L$ \qquad \qquad ...(1.12(g))

Therefore, (1.12(f) implying (1.12(g) gives us the condition for the consumer to be better off in the current period over the base period. Let us now consider the following case:

$$q_{t1}p_{01} + q_{t2}\,p_{02} < E_0 \qquad \qquad ...(1.12(h))$$

This means that the cost of the current year bundle at the base year prices is less than the base year expenditure. This implies that the consumer might have bought the current year basket in the base year, but he chose not to buy this basket.

Thus, he preferred the base year basket and was better off in the base period over the current period. In other words, he is worse off in the current year than in the base year. Dividing both sides of (1.12(h)) by E_t, we have

$$\frac{p_{01}q_{t1}}{p_{01}q_{t1}} + \frac{p_{02}q_{t2}}{p_{t2}q_{t2}} < \frac{E_0}{E_t} = \frac{1}{E}$$

$$\Rightarrow \frac{1}{P} < \frac{1}{E} \Rightarrow P < E \Rightarrow E < P \qquad \qquad …(1.12(i))$$

Therefore, (1.12(h)) implying (1.12(i)) gives us the condition for the consumer to be better off in the base period or worse off in the current period.

Next, let us suppose

$$p_{t1}q_{01} + p_{t2}q_{02} > E_t \qquad \qquad …(1.12(j))$$

Dividing both sides of inequality (1.12(j)) by E_0, we get

$$\frac{p_{t1}q_{01}}{p_{01}q_{01}} + \frac{p_{t2}q_{02}}{p_{02}q_{02}} < \frac{E_t}{E_0} = E$$

$$\Rightarrow L < E, \Rightarrow E < L \qquad \qquad …(1.12(k))$$

From (1.12(j)) implying (1.12(k)), we obtain that the cost of the base year basket at current year prices is greater than the current year expenditure. Therefore, the base year basket is not available to the consumer in the current year, i.e., he purchases the current year basket not because he prefers it to the base year basket, but because it is cheaper. Therefore, we cannot say that the consumer is better off in the current year over the base year.

Similarly, if we suppose:

$$p_{01}q_{t1} + p_{02}q_{t2} > E_0 \qquad \qquad …(1.12(i))$$

and divide (1.12(i)) by E_t to obtain

$$\frac{p_{01}q_{t1}}{p_{t1}q_{t1}} + \frac{p_{02}q_{t2}}{p_{t2}q_{t2}} < \frac{E_0}{E_t}$$

$$\Rightarrow \frac{1}{P} > \frac{1}{E}$$

$$\Rightarrow E > P \qquad\qquad\qquad\qquad\qquad …(1.12(m))$$

From (1.12(l)) implying (1.12(m)), we obtain that the cost of current year basket in the base year is greater than the base year income. Therefore, the consumer buys the base year basket in the base year not because he prefers it, but because it is cheaper than the current year basket. Therefore, here we cannot say that he is better off in the base year over the current year, or, worse off in the current year over the base year.

What we have obtained above is that if $E > L$ as given by condition (1.12(k)), the consumer is better off in the current year over the base year. On the other hand, if $E < P$ as given by (1.12(i)), the consumer is better off in the base year than in the current year.

We may use the indifference curves of the consumer to illustrate these points. Fig. 1.26 illustrates the first case, i.e., the consumer is better off in the current year than in the base year.
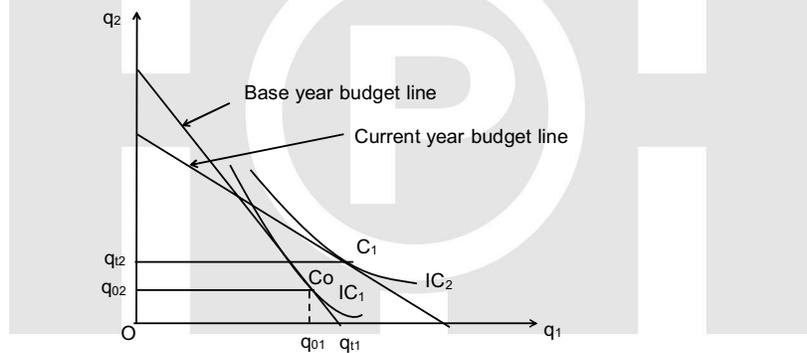
Fig. 1.26: The Consumer is Better Off in the Current Year
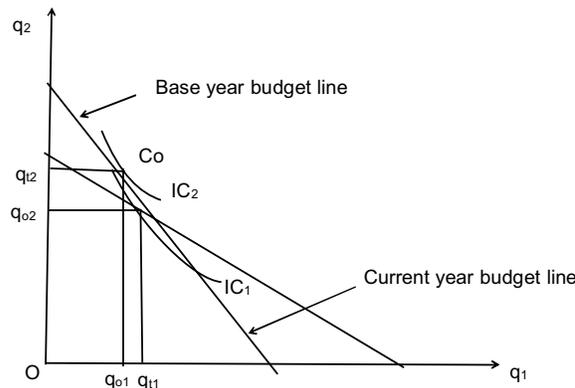
Fig. 1.27: The Consumer is Better Off in the Base Year

Here, in the current year, the consumer buys at the point C_t on the current year budget line and he buys in the base year at the point C_0 on the base year budget line. It is seen in Fig. 1.26 that C_t lies on the higher IC, *viz*., IC_2 and C_0 lies on the lower IC, *viz*., IC_1.

Similarly, Fig. 1.27 illustrates the second case, i.e., the consumer is better off in the base year than in the current year. It is seen in Fig. 1.27 that C_0 lying on the base year budget line is placed on a higher IC, *viz*., IC_2 and C_t lying on the current year budget line is placed on a lower IC, *viz*., IC_1.

From the above analysis, especially from the inequalities (1.12(g)), (1.12(i)), (1.12(k)) and (1.12(m)), we may distinguish between four cases:

1. E is greater than both L and P (E > L, E > P). Here, by (1.12(g)), i.e., E > L, the consumer is better off in the current year over the base year. On the other hand, by (1.12(m)), i.e., E > P, the standard of living does not fall in the current year. Hence, the individual is definitely better off in the current period.

2. E is less than both P and L (E < P, E < L). Here, it follows from (1.12(i)) that if E < P, the consumer would be better off in the base year, and it follows from (1.12(k)) that if E < L, the consumer would not be better off in the current period. Again, we obtain an unequivocal answer that if E < P and E < L, then the consumer would be better off in the base period, i.e., his standard of living falls in the current period from what it was in the base period.

3. L > E > P. If L > E or E < L, then by (1.12(k)), it cannot be said whether the consumer would be better off or worse off in the current period over the base period, and if E > P, then by [1(m)], we cannot say that he would be better off in the base year. Consequently, in this case, no definite conclusion can be drawn in respect of improvement or deterioration in the standard of living of the consumer between the two periods.

4. P > E > L. If P > E or, E < P, then by (1.12(i)), the consumer's standard of living falls in the current period, since he prefers the base year basket to the current year one, and if E > L, then by (1.12(g), the consumer's standard of living increases in the current year, since he prefers the current year basket to that of the base year.

Therefore, in this case also, we cannot draw any definite conclusion regarding a change in the consumer's welfare, and this is the situation where the weak axiom of the revealed preference theory has been violated.
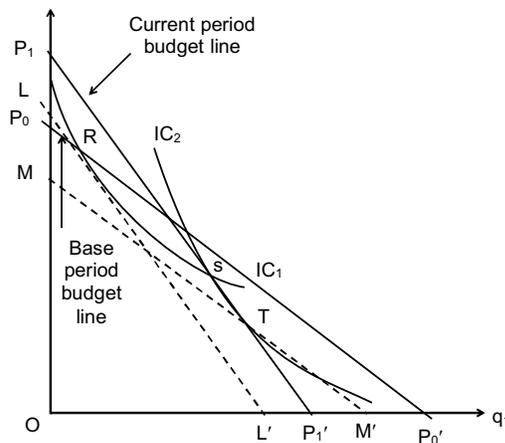
Fig. 1.28: Violation of WARP

This situation is illustrated in Fig. 1.28. Here, the base period budget line is $P_0P'_0$ and the current period budget line is $P_1P'_1$. Let us suppose that the consumer chose $R(q_{01}, q_{02})$ on IC_1 when the budget line was $P_0P'_0$ and $T(q_{t1}, q_{t2})$ on IC_2 when the budget line was P_1P_1'. Since LL' lies below P_1P_1' and is parallel to it and since R is on LL' and T is on P_1P_1', it must be true that expenditure at R (p_{t1}, p_{t2}) must be less than that at T (p_{t1}, p_{t2}), i.e., we would have

$$p_{t1}q_{01} + p_{t2}q_{02} < p_{t1}q_{t2} + p_{t2}q_{t2}$$

So that, $\dfrac{p_{t1}q_{01}}{p_{01}q_{01}} + \dfrac{p_{t2}q_{02}}{p_{02}q_{02}} < \dfrac{p_{t1}q_{t1}}{p_{01}q_{01}} + \dfrac{p_{t2}q_{t2}}{p_{02}q_{02}}$

$\Rightarrow L < E$

Also, since the point $T(q_{t1}, q_{t2})$ is on MM' which is parallel to $p_0p'_0$ but lies below it, T has the same prices as $p_0p'_0$ but has less expenditure than the point $R(q_{01}, q_{02})$ which lies on $P_0P'_0$, i.e., we have

$$p_{01}q_{01} + p_{02}q_{02} > p_{01}q_{t1} + p_{02}q_{t2}$$

$\Rightarrow \dfrac{p_{01}q_{01}}{p_{t1}q_{t1}} + \dfrac{p_{02}q_{02}}{p_{t2}q_{t2}} < \dfrac{p_{01}q_{t1}}{p_{t1}q_{t1}} + \dfrac{p_{02}q_{t2}}{p_{t2}q_{t2}}$

$\Rightarrow \dfrac{1}{E} > \dfrac{1}{P} \Rightarrow E < P$

Thus, we have $P > E > L$. But in this case, there is inconsistency. This is also obvious from Fig. 1.28. The consumer could have purchased T in the base period, since T lies below the base period budget line $p_0p'_0$, but he actually chose R, implying that he prefers R to T.

But in the current period, he could have had R, since R lies below the current period budget line P_1P_1', but he chose T, implying that he prefers T to R.

This is inconsistent if his tastes remain unchanged between the base period and the current period, and the weak axiom of revealed preference is not complied with. This inconsistency is also reflected in the fact that the ICs through R and T, *viz.*, IC_1 and IC_2 have not been obtained to be non-intersecting—they have intersected at the point S.

We have seen, therefore, that it is sometimes possible to determine whether the consumer's standard of living has increased or decreased by means of index number comparisons. However, there may be situations where we cannot arrive at any definite conclusions or where the results may be contradictory.

Example 1: When two commodity baskets are purchased by the consumer at two different points in time, explain how price-weighted quantity indices may be used to verify the weak axiom of revealed preference.

Solution: We have to explain how price-weighted quantity indices may be used to verify the weak axiom of revealed preference. Let us suppose that in the base period '0', a consumer is observed to purchase the combination $q_0(q_{01}, q_{02})$ of two goods Q_1 and Q_2 at the price set $p_0(p_{01}, p_{02})$ and in the current period 't', he is observed to purchase the combination $q_t(q_{t1}, q_{t2})$ of the goods at the price set $p_t(p_{t1}, p_{t2})$.

Therefore, the costs of purchasing the combination q_0 at the price set p_0 and p_t are:

$$q_0p_0 = q_{01}p_{01} + q_{02}p_{02} \qquad\qquad …(1.13)$$

$$\text{and } q_0p_t = q_{01}p_{t1} + q_{02}p_{t2} \qquad\qquad …(1.14)$$

Again, the costs of purchasing the combination q_t at the price set p_0 and p_t are

$$q_t p_0 = q_{t1} p_{01} + q_{t2} p_{02} \qquad \ldots(1.15)$$

$$\text{and } q_t p_t = q_{t1} p_{t1} + q_{t2} p_{t2} \qquad \ldots(1.16)$$

In the base period, the consumer purchases the quantity set q_0 at the price set p_0. If he happens to prefer q_0 to q_t, then by definition, the cost of the quantity set q_{02} must be less than or (at most) equal to that of purchasing q_0 at p_0, i.e.,

$$q_t p_0 \leq q_0 p_0$$

$$\Rightarrow q_{t1} p_{01} + q_{t2} p_{02} \leq q_{01} p_{01} + q_{02} p_{02}$$

$$\Rightarrow \frac{q_{t1} p_{01}}{q_{01} p_{01}} + \frac{q_{t2} p_{02}}{q_{02} p_{02}} \times 100 \leq 100 \qquad \ldots(1.17)$$

Since the left hand side of (1.17) is, by definition, the Laspeyre's base year price-weighted quantity index (L), we obtain the condition for q_0 at p_0 to be preferred by the consumer to q_0 at p_0 as

$$L \leq 100 \qquad \ldots(1.18)$$

Again, in the current period, the consumer is observed to purchase the combination q_t at price p_t. However, if the weak axiom of revealed preference is to be satisfied, then he must not prefer q_t at p_t to q_0 at p_t. Therefore, we may conclude that he purchases q in the current period because it is cheaper than q_0, i.e.,

$$q_t p_t < q_0 p_t$$

$$\Rightarrow q_{t1} p_{t1} + q_{t2} p_{t2} < q_{01} p_{t1} + q_{02} p_{t2} \qquad \ldots(1.19)$$

$$\Rightarrow \frac{q_{t1} p_{t1}}{q_{01} p_{t1}} + \frac{q_{t2} p_{t2}}{q_{02} p_{t2}} \times 100 < 100 \qquad \ldots(1.20)$$

Since the left-hand side of (1.20) is by definition the Paasche's current year price weighted quantity index (P), we obtain the condition for p_t at q_t to be cheaper than p_0 at q_t as

$$P < 100 \qquad \ldots(1.21)$$

(1.19) and (1.21) give us that the weak axiom of revealed preference would be satisfied if the Laspeyre's and Passche's quantity indices both are less than 100. Of course, L may be at most 100. Here, 100 is the base period index numbers for both the formulas.

Example 2: A consumer is observed to purchase $x_1 = 20$, $x_2 = 10$ at the prices $p_1 = 2$ and $p_2 = 6$. He is also observed to purchase $x_1 = 18$ and $x_2 = 4$ at the prices $p_1 = 3$ and $p_2 = 5$. Is his behaviour consistent with the weak axiom of revealed preference?

Solution: From the given data, we obtain:

(a) The cost of the combination ($x_1 = 20$, $x_2 = 10$) at the prices ($p_1 = 2$, $p_2 = 6$) is

$$E_1 = 20 \times 2 + 10 \times 6 = 100$$

(b) The cost of ($x_1 = 18$, $x_2 = 4$) at the prices ($p_1 = 2$, $p_2 = 6$) is

$$E_2 = 18 \times 2 + 4 \times 6 = 60$$

(c) The cost of ($x_1 = 18$, $x_2 = 4$) at the prices ($p_1 = 3$, $p_2 = 5$) is

$$E_3 = 18 \times 3 + 4 \times 5 = 74$$

(d) he cost of ($x_1 = 20$, $x_2 = 10$) at the prices ($p_1 = 3$, $p_2 = 5$) is

$$E_4 = 20 \times 3 + 10 \times 5 = 110$$

From above, it is obtained that the consumer buys the first set of goods (20, 10), not because it is cheaper than the second set but because he prefers it to the second set, since the cost of the former, E_1 = 100, is greater than the cost of the latter, i.e., E_2 = 60.

However, when he purchases the second set, not the first one, at the prices (p_1 = 3, p_2 = 5), he does this because it is cheaper than the first set, not because he prefers this set to the first set, since the cost of the second set, i.e., E_3 = 74, is less than that of the first set, i.e., E_4 = 110.

Therefore, the consumer's behaviour is consistent with the weak axiom of revealed preference.

Convexity and Concavity

Convex and Concave Functions
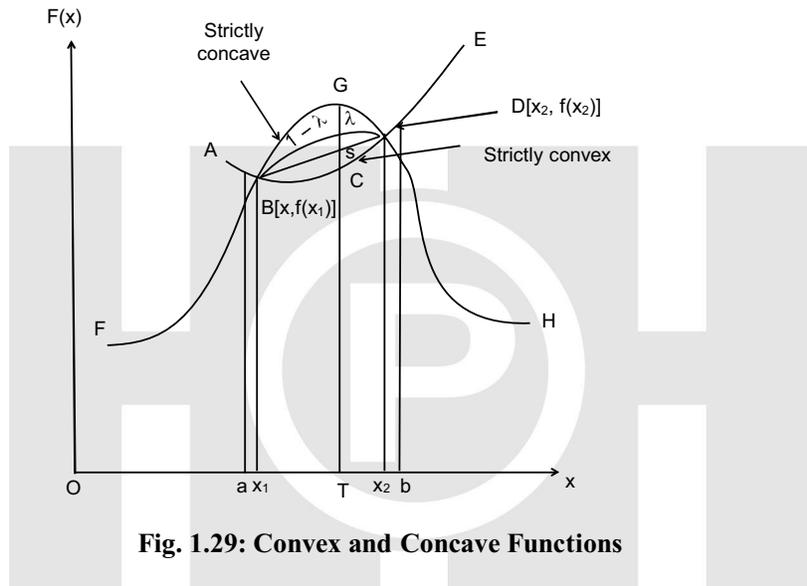
Fig. 1.29: Convex and Concave Functions

Let us refer to Fig. 1.29. A function $f(x)$ represented by the curve ABCDE is convex over the interval (a, b) if we have

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2) \qquad \ldots(1.12(n))$$

for all $a \leq x_1, x_2 \leq b$ and all $0 \leq \times \leq 1$

In Fig. 1.29, point S has divided the line segment BD in the ratio $1 - \lambda : \lambda$. Therefore, the x and y coordinates of point S are

$$OT = \lambda x_1 + (1 - \lambda)x_2$$

and $ST = \lambda f(x_1) + (1 - \lambda)f(x_2)$

The function $f(x)$ is said to be strictly convex over the interval (a, b) if strict inequality holds in [1(n)] for all $0 < \lambda < 1$.

Let us again refer to Fig. 1.29. A function $f(x)$, now represented by the curve FBGDH is concave over the interval (a, b), if we have

$$f[\lambda x_1 + (1 - \lambda)x_2] \geq \lambda f(x_1) + (1 - \lambda)f(x_2) \qquad \ldots(1.12(o))$$

and the function is strictly concave if strict inequality holds in (1.12(o)) for $0 < \lambda < 1$.

Quasi-convex and Quasi-concave Functions

By definition, a function f(x) is quasi-convex over the interval (a, b) if we have

$$f[\lambda x_1 + (1 - \lambda)x_2] \leq \max[f(x_1), f(x_2)] \qquad\qquad ...(1.12(p))$$

for all x_1 and x_2 in the interval and all $0 \leq \lambda \leq 1$. The function f(x) is strictly quasi-convex if strict inequality holds in (1(p)) for $0 < \lambda < 1$.

In Fig. 1.30, the curve A′BC′DE′ represents, by definition, a quasi-convex function over the interval (a, b).

Let us now come to quasi-concavity. A function f(x) is quasi-concave over an interval (a, b) if we have

$$f[\lambda x_1 + (1 - 1)x_2] \geq \min[f(x_1), f(x_2)] \qquad\qquad ...(1.12(q))$$

for all x_1 and x_2 in the interval (a, b) and for all $0 \leq \lambda \leq 1$. The function is strictly quasi-concave if strict inequality holds in (1(q)) for $0 < \lambda < 1$. In Fig. 1.30, the curve F′BG′DH′ represents, by definition, a quasi-concave function over the interval (a, b).

Fig. 1.30: Quasi-convex and Quasi-concave Functions

At the end of our discussion of convex and concave curves, let us note that, as per the definitions, a convex function is also quasi-convex for the former also satisfies (1.12(p)), but a quasi-convex function cannot be a convex function for it does not satisfy (11.12(n)). Similarly, a concave function is also quasi-concave for it also satisfies (11.12(q)), but a quasi-concave function cannot be concave for it does not satisfy (1.12(o).

Geometrical Illustrations

From our discussions above, we obtain the following with illustrations in Fig. 1.29:

1. The curve, ABCDE, representing a function, f(x), is convex over a certain interval (a, b) if the line segment, BD, joining any two points, B and D, on the function in the said interval lies on or above the curve; and if the line segment lies throughout above the curve, it is said that the function is strictly convex.

2. On the other hand, a function f(x), *viz.*, FGDH is concave over a certain interval (a, b) if the line segment joining any two points, B and D, on the function in the said interval lies on or

below the curve; and if the line segment lies throughout below the curve, it is said that the function is strictly concave.

We also obtain the following with illustrations in Fig. 1.30.

3. A function f(x), *viz.*, A′BC′DE′, is quasi-convex over a certain range between x = a and x = b, if at any x = h in the range, we have f(h) ≤ max [f(a), f(b)], and if the strict inequality holds, the function is said to be strictly quasi-convex.

    It may be noted that a convex function is also quasi-convex, but a quasi-convex function cannot be convex, for some quasi-convex functions, like A′BC′DE′, may lie above the line segment joining the points on the function at x = x_1 and x = x_2, which a convex function cannot.

4. Lastly, a function f(x), like F′BG′DH′, is quasi-concave over a certain range between x = x_1 and x = x_2, if at any x = h in the range, we have f(h) ≥ min [f(x_1), f(x_2)]; and if the strict inequality holds, the function is said to be strictly quasi-concave. It may be noted here that a concave function is also quasi-concave.

But a quasi-concave function cannot be concave, for some quasi-concave functions, like F′BG′DH′, may lie below the line segment joining the points on the function at x = x_1 and x = x_2, which a concave function cannot.

Utility Function for Strictly Convex Indifference Curves

Our question here is what types of utility function will produce strictly convex Indifference Curves (ICs) and thus satisfy the second-order condition. Two functions that may be accepted as such utility functions have been shown in Fig. 1.31. Part (a) of the Fig. 1.31 gives us a smooth strictly concave function.
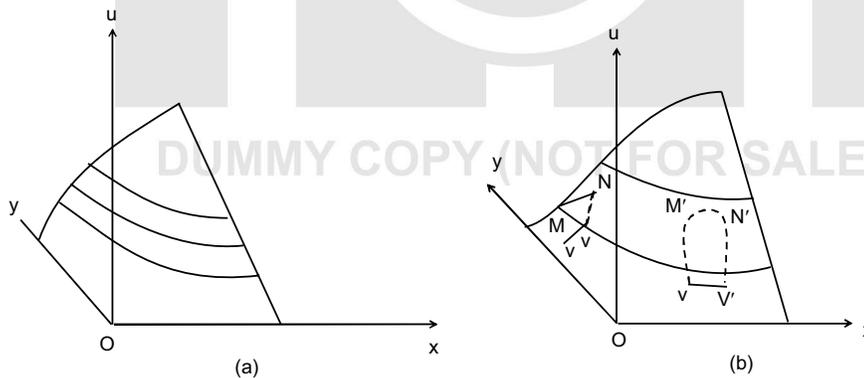
Fig. 1.31: Utility Functions for Strictly Convex Indifference Cures

Because of the assumption of positive marginal utilities, we have only shown the ascending portion of the dome-shaped surface. When this surface is cut with a plane parallel to the xy-plane, we obtain for each such cut a curve which will become a strictly convex downward sloping IC with respect to the xy-plane.

Strict concavity in a smooth utility function is, therefore, sufficient to fulfill the second-order condition (SOC) for utility maximisation. However, if we examine Part (b) of Fig. 1.31, it would be evident that strict concavity is not necessary for the SOC. This is because the strictly convex ICs can

also be obtained from the utility function given in Part (b) of the figure, which is not strictly concave—in fact, not even concave.

The function in Fig. 1.31 is generally shaped like a bell. Of course, we have shown here only the ascending portion of the bell. The surface of this function is called strictly quasi-concave.

The geometric property of this function is that, for any pair of distinct points u and v in its domain, if the line segment uv (which is assumed to lie entirely in the domain) gives rise to the arc MN on the surface, and if M is lower than or equal in height to N, then all the points on arc MN other than M and N must be higher than M.

Algebraically, a function f is said to be strictly quasi-concave if, for any two distinct points in its domain like u and v, and for all values of λ, $0 < \lambda < 1$, we would have:

$$f(u) < f(v) \Rightarrow f[\lambda u + (1 - \lambda)v] > f(u)$$

i.e., $f[\lambda u + (1 - \lambda)v] > \min f(u), f(v)]$

The quasi-concavity of the function in Fig. 1.31 may be verified by examining such arcs as MN (N higher than M) and M′N′ (M′ and N′ being of equal height). We have to note here that in the case of arc M′N′, it is the dotted arch that lies directly above the line segment uv, not the solid curve, which possesses the property of a quasi-concave function.

The interesting thing, however, is that the strictly concave function in Fig. 1.31(a) is also strictly quasi-concave.

From what we have obtained, we may conclude that only a smooth, increasing, strictly quasi-concave utility function would generate strictly convex ICs. Such a function may have convex as well as concave portions, as shown in Fig. 1.31(b) so that the marginal utilities may be either increasing or diminishing.

From this, it follows that strict convexity of ICs does not imply diminishing MUs. However, if we accept the stronger assumption of a strictly concave utility function, then we may have the features of both diminishing MU and strictly convex ICs at the same time.

Geometry of Reveal Preference

As we know the classical preference theory provides a simple, intuitive and non-paramatric way of tasting the most basic assumption of economics – that agents are rational. Here observed choice by individual reveal a preference relation are the set of consumption bundles. It is well known that the necessary and sufficient condition for the observed choice of an individual to be consistent with utility maximization is that the performance relation revealed by the choices should be acyclic or equivalently should satisfy the Generalized Axiom of Revealed Preference (GARP). Revealed preference in the standard consumption setting is a geometric property – a choose bundle is revealed preferred to all the bundles that were affordable but not chooses. Put differently revealed preference is determined by where points corresponding to choices lie is the consummation space relative to the planes determined by the budgets. A bundle is revealed preferred to another if the taller line underneath the budget plane on which the former lies. Here, we shall examine how this geometry underlying revealed preference determines the set of possible preference relations. That can be revealed by choice. This provides ex-ante information about the preference relations that are possible given the number of goods and observations in the data.

Suppose we are given a set of $\{1, ..., N\}$ with a relation $>$ defined on it. Further, whether there exists a price consumption data set $\{p_i, x_i\}^N$ consisting of K goods such that for all $i \neq j$, $p_i x_i > p_i x_j$

whenever, $i > j$ and $p_i x_i > p_i x_j$ whenever, $i \not> j$. The analysis of the above question involves examining whether budget sets can be chosen to interest in appropriate ways to allow for choices of consumption bundles which will generate relation $>$ through revealed preference.

All possible relations may not arise from reveled preference as it may not be possible to separate then consumption space into the required number of regions using down ward sloping budget planes. Hence, the set of possible revealed preference relations may depend on the dimension k of the consumption space. A higher dimensional consumption space may allow for the budget plans to separate more regions in the space potentially leading to a larger set of revealed preference relations. When there are only two goods (k = 2), the Weak Axiom of Revealed Preference (WARP) is equivalent to GARP. It implies that, when k = 2. Then can not be choice that satisfy WARP but violate Generalized Axiom of Revealed Preference (GARP) which is possible for $k \geq 3$. For instance, a relation which is cycle consisting of there elements can never be generated by revealed preference when k = 2, the Weak Axion Revealed Preference (WARP) is equivalent to GARP.

The number of goods is denoted by k. Let us denote price vector by p and consumption bundles by x. An observed date set is a finite set of price, consumption vectors.

$D = \{(p_i, x_i)\}_{i=1}^{N}$ where, $(p_i\ n_i) \in R_{++}^{k} \times (R_{+}^{k} \setminus \{0\})$ and $1 \leq M \leq \infty$. p_i^k and n_i^k denote the price and quantity consumed of the k the good in the observation respectively utility functions are given by U: $R_{+}^{k} \setminus \{0\} \to R_{+}$.

Given a data set D, a consumption bundles xi is said to be revealed preferred to another bundles xj, if the letter was affordable under price pi but was not chosen. Formally, xi is inequality is strict. Then revealed preference is strict, then revealed preference is said to be strict. The Generalized Axiom of Revealed Preference (GARP) of Varian (1982) the inquires the revealed preference relation to be acyclic.

GARP: Suppose we are given an arbitrary data set $D = \{(p_i, x_i)\}_{i=1}^{N}$. For any two consumption bundles x_i, x_j we say that $x_i\ R_0\ x_j$, if x_i is revealed preferred to x_j. We say $x_i\ Rx_j$ if for some sequence of observations $(x_1, x_2, \ldots, x_m)$, we have $x_i\ R°n$, , x, $R°x_2$, ..., $x_m R°x_j$ (R is the transitive do sure of R_0). The data not D satisfies GARP if $x_i Rx_j \Rightarrow p_j x_j \leq p_j x_i\ \forall i, \in \{1, \ldots, N\}$ where $i \neq j$ we use $>$ to denote an arbitary binary relation defined on a set $\{1, \ldots, N\}$. We now define formally what is meant by relation being generated by revealed preference.

Generating a Relation: A data set $\{(p_i, n_j)\}_{i=1}^{N}$ is said to generated solution $>$ defined an a set $\{1, \ldots, N\}$ if for all $i \neq j\ p_i x_i > p_i x_j$ if $i > j$ and $p_i x_i < p_i x_j$ if $i \not> j$.

Here, we require x_i to be strictly revealed preference to x_j wherever, $i > j$. Our choice of the strict inequality reflects the fact that in real data it is almost never the case that two different choice bundles cost exactly the same amount.

Theorem 1: (Varian, 1982).

Let $D = \{(p_i, x_i)\}_{i=1}^{N}$ be a price consumption data set.

The following are equivalent:

1. Data set D is consistent with utility maximization. In other words, there exists a non-satiated utility/function U such that for each observation i.

   $U(x') \leq U(x_i)$, for all x′ satisfying $p_i x' \leq p_i x_i$

2. Data set D satisfies GARP.

The revealed preference can generate any binary relation as long as the number of goods is at least as many as the number of observations.

Theorem 2: Given are arbitrary relation $>$ defined on a set $\{1, \ldots, 4\}$. There exists a date set $\{(p_i, x_i)\}_{i=1}^{N}$ consisting of N goods (K = N) which generates $>$.

Proof: We label each goods is a bundle by using a superscript. Thus, the jth good in the observation is represented by x_c^j and the price of the jth good is the ith observation by p_c^j, where, $1 \leq i, j \leq N$.

We low construct the date set $D = \{(p_i, x_i)\}_{i=1}^{N}$ as follows

$$p_i^i = 1 \quad x_i^i = 1$$

$$p_i^j = 0 \quad x_i^j = 0 \quad \text{if } j \neq i \text{ and } j > i \qquad \qquad \ldots(1.22)$$

$$p_i^j = 0 \quad x_i^j = 0 \quad \text{if } j \neq i \text{ and } j \not> i$$

We now check to see if this data set included generates $>$, for an arbitrary $i \neq j$, if we have $i > j$ then

$$p_i x_i = p_i^i x_i^i = 1$$

$$p_i x_j = p_i^i x_j^i = 0$$

$$\Rightarrow p_i x_i > p_i x_j$$

Similarly, if we have $i \not> j$ then

$$p_i x_i = p_i^i x_i^i = 1$$

$$p_i x_j = p_i^i x_j^i = 2$$

$$\Rightarrow p_i x_i < p_i x_j$$

Clearly the above data set generates the relation $>$. However, the proof is not complete, as we do not allow the observed data to contain zero prices. But of course, it is easy to replace every instance of a 0 price by a small enough positive $\in > $ our Eq(1.22). Since the above inequalitive are strict, for e small enough $\in$, they will not be violated and this complete the proof.

Theory 2, above show, that any relation can be generated by revealed preference, however, it requires the number of goods in the data set to be increasing in the size of the panel. It does not claim that N is the minimum number of goods required to generate any relation defined as a set of size N. This immediately leads to two questions. What is the minimum number of goods required to generate any relation and does this minimum number depend on N? White we do not have an answer to the first question we can provided an answer to the latter. Further, second result shows that for away K, there are relations $>$ defined on a set of size $N > K$ which can not be generated where the observed data has k goods can not be generated where the observed data has K goods.

Theorem 3: For any $K \geq 2$, there is a relation $>$ defined as a set of size $N = 0(2^K)$ such that no data set consisting of M observations and K good can generate relation $>$. Moreover, relation $>$ can be chosen to be acyclic.

Appendix: Proof of Theorem 3

Suppose, we are given a relation $>$ and N consumption bundles $\{(x_i)\}_{i=1}^{N}$ each with K goods. Where are there prices $\{(p_i)\}_{i=1}^{N}$, $p_i \in R_{++}^{K}$, such that $\{(p_i, x_i)\}_{i=1}^{N}$ generates exactly the relation $>$? This is essentially the problem we are studying here, but it assumes that consumption bundles are observed.

Mathematically price that generate $>$ will exist if the following system of linear inequalities has a solution for all i, j, $\in \{i, \ldots.N\}$ when $i \neq j$.

$$p_i \cdot (x_i - x_j) > 0 \text{ when } i > j \qquad \qquad \qquad \text{…(1.23)}$$

$$p_i \cdot (x_j - x_j) > 0 \text{ when } i \not> j \qquad \qquad \qquad \text{…(1.24)}$$

$$p_i \gg 0 \qquad \qquad \qquad \text{…(1.25)}$$

We will new use version of Farkar Lemma which will allows us to examine the Farkar alternative of the above system. This Lemma is stated below:

Lemma 1: For any matrix $A \in R^{m \times n}$, either there exits $ay \in R^n$ such that

$$Ay = \begin{bmatrix} 0 \\ . \\ . \\ . \\ 0 \end{bmatrix}, [1 \ldots\ldots 1]y = 1, y \geq \begin{bmatrix} 0 \\ . \\ . \\ . \\ 0 \end{bmatrix}$$

or there exists a $x \in R^m$ such that,

$$x^T A \gg [0 \ldots\ldots\ldots 0]$$

Proof: Let $i \in R^n$, $i = \begin{bmatrix} 1 \\ 1 \\ . \\ . \\ 1 \\ 1 \end{bmatrix}$, $0 \leftarrow R^m$, $0 = \begin{bmatrix} 0 \\ 0 \\ . \\ . \\ 0 \\ 0 \end{bmatrix}$. The original system can be written as,

$$\begin{bmatrix} A \\ i \end{bmatrix} T \, y = \begin{bmatrix} 0 \\ i \end{bmatrix}, y \geq 0.$$

By the Farkar Lemma, either the system above is feasible or there exists a solution $\tilde{x} \in R^{m+1}$ to the system:

$$\tilde{x}^T \begin{bmatrix} A \\ i \end{bmatrix} T \geq [0 \ldots\ldots 0] \qquad \qquad \qquad \text{…(1.26)}$$

$$\tilde{x}^T \begin{bmatrix} 0 \\ 1 \end{bmatrix} < 0 \qquad \qquad \qquad \text{…(1.27)}$$

Rewriting $\tilde{x}^T = [x, \tilde{x}]$ where, $x \in R^m$, $\hat{x} t R$, we see that (1.27) is equivalent to $\tilde{x} < 0$. Substituting this into (e), we get

$x^T A \geq [-\hat{x} - \hat{x} \ldots.. - \hat{x}] \gg [0 \ldots..0]$, which concludes the proof.

We new use Lemma 1 to take the Farkar alternative for the above system (b) – (c). We denote the dual variable for the inequality corresponding to the directed pair i > j as jij and the dual variable corresponding to the price of the kth good in the ith observations p_i^k as η_i^k. The Farmar alternative is:

$$\sum_{i=1}^{M}\sum_{j\neq i}yij + \sum_{i=1}^{M}\sum_{k=1}^{k}\eta_i^k = 1$$

$$\sum_{\{j/i>j\}}yij(x_i^k - x_i^k) + \sum_{\{j/i>j\}}yij(x_i^k - x_i^k) + \eta_i^k = 0$$

For all $1 \leq i \leq M$, $1 \leq k \leq k$

yij, $\eta_i^k \geq 0$ for all $1 \leq i \neq j \leq N$, $1 \leq k \leq k$.

We eliminate the η variable to get an equivalent system which consist of fewer unknowns.

$$\sum_{i=1}^{N}\sum_{j\neq i}yij > 0$$

$$\sum_{\{j/i>j\}}yij(x_i^k - x_j^k) + \sum_{\{j/i>j\}}yij(x_j^k - x_i^k) \leq 0$$

For all $1 \leq i \leq N$, $1 \leq k \leq k$

yij ≥ 0 for all $1 \leq i \neq j \leq N$.

The intution of this elimination is straight for ward. There can be no solution to the original system where $\sum_{i=1}^{M}\sum_{j\neq i}yij = 0$. This is because if any η_i^k is positive then at least are yij must be positive to satisfy the second equation.

Therefore, given consumption bundles $\{x\}_{i=1}^{M}$, there exist prices which generate the relation > if and only if the above system has no solution. Going one step further, therefore given a relation >, it the above system has no solution. Going one step further, therefore given a relation >, if the above system has a solution for all choice of x, then there is no data set with k goods which can generate > we collect his observation in the following Lemma.

Lemma 2: Suppose we are given relation >. There does not exist a data set $\{(p_i, x_i)\}_{i=1}^{N}$, generating the relation > if and only if the following system has a solution for every $\{x_i\}_{i=1}^{M\in R}$.

$$\sum_{i=1}^{M}\sum_{j\neq i}yij > 0$$

$$\sum_{\{j/i>j\}}yij(x_i^k - x_j^k) + \sum_{\{j/i>j\}}yij(x_j^k - x_i^k) \leq 0$$

For all $1 \leq i \leq N$, $1 \leq k \leq k$

yij ≥ 0 for all $1 \leq i \neq j \leq N$.

We now consult a relation > while can not be generated based on the following two observations.

Observation 1: The Farkar system have a solution if one only if the following subsystem have a solution for some i:

$$\sum_{j \neq i} yij > 0 \qquad \qquad \qquad \text{...(1.28)}$$

$$\sum_{\{j/i > j\}} yij(x_i^k - x_j^k) - \sum_{\{j/i \not> j\}} yij(x_j^k - x_i^k) \leq 0$$

For all $1 \leq k \leq k$ ...(1.29)

$yij \geq 0$ for all $1 \leq j \neq i \leq N$. ...(1.30)

Proof: When the above system has a solution, we can set the remaining $yi'j$'s for all other $i' \neq i$ which would yield a solution to the original problem.

Observation 2: Given k vectors $\{v_i, \ldots, v_k\}$ each in R^k, the following system has a non-zero solution for λ_i's (equivalently $\{\lambda i\}_{i=1}^k \in R^k \backslash \{0\}$): $\sum_{j \neq i}^k \lambda_i v_i \leq 0$

Proof: If the given vectors are linearly independent, any vector in the negative orthant can be generated by non-trivial linear combinations. It the given vector are linearly dependent. There will be a non-trivial solution for the above expression holding with equality.

Proof of Theorem 3: Let us now construct a relation > defined an a set of size $m = 2^{k+1} + k + 1$ and we will show that it can not be generated by any data set consisting of N observations each with K goods.

In the proof, we will describe the essential pairs of the relation > for all remaining pairs $i, j \in \{1, \ldots, N\}$ we can either assign $i \not> j$ for all remaining i, j then the relation > will be acyclic.

The construction is as follows

We describe how the last 2^{k+1} elements $\{k + 2, \ldots N\}$ are related to the first $K + 1$ elements $\{1, \ldots, K + 1\}$.

Consider, each vector $e = (d_1, \ldots, e_k) \in \{0, 1\}^k$. This vector can be thought of as a binary representation of an integer and we denote the integer represented by the binary number e as E. We can now system the crucial part of the relation > corresponding to each e.

$K + 2E + 2 > 1$

$K + 2E + 3 > 1$

For each $2 \leq i \leq K + 1$:

– If $e_i = 0$: $K + 2E + 2 > i$ and $K + 2E + 3 > i$

– If $e_{ii} = 1$: $K + 2E + 2 \not> i$ and $K + 2E + 3 \not> ii$

We new show that given the relation, the system (5a) – (5c) will have a solution for some $i \in \{K + 2, \ldots, 2^k + K + 1\}$ for away set of consumption bundles $\{x_i\}_{i=1}^N$. We first define for all $j = 1, \ldots, K$.

$vj = (x_1 - x_j + 1)$

By our construction of relation >, we have ensured that every possible combination of vectors $\{(-1)^{ej} - v_j\}_{i=1}^K$ where $e = (e_1, \ldots, e_k) \in \{0, 1\}^K$ is present on the left ride of inequality (1.29). We can then use the observation? Which samp that there is a non-trivial solution for λ to the following use quality:

$$\sum_{i=1}^{K} \lambda_i v_i \le 0$$

We new use the signs of the λ's to choose an e for all $1 \le i \le k$

$$e_i = \begin{cases} 0 & \text{if} & \lambda i \ge 0 \\ 1 & \text{if} & \lambda i < 0 \end{cases}$$

Recall, E is the integer corresponding to binary number e defined above. If $\sum_{j=1}^{k} \lambda_i < 0$ we take i = k + 2E + 2 and sign variables s = 0, else we take i = k + 2E+ 3 and sign variable s = 1. We define yij′ for $1 \le j \le k + 1$ as

$$y_{ij} = \left| \sum_{i=1}^{k} \lambda_i \right|,$$

$y_{ij}' = | \lambda j' - 1 |$ for all $2 \le j' \le k + 1$

We set all remaining y's to u. Formally

$y_{ij} = 0$ for all $j > k + 1$. and

$y_{1n} = 0$ for all $L \ne i$, $1 \le n \le N$.

We new show that this choice of y leads to a solution of (1.28) – (1.30) for the above choice i. inequalities (1.30) are satisfied as the chosen y's are non-negative and inequality (1.1) is satisfied due to observation 2. It remains to be shown that inequality (1.29) is satisfied.

We simplify the left of inequality (1.29) for can choice of i and an arbitrary $1 \le k \le k$ an follows:

$$\sum_{\{j/i > j\}} y_{ij}(x_i^k - x_j^k) + \sum_{\{j/i \not> j\}} y_{ij}(x_j^k - x_i^k)$$

$$= (-1)_{y_{i1}}^{S}(x_i^k - x_1^k + \sum_{\{l \backslash e_i = 0\}} y_{il} + 1(x_i^k - x_{l+1}^k) + \sum_{\{l \backslash e_i = 1\}} y_{il} + 1(x_{l+1}^k - x_1^k)$$

$$= (-1)_{y_{i1}}^{S}(x_i^k - x_1^k) + \sum_{\{l/e_i = 0\}} y_{il} + 1(x_i^k - x_1^k + x_1^k - x_{l+1}^k + \sum_{\{l/e_i = 1\}} y_{il} + 1(x_{l+1}^k - x_1^k + x_1^k - x_i^k$$

$$= (-1)_{y_{i1}}^{S}(x_i^k - x_1^k) + \sum_{\{l/e_i = 0\}} y_{il} + 1(x_i^k - x_1^k + v_l^k) + \sum_{\{l/e_i = 1\}} y_{il} + 1 - (-v_l^k + x_1^k - x_1^k)$$

$$= (-1)^{S} y_{i1}(x_i^k - x_1^k) + \sum_{\{l/e1 = 0\}} y_{il} + 1(x_i^k - x_1^k + v_1^k) + \sum_{\{l/e1 = 1\}} y_{il} + 1 - (-v_i^k + x_1^k + x_1^k)$$

$$= (-1)^{S} \left| \sum_{l=1}^{k} \lambda_1 \right| (x_i^k - x_1^k) + \sum_{\lambda=1}^{k} [\lambda_1(0_i^k - 0_l^k) + \lambda_1 v_1^k]$$

$$= \sum_{\lambda=1}^{k} \lambda_1^v {}_1^k \le 0$$

Thus, for any choice of $\{x_1\}_{i=1}^{N}$, we can find an i such that (1.28) – (1.30) has a solution. This > can not be generated with K goods and this completes the proof.

Examples 1.

For the total utility function $U = 20x^4 + 7x^3 + 12x^2 + 12x + 9$ Compute Marginal Utility.

Solution:

$$\text{Marginal Utility} = \frac{du}{dx} = \frac{d}{dx}(u)$$

$$= \frac{d}{dx}(20x^4 + 7x^3 + 13x^2 + 12x + 9)$$

$$= 80x^3 + 21x^2 + 26x + 12$$

Example 2.

Compute Marginal utility for the utility function

$u = 8x^3 - 5x^2 - 4x + 4.$

Solution:

$$\text{Marginal utility} = \frac{du}{dx} = \frac{d}{dx}(u)$$

$$= \frac{d}{dx}(8x^3 - 5x^2 - 4x + 4)$$

$$= 24x^2 - 10x - 4$$

Example 3.

Find out Marginal utility for the total utility function

$$u = 8x^5 + \frac{1}{x^4} - 3x^{-3} - 2x^2 - x + 9$$

Solution:

$$\text{Marginal utility} = \frac{du}{dx} = \frac{d}{du}(u)$$

$$= \frac{d}{dx}\left(8x^5 + \frac{1}{4} - 3x^{-3} - 2x^2 - x + 9\right)$$

$$= \frac{d}{dx}\left(8x^5 + x^{-4} - 3x^{-3} - 2x^2 - x + 9\right)$$

$$= 40x^4 - 4x^{-5} + 9x^{-4} - 4x - 1$$

Example 4.

Compute Marginal utilities of x and y for the utility function $u = 5xy - y^2$

Solution:

(a) Marginal utility of $x = \dfrac{du}{dx} = 5y$

(b) Marginal utility of $y = \dfrac{du}{dx} = 5x - 2y$

Example 5.

Find out marginal utility of x and y for the utility function u = x^2 – y^2 – 2x^{2}y.

Solution:

Marginal utility of $x = \dfrac{du}{dx} = 2x - 4xy$

Marginal utility of $y = \dfrac{du}{dy} = -2y - 2x^2$

1.6 SUMMARY

1. The price of a product depends upon the demand for and the supply of it.

2. In Slutsky's version of substitution effect when the price of good changes and consumer's real income or purchasing power increases, the income of the consumer is changed by the amount equal to the change in its purchasing power which occurs as a result of the price change.

3. Prof. J.R. Hicks points out that the method of adjusting the level of money income by the compensating variation has the merit that on this interpretation, the substitution effect measures the, effect of change in relative price, with real income constant, the income effect measures the, effect of the change in real income. Thus, the analysis which is based upon the compensating variation is a resolution of the price change into two fundamental economic 'directions', we shall not encounter a more fundamental distinction upon any other route.

4. The Rybczynski Theorem (RT) says that if the endowment of some resource increases, the industry that uses that resource most intensively will increase its output while the other industry will decrease its output. The relative factor intensity is measured by the ratio of factor use in each industry.

5. Prof. Samuelson has invented an alternative approach to the theory of consumer behaviour which, in principle, does not require the consumer to supply any information about himself.

6. If his tastes do not change, this theory, known as the Revealed Preference Theory (RPT), permits us to find out all we need to know just by observing his market behaviour, by seeing what he buys at different prices, assuming that his acquisitions and buying experiences do not change his preference patterns or his purchase desires.

1.7 SELF ASSESSMENT QUESTIONS

1. 'The Hicksian Version of Consumer surplus is superior to Marshall.' Comment.

2. Explain in detail about Law of Diminishing Marginal Utility.

3. Explain in detail about Law of Equi-marginal Utility.

4. Explain in detail about' Slutsky Substitution Effect for a fall in Price'.

5. Explain in detail about' Slutsky Substitution Effect for a rise in Price'.

<p align="center">৵৻৵❖৵৻৵</p>

UNIT – II

Chapter	
2	**THEORY OF THE FIRM**

Objectives

The objectives of this lesson are to learn:

- Production function
- Cost function
- Euler's theorem
- Cobb-Douglas production function
- CES production function
- Duality relationship
- Comparative static result
- Joint production

Structure:

2.1 PRODUCTION AND COST FUNCTION: MEANING, DEFINITIONS AND FEATURES

Production is the result of co-operation of four factors of production, *viz*., land, labour, capital and organisation. This is evident from the fact that no single commodity can be produced without the help of any one of these four factors of production. Therefore, the producer combines all the four factors of production in a technical proportion. The aim of the producer is to maximize his profit. For this sake, he decides to maximise the production at minimum cost by means of the best combination of factors of production. The producer secures the best combination by applying the principles of equi-marginal returns and substitution. According to the principle of equi-marginal returns, any producer can have maximum production only when the marginal returns of all the factors of production are equal to one another. For instance, when the marginal product of the land is equal to that of labour, capital and organisation, the production becomes maximum.

Meaning of Production Function

In simple words, production function refers to the functional relationship between the quantity of a good produced (output) and factors of production (inputs). According to Prof. Koutsoyiannis, "The production function is purely a technical relation which connects factor inputs and output." Prof. Watson defined production function as "the relation between a firm's physical production (output) and the material factors of production (inputs)." In this way, production function reflects how much output we can expect if we have so much of labour and so much of capital as well as of labour etc. In other words, we can say that production function is an indicator of the physical relationship between the inputs and output of a firm.

The reason behind physical relationship is that money prices do not appear in it. However, here one thing that becomes most important to quote is that, like demand function, a production function is for a definite period. It shows the flow of inputs resulting into a flow of output during some time. The production function of a firm depends on the state of technology. With every development in technology the production function of the firm undergoes a change. The new production function brought about by developing technology displays same inputs and more output or the same output with lesser inputs. Sometimes, a new production function of the firm may be adverse as it takes more inputs to produce the same output.

Mathematically, such a basic relationship between inputs and outputs may be expressed as:

$$Q = f(L, C, N)$$

where Q = Quantity of output

L = Labour

C = Capital

N = Land

Hence, the level of output (Q) depends on the quantities of different inputs (L, C, N) available to the firm. In the simplest case, where there are only two inputs, labour (L) and capital (C) and one output (Q), the production function becomes

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

Definitions

"The production function is a technical or engineering relation between input and output. As long as the natural laws of technology remain unchanged, the production function remains unchanged." **– Prof. L.R. Klein**

"Production function is the relationship between inputs of productive services per unit of time and outputs of product per unit of time." **– Prof. George J. Stigler**

"The relationship between inputs and outputs is summarised in what is called the production function. This is a technological relation showing for a given state of technological knowledge how much can be produced with given amounts of inputs." **– Prof. Richard J. Lipsey**

Thus, from the above definitions, we can conclude that production function shows for a given state of technological knowledge, the relation between physical quantities of inputs and outputs achieved per period of time.

Features of Production Function

Following are the main features of production function:

1. **Substitutability:** The factors of production or inputs are substitutes of one another which make it possible to vary the total output by changing the quantity of one or a few inputs, while the quantities of all other inputs are held constant. It is the substitutability of the factors of production that gives rise to the laws of variable proportions.

2. **Complementarity:** The factors of production are also complementary to one another, i.e., two or more inputs are to be used together as nothing will be produced if the quantity of either of the inputs used in the production process is zero.

   The principles of returns to scale is another manifestation of complementarity of inputs as it reveals that the quantity of all inputs are to be increased simultaneously in order to attain a higher scale of total output.

3. **Specificity:** It reveals that the inputs are specific to the production of a particular product. Machines and equipment specialised workers and raw materials are a few examples of the specificity of factors of production. The specificity may not be complete as factors may be used for production of other commodities too. This reveals that in the production process none of the factors can be ignored, and in some cases, ignorance to even slightest extent is not possible if the factors are perfectly specific.

Production involves time; hence, the way the inputs are combined is determined to a large extent by the time period under consideration. The greater the time period, the greater the freedom the producer has to vary the quantities of various inputs used in the production process. In the production function, variation in total output by varying the quantities of all inputs is possible only in the long run whereas the variation in total output by varying the quantity of single input may be possible even in the short run.

Production Function

With the definitions of the technology set, input set and output set presented in the above section in place, it is now possible to give a moral formal and precise definition of a production function than the definition associated with the impirical production function. Following definition presuppose that y is a scalar (an output), with x is a scalar q vector of input:

A production function f is defined on

$$f(x) = \max \{y : y \in y(x)\} \qquad \qquad \qquad ....(2.1)$$

The production function could use be defined as

$$f(x) = \max \{y : y \in T(x,y)\} \qquad \qquad \qquad ....(2.2)$$

Hence, a production function is defined as the maximum amount of output that can be produced (through the use of a given production technology) with a given amount of output.

Similarly, isoquants and production possibility curves can be given formal definitions. An isoquant is defined as the border of the input set, i.e., as the x's for which the following is there:

$$G(y) = \{x : x \in X(y) | \quad x^k \notin X(y) \text{ for } x^k \leq x\} \qquad \qquad ....(2.3)$$

In which $x^k \leq x$ isto be understood as: None of the elements (x_i) with vector x^k an greater than the corresponding elements is the vector x, and at least one of the elements is x^k is smaller than the similar elements is x.

If the possibility of production of multiple outputs exists, then the production possibility curve is defined similarly as:

$$P(x) = \{y : y \in y(x) | y^k \notin y(x) \text{ for } y^k \geq y\} \qquad \qquad ....(2.4)$$

In which $y^k \geq y$ is to be understood as: Now, of the elements (y) in the vector y^k as smaller than the corresponding element in the vector y, and at least one of the element in y^k is greater than the similar element in y.

Mathematical Representation of the Production Function

The formal mathematical representation of the production function for the production of one output can be written as:

$$y = f(x) \qquad \qquad \qquad ....(2.5)$$

In which y is a scalar (the amount of the product y), f is the production function and x is a vector of inputs.

The production function

$$Y = f(x_1) \qquad \qquad \qquad ....(2.6)$$

Upprises the production of y only as a function of the variable input x_1. If it is appropriate to explicitly express that the production of output y is a function of the variable input x_1 and the fixed inputs $x_2, ..., x_n$, then the function should be written as $y = f(x_1 |, x_2, ..., x_n)$. Normally, fixed inputs are not included when writing the production function. It is however, important to keep in mind that the production may depend as considerably more inputs than specified in the actual production function. We can write $y = f(x_1, x_2) | x_3, ..., x_n)$ if production is a function of two variable inputs.

There is no given mathematical, functional form for a production function. All the functional forms that have been used to describe the production have historically been used or more or less subjective choices. The best known of these function forms is the so called Cobb-Dougles production function which, with the variable inputs, has the form

$$y = A_{x_1 x_2}^{b_1 b_2} \qquad \qquad \qquad ....(2.7)$$

In which, A, b_1 and b_2 are predetermined parameters (constants). The choice of functional form and the subsequent utimation of the parameters of the function is a computerise science is itself, which is not discussed here - as it is referral to studies within the raligit area of Econometrics.

Mathematical Treatment of Production theory choice of Input: Cost minimization for a Given Output

Consider a firm that use labour (L) and capital (K) to produce output (Q). Let it is the price of labour that is, wage rate and r is the price of capital and C is the cost incurred to produce a level of output is given by

$$C = wL + rk \qquad \qquad ...(2.8)$$

The objective of the firm is to minimise cost for producing a given level of output. Let the production function is given by the following.

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

In general there are rural labour capital combinations to produce a given level of output. Which combination of factors a firm should choose that will minimise its total cost of production.

Minimise $C = wL + rK$

Subject to produce a given level of output, say Q_1, that satisfice the following production.

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

The device of an optimal factor combination can be obtained through using Lagrangian method. Let us first form the Lagrangian function which is given below

$$Z = wL + rk + \lambda[Q_1 - f(L, k)] \qquad \qquad ...(2.9)$$

Which, λ is Lagrangian multiplier.

For minimisation of cost it is necessary that partial derivatives of Z with respect to L, k and λ zero

$$\frac{\partial Z}{\partial L} = w - \frac{\lambda \partial f(L, L)}{\partial L} = 0 \qquad \qquad ...(2.10)$$

$$\frac{\lambda \partial Z}{\partial K} = r - \frac{\lambda \partial f(L, K)}{\partial K} = 0 \qquad \qquad ...(2.11)$$

$$\frac{\partial Z}{\partial \lambda} = Q_1 - f(L, K) = 0 \qquad \qquad ...(2.12)$$

Note that $\dfrac{\partial f(L, K)}{\partial L}$ and $\dfrac{\partial f(L, K)}{\partial K}$ are the marginal physical products of labour and capital respectively. It will also be noticed that equation (2.12) is the given production function. Rewriting the above equation we have

$$w - \lambda\, MP_L \;=\; 0 \qquad \qquad ...(2.13)$$

$$r - \lambda\, MP_K \;=\; 0 \qquad \qquad ...(2.14)$$

$$Q_1 \;=\; f(L, K) \qquad \qquad ...(2.15)$$

By combining the two equations (2.13) and (2.14), we have

$$\frac{w}{r} = \frac{MP_L}{MP_K} \qquad\qquad\qquad \dots(2.16)$$

Equation (2.16) shows that total cost is minimised where the factor price ratio $\frac{w}{r}$ equals the ratio of marginal physical products of labour and capital. Since the ratio of marginal products of the two factor equals the marginal rate of technical substitution (MRT_{xy}), capital or cost-minimising quantities of the two factor, are obtained when we have

$$\frac{w}{r} = \frac{MP_L}{MP_K} = MRT\ S_{LK}.$$

By rearranging the equation (2.16) an get the following condition for choice of optimal factor combination:

$$\frac{MP_L}{w} = \frac{MPk}{r} \qquad\qquad\qquad \dots(2.17)$$

From the above equation (2.17) it follows that when price of capital (r) falls, wage rate remaining constant, $\frac{MP_K}{r} > \frac{MP_L}{w}$. This will induce the firm to substitute capital for labour so that MP_k falls and in this way $\frac{MPk}{r}$ becomes equal to $\frac{MP_K}{w}$ again to obtain a new cost minimising factor combination.

Finally, we combine the two equations (2.13) and (2.14) in an alternatives way to obtain the value of Lagrangian multiplier λ. Thus,

$$\lambda = \frac{w}{MP_L} \text{and } \lambda = \frac{r}{MP_K}$$

$$\lambda = \frac{w}{MP_L} \text{and } \frac{r}{MP_K}$$

Suppose, output is increand by one unit. Since, the marginal product of labour measures the extra output obtained by using an additional unit of labour, $\frac{1}{MP_L}$ inpresent the extra labour required to produce, One unit of output, $\frac{w}{MP_L}$ measure the extra or marginal cost of producing an extra unit of output by using an additional unit of capital. It is thus, clear that Lagrange multiplier (λ) equals marginal cost of output because it tells us how much addition to cost occurs when an extra unit of output is produced by using an additional unit of labour or capital.

Duality of Cost-Minimisation Problem: Output-Maximisaton write a Given Cost

In the field of both consumer theory and production theory, for every minimisation problem there is corresponding maximisation problem and *vice versa*. In other words, for every minimisation problem there is corresponding dual problem of maximisation and *vice versa*. The rotation of both the primal problem and its dual is the same. Let us explain the dual nature of optimal choice of factor combination used for production of a commodity.

The dual of the primal problem of cost-minimisation for producing a given level of output is maximisation of output for a given cost. In other words, for a given iso-cost line the dual problem for

a firm is to choose a factor combination that enables it to reach the highest production isoquant which is tangent to the given iso-cost line. Thus the dual problem of cost minimisation is of the following type

Maximise $Q = f(L, K)$

Subject of cost constant: $Lo = rL + rK$.

Lagrangian impression for the above constrained maximisation problem is

$Z = f(L, K) + \lambda(C_o = wL - rK)$

Where, C_o with given amount of cost, λ is the Lagrangian multiplier for maximisation of output with a given cost, the necessary conditions are that partial derivatives of Lagrangian function with respect to L, K and λ be zero.

Thus,

$$\frac{\partial Z}{\partial L} = \frac{\partial f(L, K)}{\partial L} - \lambda w = 0 \qquad \qquad \text{…(2.18)}$$

$$\frac{\partial Z}{\partial K} = \frac{\partial f(L, K)}{\partial K} - \lambda r = 0 \qquad \qquad \text{…(219)}$$

$$\frac{\partial Z}{\partial \lambda} = C_o - wL - rk = 0 \qquad \qquad \text{…(2.20)}$$

Note that $\frac{\partial f(L, K)}{\partial L} = MP_L$ and $\frac{\partial f(L, K)}{\partial K} = MP_K$. Further note that the equation represents the budget constraints (that is, the equation of the iso-cost line). Therefore, rewriting the above equations we have

$MP_L - \lambda w = 0$ …(2.21)

$MP_K - \lambda r = 0$ …(2.22)

$wL + rK = C_o$ …(2.23)

Combining the equations (2.21) and (2.22), we get,

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

It will be size from equation (2.16) that necessary condition for output maximisation with a given cost is the same for or the one we obtained for cost-minimisation problem for a given level of output.

Cost Function: Concept and Importance

Concept of Cost Function

The relationship between output and costs is expressed in terms of cost function. By incorporating prices of inputs into the production function, one obtains the cost function since cost function is derived from production function. However, the nature of cost function depends on the time horizon. In microeconomic theory, we deal with short-run and long-run time.

A cost function may be written as:

$C_q = f(Q_f \, P_f)$

where C_q is the total production cost, Q_f is the quantities of inputs employed by the firm, and P_f is the prices of relevant inputs. This cost equation says that cost of production depends on prices of inputs and quantities of inputs used by the firm.

Importance of Cost Function

The study of business behaviour concentrates on the production process—the conversion of inputs into outputs—and the relationship between output and costs of production. We have already studied a firm's production technology and how inputs are combined to produce output. The production function is just a starting point for the supply decisions of a firm. For any business decision, cost considerations play a great role. Cost function is a derived function. It is derived from the production function which captures the technology of a firm. The theory of cost is a concern of managerial economics. Cost analysis helps allocation of resources among various alternatives. In fact, knowledge of cost theory is essential for making decisions relating to price and output.

Whether production of a new product is a wiser one on the part of a firm greatly depends on the evaluation of costs associated with it and the possibility of earning revenue from it. Decisions on capital investment (e.g., new machines) are made by comparing the rate of return from such investment with the opportunity cost of the funds used. The relevance of cost analysis in decision-making is usually couched in terms of short and long periods of time by economists. In all market structures, short-run costs are crucial in the determination of price and output. This is due to the fact that the basis for cost function is production and the prices of inputs that a firm pays.

On the other hand, long-run cost analysis is used for planning the optimal scale of plant size. In other words, long-run cost functions provide useful information for planning the growth as well as the investment policies of a firm. Growth of a firm largely depends on cost considerations. The position of the U-shaped long-run AC of a firm is suggestive of the direction of the growth of a firm. That is to say, a firm can take a decision whether to build up a new plant or to look for diversification in other markets by studying its existence on the long-run AC curve. Further, it is the cost that decides the merger and takeover of a sick firm. Non-profit sector or the government sector must also have a knowledge of cost function for decision-making. Whether the Narmada Dam is to be built or not, it should evaluate the costs and benefits 'flowing' from the dam.

Numerical Problems on Cost Function

Problem 1: Suppose a firm face a cost function of $C = 8 + 4q + q^2$:

(i) What is the firm's fixed cost?

(ii) Derive an expression for the firm's average variable cost and marginal cost.

Solution 1: As fixed cost of the firm does not very with output, the terms in the given cost function which has no output (q) term will be the fixed cost. From the given cost function it is evident that fixed cost in 8.

(ii) Total variable cost (TVC) = TC – TFC

$$= (8 + 4q + q^2) - 8 = 4q + q^2$$

And AVC $= \dfrac{\text{TVC}}{q} = \dfrac{4q + q^2}{q} = 4 + q$

Marginal cost is the first derivative of total cost function or total variable cost function.

Thus, $MC = \dfrac{\Delta TVC}{\Delta q} = 4 + 2q$

Problem 2: A biscuit producing company has the following variable cost function

$TVC = 200Q + 9Q^2 + 0.25Q^3$

If the company's fixed costs an equal to ₹ 150 lakhs find out:

(a) Total cost function

(b) Marginal cost function

(c) Average variable cost function

(d) Average total cost function

(e) At what output levels average variable cost and marginal cost will be minimum

Solution: Since total cost is the sum of total fixed cost and total variable cost (TC = TFC + TVC), we get the total cost function as under:

$TC = 150 + 200Q - 9Q^2 + 0.25Q^3$

To determine the marginal cost in take the first derivative of the total variable cost function with respect to output θ. Thus,

$MC = \dfrac{d(TC)}{do} = 200 - 180 + 0.75Q^2$

To derive the average total cost and average variable cost an divide the respective total cost by the output level

$AC = \dfrac{TC}{Q} = \dfrac{150}{Q} + \dfrac{200Q}{Q} - \dfrac{qQ^2}{Q} + \dfrac{0.25Q^3}{Q}$

$= \dfrac{TC}{Q} = \dfrac{150}{Q} + 200 - qQ + 0.25Q^2$

And $AVC = \dfrac{TVC}{Q} = 200 = 9Q + 0.25Q2$

It is also useful to knew at what level of output, average variable cost taken on its minimum value. To determine the level of output at which average variable cost is minimum, we have to take first derivatives of the average variable cost (AVC) function and set this derivative equal to zero.

Thus, taking the first derivatives of the AVC function (AVC = 200 + 9Q + 0.25 Q^2), we have

$\dfrac{d(AVC)}{dq} = -9 + 0.5OQ.$

Seating it equals to zero, we have

$-9 + 0.50\,Q \;=\; 0$

$\Rightarrow 0.50\,Q \;=\; q$

$\dfrac{1}{2} \;=\; 9$

$Q \;=\; 18$

Output at which MC Function is Minimum

$MC = 200 - 18Q + 0.75\ Q^2$

To find the output level at which MC is minimum, we have to set the first derivative of MC function equal to zero. The first derivative of MC function is

$$\frac{d(MC)}{dQ} = -18 + 1.50\ Q.$$

Setting $\frac{d(MC)}{dQ}$ equals to zero, we have

$-18 + 1.50\ Q = 0$

$\Rightarrow 150\ Q = 18 \Rightarrow Q = -18 \times \frac{10}{15} = 12$

Thus, at output level 12, MC is minimum.

It is thus clear from above that marginal cost taken as the minimum value at an output level smaller than that at which AVC is minimum.

Problem 3: A firm producing sticks have a production function given by $Q = 2\sqrt{KL}$. In the short run, firm's amount of capital equipment is fixed at K = 100. The rental rate for K is ₹ 1 and the wage rate is ₹ 4:

 (i) Calculate the firm's short run total and average costs.

 (ii) What are STC, SAC and SMC for producing 25 sticks.

Solution: The given production function of the firm is

$Q = 2\sqrt{KL}$

Write K = 100 with short run, the short-run production function is

$$Q = 2\sqrt{100L} = 2 \times 10\sqrt{L} = 20\sqrt{L}$$

Cost, $C = wL + rK$

Given that $w = 4$ and $r = 1$

$C = 4L + 1K$

Curts the given $K = 100$

$$C = 4L + 100 \qquad\qquad ....(2.24)$$

The short run production function when K = 100 as obtained above is:

$Q = 20\sqrt{L}$

Taking square of both sides we have

$$Q^2 = 400L$$

$$q\ \frac{Q^2}{400} = L \qquad\qquad ....(2.25)$$

Substituting as in (2.24) we have

$$C = 100 + 4. \; 100 + 4.\frac{Q^2}{100}$$

$$\Rightarrow \quad C = 100 + \frac{Q^2}{100} \qquad \qquad ....(2.26)$$

The above equation (2.26) represents the short-run total cost function.

To get the short-run average cost function we divide the short-run total cost function is (2.26) by output (Q), Thus,

$$SAC = \frac{1w + Q^2/100}{Q} = \frac{100}{Q} + \frac{Q}{100}$$

Short-run Marginal Cost Function

Short-run marginal cost function can be obtained by taking the first derivative of the short-run total cost function. Short run total cost function an found above is

$$C = 100 + \frac{Q^2}{100}$$

$$SMC = \frac{dC}{dQ} = \frac{2Q}{100} = \frac{Q}{50}$$

(ii) If output of hockey sticks = 25, then,

$$STC = 100 + \frac{(25)^2}{100} = 100 + \frac{625}{100}$$

$$STC = 106.25$$

$$SAC = \frac{STC}{Q} = \frac{106.25}{25} = 4.25$$

$$MMC = \frac{Q}{50} = \frac{25}{50} = 0.5$$

Problem 4: If $Q_3 = A(KL)^{0.5}$, what is short-run cost function when K = 100? What is MC function?

Solution: Write K = 100, the short-run production function can be written as

$Q = A(100L)^{0.5} = 10A(L)^{0.5}$

Squaring both sides, we have

$Q^2 = 100A^2L \qquad \qquad ....(2.27)$

Now, the short-run cost function is

$C = TFC + TVC$

$TFC = Kr = 100r$ and $TVC = wL$

Where r is the rental price of capital and to is wage rate of labour and given K = 100

Therefore, $C = 100r + wL \qquad \qquad ....(2.28)$

From equation (2.24) we have

$$L = \frac{QL}{100A^2}$$

Substituting the value of L is (ii) all get the following short-run cost function:

$$C = 100\, r + W.\frac{Q^2}{100A^2} \hspace{3cm} ....(2.29)$$

Note that total variable cost function is w $\frac{Q^2}{100A^2}$. Differentiating the Total Variable Cost (TVC) function with respect to output (Q) we have the following marginal cost function:

$$MC = \frac{dTVC}{dQ} = \frac{200Q}{100A^2}$$

2.2 HOMOGENEOUS PRODUCTION FUNCTIONS

A function is said to be homogeneous of degree n if the multiplication of all the independent variables by the same constant, say λ, results in the multiplication of the dependent variable by λ^n. Thus, the function

$$Y = X^2 + Z^2$$

is homogeneous of degree 2 since

$$(\lambda X)^2 + (\lambda Z)^2 = \lambda^2(X^2 + Y^2) = \lambda^2 Y$$

A function which is homogeneous of degree 1 is said to be linearly homogeneous, or to display linear homogeneity. A production function which is homogeneous of degree 1 displays constant returns to scale since doubling all inputs will lead to an exact doubling of output. So, this type of production function exhibits constant returns to scale over the entire range of output. In general, if the production function $Q = f(K, L)$ is linearly homogeneous, then

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

for any combination of labour and capital and for all values of λ. If λ equals 3, then a tripling of the inputs will lead to a tripling of output.

There are various examples of linearly homogeneous functions.

Two such examples are the following:

$$Q = aK + bL$$

and $Q = AK^\alpha L^{1-\alpha}\ 0 < \alpha < 1$

The second example is known as the Cobb-Douglas production function. To see that it is, indeed, homogeneous of degree one, suppose that the firm initially produces Q_0 with inputs K_0 and L_0 and then doubles its employment of capital and labour.

The resulting output would equal:

$$Q = A(2K_o)^\alpha (2L_o)^{1-\alpha}$$

$$= 2^\alpha\, 2^{\alpha-1}\, A\; K_0^\alpha L_0^{1-\alpha}$$

$$= 2Q^o$$

Since $Q_o = AK_0^\alpha \, L_0^{1-\alpha}$

This shows that the Cobb-Douglas production function is linearly homogeneous.

Properties

Homogenous Production Function

For homogeneous production return to scale could easily be defined.

$f(\in x_1, \in x_2) = e^k f(x_1, x_2)$

$K > 1$, increasing return to scale for small range

$K = 1$, constant return to scale

$K < 1$, decreasing return to scale

Just like the utility function, the homogeneous production function contain linear expansion path which means

$RTSx_1 x_2 = f(x_2/x_1)$

The homothetic production function is a increasing transformation function of a linear homogeneous production function. As it is proved for the homothetic production utility function, for the homothetic production function the average cost function is independent from the level of production, and it is only a function the ratio of input price levels.

This can be shown for the cobb: Douglas production function.

$$q = f(x_1, x_2)$$

If $f(tx_1, tx_2) = t^k f(x_1, x_2)$, then $x_1 f_1 + x_2 f_2 = k f(x_1, x_2)$

If $K = 1$, than $x_1 f_1 + x_2 f_2 = q = f(x_1, x_2)$

$(x_1 f_1)/q + (x_2 f_2)/q = 1$

$(\delta q/\delta x_1)(x_1/q) + (\delta q/\delta x_2)(x_2/q) = 1$

$\in x_1, q + \in x_2 q = 1$ exhaustion Theorem, or

$x_1 f_1 + x_2 f_2 = q$ Marginal Productivity theory of distribution.

Two Steps

1. Each factor should receive its marginal productivity

2. All output should be exhausted.

For $k = 1$, long run profit equal to zero.

$\pi = pq - r_1 x_1 - r_2 x_2 = pq - pf_1 x_1 - pf_2 x_2 = pq - p(f_1 x_1 + f_2 x_2) = pq - pq = 0.$

$\pi(t) = pf(tx_1, tx_2) - r_1 tx_1 - r_2 tx_2 = tpf(x_1, x_2) - tr_1 x_1 - tr_2 x_2 = t\pi$

Profit function is homogeneous of degree one with respect to scale of production.

If each factor is paid according to its value of marginal product, profit will be zero regardless of its scale of production.

So, when production function is homogeneous of degree we, the scale of production is not defined'

If $\pi > 0$, t (scale of production) can be increased forever.

If $\pi < 0$ the firm will go out of business

If π U scale can not be defined.

Solution: In order to use the exhaustion theorem results, considering the above difficulties:

1. Production function defined not as homogeneous of degree are.

2. First and second order condition for profit maximisation should exist.

3. Maximum profit should equal to zero.

     $\pi = pq - r_1x_1 - r_2x_2 = 0$, $r = pf_1$, $r_2 = pf_2$

     $\pi = pq - pf_1x_1 - pf_2x_2 = 0$, $q = f_1x_1$, $+ f_2x_2$ (exhaustion theorem)

Undefined scale of production, actually means non existence of the second order condition for profit maximization.

$x_1f_1 + x_2f_2 = q$

$(f_1 + x_1f_{11} + x_2f_{21}) dx_1 + (f_2 + x_1f_{12} + x_2f_{22}) dx_2 = dq$

$dq/dx_1 (dx_2 = 0) = f_1 + f_{11}x_1 + x_2f_{21} = f_1$, $f_{11} = (- x_2/x_1)f_{21}$

$dq/dx_2 (dx_1 = 0) = f_2 + f_{22} x_2 fx_1 f_{12} = f_2$, $f_{22} = (- x_1/x_2)f_{12}$

$f_{11}f_{12} - f_{12}^2 = f_{12}^2 - f_{12}^2 = 0$ (straight line. It should be greater than zero).

However, constant return to scale assumption is needed in many cases, what should be done; we assume that:

1. The whole industry has a constant – return to scale production function but the individual firm does not

2. Scale of production infinte, in such a away that equals demand with supply in the whole industry.

The long run cost function will have a social shape when production function is homogeneous. Suppose that x_2^1 and x_2^0, rule ths to the one unit of the production level.

$f(tx_1^0, tx_2^0) = t^k f(x_1^0, x_2^0) = t^k$ So, $q = t^k$

$C = r_1x_1^0 + r_2x_2^0 = a$ Cost of producing one unit.

$C = at$ total cost of producing q untis.

$q = t^k$ production function

$C = aq^{1/k}$ total cost function, $AL = TC/q = aq(1 - k)/k$

$$ML = (a/k)q(1 - k)/k$$

There are various interesting properties of linearly homogeneous production functions. First, we can express the function, $Q = f(K, L)$ in either of two alternative forms:

1. $Q = Kg(L/K)$ or,

2. $Q = Lh(K/L)$

This property is often used to show that marginal products of labour and capital are functions of only the capital-labour ratio.

In particular, the marginal products are as follows:

$MP_k = g(L/K) - (L/K) g'(L/K)$

and $MP_L = g'(L/K)$

where g'(L, K) denotes the derivative of g(L/K). The significance of this is that the marginal products of the inputs do not change with proportionate increases in both inputs. Since the marginal rate of technical substitution equals the ratio of the marginal products, this means that the MRTS does not change along a ray through the origin, which has a constant capital-labour ratio. Since the MRTS is the slope of the isoquant, a linearly homogeneous production function generates isoquants that are parallel along a ray through the origin.

Expansion Path

If a firm employs a linearly homogeneous production function, its expansion path will be a straight line. To verify this point, let us start from an initial point of cost minimisation in Fig. 2.1, with an output of 10 units and an employment (usage) of 10 units of labour and 5 units of capital. Now, suppose, the firm wants to expand its output to 15 units. Since input prices do not change, the slope of the new isoquant must be equal to the slope of the original one.
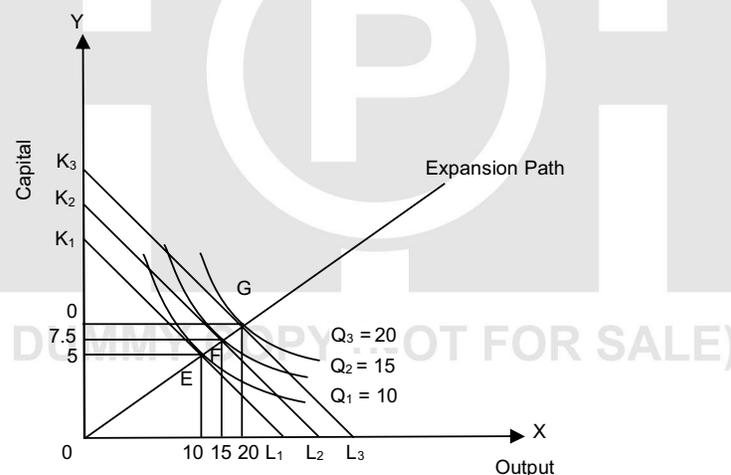
Fig. 2.1: Expansion Path

But, the slope of the isoquant is the MRTS, which is constant along a ray from the origin for linearly homogeneous production function. Consequently, the cost minimising capital-labour ratio will remain constant. Since output has increased by 50%, the inputs will also increase by 50% from 10 units of labour to 15 and from 5 units of capital to 7.5. Thus, the expansion path is a straight line.

Production functions may take many specific forms. Typically, economists and researchers work with homogeneous production function. A function is said to be homogeneous of degree n if the multiplication of all of the independent variables by the same constant, say λ, results in the multiplication of the independent variable by λ^n. Thus, the function:

$Q = K^2 + L^2$

is homogeneous of degree 2 since

Notes

$$(\lambda K)^2 + (\lambda L)^2 = \lambda^2(K^2 + L^2) = \lambda^2 Q$$

A function which is homogeneous of degree 1 is said to be linearly homogeneous, or to display linear homogeneity. A production function which is homogeneous of degree 1 displays constant returns to scale since doubling all inputs will lead to a doubling of output.

A production function is homogeneous of degree n when inputs are multiplied by some constant, say, α, the resulting output is a multiple of a^2 times the original output.

That is, for a production function:

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

then if and only if

$$Q = f(\alpha K, \alpha L) = \alpha^n f(K, L)$$

is the homogeneous function. The exponent, n, denotes the degree of homogeneity. If n = 1, the production function is said to be homogeneous of degree one or linearly homogeneous (this does not mean that the equation is linear). A linearly homogeneous production function is of interest because it exhibits CRS (Constant Returns to Scale).

This is easily seen since the expression $\alpha^n.f(K, L)$ when n = 1 reduces to α.(K, L) so that multiplying inputs by a constant simply increases output by the same proportion. Examples of linearly homogeneous production functions are the Cobb-Douglas production function and the Constant Elasticity of Substitution (CES) production function. If n > 1, the production function exhibits IRS. If n < 1, DRS (Diminishing Returns to Scale) prevails.

Cobb-Douglas Production Function

Economists have at different times examined many actual production functions and a famous production function is the Cobb-Douglas production function. Such a function is an equation showing the relationship between the input of two factors (K and L) into a production process, and the level of output (Q), in which the elasticity of substitution between two factors is equal to one.

As applied to the manufacturing production, this production function, roughly speaking, states that labour contributes about three-quarters of the increases in manufacturing production and capital the remaining one-quarter.

Suppose, the production function is of the following type

$$Q = AK^\alpha L^\beta$$

where Q is output, A is constant, K is capital input, L is labour input, and α and β are the exponents of the production function. This is known as the Cobb-Douglas production function. It has an important property.

The sum of the two exponents indicates the returns to scale:

1. If α + β > 1, the production function exhibits increasing returns to scale.

2. If α + β = 1, there are constant returns to scale.

3. Finally, if α + β < 1, there are decreasing returns to scale.

Suppose, the production is of the following type:

$$Q = AK^{0.+75}L^{0.25}$$

It exhibits constant returns to scale because α = 0.75 and β = 0.25 and α + β = 1.

2.3   EULER'S THEOREM

In number theory, Euler's theorem states that if n and a are coprime positive integers, then

$a^{\phi(n)} \equiv 1 \pmod{n}$

where $\phi(n)$ is Euler's totient function.

Euler's theorem is sometimes cited as forming the basis of the RSA encryption system. However it is insufficient (and unnecessary) to use Euler's theorem to certify the validity of RSA [RSA (Rivest-Shamir-Adleman) is one of the first public-key cryptosystems and is widely used for secure data transmission.] encryption.

In RSA, the net result of first encrypting a plaintext message, then later decrypting it, amounts to exponentiating a large input number by $k\phi(n) + 1$ for some positive integer **k.** In the case that the original number is relatively prime to **n,** Euler's theorem then guarantees that the decrypted output number is equal to the original input number, giving back the plaintext. However, because **n** is a product of two distinct primes, **p** and **q,** when the number encrypted is a multiple of **p** or **q,** Euler's theorem does not apply and it is necessary to use the uniqueness provision of the Chinese Remainder Theorem. The Chinese Remainder Theorem also suffices in the case where the number is relatively prime to **n,** and so Euler's theorem is neither sufficient nor necessary.

As we saw, in the case f(x,y) = f(x) the Euler method corresponds to a Riemann sum approximation for an integral, using the values at the left endpoints.

$$Y_n - y_0 = \int_{r0}^{xn} f(t)dt \approx h\sum_{j=0}^{h-1} f(xj)$$

A better method of numerical integration would be Trapezoid Rule:

$$\int_{r0}^{xn} f(t)dt \approx h\sum_{j=0}^{n-1} (f(x_j) + f(x_{j+1}))/2$$

This would correspond to an ineration formula $y_{j+1} = y_j + f(x_j) + f(x_{j+1})/2$

The obvious generalization to the case where f can depend or both x and y is

$$Y_{j+1} = y_j + \frac{n}{2}(f(x_j, y_j) + f(x_{j+1}, y_{j+1}))$$

The tremble with this is that y_{j+1}, which is what we are trying to computs, appears on both sides of the equation. But we will replace the y_{j+1} can the right side by the Euler approximation for y_{j+1}

$$Y_{j+1} = y_j + \frac{h}{2}(f(x_j, y_j) + f(x_{j+1}, y_j + hf(x_j, y_j)))$$

This is the itration formula for the Improved Euler Method, also known as Heean's method. It looks a bit complicated we would actually compute it is there steps:

1. $m_1 = f(x_j, y_j)$

2. $m_2 = f(x_{j+1}, y_j + h_{mi})$

3. $y_{j+1} = y_j + h(m_1 + m_2)/2$

Let's try the same example we used for the Euler method:

$y_1 = 2(y^2 + 1)/(x^2 + 4)$, y(0) = 1

With step size h = .1 with $x_0 = 0$, $y_0 = 1$, we have $m_1 = f (0, 1) = 1$, $m_2 = f(.1, 1.1) = 1.02244389$,

$y_1 = 1 + .1 (m_1 + m_2)/2 = 1.105 112219$. The true relation had $\phi(.1) = 1.1.5263158$, so the error is .000150939, compound to .005263158 for the Euler method.

Here is a table of the results of the first 10 steps for the improved Euler and Euler method with h = .1, and their respective errors

Table 2.1

x_i	Hcan y_i	Euler y_i	$\phi (x_i)$	Haun Error	Euler Error
0.0	1.00000000	1.00000000	1.00000000	0.0	0.0
0.1	1.10511222	1.10000000	1.10526316	0.00015094	0.00526316
0.2	1.22185235	1.21022444	1.1.22222222	0.00036987	0.01199778
0.3	1.35225607	1.33223648	1.35294118	0.00068510	0.02070470
0.4	1.49886227	1.46792616	1.50000000	0.00113773	0.03207384
0.5	1.66487828	1.61959959	1.66666667	0.00178838	0.04706708
0.6	1.85441478	1.79009854	1.85714286	0.00272808	0.06704432
0.7	2.07282683	1.98296335	2.07692308	0.00409625	0.09395973
0.8	2.32722149	2.20265794	2.33333333	0.00611184	0.13067539
0.9	2.62723508	2.45488648	2.63636364	0.00912856	0.18047716
1.0	2.98626232	2.74704729	3.00000000	0.01373768	0.25295271

Step Size 0.1

Charly, in this example the Improved Euler Method is much more accurate than the Euler method: about 18 times more accurate at x = 1. Now if the order of the method is better, improved Euler's relative advantage should be turn greater at a smaller step size. Hare is the table for h = .05

Table 2.2

x_i	Heun y_i	Euler' y_i	$\phi (x_i)$	Heun Error	Euler Error
0.0	1.00000000	1.00000000	1.00000000	0.0	0.0
0.1	1.10522508	1.10252967	1.0526316	0.00003808	0.00273349
0.2	1.22212855	1.21596496	1.22222222	0.00009367	0.00625726
0.3	0.35276701	1.34209198	1.35294118	0.00017417	0.01084920
0.4	1.49970962	1.48310373	1.50000000	0.00029038	0.01689627
0.5	1.66620837	1.64172213	1.66666667	0.00045830	0.02494454
0.6	1.85644079	1.82136643	1.85714286	0.00070207	0.03577643
0.7	2.07586420	2.02638978	2.07692308	0.00105887	0.05053330
0.8	2.33174590	2.26241822	2.33333333	0.00158743	0.07091511
0.9	2.63398036	2.53684738	2.63636364	0.00238328	0.09951625
1.0	2.99639263	2.85958887	3.00000000	0.00360737	0.1404113

Step Size 0.05

So, in the improved Eular method, the 1/2 becomes 1/4 (the actual ration is from .252 to .063). This supports the idea that Improved Euler's global error is $0(h^2)$.

Euler's Homogenous Function Theorem

Let f(x, y) be a homogenous function of order n so that

$f(tx, ty) = t^n f(x, y)$

Then define $x' \equiv xt$ and $y' \equiv yt$

Then

$$nt^{n-1}f(x, y) = \frac{\partial f}{\partial x'}\ \frac{\partial x'}{\partial t} + \frac{\partial f}{\partial y'}\ \frac{\partial y'}{\partial t}$$

$$= x\frac{\partial f}{\partial x'} + y\frac{\partial f}{\partial y'}$$

$$= x\frac{\partial f}{\partial (x\,t)} + y\frac{\partial f}{\partial (y\,t)}$$

Let t = 1, then

$$x\frac{\partial f}{\partial x} + y\ \frac{\partial f}{\partial y} = nf(x, y)$$

This can be generalised to an arbitrary number of variables

$$x_i\ \frac{\partial f}{\partial x_i} = nf(x)$$

2.4 COBB-DOUGLAS PRODUCTION FUNCTION

The Cobb-Douglas production function is based on the empirical study of the American manufacturing industry made by Paul H. Douglas and C.W. Cobb. It is a linear homogeneous production function of degree one which takes into account two inputs, labour and capital, for the entire output of the manufacturing industry.

The Cobb-Douglas production function is expressed as:

$Q = AL^\alpha\, C^\beta$

where Q is output, and L and C are inputs of labour and capital respectively. A, α and β are positive parameters where $= \alpha > 0, \beta > 0$.

The equation tells that output depends directly on L and C, and that part of output which cannot be explained by L and C is explained by A which is the 'residual', often called technical change.

The production function solved by Cobb-Douglas had 1/4 contribution of capital to the increase in manufacturing industry and 3/4 of labour so that the C-D production function is

$Q = AL^{3/4}\, C^{1/4}$

which shows constant returns to scale because the total of the values of L and C is equal to one: (3/4 + 1/4), i.e., (a + β = 1). The coefficient of labourer in the C-D function measures the percentage increase in Q that would result from a 1 per cent increase in L, while holding C as constant.

Similarly, B is the percentage increase in Q that would result from a 1 per cent increase in C, while holding L as constant. The C-D production function showing constant returns to scale is depicted in Fig. 2.2. Labour input is taken on the horizontal axis and capital on the vertical axis.

To produce 100 units of output, OC units of capital and OL units of labour are used. If the output were to be doubled to 200, the inputs of labour and capital would have to be doubled. OC is exactly double of OC_1 and of OL_2 is double of OL_2.

Similarly, if the output is to be raised three-fold to 300, the units of labour and capital will have to be increased three-fold. OC_3 and OL_3 are three times larger than OC_1 and OL_1 respectively. Another method is to take the scale line or expansion path connecting the equilibrium points Q, P and R. OS is the scale line or expansion path joining these points.

It shows that the isoquants 100, 200 and 300 are equidistant. Thus, on the OS scale line, OQ = QP = PR which shows that when capital and labour are increased in equal proportions, the output also increases in the same proportion.

Essential Features of Cobb-Douglas Production Function

Let we explain some of the essential features of linear homogeneous production function with special reformer to Cobb-Douglas production which is an important example of linear homogenous production function.

1. Average Product of Factors in Cobb-Douglas Production Function: The first important features of Cobb-Douglas production function when the sum of its exponents is equal to one (i.e., $\alpha + \beta = 1$) and therefore it is Linearly homogeneous production function that. The average and marginal products of factors depend upon the ratio in which factors or combined for the production of a commodity.

Linear Cobb. Douglas function can be written as

$Q = AL^{\alpha}C^{1-\alpha}, (\alpha + 1 - \alpha = 1)$

Average product of labour can be obtained from dividing the production function by the amount of Labour L.

Thus,

Average Product of labour $= \dfrac{AL^{\alpha}C^{1-\alpha}}{L} = \dfrac{AC^{1-\alpha}}{L^{1-\alpha}} = A\left(\dfrac{C}{L}\right)^{1-\alpha}$

Since, A and α are constants, average product of labour will depend on the ration of factor $\left(\dfrac{K}{L}\right)$ and will not depend upon the absoluts quantities of the factors used.

Let us take a numerical example. Suppose the constant term A in the Cobb-Douglas production is equal to 50 and constant exponent a is equal to $\dfrac{1}{2}$ and the quantity of capital and is 8 units and quantities of labour and is 2 so that can get the capital labour ratio equal to 4:1. The average product of labour will be equal to

$AP_L = A\left(\dfrac{C}{L}\right)^{1-\alpha} = 50\left(\dfrac{8}{2}\right)^{1-\frac{1}{2}} = 50 \times (4)^{1/2} = 50.\sqrt{4} = 100$

Now, f the quantities of factors used are increase to C = 400 and L = 100, keeping the factor-ratio constant at 4:1, the average product of labour will still remain equal to 100.

$$AP_L = A\left(\frac{C}{L}\right)^{1-\alpha} = 50\left(\frac{400}{100}\right)^{1/2} = 50.\sqrt{4} = 100$$

2. Marginal Product of Factors and Cobb-Douglas Production Function: Like the average product of a factor, the marginal product of a factor of a linear Cobb-Douglas production function also depends upon the ratio of the factors and is independent of the absolutes quantities of the factors used. Note that marginal product of a factor say labour, is the first derivatives after production function, with respect to labour. The marginal product of labour from linear Cobb-Douglas production can be obtained as under:

$$Q = AL^{\alpha}C^{1-\alpha}$$

Marginal Product of Labour $\dfrac{dQ}{dL} = AaL^{1-a}C^{1-a}$

$$= \frac{AaL^{\alpha}C^{1-\alpha}}{L} = \frac{A\alpha L^{a-\alpha}C^{1-\alpha}}{L^{1-\alpha}}$$

$$= \frac{AaC^{1-\alpha}}{L^{1-\alpha}} = A\alpha\left(\frac{C}{L}\right)^{1-\alpha}$$

Since, A and α are constants marginal product of labour will depend on capital labour ratio, i.e., capital per worker and is independent of the magnitudes of the factors employed.

3. Cobb-Douglas Production Function and Marginal Returns to a Variable Factor: In case of linear-homogeneous Cobb-Douglas production function such as $Q = AL^{0.75} C^{0.25}$ marginal returns to a variable factor, say labour, diminishes or more labour is used. This will hold if the second derivatives of the given Cobb-Douglas production function ($Q = AL^{0.75} C^{0.25}$ is be negative.

The first derivative $\dfrac{dQ}{dL}$ (i.e., MP_L) = 0.75AL$^{-0.25}$ C$^{0.25}$

The Second Derivative $\dfrac{dQ^2}{d^2L}$ $= -0.25 \times 0.75\, AL^{-1.25}\, C^{0.25}$

$$= -1875\, AL^{-1.25}\, C^{0.25}$$

Since, the second derivatives is negative, MP_L will diminish with the increase in the amount of Labour.

4. Cobb-Douglas Production Function and Marginal Rate of Substitution: Marginal rate of substitution is an important concept which is exclusively used in the analysis of Cost-minimizing choice of inputs for producing a given level of output of a commodity. As has been shown above, marginal rate of substitution between factors is equal to the ratio of the marginal physical products of the factors. Therefore, inorder to derive marginal rate of substitution from Cobb-Douglas production function we need to obtain the marginal physical products of two factors from the Cobb-Douglas production function. We now proceed to drive them below:

Differentiating this with respect to L we have

$$\frac{dQ}{dL} = aAL^{a-1}C^{\beta} = \frac{a(AL^{a}C^{\beta})}{L}$$

Now, $AL^aC^\beta = Q$. Therefore,

$$\frac{dQ}{dL} = a\left(\frac{Q}{L}\right)$$

$\frac{dQ}{dL}$ represent marginal product of labour and $\frac{Q}{L}$ stands for the average product of labour. Thus,

$$MP_L = a(AP_L) \qquad\qquad …(2.30)$$

Similarly, by differentiating Cobb-Douglas production function with respect to capital we can show that marginal product of capital

$$\frac{dQ}{dL} = \beta\ \frac{Q}{C}\ q,\ MP_C = \beta(AP_C) \qquad\qquad …(2.31)$$

It follows from (2.30) and (2.31) above, that $MRS_{L,C} = \dfrac{MP_L}{MP_C} = \dfrac{\alpha.Q/L}{O\beta Q/C} = \dfrac{\alpha}{\beta}.\dfrac{C}{L}$ …(2.32)

5. Cobb-Douglas Production Function and Elasticity of Substitution: Now, we can show that in Cobb-Douglas production function elasticity of factor substitution $e_s\ \alpha_a$ is equal to unity.

$$es\alpha_a = \frac{\text{Proportionate change in capital}-\text{Labour Ratio}\left(\dfrac{C}{L}\right)}{\text{Proportionate change in } MRS_{LK}}$$

$$= \frac{\alpha\dfrac{C}{L}/\dfrac{C}{L}}{\alpha(MRS_{LC})/MRS_{LC}}$$

and substituting the value of marginal rate of substitution obtained in (2.32) above, we here

$$es\alpha_a = \frac{d\left(\dfrac{C}{L}\right)/\left(\dfrac{C}{L}\right)}{\alpha\left(\dfrac{\alpha}{\beta}.\dfrac{C}{L}\right)/\left(\dfrac{\alpha}{\beta}\dfrac{C}{L}\right)}$$

Since, $\dfrac{\alpha}{\beta}$ is constant and would not affect the derivatives it can therefore be factored out. Thus,

$$\alpha = \frac{d\left(\dfrac{C}{L}\right).\dfrac{\alpha}{\beta}}{\dfrac{\alpha}{\beta}.d\left(\dfrac{C}{L}\right)} = 1$$

6. Cobb-Douglas Production function in the Extended form: Cobb-Douglas production function can be extended to include more than two factors. For example, agricultural production depends not only as labour and capital used but also on there of other inputs such as land, irrigation, fertilisers. Incorporating these inputs is the Cobb-Douglas production function we have:

$$Q = AL^aC^{b1}D^{b2}G^{b3}\ F^{b4}$$

When, Q stands for output, L and C for labour and capital respectively, a and b are exponents of labour and capital respectively. D stands for land, G stands for irrigation, f for fertiliser and b_2, b_3 are exponents of land, irrigation and fertilisers respectively.

The above Cobb-Douglas production can be ultimated by regression analysis by first converting it into the following log form.

Log\Q = log A + a Log L + b_1 log C + $b_2$log D + $b_3$log G + b_4 log F Cobb-Douglas production function is a linear function and therefore it can be ultimated by least squares regression techniques.

Criticisms of C-D Production Function

The C-D production function has been criticised by Arrow, Chenery, Minhas and Solow as discussed below:

1. The C-D production function considers only two inputs, labour and capital, and neglects some important inputs, like raw materials, which are used in production. It is, therefore, not possible to generalize this function to more than two inputs.

2. In the C-D production function, the problem of measurement of capital arises because it takes only the quantity of capital available for production. But the full use of the available capital can be made only in periods of full employment. This is unrealistic because no economy is always fully employed.
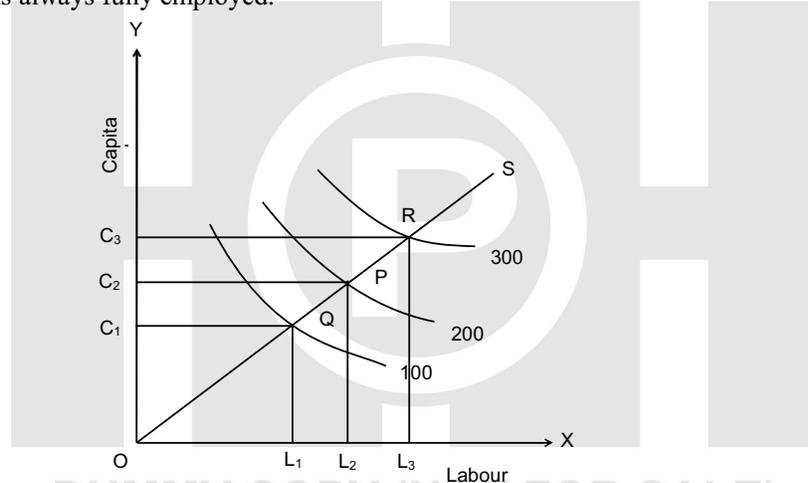
Fig. 2.2: Cobb-Douglas Production Function

3. The C-D production function is criticised because it shows constant returns to scale. But constant returns to scale are not an actuality, for either increasing or decreasing returns to scale are applicable to production.

It is not possible to change all inputs to bring a proportionate change in the outputs of all the industries. Some inputs are scarce and cannot be increased in the same proportion as abundant inputs. On the other hand, inputs like machines, entrepreneurship, etc. are indivisible. As output increases due to the use of indivisible factors to their maximum capacity, per unit cost falls.

Thus, when the supply of inputs is scarce and indivisibilities are present, constant returns to scale are not possible. Whenever the units of different inputs are increased in the production process, economies of scale and specialisation lead to increasing returns to scale.

In practice, however, no entrepreneur will like to increase the various units of inputs in order to have a proportionate increase in output. His endeavour is to have more than proportionate increase in output, though diminishing returns to scale are also not ruled out.

4. The C-D production function is based on the assumption of substitutability of factors and neglects the complementarity of factors.

5. This function is based on the assumption of perfect competition in the factor market which is unrealistic. If, however, this assumption is dropped, the coefficients α and β do not represent factor shares.

6. One of the weaknesses of C-D function is the aggregation problem. This problem arises when this function is applied to every firm in an industry and to the entire industry. In this situation, there will be many production functions of low or high aggregation. Thus, the C-D function does not measure what it aims at measuring.

Conclusion

Thus, the practicability of the C-D production function in the manufacturing industry is a doubtful proposition. This is not applicable to agriculture where for intensive cultivation, increasing the quantities of inputs will not raise output proportionately. Even then, it cannot be denied that constant returns to scale are a stage in the life of a firm, industry or economy. It is another thing that this stage may come after some time and for a short while.

Its Importance

Despite these criticisms, the C-D function is of much importance:

1. It has been used widely in empirical studies of manufacturing industries and in inter-industry comparisons.

2. It is used to determine the relative shares of labour and capital in total output.

3. It is used to prove Euler's Theorem.

4. Its parameters α and β represent elasticity coefficients that are used for inter-sectoral comparisons.

5. This production function is linear homogeneous of degree one which shows constant returns to scale. If $\alpha + \beta = 1$, there are increasing returns to scale and if $\alpha + \beta < 1$, there are diminishing returns to scale.

6. Economists have extended this production function to more than two variables.

2.5   CES PRODUCTION FUNCTION

Arrow, Chenery, Minhas and Solow in their new famous paper of 1961 developed the Constant Elasticity of Substitution (CES) function. This function consists of three variables Q, C and L, and three parameters A, α and θ.

It may be expressed in the form

$$Q = A[\alpha C^{-\theta} + (1 - \alpha)L^{-\theta}]^{-1/\theta}$$

where Q is the total output, C is capital, and L is labour. A is the efficiency parameter indicating the state of technology and organisational aspects of production.

It shows that with technological and/or organisational changes, the efficiency parameter leads to a shift in the production function, α (alpha) is the distribution parameter or capital intensity factor coefficient concerned with the relative factor shares in the total output, and θ (theta) is the substitution parameter which determines the elasticity of substitution.

And A > 0; 0 < α < 1; θ > –1.

Production Function

It may be noted that a change in efficiency parameter A cause a shift in the production function that can occur as a result of technological or organisational changes. The distribution parameter α indicates the relative importance of capital (C) and Labour (L) in various production processes.

Lastly, the substitution parameter θ indicates that substitution possibilities in the production processer. The elasticity of substitution between factors (α) for this production function depends upon this parameter. Thus, for this production function.

$$\alpha = \frac{1}{1+\theta}$$

When, θ = 0, q = 1 as $\alpha = \frac{1}{1+\theta} = \frac{1}{1+0} = 1$

When θ = α, q = 0 as $\alpha = \frac{1}{1+\theta} = \frac{1}{1+\alpha} = 0$

When, Q = – 1, q = α, as $\alpha = \frac{1}{1+\theta} = \frac{1}{1+(-1)} = \frac{1}{\theta} = \alpha$

It is evident from above that CES production is quite general and it includes α = 1, α = 0, α = α as special cases.

Isoquent of CES production are also of normal cervix shape further. CES production also exhibits constant returns to scale which can be proved as under

$$Q = A\left[\alpha C^{-\theta} + (1-\alpha)L^{-Q}\right]^{-\frac{1}{Q}}$$

Let, Capital (C) and Labour (L) inputs be increased by α positive number m. Thus,

$$Q = f(mc, mL) = A\left[\alpha(mc)^{-\theta} + (1-\alpha)(mL) - Q\right]^{-\frac{1}{Q}}$$

In this $\left[m^{-Q}\right]^{-\frac{1}{Q}}$ can be factored out and let take

$$Q = f(mc, mL) = A\,[m^{-Q}] - \frac{1}{Q}\,[\alpha C^{-Q} + (1-\alpha)L^{-Q}] - \frac{1}{Q}$$

$$= m\,f(C, L)$$

Its Properties

The CES production function possesses the following properties:

1. The CES function is homogeneous of degree one. If we increase the inputs C and L in the CES function by n-fold, output Q will also increase by n-fold.

    Thus, like the Cobb-Douglas production function, the CES function displays constant returns to scale.

2. In the CES production function, the average and marginal products in the variables C and L are homogeneous of degree zero like all linearly homogeneous production functions.

3. From the above property, the slope of an isoquant, i.e., the MRTS of capital for labour can be shown to be convex to the origin.

4. The parameter (theta) in the CES production function determines the elasticity of substitution. In this function, the elasticity of substitution, $\sigma = 1/1 + \theta$.

This shows that the elasticity of substitution is a constant whose magnitude depends on the value of the parameter θ. If $\alpha = 0$, then a = 1. If $\theta = \infty$, then a = 0. If $\theta = -0$, then a = ∞. This reveals that when a = 1, the CES production function becomes the Cobb-Douglas production function.

If Q < 0, then $\alpha = -0$; and if $\beta < \infty$, then $\alpha < 1$. Thus of the isoquants for the CES production function range from right angles to straight lines as the elasticity of substitution ranges from 0 to ∞.

5. As a corollary of the above, if L and C inputs are substitutable ∞ for each other, an increase in C will require less of L for a given output. As a result, the MP of L will increase. Thus, the MP of an input will increase when the other input is increased.

Merits of CES Production Function

The CES function has the following merits:

1. CES function is more general.
2. CES function covers all types of returns.
3. CES function takes account of a number of parameters.
4. CES function takes account of raw materials among its inputs.
5. CES function is very easy to estimate.
6. CES function is free from unrealistic assumptions.

CES Function vs. CD Function

There are some fundamental differences between the CES function and the CD production function:

1. The CD function is based on the observation that the wage rate is a constant proportion of output per head. On the other hand, the CES function is based on the observation that output per head is a changing proportion of wage rate.

2. The CES production function is based on larger parameters than the CD production function and as such allows factors to be either substitutes or complements. The CD function is, on the other hand, based on the assumption of substitutability of factors and neglects the complementarity of factors. Thus, the CES function has wider scope and applicability.

3. The CES production function can be extended to more than two inputs, unlike the CD function which is applicable to only two inputs.

4. In the CES function, the elasticity of substitution is constant but not necessarily equal to unity. It ranges from 0 to ∞. But the CD function is related to elasticity equal to unity. Thus, the CD function is a special case of the CES function.

5. The CES function covers constant, increasing and decreasing returns to scale, while the CD function relates to only constant returns to scale.

Limitations of CES Production Function

But the CES function has certain limitations:

1. The CES production function considers only two inputs. It can be extended to more than two inputs. But it becomes very difficult and complicated mathematically to use it for more than two inputs.

2. The distribution parameter or capital intensity factor coefficient, α is not dimensionless.

3. If data are fitted to the CES function, the value of the efficiency parameter A cannot be made independent of 0 or of the units of Q, C and L.

4. If the CES function is used to describe the production function of a firm, it cannot be used to describe the aggregate production function of all the firms in the industry. Thus, it involves the problem of aggregation of production function of different firms in the industry.

5. It suffers from the drawback that elasticity of substitution between any part of inputs is the same which does not appear to be realistic.

6. In estimating the parameters of CES production function, we may encounter a large number of problems like choice of exogenous variables, estimation procedure and the problem of multi-collinear ties.

7. There is little possibility of identifying the production function under technological change.

Conclusion

Despite these limitations, the CES production function is useful in its application to prove Euler's theorem, to exhibit constant returns to scale, to show that average and marginal products of C and L are homogeneous of degree zero, and to determine the elasticity of substitution.

2.6 CHARACTERISTICS OF PRODUCTION POSSIBILITY SETS

A production possibility set and the surface of a production possibility set called a product transformation curve (or surface) is a very popular concept employed often in undergraduate textbooks and in applied economics. It provides us with a summary of production sectors incorporating the initial endowment of resources and technological constraints. The usefulness of the concept of a production possibility set originates in two basic assumptions or restrictions on an economy:

(a) That commodities can be classified into producible commodities and non-producible commodities called primary factors of production, and

(b) That the supply of primary factors are inelastic or fixed. Thus, given the supply of primary factors of production and given technology sets of production sectors, the set of feasible production can be projected onto the space of producible goods and this generates a production possibility set summarising the endowment of primary factors and technological restrictions in the space of producible goods. This reduction of the endowment of primary factors and production sectors into a production possibility set enables us to lower the dimension of commodities to consider and also it enables us to link prices of producible

goods to prices of primary factors or factor prices. The surface of a production possibility set is sometimes called either a production possibility frontier or a product transformation curve (or surface).

Production Possibility Frontier (PPF)

A production Possibility Frontier (PPF) or Production Possibility Curve (PPC) is a curve which shows various combinations of set of two goods which can be produced with the given resources and technology where the given resources are fully and efficiently utilised per unit time. One good can only be produced by diverting resources from other goods, and so by producing less of them. This trade-off is usually considered for an economy, but also applies to each individual, household, and economic organisation.

Graphically bounding the production set for fixed input quantities, the PPF curve shows the maximum possible production level of one commodity for any given production level of the other, given the existing state of technology. By doing so, it defines productive efficiency in the context of that production set: a point on the frontier indicates efficient use of the available inputs (such as points B, D and C in the graph), a point beneath the curve (such as A) indicates inefficiency, and a point beyond the curve (such as X) indicates impossibility.

PPFs are normally drawn as bulging upwards or outwards from the origin ("concave" when viewed from the origin), but they can be represented as bulging downward (inwards) or linear (straight), depending on a number of assumptions. A PPF illustrates several economic concepts, such as scarcity of resources (the fundamental economic problem that all societies face), opportunity cost (or marginal rate of transformation), productive efficiency, allocative efficiency, and economies of scale.

An outward shift of the PPC results from growth of the availability of inputs, such as physical capital or labour, or from technological progress in knowledge of how to transform inputs into outputs. Such a shift reflects, for instance, economic growth of an economy already operating at its full productivity (on the PPF), which means that more of both outputs can now be produced during the specified period of time without sacrificing the output of either good. Conversely, the PPF will shift inward if the labour force shrinks, the supply of raw materials is depleted, or a natural disaster decreases the stock of physical capital.

However, most economic contractions reflect not that less can be produced but that the economy has started operating below the frontier, as typically, both labour and physical capital are underemployed, remaining therefore idle.
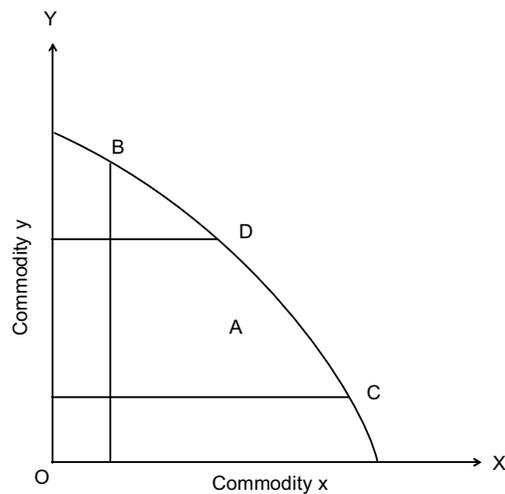
Fig. 2.3: Production Possibility Frontier (PPF)

Characteristics of PPF

The two basic characteristics or features of PPF are:

1. **PPF slopes downwards:** PPF shows all the maximum possible combination of two goods, which can be produced with the available resources and technology. In such a case, more of one good can be produced only by taking resources away from the production of another good. As there exists an inverse relationship between changes in quantity of one commodity and change in quantity of the other commodity, PPF slopes downwards from left to right (Fig. 2.3).

2. **PPF is concave-shaped:** PPF is concave-shaped because of increasing marginal opportunity costs, i.e., more and more units of one commodity are sacrificed to gain an additional unit of another commodity.

2.7 DUALITY RELATIONSHIP BETWEEN PRODUCTION AND COST FUNCTIONS

Duality of Production and Cost Functions Using the Implicit Function Theorem

The theory of duality links the production function models to the cost function models by way of a minimisation or maximisation framework. The cost function is derived from the production function by choosing the combination of factor quantities that minimise the cost of producing levels of output at given factor prices. Conversely, the production function is derived from the cost function by calculating the maximum level of output that can be obtained from specified combinations of inputs:

I. Profit (Wealth) Maximizing Firm

Production and cost functions (and profit functions) can be used to model how a profit (wealth) maximising firm hires or purchases inputs (factors), such as labour, capital (structures and machinery), and materials and supplies, and combines these inputs through its production process to produce the products (outputs) that the firm sells (supplies) to its customers.

Notes

II. The Production Function

The production function describes the maximum output that can be produced from given quantities of factor inputs with the firm's existing technological expertise. Let the variables q, L, K, and M represent the quantity of output, and the input quantities of labour, capital, and materials and supplies, respectively.

Mathematically, the production function, f, relates output, q, to inputs, L, K, and M, written as:

q = f(L, K, M)

with the function f having certain desirable properties.

III. Example: The CES Production Function

$q = A * [alpha* (L^{-rho}) + beta* (K^{-rho}) + gamma*(M^{-rho})]^{(-nu/rho)} = f(L, K, M)$

The coefficients of the production function, A, alpha, beta, gamma, nu, and rho are positive, real numbers. The production function's inputs, L, K, and M, are non-negative real numbers.

IV. The Total Cost of Production

Let the variables wL, wK, and wM represent the unit prices of the factors L, K and M, respectively. For any given combination of factor inputs, L, K and M, the total cost of using these inputs is:

TC = wL * L + wK * K + wM * M

i.e., the sum of the quantities of factor inputs weighted by their respective factor prices.

V. The Cost Function

The cost function describes the total cost of producing any given output quantity, using the cost minimising quantity of inputs.

Mathematically, the cost function, C, relates the total cost, TC, to output, q, and factor prices wL, wK, and wM, if the cost minimising combination of factor inputs is used, written as:

TC = C(q; wL, wK, wM)

with the function C having certain desirable properties.

VI. Example: The CES Cost Function

$C(q; wL, wK, wM) = h(q) * c(wL, wK, wM) = (q/A)^{1/nu} * [alpha^{(1/(1 + rho))} * wL^{(rho/(1 + rho))} + beta^{(1/(1 + rho))} * wK^{(rho/(1 + rho))} + gamma^{(1/(1 + rho))} * wM^{(rho/(1 + rho))}]^{((1 + rho)/rho)}$

The cost functions factor prices, wL, wK and wM, are positive real numbers.

VII. The Least-cost Combination of Inputs: Production Function to Cost Function

The entrepreneur, management and employees of the profit maximising firm choose the factor proportions and quantities, and output levels given the prices of factor inputs and products. For any specified combination of positive factor prices, wL, wK and wM, what combination of factor inputs, L, K, and M, will minimise the cost of producing any given level of positive output, q?

$C^*(q; wL, wK, wM) = \min_{L,K,M}\{wL * L + wK * K + wM * M : q - f(L, K, M) = 0, q > 0, wL > 0; wK > 0, \text{ and } wM > 0\}$

VIII. Constrained Optimization (Minimum): The Method of Lagrange

Define the Langrangian function, G, of the least-cost problem of (VII):

$G(q; wL, wK, wM, L, K, M, \mu) = wL * L + wK * K + wM * M + \mu * (q – f(L, K, M))$

where the new variable, μ, is called the Lagrange multiplier.

(a) First Order Necessary Conditions

0. $G_\mu(q; wL, wK, wM, L, K, M, \mu) = q – f(L, K, M) = 0$

1. $G_L(q; wL, wK, wM, L, K, M, \mu) = wL – \mu * f_L(L, K, M) = 0$

2. $G_K(q; wL, wK, wM, L, K, M, \mu) = wK – \mu * f_K(L, K, M) = 0$

3. $G_M(q; wL, wK, wM, L, K, M, \mu) = wL – \mu * f_L(L, K, M) = 0$

(b) Solution Functions

We want to solve, simultaneously, these four equations for the variables L, K, M and μ as continuously differentiable functions of the variables q, wL, wK and wM, and the parameters of the production function.

$L = L(q; wL, wK, wM)$

$K = K(q; wL, wK, wM)$

$M = M(q; wL, wK, wM)$

$\mu = \mu(q; wL, wK, wM)$

Suppose that the point $Z = (q, wL, wK, wM, L, K, M, \mu)$ satisfies equations 0 to 3, and that each of the functions G_μ, G_L, G_K and G_M has continuous partial derivatives with respect to each of the variables q, wL, wK, wM, L, K, M and μ at the point Z. Also, suppose the determinant of the Jacobian matrix, defined below, when evaluated at the point Z is not equal to zero.

According to the Implicit Function Theorem, functions L, K, M and μ exist that express the variables L, K, M and μ as continuously differentiable functions of the variables q, wL, wK and wM.

Moreover:

$L = L(q; wL, wK, wM)$

$K = K(q; wL, wK, wM)$

$M = M(q; wL, wK, wM)$

$\mu = \mu(q; wL, wK, wM)$

and

0'. $G_u(q; wL, wK, wM, L, K, M, \mu) = q – f(L(q; wL, wK, wM),$

   $K(q; wL, wK, wM), M(q; wK, wM)) = OwL$

1'. $G_L(q; wL, wK, wM, L, K, M, \mu) = wL – \mu(q; wL, wK, wM) *$

   $f_L(L(q; wL, wK, wM), K(q; wL, wK, wM), M(q, wL, wK, wM)) = 0$

2'. $G_K(q; wL, wK, wM, L, K, M, \mu) = wK – \mu(q; wL, wK, wM) * f_K(L(q; wL, wK, wM), K(q; wL, wK, wM), M(q; wL, wK, wM)) = 0$

3'. $G_M(q; wL, wK, wM, L, K, M, \mu) = wM – \mu(q; wL, wK, wM) *$

f_M(L(q; wL, wK, wM), K(q; wL, wK, wM),

M(q; wL, wK, wM)) = 0

for points (q; wL, wK, wM) in a neighborhood of the point (q; wL, wK, wM), i.e., the points (q, wL, wK, wM, L(q; wL, wK, wM), K(q; wL, wK, wM), M(q; wL, wK, wM), μ(q; wL, wK, wM)) also satisfy the first order conditions.

(c) Jacobian Matrices

The Jacobian matrix (bordered Hessian of G) of the four functions, G_μ, G_L, G_K and G_M, with respect to the choice variables, μ, L, K and M:

$$J_3 = \begin{matrix} G_{\mu\mu} & G_{\mu L} & G_{\mu K} & G_{\mu M} \\ G_{L\mu} & G_{LL} & G_{LK} & G_{LM} \\ G_{K\mu} & G_{KL} & G_{KK} & G_{KM} \\ G_{M\mu} & G_{ML} & G_{MK} & G_{MM} \end{matrix} = \begin{matrix} 0 & -f_L & -f_K & -f_M \\ -f_L & -\mu*f_{LL} & -\mu*f_{LK} & -\mu*f_{LM} \\ -f_K & -\mu*f_{KL} & -\mu*f_{KK} & -\mu*f_{KM} \\ -f_M & -\mu*f_{ML} & -\mu*f_{MK} & -\mu*f_{MM} \end{matrix} = J_{u,L,K,M}$$

The bordered principal minor of the bordered Hessian of the Langrangian function, G:

$$J_2 = \begin{matrix} 0 & -f_L & -f_K \\ -f_L & -\mu*f_{LL} & -\mu*f_{LK} \\ -f_K & -\mu*f_{KL} & -\mu*f_{KK} \end{matrix}$$

(d) Second Order Necessary Conditions

The second order necessary conditions require that the Jacobian matrix (bordered Hessian of G) be positive definite at Z = (q, wL, wK, wM, L, K, M, μ). The Jacobian matrix, J_3, is a positive definite matrix if the determinants of J_2 and J_3 are both negative.

(e) Sufficient Conditions

If the second order necessary conditions are satisfied, then the first order necessary conditions are sufficient for a minimum at Z. Therefore:

C*(q; wL, wK, wM) = wL * L(q; wL, wK, wM) + wK * K(q; wL, wK, wM) + wM * M(q; wL, wK, wM)

Moreover, since the sixteen functions that comprise the components of J_3 are continuous, the Jacobian matrix, J_3, is a positive definite matrix at points (q; wL, wK, wM) in a neighbourhood of the point (q; wL, wK, wM). That is, the points (q, wL, wK, wM, L(q; wL, wK, wM), K(q; wL, wK, wM), M(q; wL, wK, wM), μ(q; wL, wK, wM)) also satisfy the second order conditions.

Consequently, for points (q; wL, wK, wM) in a neighbourhood of the point (q; wL, wK, wM):

C*(q; wL, wK, wM) = wL * L(q; wL, wK, wM) + wK * K(q; wL, wK, wM) + wM * M(q; wL, wK, wM) = G(q; wL, wK, wM, L(q; wL, wK, wM), K(q; wL, wK, wM), M(q; wL, wK, wM), μ(q; wL, wK, wM))

IX. Marginal Cost Function

MC*(q; wL, wK, wM) = ∂C*(q; wL, wK, wM)/∂q = ∂G(q; wL, wK, wM, L, K, M, μ)/∂q = μ(q; wL, wK, wM)

Example: CES Marginal Cost

MC(q; wL, wK, wM) = ∂C(q; wL, wK, wM)/∂q = h'(q)*c(wL, wK, wM) = (1/nu)*(1/A)*(q/A)^(1–nu)/nu *c(wL, wK, wM)

X. Factor Demand Functions

If the cost function $C*(q; wL, wK, wM)$ satisfies certain properties (Hotelling-Shephard's lemma), properties derived from the properties of the production function $f(L, K, M)$:

$L(q; wL, wK, wM) = \partial C*(q; wL, wK, wM)/\partial wL = L$

$K(q; wL, wK, wM) = \partial C*(q; wL, wK, wM)/\partial wK = K$

$M(q; wL, wK, wM) = \partial C*(q; wL, wK, wM)/\partial wM = M$

where (L, K, M) are the factor proportions that minimise the cost of producing q units of output at specific factor prices (wL, wK, wM) for a given output quantity q.

Example: CES Factor Demand Functions

$L(q; wL, wK, wM) = h(q)*l(wL, wK, wM) = h(q)*[(alpha/wL)*c(wL, wK, wM)]^{(1/(1 + rho))}$

$K(q; wL, wK, wM) = h(q)*m(wL, wK, wM) = h(q)*[(beta/wK)*c(wL, wK, wM)]^{(1/(1 + rho))}$

$M(q; wL, wK, wM) = h(q)*m(wL, wK, wM) = h(q)*[gamma/wM)*c (wL, wK, wM)]^{(1/(1 + rho))}$

If the cost function $C*(q; wL, wK, wM)$ satisfies the properties required by the Hotelling-Shephard's lemma, then the factor demand functions satisfy:

$wL * \partial L(q; wL, wK, wM)/\partial wL + wK * \partial K(q; wL, wK, wM)/\partial wL + wM * \partial M(q; wL, wK, wM)/\partial wL = 0$

$wL * \partial L(q; wL, wK, wM)/\partial wK + wK * \partial K(q; wL, wK, wM)/\partial wK + wM * \partial M(q; wL, wK, wM)/\partial wK = 0$

$wL * \partial L(q; wL, wK, wM)/\partial wM + wK * \partial K(q; wL, wK, wM)/\partial wM + wM * \partial M(q; wL, wK, wM)/\partial wM = 0$

If the first-order conditions are sufficient for minimum, then;

$q \equiv f(L(q; wL, wK, wM), K(q; wL, wK, wM), M(q; wL, wK, wM)) \rightarrow$

$0 \equiv f_L * \partial_L(q; wL, wK, wM)/\partial wL + f_M * \partial K(q; wL, wK, wM)/\partial wL + f_M * \partial M(q; wL, wK, wM/\partial wL$, and

$wL = \mu(q; wL, wK, wM) * f_L(L(q; wL, wK, wM), K(q; wL, wK, wM), M(q; wL, wK, wM))$

$wK = \mu(q; wL, wK, wM) * f_K(L(q; wL, wK, wM), K(q; wL, wK, wM), M(q; wL, wK, wM))$

$wM = \mu(q; wL, wK, wM) * f_M(L(q; wL, wK, wM), K(q; wL, wK, wM)), M(q; wL, wK, wM))$

So, $\partial c*(q; wL, wK, wM)/\partial wL = \partial\{wL * L(q; wL, wK, wM) + wK * K(q; wL, wM) + wM * M(q; wL, wK, wM)\}/\partial wL$

$= L(q; wL, wK, wM) + wL * \partial L(q; wL, wK, wM)\partial wL + wM * \partial M(q; wL, wK, wM)/\partial wL$

$= L(q; wL, wK, wM) + \mu(q; wL, wK, wM) * (f_K * \partial L(q; wL, wK, wM)/\partial wL + f_M * \partial M(q; wL, wK, wM)/\partial wL)$

$= L(q; wL, wK, wM)$

XI. Example. Duality: CES Production Function to CES Cost Function

XII. The Solution Functions' Comparative Statics.

The Jacobian matrix of the four functions, G_μ, G_L, G_K and G_M, with respect to the variables, q, wL, wK and wM:

$$
J_{q, wL, wK, wM} = \begin{matrix} G_{Mq} & G_{MWL} & G_{\mu WK} & G_{\mu WM} \\ G_{Lq} & G_{LWL} & G_{LWK} & G_{LWM} \\ G_{Kq} & G_{KWL} & G_{KWK} & G_{KWM} \\ G_{Mq} & G_{MWL} & G_{MWK} & G_{MWM} \end{matrix} = \begin{matrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{matrix}
$$

The Jacobian matrix of the four solution functions, $\phi = \{\mu, L, K, M\}$, with respect to the variables, q, wL, wK and wM:

$$
J_\phi = \begin{matrix} \mu_q & \mu_{wL} & \mu_{wM} & u_{wK} \\ L_q & L_{wL} & L_{wM} & L_{wK} \\ K_q & K_{wL} & K_{wM} & K_{wK} \\ M_q & M_{wL} & M_{wM} & M_{wK} \end{matrix}
$$

From the Implicit Function Theorem:

$J_{q, wL, wK, wM;} + J_{\mu, L, K, M} * J_\phi = 0$ (zero matrix) $\rightarrow J_\phi = - (J_{\mu, L, K, M})^{-1}$

for points (q; wL, wK, wM) in a neighborhood of the point (q; wL, wK, wM), and L = L(q; wL, wK, wM), K = K(q; wL, wK, wM), M = M(q; wL, wK, wM), and $\mu = \mu$(q; wL, wK, wM).

XIII. Numerical Example: Production Function to Cost Function.

The CES production function as specified:

f(L, K, M) = 1 * [0.35*($L^{\wedge -0.17647}$) + 0.4 * ($K^{\wedge -0.17647}$) + 0.25* ($M^\wedge - 0.17647$)]$^{\wedge(-1/0.17647)}$

with A = 1, alpha = 0.35, beta = 0.4, gamma = 0.25, rho = 0.17647 (sigma = 0.85) and nu = 1. The CES production function has continuous first and second order partial derivatives with respect to its arguments.

f_L(L, K, M) = 1 * 0.35* $1^{\wedge(-0.17647/1)}$ * $L^{\wedge -(1 + 0.17647)}$ * f(L, K, M)$^{\wedge(1 + 0.1764/1)}$

f_K(L, K, M) = 1 * 0.4 * $1^{\wedge(-0.17647)/1}$ * $K^{\wedge -(1 + 0.17647)}$ * f(L, K, M)$^{\wedge(1 + 0.17647/1)}$

f_M(L, K, M) = 1 * 0.25 * $1^{\wedge(-0.1764777/1)}$ * $M^{\wedge -(1 + 0.17647)}$ * f(L, K, M)$^{\wedge (1 + 0.17647/1)}$

The factor pries as set: wL = 7, wK = 13 and wM = 6. Output is set: q = 30.

The Least-cost Combination of Inputs: Production Function to Cost Function

The dual cost function, C*, is obtained from the production function, f, by:

C*(q; wL, wK, wM) = min$_{L,K,M}$\{wL * L + wK * K + wM * M : q – f(L, K, M) = 0, q > 0, wL > 0; wK > 0, and wM > 0\}

C*(30; 7, 13, 6) = min$_{L,K,M}$\{7 * L + 13 * K + 6 * M : 30 – f(L, K, M) = 0\}

Constrained Optimisation (Minimum): The Method of Lagrange

G(q; wL, wK, wM, L, K, M, μ) = wL * L + wK * K + wM * M + μ * (q – f(L, K, M))

G(30; 7, 13, 6, L, K, M, μ) = 7 * L + 13 * K + 6 * M + μ * (30 – f(L, K, M))

First Order Necessary Conditions

    0. $G_\mu(30; 7, 13, 6, L, K, M, \mu) = 30 - f(L, K, M) = 0$

    1. $G_L(30; 7, 13, 6, L, K, M, \mu) = 7 - \mu * f_L(L, K, M) = 0$

    2. $G_K(30; 7, 13, 6, L, K, M, \mu) = 13 - \mu* f_K(L, K, M) = 0$

    3. $G_\mu(30; 7, 13, 6, L, K, M, \mu) = 16 - \mu * f_M(L, K, M) = 0$

Solve these four equations simultaneously (say, using Newton's Method) for L, K, M and μ. With L = 36.89, K = 24.42, M = 31.59 and μ = 25.506,

    $f(L, K, M) = 30$, $f_L(L, K, M) = 0.2744$, $f_K(L, K, M) = 0.5097$, $f_M(L, K, M) = 0.2352$

    0. $G_\mu(30; 7, 13, 6, L, K, M, \mu) = 30 - 30 = 0$

    1. $G_L(30; 7, 13, 6, L, K, M, \mu) = 7 - 25.506 * 0.2744 = 0$

    2. $G_K(30; 7, 13, 6, L, K, M, \mu) = 13 - 25.506 * 0.5697 = 0$

    3. $G_M(30; 7, 13, 6, L, K, M, \mu) = 6 - 25.506 * 0.2352 = 0$

Second Order Necessary Conditions

The second order necessary conditions require that the Jacobian matrix (bordered Hessian of G) be positive definite at Z = (q, wL, wK, wM, L, K, M, μ) = (30, 7, 13, 6, 36.89, 24.42, 31.59, 25.506). The Jacobian matrix, J_3, is a positive definite matrix if the determinants of J_2 and J_3 are both negative.

$$J_{\mu, L, K, M} = J_3 = \begin{bmatrix} 0 & -f_L & -f_K & -f_M \\ -f_L & -\mu * f_{LL} & -\mu * f_{LK} & -\mu * f_{LM} \\ -f_K & -\mu * f_{KL} & -\mu * f_{KK} & -\mu* f_{KM} \\ -f_M & -\mu * f_{ML} & -\mu * f_{MK} & -\mu* f_{MM} \end{bmatrix} = \begin{bmatrix} 0 & -0.274 & -0.51 & -0.235 \\ -0.274 & 0.148 & -0.14 & -0.065 \\ -0.51 & -0.14 & -0.307 & -0.12 \\ -0.235 & -0.065 & -0.12 & 0.168 \end{bmatrix}$$

[Determinant (J_3) = 0.0.12]

$$J_2 = \begin{bmatrix} 0 & -f_L & -f_K \\ -f_L & -\mu f_{LL} & -\mu * f_{LK} \\ -f_K & -\mu f_{KL} & -\mu * f_{KK} \end{bmatrix} = \begin{bmatrix} 0 & -0.274 & -0.51 \\ -0.274 & 0.148 & -0.14 \\ -0.51 & -0.14 & 0.367 \end{bmatrix}$$

[Determinant (J_2) = –0.1052]

The Least-cost Combination of Inputs

$C*(30; 7, 13, 6) = 7 * L + 13 * K + 6 * M = 7 * 36.89 + 13 * 24.42 + 6 * 31.59 = 765.17$

The Solution Functions' Comparative Statics

From the Implicit Function Theorem:

$J_\phi = -(J_{\mu, L, K, M})^{-1}$

$$J_\phi = \begin{bmatrix} \mu_q & \mu_{wL} & \mu_{wK} & \mu_{wM} \\ L_q & L_{wL} & L_{wK} & L_{wM} \\ K_q & K_{wL} & K_{wK} & K_{wM} \\ M_q & M_{wL} & M_{wK} & M_{wM} \end{bmatrix} = \begin{bmatrix} 0 & 1.23 & 0.814 & 1.053 \\ 1.23 & -2.968 & 1 & 1.295 \\ 0.814 & 1 & -0.934 & 0.857 \\ 1.053 & 1.295 & 0.857 & -3.367 \end{bmatrix}$$

XIV. Maximum Output: Cost Function to Production Function

The entrepreneur, management and employees of the profit maximising firm can investigate the technology (production function) available in the firm's cost function, C(q; wL, wK, wM), by

determining the factor prices, wL, wK and wM, consistent with the maximum level of output, q, for a given combination of factor inputs, L, K and M.

$f^*(L, K, M) = \max_q \{q : C(q; wL, wK, wM) <= wL * L + wK * K + wM * M, L > 0, K > 0, M > 0, \text{ for all } wL >= 0, wK >= 0, wM >= 0\}$

Question: $f^* = f$(original production function)?

Consider the case where the cost function, C(q; wL, wK, wM), factors:

$C(q; wL, wK, wM) = q^{1/nu} * c(wL, wK, wM)$

Setting:

$f^*(L, K, M) = \max_q\{q : q^{1/nu} * c(wL, wK, wM) <= wL * L + wK * K + wM * M, L > 0, K > 0, M > 0, \text{ for all } wL >= 0, wK >= 0, wM >= 0\}$

With c(wL, wK, wM) and wL * L + wK * K + wM * M linear homogeneous:

$f^*(L, K, M) = \max_q\{q : q^{1/nu} * c(wL, wK, wM) <= 1, L > 0, K > 0, M > 0, wL * L + wK * K + wM * M = 1\}$

$f^*(L, K, M) = \max_q \{q : q^{1/nu} <= 1/c(wL, wK, wM), L > 0, K > 0, M > 0, wL * L + wK * K + wM * M = 1\}$

Rewrite this as (Diewert, 1974, 157):

$f^*(L, K, M)^{1/nu} = \min_{wL, wK, wM} \{1/c(wL, wK, wM) : wL * L + wK * K + wM * M = 1, wL >= 0, wK >= 0, wM >= 0\}$

$f^*(L, K, M)^{1/nu} = 1/\max_{wL, wK, wM} \{c(wL, wK, wM) : wL * L + wK * K + wM * M = 1, wL >= 0, wK >= 0, wM >= 0\}$, since c(wL, wK, wM) >= 0

XV. Constrained Optimisation (Maximum): The Method of Lagrange

Define the Langrangian function, H, of the output maximisation problem of (XIV):

$H(L, K, M, wL, wK, wM, \lambda) = c(wL, wK, wM) + \lambda * (1 - (wL * L + wK * K + wM * M))$

where the new variable, λ, is called the Lagrange multiplier.

(a) First Order Necessary Conditions

0. $H_\lambda(L, K, M, wL, wK, wM, \lambda) = 1 - (wL*L + wK * K + wM * M) = 0$

1. $H_{wL}(L, K, M, wL, wK, wM, \lambda) = \partial c(wL, wK, wM)/\partial wL - \lambda * L = 0$

2. $H_{wK}(L, K, M, wL, wK, wM, \lambda) = \partial c(wL, wK, wM)/\partial wK - \lambda * K = 0$

3. $H_{wM}(L, K, M, wL, wK, wM, \lambda) = \partial c(wL, wK, wM)\partial wM - \lambda * M = 0$

(b) Solution Functions

We want to solve, simultaneously, these four equations for the variables wL, wK and wM, and λ as continuously differentiable functions of the variables L, K and M, and the parameters of the cost function.

$\lambda = \lambda(L, K, M)$

$wL = wL(L, K, M)$

$wK = wK(L, K, M)$

$wM = wM(L, K, M)$

Suppose that the point W = (L, K, M, wL, wK, wM, λ) satisfies equations 0 to 3, and that each of the functions H_λ, H_{wL}, H_{wK} and H_{wM} has continuous partial derivatives with respect to each of the variables L, K, M, wL, wK, wM and λ at the point W. Also, suppose the determinant of the Jacobian matrix, defined below, when evaluated at the point W is not equal to zero.

According to the Implicit Function Theorem, functions wL, wK, wM and λ exist that express the variables wL, wK, wM and λ as continuously differentiable functions of the variables L, K and M.

Moreover:

λ   =   λ(L, K, M)

wL =   wL(L, K, M)

wK =   wK(L, K, M)

wM =   wM(L, K, M)

and:

0′. H_λ(L, K, M, wL, wK, wM, λ) = 1 – (wL(L, K, M) * L + wK(L, K, M) * K + wM(L, K, M) * M) = 0

1′. H_{wL}(L, K, M, wL, wK, wM, λ) = ∂c(wL(L, K, M), wK(L, K, M), wM(L, K, M)/∂wL – λ(L, K, M) * L = 0

2′. H_{wK}(L, K, M, wL, wK, wM, λ) = ∂c(wL(L, K, M), wK(L, K, M), wM(L, K, M)/∂wK – λ(L, K, M)* K = 0

3′. H_{wM}(L, K, M, wL, wK, wM, λ) = ∂c(wL(L, K, M), wK(L, K, M), wM(L, K, M)/∂wM – λ(L, K, M) * M = 0.

for points (L, K, M) in a neighbourhood of the point (L, K, M). i.e., the points (L, K, M, μ) wL(L, K, M), wK(L, K, M), wM(L, K, M), λ(L, K, M)) also satisfy the first order conditions.

(c) Jacobian Matrices

The Jacobian matrix (bordered Hessian of H) of the four functions, H_λ, H_{wL}, H_{wK} and H_{wM} with respect to the choice variables, λ, wL, wK and wM:

$$J_3 = \begin{matrix} H_{\lambda\lambda} & H_{\lambda wL} & H_{\lambda wK} & H_{\lambda wM} \\ H_{wL\lambda} & H_{wLwL} & H_{wLwK} & H_{wLwM} \\ H_{wK\lambda} & H_{wKwL} & H_{wKwK} & H_{wKwm} \\ H_{wM\lambda} & H_{wMwL} & H_{wMwM} & H_{wMwM} \end{matrix} = \begin{matrix} 0 & -L & -K & -M \\ -L & C_{wLwL} & C_{wLwK} & C_{wLwM} \\ -K & C_{wKwL} & C_{wKwK} & C_{wKwM} \\ -M & C_{wMwL} & C_{wMwK} & C_{wMwM} \end{matrix} = J_{\lambda, wL, wK, wM}$$

The bordered principal minor of the bordered Hessian of the Langrangian function, H:

$$J_2 = \begin{matrix} 0 & -L & -K \\ -L & C_{WLWL} & C_{WLWK} \\ -K & C_{WKWL} & C_{WKWK} \end{matrix}$$

(d) Second Order Necessary Conditions

The second order necessary conditions require that the Jacobian matrix (bordered Hessian of H) be negative definite at W = (L, K, M, wL, wK, wM, λ). The Jacobian matrix, J_3, is a negative definite matrix if the determinant of J_2 is positive, and the determinant of J_3 is negative.

(e) Sufficient Conditions

If the second order necessary conditions are satisfied, then the first order necessary conditions are sufficient for a maximum at W. Therefore:

$f^*(L, K, M)^{\wedge 1/nu} = q(L, K, M)^{\wedge 1/nu} = 1/c(wL(L, K, M, wK(L, K, M), wM(L, K, M))$

Moreover, since the sixteen functions that comprise the components of J_3 are continuous, the Jacobian matrix, J_3, is a negative definite matrix at points (L, K, M) in a neighbourhood of the point (L, K, M), i.e., the points (L, K, M, wL(L, K, M), wK(L, K, M), wM(L, K, M), λ(L, K, M)) also satisfy the second order conditions.

Consequently, for points (L, K, M) in a neighbourhood of the point (L, K, M):

$f^*(L, K, M)^{\wedge 1/nu} = q(L, K, M)^{\wedge 1/nu} = 1/H(L, K, M, wL(L, K, M), wK(L, K, M), wM(L, K, M),$ λ(L, K, M))

XVI. The Lagrange Multiplier

The first order necessary conditions (XV. a. 1– 3.) imply:

$λ(wL, wK, wM) = c(wL, wK, wM)/∂wL/L = c(wL, wK, wM)/∂wK/K = c(wL, wK, wM)/∂wM/ M$

where:

wL = wL(L, K, M)

wK = wK(L, K, M)

wM = wM(L, K, M)

for points (L, K, M) in a neighbourhood of the point (L, K, M).

XVII. Factor Demand Functions

If the specified (or derived) cost function, $C(q; wL, wK, wM) = {}^{q^{\wedge 1/nu}\,*} c(wL, wK, wM)$, satisfies the Hotelling-Shephard properties, then the factor demand functions are given by:

$L(q; wL, wK, wM) = ∂C(q; wL, wK, wM)/∂wL = q^{\wedge 1/nu} * ∂c(wL, wK, wM)/∂wL = q^{\wedge 1/nu} *$ l(wL, wK, wM)

$K(q; wL, wK, wM) = ∂C(q; wL, wK, wM)/∂wK = q^{\wedge 1/nu} * ∂c(wL, wK, wM)/∂wK = q^{\wedge 1/nu} *$ k(wL, wK, wM)

$M(q; wL, wK, wM) = ∂C(q; wL, wK, wM)/∂wM = q^{\wedge 1/nu} * ∂c(wL, wK, wM)/∂wM = q^{\wedge 1/nu} *$ m(wL, wK, wM)

XVIII. Example. Duality: CES Cost Function to CES Production Function

XIX. The Solution Functions' Comparative Statics

The Jacobian matrix of the four functions, $H_λ$, H_{wL}, H_{wK} and H_{wM}, with respect to the variables, L, K and M:

		$H_{λL}$	$H_{λK}$	$H_{λM}$		− wL	− wK	− wM
		H_{WLL}	H_{WLK}	H_{WLM}		− λ	0	0
$J_{L, K, M}$	=	H_{WKL}	H_{WKK}	H_{WFM}	=	0	–λ	0
		H_{WML}	H_{WMK}	H_{WMM}		0	0	− λ

The Jacobian matrix of the four solution functions, $\phi = \{\lambda, wL, wK, wM\}$, with respect to the variables, L, K and M:

$$J_\phi = \begin{matrix} \lambda_L & \lambda_K & \lambda_M \\ WL_L & WL_K & WL_M \\ WK_L & WK_K & WK_M \\ WM_L & WM_K & WM_M \end{matrix}$$

From the Implicit Function Theorem:

$J_{L, K, M} + J_{\lambda, wL, wK, wM, q} * J_\phi = 0$ (zero matrix) $\rightarrow J_\phi = -(J_{\lambda, wL, wK, wM, q})^{-1} * J_{L, K, M}$

for points (L, K, M) in a neighbourhood of the point (L, K, M), and wL = wL(L, K, M), wK = wK(L, K, M), wM = wM(L, K, M) and $\lambda = \lambda$(L, K, M)

with:

$f^*(L, K, M) = (1/c(wL(L, K, M), wK(L, K, M), wM(L, K, M))^{\wedge nu}$

XX. Numerical Example: Cost Function to Production Function

The CES cost function as specified:

$c(q; wL, wK, wM) = \lambda(q) * c(wL, wK, wM) = (q/1)^{\wedge 1/11} * [0.35^{\wedge(1/(1 + 0.17647)))} * wL^{\wedge(0.17647)/(1 + 0.17647)))} + 0.4^{\wedge(1/(1 + 0.17647)))} * wK^{\wedge(0.17647)/(1 + 0.17647)))} + 0.25^{\wedge(1/(1 + 0.17647)))} * wM^{\wedge(0.17647)/(1 + 0.17647)))}]^{\wedge((1 + 0.17647))/0.17647))}$

With A = 1, alpha = 0.35, beta = 0.5, gamma = 0.25, rho = 0.17647 (sigma = 0.85) and ru = 1. The CES function has continuous first and second order partial derivatives with respect to their arguments.

$\partial c(q; wL, wK, wM)/\partial wL = h(q) * 0.35^{\wedge 1/(1 + 0.17647))} * wK^{\wedge(-1/(1 + 0.17647))} * c(wL, wk, wM)^{\wedge(1/(1 + 0.17647))}$;

$\partial c(q; wL, wK, wM)/\partial wK = h(q) * 0.4^{\wedge(1/1 + 0.17647))} * wK^{\wedge(-1/(1 + 0.17647))} * c(wL, wK, wM)^{\wedge(1/(1 + 0.17647))}$;

$\partial c(q; wL, WK, wM)/\partial wM = h(q) * 0.25^{\wedge(1/1 + 0.17647))} * wM^{\wedge(-1/(1 + 0.17647))} * c(wL, wK, wM)^{\wedge(1/1 + 0.17647))}$

The factor inputs are set: L = 36.89, K = 24.42, M = 31.59

Maximum Output: Cost Function to Production Function

$f^*(L, K, M)^{\wedge 1/nu} = 1/\max_{wL, wK, wM} \{c(wL, wK, wM): wL * L + wK * K + wM * M = 1, wL >= 0, wK >= 0, wM >= 0\}$

$f^*(36.89, 24.42, 31.59)^{\wedge 1/1} = 1/\max_{wL, wK, wM} \{c(wL, wK, wM): wL * 36.89 + wK * 24.42 + wM * 31.59 = 1, wL >= 0, wK >= 0, wM >= 0\}$

Constrained Optimisation (Maximum): The Method of Lagrange

$H(L, K, M, wL, wK, wM, \lambda) = c(wL, wK, wM) + \lambda * (1 - (wL * L + wK * K + wM * M))$

$H(36.89, 24.42, 31.59, wL, wK, wM, \lambda) = c(wL, wK, wM) + \lambda * (1 - (wL * 36.89 + wK * 24.42 + wM * 31.59))$

First Order Necessary Conditions

0. $H_\lambda(36.89, 24.42, 31.59, wL, wK, wM, \lambda) = 1 - (wL * 36.89 + wK * 24.42 + wM * 31.59) = 0$

1. $H_{wL}(36.89, 24.42, 31.59, wL, wK, wM, \lambda) = \partial c(wL, wK, wM)/\partial wL - \lambda * 36.89 = 0$

2. $H_{wK}(36.89, 24.42, 31.59, wL, wK, wM, \lambda) = \partial c(wL, wK, wM)/\partial wK - \lambda * 24.42 = 0$

3. $H_{wM}(36.89, 24.42, 31.59, wL, wK, wM, \lambda) = \partial c(wL, wK, wM)/\partial wM - \lambda * 31.59 = 0$

Solve these four equations simultaneously (say, using Newton's Method) for wL, wK, wM and λ with wL = 0.00915, wK = 0.01699, wM = 0.00784 and λ = 0.0333.

Second Order Necessary Conditions

The second order necessary conditions require that the Jacobian matrix (bordered Hessian of H) be negative definite at W = (L, K, M, wL, wK, wM, λ) = (36.89, 24.42, 31.59, 0.00915, 0.01699, 0.00784, 0.0333). The Jacobian matrix, J_3, is a negative definite matrix if the determinant of J_2 is positive, and the determinant of J_3 is negative.

$$J_\lambda W_L, W_K, W_M = J_3 =$$

0	$-L$	$-K$	$-M$		0	-36.888	-24.415	-31.592
$-L$	C_{wLwL}	C_{wLwK}	C_{wLwM}	=	-36.888	-75.693	25.518	33.019
$-K$	C_{wKwL}	C_{wKwK}	C_{wKwM}		-24.415	25.518	-23.827	21.855
$-M$	C_{wMwL}	C_{wMwK}	C_{wMwM}		-31.592	33.019	21.855	-85.874

[Determinant (J_3) = -18741780.5534]

$$J_2 = \begin{matrix} 0 & -36.888 & -24.415 \\ -36.888 & -75.693 & 25.518 \\ -24.415 & 25.518 & -23.827 \end{matrix}$$

[Determinant (J_2) = 123508.5984]

Maximum Output

With L = 36.89, K = 24.42, M = 31.59, set wL = 0.00915, wK = 0.01699, wM = 0.00784,

$f^*(L, K, M) = (1/c(wL, wK, wM))^{\wedge(nu)} \rightarrow$

$f^*(36.89, 24.42, 31.59) = (1/c(0.00915, 0.01699, 0.00784))^{\wedge 1} = (1/0.0333)^{\wedge 1} = 30$

Compare $f^*(L, K, M)$ with f(L, K, M) at L = 36.89, K = 24.42, M = 31.59:

f(36.89, 24.42, 31.59) = 30

The Solution Functions Comparative Statics

$$J_{L, K, M} = \begin{matrix} -wL & -wK & -wM \\ -\lambda & 0 & 0 \\ 0 & -\lambda & 0 \\ 0 & 0 & -\lambda \end{matrix} = \begin{matrix} -0.00915 & -0.01699 & -0.0078 \\ -0.0333 & 0 & 0 \\ 0 & -0.0333 & 0 \\ 0 & 0 & -0.0333 \end{matrix}$$

From the Implicit Function Theorem:

$J_\phi = -(J_{\lambda, wL, wK, wM, q})^{-1} * J_{L, K, M}$

	λ_L	λ_K	λ_M		-0.000305	-0.000566	-0.000261
$J_\phi =$	WL_L	w_{LK}	w_{LM}	=	-0.00028	$3.0E-5$	$1.0E-5$
	WK_L	w_{KK}	w_{KM}		$-3.0E-5$	-0.00077	$20E-5$
	WM_L	w_{MK}	w_{MM}		$1.0E-5$	$2.0E-5$	-0.00028

The Partial Derivates of f*(L, K, M)

Since $f^*(L, K, M) = (1/c(wL, wK, wM))^{\wedge}nu = (1/\lambda)^{\wedge}nu$

$f^*_L(L, K, M) = nu * (-\lambda_L/\lambda^{\wedge}2) * (1/\lambda)^{\wedge(nu-1)}$

$f^*_K(L, K, M) = nu * (-\lambda_K/\lambda^{\wedge}2) * (1/\lambda)^{\wedge(nu-1)}$

$f^*_M(L, K, M) = nu * (-\lambda_M/\lambda^{\wedge}2) * (1/\lambda)^{\wedge(nu-1)} \rightarrow$

$f^*_L(36.89, 24.42, 31.59) = 1 * (0.000305/0.0333^{\wedge}2) * (1/0.0333)^{\wedge(1-1)} = 0.274$

$f^*_K(36.89, 24.42, 31.59) = 1 * (0.000566/0.0333^{\wedge}2) * (1/0.0333)^{\wedge(1-1)} = 0.51$

$f^*_M(36.89, 24.42, 31.59) = 1 * (0.000261/0.0333^{\wedge}2) * (1/0.0333)^{\wedge(1-1)} = 0.235$

Partial Derivatives of f(L, K, M)

$f_L(36.89, 24.42, 31.59) = 0.274$

$f_K(36.89, 24.42, 31.59) = 0.51$

$f_M(36.89, 24.42, 31.59) = 0.235$

Conclude: f* = f.

CES Production/Cost Functions Numerical Example

CES Production Function

$q = A * [alpha * (L^{\wedge}{-rho}) + beta * (K^{\wedge}{-rho}) + gamma *(M^{\wedge}{-rho})]^{\wedge(-nu/rho)} = f(L, K, M)$

where L = labour, K = capital, M = materials and supplies, and q = product. The parameter nu is a measure of the economies of scale, while the parameter rho yields the elasticity of substitution:

$sigma = 1/(1 + rho)$.

The CES Cost Function

$C(q; wL, wK, wM) = h(q) * c(wL, wK, wM) = (q/A)^{\wedge}{1/nu} * [alpha^{\wedge(1/(1+rho))} * wL^{\wedge(rho/(1+rho))} +$
$beta^{\wedge(1/(1+rho))} * wK^{\wedge(rho/(1+rho))} + gamma^{\wedge(1/(1+rho))} * wM^{\wedge(rho/(1+rho))}]^{\wedge((1+rho)/rho)}$

The cost function's factor prices, wL, wK and wM, are positive real numbers.

Set the parameters below to re-run with your own CES parameters.

Restrictions: .7 < nu < 1.3; .5 < sigma < 1.5;

.25 < alpha < .45, .3 < beta < .5, .2 < gamma < .35

sigma = 1 → nu = alpha + beta + gamma (Cobb-Douglas)

sigma < 1 → inputs complements; sigma > → inputs substitutes

15 < q < 45; 4 <= wL* <= 11, 7 <= wK* <= 16, 4 <= wM* <= 10

The CES production function as specified:

$f(L, K, M) = 1 * [0.35L^{\wedge-0.17647} + 0.4 * (K^{\wedge-0.17647}) + 0.25(M^{\wedge-0.17647})]^{\wedge(-1/0.17647)}$

The CES cost function as specified:

$C(q; wL, wK, wM) = h(q) * c(wL, wK, wM) = (q/1)^{\wedge}{1/1} *$

$[0.35^{\wedge(1/1+0.17647)))} * wL^{\wedge(0.17647)/(1+0.17647)))} +$

$0.4^{\wedge(1/1+0.17647)))} * wK^{\wedge(0.17647)/(1+0.17647)} +$

$0.25^{\wedge(1/1 + 0.17647)))} * wM^{\wedge(0.17647)/(1 + 0.17647)))]\wedge(1 + 0.17647))/0.17647))}$

Both functions have continuous first and second order partial derivatives with respect to their arguments.

Curvature

The CES production function, $q = f(L, K, M)$ is (quasi) concave to the origin of the 3-dimensional (L, K, M) space if its Hessian matrix, F is negative (semi) definite.

At the specified parameters of the production function, with $q = 30$, $wL = 7$, $wK = 13$, $wM = 6$ and with $L = 36.89$, $K = 24.42$, $M = 31.59$.

$$
F = \begin{matrix} f_{LL} & f_{LK} & f_{LM} \\ f_{KL} & f_{KK} & f_{KM} \\ f_{ML} & f_{MK} & f_{MM} \end{matrix} = \begin{matrix} -0.0058 & 0.0055 & 0.0025 \\ 0.0055 & -0.0144 & 0.0047 \\ 0.0025 & 0.0047 & -0.0066 \end{matrix}
$$

The Hessian matrix, F, is negative definite if its eigenvalues are negative; negative semi-definite its eigenvalues or non-positive. If one or more eigenvalues of F are positive, $f(L, K, M)$ is not concave.

The eigenvalues of F are $e_1 = -0.018$, $e_2 = -0.00876$, $e_3 = -0$.

The CES cost function, $C(q; wL, wK, wM)$ is (quasi) concave for the origin of the 3-dimensional (wL, wK, wM) space if the Hessian matrix, C of its unit cost function, $c(wL, wK, wM)$ is negative (semi) definite.

At the specified parameters of the cost function, with $q = 30$, $wL = 7$, $wK = 13$, $wM = 6$:

$$
C = \begin{matrix} c_{wLwL} & c_{wLwK} & c_{wLwM} \\ c_{wKwL} & c_{wKwK} & c_{wKwM} \\ c_{wMWL} & c_{wMwK} & c_{wMwM} \end{matrix} = \begin{matrix} -0.1989 & 0.0333 & 0.0432 \\ 0.0333 & -0.0311 & 0.028 \\ 0.0432 & 0.0286 & -0.112 \end{matrix}
$$

The Hessian matrix, C, is negative definite; if its eigenvalues are negative; negative semi-definite its eigenvalue are non-positive. If one or more eigenvalues fo C are positive, $C(q; wL, wK, wM)$ is not concave.

The eigenvalues of C are $e_1 = -0.14924$, $e_2 = -0.09305$, $e_3 = -0$.

Mathematical Notes

Implicit Function Theorem

Given $N + M$ variables, $x_1, ..., x_N, y_1, ..., y_M$, a system of N equations expressed as:

$F_1(x_1, ..., x_N, y_1, ..., y_M) = 0$,

$F_2(x_1, ..., x_N, y_1, ..., y_M) = 0$,

. . . .

$F_N(x_1, ..., x_N, y_1, ..., y_M) = 0$,

and a vector (point) $Z = (a_1, ..., a_N, b_1, ..., b_M)$ that satisfies the system of equations.

Under what conditions can the system of equations be solved for the M variables $y_1, ..., y_M$ as continuously differentiable functions of the N variables $x_1, ..., x_N$ in a neighbourhood of Z:

$y_1 = \phi_1(x_1, ..., x_N)$,

$y_2 = \phi_2(x_1, ..., x_N)$,

. . . .

$y_M = \phi_M(x_1, ..., x_N),$

such that

$b_1 = \phi_1(a_1, ..., a_N),$

$b_2 = \phi_2(a_1, ..., a_N),$

. . . .

$b_M = \phi_M(a_1, ..., a_N),$

and such that the equations

$F_1(x_1, ..., x_N, \phi_1(x_1, ..., x_N), ..., \phi_M(x_1, ..., x_N)) = 0,$

$F_2(x_1, ..., x_N, \phi_1(x_1, ..., x_N), ..., \phi_M(x_1, ..., x_N)) = 0,$

. . . .

$F_N(x_1, ..., x_N, \phi_1(x_1, ..., x_N), ..., \phi_M(x_1, ..., x_N)) = 0,$

are satisfied for all $(x_1, ..., x_N)$ in a neighbourhood of $(a_1, ..., a_N)$?

The M continuously differentiable functions, $\phi_1, \phi_2, ..., \phi_M$, exist if each of the N functions, F_1, $F_2, ... , F_N$, has continuous partial derivatives with respect to each of the $N + M$ variables, $x_1, ..., x_N$, $y_1, ..., y_M$, near Z, and if the Jacobian determinant of the N functions $F_1, F_2, ..., F_N$ with respect to the M variables, $y_1, ..., y_M$, is not equal to zero when evaluated at Z.

The Jacobian determinant of the N functions $F_1, F_2, ... , F_N$ with respect to the M variables, $y_1, ..., y_M$ is the determinant of the Jacobian matrix, J_y, of partial derivatives of $F_1, F_2, ..., F_N$ with respect to $y_1, ..., y_M$. This is written as the matrix:

$$
J_y = \begin{matrix}
\partial F_1/\phi y_1, & \partial F_1/\partial y_2, & ..., & \partial F_1/\partial y_M \\
\partial F_2/\phi y_1, & \partial F_2/\partial y_2, & ...., & \partial F_2/\partial y_M \\
.... & & & \\
\partial F_N/\phi y_1, & \partial F_N/\partial y_2, & ...., & \partial F_N/\partial y_M
\end{matrix}
$$

The Jacobian matrix, J_x, of the N functions $F_1, F_2, ..., F_N$ with respect to the N variables, $x_1, ..., x_N$ is the matrix of partial derivatives of $F_1, F_2, ..., F_N$ with respect to $x_1, ..., x_N$. This is written as the matrix:

$$
J_\phi = \begin{matrix}
\partial F_1/\partial x_1, & \partial F_1/\partial x_2, & ..., & \partial F_1/\partial x_N \\
\partial F_2/\partial x_1 & \partial F_2/\partial x_2, & ...., & \partial F_2/\partial x_N \\
.... & & & \\
\partial F_N/\partial x_1 & \partial F_N/\partial x_2, & ...,. & \partial F_N/\partial x_N
\end{matrix}
$$

The Jacobian matrix, J_ϕ, of the M functions $\phi_1, \phi_2, ..., \phi_M$ with respect to the N variables, $x_1, ..., x_N$ is the matrix of partial derivatives of $\phi_1, \phi_2, ..., \phi_M$ with respect to $x_1, ..., x_N$. This is written as the matrix:

$$
J_x = \begin{matrix}
\partial \phi_1/\partial x_1, & \partial \phi_1/\partial x_2, & ..., & \partial \phi_1/\partial x_N \\
\partial \phi_2/\partial x_1, & \partial \phi_2/\partial x_2, & ...., & \partial \phi_2/\partial x_N \\
.... & & & \\
\partial \phi_M/\partial x_1, & \partial \phi_M/\partial x_2, & ...., & \partial \phi_M/\partial x_N
\end{matrix}
$$

Moreover, the Jacobian matrices, J_y, J_x, and J_ϕ satisfy

$J_x + J_y * J_\phi = 0$ (zero matrix)

for $(x_1, ..., x_N)$ in a neighbourhood of $(a_1, ..., a_N)$, and $y_1 = \phi_1(x_1, ..., x_N)$, $y_2 = \phi_2(x_1, ..., x_N)$, ... , and $y_M = \phi_M(x_1, ..., x_N)$.

Continuously Differentiable Functions

A function, $f(x_1, ..., x_N)$, is called continuously differentiable if its partial derivatives, $\partial f/\partial x_1, ..., \partial f/\partial x_N$, are continuous functions.

Young's Theorem

If it is possible to interchange the order of taking the first two partial derivatives of a function, the function satisfies Young's Theorem.

If the function, $f(x_1, ..., x_N)$, has continuous second-order derivatives, it satisfies Young's Theorem. Thus,

$\partial^2 f/\partial x_i \partial x_j = \partial^2 f/\partial x_j \partial x_i$ $ij = 1, ..., N$

Euler's Theorem

If the function $f(x_1, ..., x_N)$ is homogeneous of degree r, then:

$\partial f/\partial x_1 * x_1 + ... + \partial f/x_N * x_N = r * f(x_1 + ... + x_N)$

Positive Definite Matrix

A symmetric, real matrix is a positive definite matrix if all of its eigenvalues are positive.

Positive Semi-definite Matrix

A symmetric, real matrix is a positive semidefinite matrix if all of its eigenvalues are non-negative.

Negative Definite Matrix

A symmetric, real matrix is a negative definite matrix if all of its eigenvalues are negative.

Negative Semi-definite Matrix

A symmetric, real matrix is a negative semi-definite matrix if all of its eigenvalues are non-positive.

2.8 COMPARATIVE STATIC RESULT

Most of economic theory consists of comparative statics analysis. Comparative Statics is the determination of the changes in the endogenous variables of a model that will result from a change in the exogenous variables or parameters of that model. A crucial bit of information is the sign of the changes in the endogenous variables.

There is very limited opportunity to establish the signs of the impacts of changes in macroeconomics or any field that does not have an explicit maximisation or minimisation operation involved. But in microeconomics, comparative statics is a powerful tool for establishing important deductions of theories.

First, consider the case without maximisation or minimisation being involved, such as occurs in macroeconomics. The simplest case is situation in which one variable y is determined by some variable x. Suppose the value of y is determined as the solution to an equation

$$f(x,y) = 0$$

This equation holds for all values of x. So, it holds that the differential dy and dx satisfy the equation

$$(\partial f/\partial y)dy + (\partial f/\partial x)dx = 0$$

or equivalently

$$f_y d_y + f_x dx = 0$$

This relationship can be solved for dy, i.e.,

$$dy = -(f_x/f_y)dx$$

In order to know the sign of the impact of a change in x on y, we need to know the signs of both derivatives, f_x and f_y.

Now, consider the case in which y is determined such as to maximise some function g(x, y), where x has a value outside of the control of the decision maker. The first order condition for g(x, y) to be a maximum with respect to y is:

$$\partial g/\partial y = 0$$

or equivalently

$$g_y = 0$$

The second order condition is that:

$$g_{yy} > 0$$

If the value of x changes, then

$$g_{yy}dy + g_{yx}dx \ 0$$

So,

$$dy = -(g_{yx}/_{yy})dx$$

We know because y is chosen so as to maximise g that the second order condition requires that $g_{yy} > 0$. The sign of the impact of a change in x on y depends only upon the sign of g_{yx}.

Example 1: Consider a comparative statics analysis of monopoly pricing for a monopolist facing a market with a demand function of the form:

$$Q = N(ay - bp)$$

where N is the population in the market area, y is the per capita disposable income and p is the price of the product. a and b are positive parameters.

The total cost C for the firm is given by:

$$C = F + cQ$$

where F is fixed cost and c is the constant marginal variable cost.

A comparative statics analysis tells how the monopoly price would be affected by changes in the exogenous variables N and y and in the parameters F and c.

From the demand function $Q = N(ay - bp)$, the inverse demand function (price as a function of quantity sold) is

$p = (a/b)y - Q/bN$

The profit function for the monopolist is then

$\Pi = pQ - F - cQ = [(a/b)y - Q/bN]Q - F - cQ$

The first order condition for a maximum is

$d\Pi/dQ = (a/b) - 2Q/bN - c = 0$

which means the critical value of Q is

$Q = N(a - bc)/2$

which says that a meaningful solution exists only if $(a - bc) > 0$.

The second order condition for a maximum is

$d^2\Pi/dQ^2 < 0$

which reduces to $-2/bN < 0$

Since b and N are positive the second order condition is automatically satisfied.

The comparative statics results can be determined in this case by simply differentiating the first order condition with respect to the parameters, i.e.,

$\partial Q/\partial a = N/2$ which is positive

$\partial Q/\partial b = -Nc/2$ which is negative

$\partial Q/\partial c = -Nb/2$ which is negative

$\partial Q/\partial F = 0$

$\partial Q/\partial N = (a - bc)/2$ which is positive

Example 2: In the above example, the second order condition was automatically satisfied. Now, suppose the cost function is

$C = F + cQ - eQ^2 + fQ^3$

This is the case of U-shaped marginal and average costs.

In this case, the first and second conditions for a profit maximum reduce to:

$(a/b) - 2Q/bN - c + 2eQ - 3Q^2 = 0$

$-2/bN + 2e - 6Q < 0$

The second order condition is satisfied only if

$Q > (e - 1/bN)/3$

The first order condition is a quadratic equation in Q. It will have two solutions. One solution will be for a profit minimum and the other for a profit maximum. The solution that is greater than $(e - 1/bN)/3$ will be for the profit maximum.

The partial derivative of the first order condition with respect to a is

$1/b - (2/bN - 2e + 6Q)\partial Q/\partial a = 0$

Thus, $\partial Q/\partial a = (1/b)/[2/bN - 2e + 6Q)] = (1/b)/[2(3Q - (e - 1/bN))]$

The denominator of the fraction involves positive and negative terms. So, without further information, it would not be possible to determine the sign of the ratio. But the second order condition tells us that $3Q > (e-1/bN)$. So the numerator has to be positive and thus the ratio is positive. Therefore, $(\partial Q/\partial a) > 0$. Likewise the signs of the effects of changes in the other parameters can be determined.

Now, consider a couple of cases in which the economic variables are not determined from an optimisation procedure.

It should be noted that when variables are not determined by the results of optimisation, less can be said about the sign of the comparative statics effects.

Example 3: An important application of comparative statics analysis is in macroeconomics. This is a non-optimising application. So, the opportunity to make theoretical deductions as to the sign of the impact of changes in the exogenous variables is more limited.

A macroeconomic model is given in terms of a set of equations. The simplest macroeconomic model is the following in which,

Y =   national output, GDP

C =   consumption

I =   investment

G =   government purchases

NX =   net exports (exports – imports).

The model is then:

$Y = C + I + G + NX$

(an equilibrium condition)

$C = a + bY$

(the consumption function, a behavioural relation)

with I, G and NX exogenous, and a and b parameters.

This model is so simple. It hardly requires any analysis. It can be solved explicitly for the endogenous variables Y and C, i.e.,

$Y = k(a + I + G + NX)$

$C = a(1 + kb) + kb(I + G + NX)$

where the multiplier $k = 1/(1 – b)$.

Despite the simplicity of the above model, it is worthwhile going through the general procedure which would have to be applied to more complicated models. First, we need the differential form of the model, which in the above case is:

$dY = dC + dI + dG + dNX$

$dC = bdY$

The next step is put all the exogenous variables, in this case the differentials of Y and C, on the left side of the equations, leaving the right side for the differentials of the exogenous variables; i.e.,

$dY – dC = dI + dG + dNX – bdY + dC = 0$

Then the necessarily linear equations for the differentials are written as a matrix equation.

$$|1 \quad . \; -1| = |dy| = |111| . \; |dI|$$
$$|-b \quad 1| \quad |000| \quad |000| \quad |dG|$$
$$\qquad\qquad\qquad\qquad\qquad\qquad |dNX|$$

If the vector of differentials of the endogenous variables is denoted as dZ and the vector of differentials of the exogenous variables as dX, then the matrix equations can be expressed in the form

$\Gamma dZ = bdX$

The solution is then

$dZ = \Gamma^{-1}bdX$

The comparative statics analysis consists of finding the elements of the matrix Γ^{-1B}.

While the matrix formulation has certain advantages for the purpose of an introduction to comparative statics, it is better to obtain the solutions to the system of equations by way of Cramer's Rule. Cramer's Rule says that the solutions for the dependent variable can be expressed as a ratio of determinants. The denominator of the ratio is the determinant of the matrix of coefficients of the dependent variables. The numerator is the determinant of the matrix constructed by replacing the column of the coefficient matrix by the column of the constants on the RHS of the system of equations.

For example, if the effect of a change in I on Y is sought, then in the above equations dG and dNX are set equal to 0. The system of equations is then

$$|1 \quad -1| . \; |dy| = |dI|$$
$$|-b \quad 1| \quad |dc| \quad |O|$$

To obtain dY in terms of dI, take the ratio of the determinants of two matrices. One matrix is the coefficient matrix $\begin{vmatrix} 1 & 1 \\ -b & 1 \end{vmatrix}$.

Note: That dY corresponds to the first column of the coefficient matrix. So, the other matrix is the above matrix with the first column replaced.

$$\begin{vmatrix} dI & 1 \\ O & 1 \end{vmatrix}$$

Their determinants are $(1 - b)$ and dI, respectively. So, the solution for dY by Cramer's Rule is

$dY = dI/(1 - b)$

and hence

$\partial Y/\partial I = 1/(1 - b)$

The value of $1/(1 - b)$ is called the multiplier.

Likewise,

$\partial Y/\partial G = 1/(1 - b)$

and

$\partial Y/\partial NX = 1/(1 - b)$

The numerator in the ratio for dC is

$$\begin{vmatrix} 1 & dI \\ -b & O \end{vmatrix}$$

and thus dC = bdI/(1 − b) and

Hence, $\partial C/\partial I = b/(1 - b)$

This is also the value for $\partial C/\partial G$ and $\partial C/\partial NX$

Example 3(a): An extension of the analysis for the above macroeconomic model is one which is the same as above except that consumption depends upon disposable income YD and disposable income is GDP minus net taxes

YD = Y − T

Where net taxes T is given by

T = − s + tY

Thus,

$dY_D = dY - dT = dY + ds - tdY - Ydt$

Which reduces to

$dY_D = (1 - t)dY + ds - Ydt$

If there are no changes in the parameters a and b, then the analysis is the same as the previous model with b replaced with b(1 − t).

Example 4: This example deals with the interesting aspect of exports and imports being money values rather than physical units. So, exports and imports are expenditures rather than quantities.

Suppose exports depend upon the exchange rate E. Let E be the number of foreign currency units per dollar, say 100 yen per dollar. Suppose the demand function for American timber by Japanese users is:

Q = a – bP

where Q is in physical units per year, say board-feet/yr, and P is the price of timber in yen, say yen per board-foot. If p is the U.S. price of timber, $ per board-foot, the price to Japanese buyers is pE. Thus, the physical quantity of timber sold as a function of E is

Q = a – bpE

But for macroeconomic analysis what is needed is the dollar value of the sales, i.e.,

$pQ = pa - bp^2E$

Thus, the dollar value of the level of exports is negatively related to E i.e.,

$X = pa - bp^2E$

The comparative statics analysis for this case gives effects on the dollar value of exports of the various variables and parameters:

$\partial X/\partial a = p$ which is positive

$\partial X/\partial b = -p^2E$ which is negative

$\partial X/\partial E = -b^2$ which is negative

$\partial X/\partial p = -2bpE$ which is negative

Example 5: Now, suppose we have the demand function for some import to the U.S., say laptops from Japan,

$Q = a - bp$

where Q is the number of laptops per year and p is the price of laptops in dollars. If the price of laptops in Japan is P yen, then the price in dollars is P/E. Thus, the relationship between physical units of imports and the exchange rate is

$Q = a - bP/E$

But again, we want the dollar value of the imports, pQ rather than physical units. Therefore, the level of imports is

$M = pQ = PQ/E = P(a - bP/E)/E = aP/E - bP/E^2$

a more complicated relationship than occured in Example 3 for exports.

Now, consider the marginal effects on the dollar value of imports M of a change in the parameters of the demand function, the price of laptops in Japan and the exchange rate E.

$\partial M/\partial a = aP/E$ which is positive

$\partial M/\partial b = -P/E^2$ which is negative

$\partial M/\partial P = a/E$ which is positive

$\partial M/\partial E = -aP/E^2 + 2bP/E^3$ which is ambiguous

Example 6: An interesting comparative statics problem can now be formulated making use of the ideas presented above. Suppose a Japanese producer has monopoly for television sets in the U.S. as well as Japan. It can set the price for TVs in Japan. Given the exchange rate E, the price for TVs in the U.S. is then determined. Let the cost function be

$C = F - cQ$

Consider the following:

- The level of profits in Yen for the TV monopolist

- The profit maximizing prices in Japan and the U.S.

- The marginal effects of a change in the exchange rate on the prices in Japan and the U.S.

- The marginal effects on prices of changes in the fixed cost and the marginal variable cost

The quintessential economics problem is constrained optimisation. Likewise, the most interesting comparative statics analysis involves constraints. Consider the problem of maximising utility with respect to the consumption of two goods, x_1 and x_1 subject to a budget constraint, $p_1x_1 + p_2x_2 = Y$. The first order conditions for such a constrained maximisation problem are:

$\partial U/\partial x_1 = \lambda p_1$ and

$\partial U/\partial x_2 = \lambda p_2$

The second order conditions are that the relevant bordered Hessian matrix is negative definite.

Now, consider changes in p_1 and p_2, say dp_1 and dp_2, and a change in consumer income y, say dY. As a result of the changes in the parameters, the rates of consumption will undergo some infintesimal changes, dx_1 and dx_2. These infinitesimal changes must satisfy the condition.

$$p_1 dx_1 + p_2 dx_2 + x_1 dp_1 + x_2 dp_2 = d_y.$$

The first order conditions must be satisfied at any values for the parameters. Thus, it is valid to differentiate the first order conditions with respect to the parameters. (In differentiating, it must be remembered that the Lagrangian multiplier λ is now also a dependent variable like x_1 and x_2, and a function of the parameteres p_1, p_2 and Y.) The result is a set of equations that must be satisfied by the infinitesimal changes, i.e.,

$$(\partial^2 U/\partial x_1{}^2)dx_1 + (\partial^2 U/\partial x_2 \partial x_1)dx_2 - p_1 d\lambda = \lambda dp_1$$

and

$$(\partial^2 U/\partial x_1 \partial x_2)dx_1 + (\partial^2 U/\partial x_2{}^2)dx_2 - p_2 d\lambda = \lambda dp_2$$

and

$$-p_1 dx_1 - p_2 dx_2 = -dY + x_1 dp_1 + x_2 dp_2$$

These equations form a system which can be represented in matrix form as:

Example 7: This is a numerical example of the general case dealt with in the previous material. Let $U = x_1 x_2$, with $p_1 = 2$, $p_2 = 1$ and $Y = 12$. The values of x_1 and x_2 and of λ can be determined which maximise utility. Values of x_1, x_2 and λ can be determined which satisfy the first order conditions. The values of the second derivatives of U at the critical level can also be determined. The second order conditions require that the principal sub-determinants of the bordered Hessian matrix made up of the second derivatives and the prices should have specified signs.

The equations satisfied by effects of changes in the parameters can be created from the first order conditions. This solutions for the effects of the changes in the parameters can be expressed in terms of Cramer's rule as the ratio of determinants. The denominator of these ratios is a determinant whose sign is known from the second order conditions. Thus, in many cases, comparative statics results can be established with the combined use of the first order and second order conditions.

2.9 JOINT PRODUCTION

Joint production is suggested as one of the conceptual foundations of ecological economics. The notion of joint production springs immediately from the application of thermodynamics, and has a long history in economic analysis. Considerations of joint production give rise to philosophical concerns relating to responsibility and knowledge. The concept of joint production is easily comprehensible, and is also constitutive and supportive of a range of concepts current in ecological economic thought.

The production of wanted goods gives rise to additional unwanted outputs (bads), which may be harmful to the environment. The fundamental economic notion describing this relationship is that of joint production. Briefly put, this means that several outputs necessarily emerge together from a single productive activity. An example is the refining of crude oil, in which gasoline, kerosene, light heating oil and other mineral oil products are produced. However, harmful sulphurous wastes and carbon dioxide emissions are also necessarily generated.

Joint Production and Thermodynamics

Why is joint production such a ubiquitous and useful notion in ecological economics? We believe that this is because joint production is intimately related to the laws of thermodynamics. The application of thermodynamics is widely recognised as an essential element in much current

ecological economics thought, since it gives rich insights into the nature of economy-environment interactions. The usefulness of thermodynamics derives from its applicability to all real production processes, which are the basis of economic activity. Thus, thermodynamics relates ecological economics to the natural sciences, such as chemistry, biology and ecology, which also facilitates interdisciplinary research.

The laws of thermodynamics lead us to recognise that the human economy is an open subsystem embedded in the larger, but finite, system of the natural environment. The strength of the concept of joint production is that it allows us to incorporate this insight about economy-environment interactions into ecological economics. This can be seen from the following argument. From a thermodynamic point of view, energy and matter are the fundamental factors of production. Every process of production is, at root, a transformation of these factors. Hence, in this view, production processes are subject to the laws of thermodynamics, which in an abbreviated form can be stated as follows:

- **First law:** Energy and matter can be neither created nor destroyed, i.e., in an isolated system matter and energy are conserved.
- **Second law:** In every real process of transformation, a positive amount of entropy is generated.

One can describe the process of production as a transformation of a certain number of inputs into a certain number of outputs, each of which is characterised by its mass and its entropy. From the laws of thermodynamics it then follows that every process of production is a joint production, i.e., it results necessarily in more than one output. In particular, production processes that generate low entropy desired goods necessarily and unavoidably jointly produce high entropy waste materials. We can represent this thermodynamic constraint on real production processes as in Fig. 2.4. For example, in the production of iron one starts from iron ore. In order to produce the desired product, iron, which has lower specific entropy than iron ore, one has to reduce the raw material's entropy. This is achieved by employing a low entropy fuel, e.g., coal, which provides the energy necessary for this process. From a thermodynamic point of view, one may therefore consider production as a shifting of high entropy from the raw material to the waste product. At the same time, it becomes apparent that the inputs are also joint in the sense that high entropy iron and low entropy fuel are complementary. Hence, the fundamental idea of joint production applies both on the input and the output side. In that sense, the concept of joint production can capture the essential thermodynamic constraints on production processes as expressed by the first and second laws, through an easy-to-use and easy-to-understand economic concept.

Low entropy fuel

Low entropy desired good

Production Process

High entropy raw material
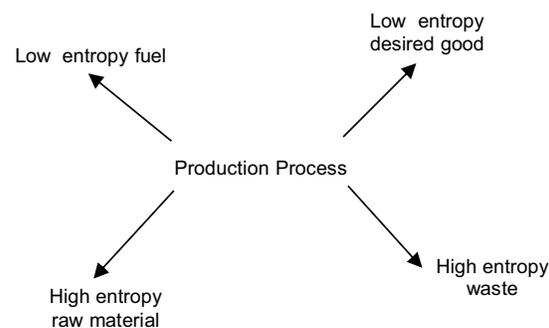
High entropy waste

Fig. 2.4: Production Processes Generating Low-entropy Desired Goods Necessarily and unavoidably Jointly Produce High-entropy Waste Materials

This holds for production in both economic systems and ecosystems. Joint production, therefore, is also a fundamental notion in ecology, even though it is not often expressed as such in that discipline. Organisms and ecosystems, as open, self-organising systems, necessarily take in several inputs and generate several outputs, just as an economy. Indeed, such natural systems are the earliest examples of joint production. The power and generality of the joint production concept can be demonstrated through the way it embraces four central issues in ecological economics: irreversibility, limits to substitution, the ubiquity of waste; and the limits to growth. Irreversibility is explicitly included within the above thermodynamic formalisation of joint production, as it is necessarily the case that the production process generates entropy and is therefore irreversible. Limits to substitution are also included, as the requirement that high-entropy material inputs must be converted into lower-entropy desired goods requires that the material inputs be accompanied by an irreducible minimum of low-entropy fuels. The ubiquity of waste can be easily derived from the thermodynamically founded joint production approach; it follows from the necessity of jointly producing high entropy, which very often is embodied in undesired material, and hence constitutes waste (e.g., CO_2, slag, etc.). The combination of the above three issues leads to the notion of limits to growth, further emphasising the power and generality of the joint production concept for ecological economics.

Joint Production in Economics

Having developed the concept of joint production as a necessary consequence of thermodynamics, we now review the way this theory has evolved in economics. This will help us to assess how far economic theory has already laid the ground for implementing this approach in ecological economic analysis. The analysis of joint production actually has a long tradition in economics. Adam Smith, Johann Heinrich von Thunen, John Stuart Mill, Wiliam Jevons, Karl Marx and Alfred Marshall — all devoted considerable effort to the study of joint production. As a matter of history, the analysis of joint production contributed to the abandonment of the classical theory of value and the establishment of the neoclassical theory of value. For ecological economists, it is very significant that several of these authors, in particular von Thunen, Marx and Jevons, emphasised that environmental pollutants come into existence as joint products of desired goods.

There is a substantial body of both theory and applications of joint production in the economics and business administration literature. In general, within this literature, two cases are distinguished: (i) all joint products are desired goods, and (ii) at least one output is undesired while at least one other is desired. While the former is the case which has received most treatment in the literature, our above thermodynamic discussion leads us to conclude that it is the second case that is of interest in ecological economics.

The theory of joint production has been extensively developed in business administration. For example, joint production is necessarily the case in chemical transformation processes, and in processes of splitting and separation. A range of computer-based models and methods has been developed to solve the resulting problems concerning the planning and cost allocation of joint production. Further, the quantitative relations between inputs and outputs in joint production can be described with input/output graphs, and one can use linear or non-linear algebraic systems generalising Koopmans' (1951) activity analysis. There are also relevant dynamic and stochastic graph theoretic models in computer science as well as models in process engineering and chemistry which are particularly important for balancing and managing the flows of material and energy. Even the problem of allocating ecological effects to joint products is being addressed. Important theoretical results about the economics of joint production include the following. Joint production of private and public goods may reduce the usual problem of under-provision of public goods in a decentralised

economy. Under joint production of goods and polluting residuals, and making the realistic assumption that the assimilative capacity of the natural environment for these pollutants is limited, a steady state growth path does not exist. A well-known problem in the theory of joint production is that, from the firm's point of view, the allocation of costs between joint products is essentially arbitrary. Perhaps as a result, with few exceptions, the modern literature on general equilibrium theory does not explicitly investigate the properties of economies characterised by joint production. Instead, it is focused on identifying the most general assumptions under which certain results hold, e.g., existence and optimality of general equilibrium. Yet, by doing so, it implicitly supplies insights into the economics of joint production. Arrow and Debreu (1954) and Debreu (1959) have shown that even in cases of joint production — be they goods or bads — under standard assumptions, there exists a general equilibrium in a competitive economy if: (i) the individual production sets are all convex and (ii) the possibility of free disposal is given, i.e., unwanted and harmful joint outputs can be disposed of at no cost. McKenzie (1959) showed the same result using a weaker assumption about disposal (disposal is possible but not necessarily free and the economy is "irreducible"), yet only for a technology characterised by constant returns to scale. Furthermore, any general competitive equilibrium, in particular under joint production, is pareto-optimal in the absence of negative externalities (Debreu, 1951). Lindahl (1919) and Pigou (1920) conceived mechanisms to internalise such externalities, thereby re-establishing optimality of the equilibrium. In the case of negative externalities exhibiting the character of public bads, however, this mechanism can only be established under very restrictive and unrealistic assumptions. In particular, every individual is assumed to reveal a personalised willingness to pay for the absence of the public bad, thereby having no incentive to act as a free-rider.

In summary, while modern economic theory has produced many interesting results concerning existence and optimality of equilibrium under joint production, in the case which is most relevant from the ecological economic point of view — joint production of bads causing public negative externalities — we are essentially left with a negative result.

Joint Production and Philosophy

The above thermodynamic foundation of joint production stresses that economic activity generally produces two types of output: the desired principal product, and the undesired waste product. We would expect, and indeed observe, that manufacturers will focus their attention and energies on the former, while the latter will be largely ignored, at least to the extent permitted by legal constraints and social mores. This inattention to the undesired products raises two issues of a philosophical nature, one relating to responsibility, i.e., ethical, and one relating to knowledge, i.e., epistemological. Turning first to ethics, the thermodynamically necessary waste products bring with them new issues of moral responsibility. This becomes obvious if we consider the hypothetical case of single production where waste products are not generated. In such an idealised world, assuming the existence of perfect markets and a fair social and legal order, the producers of a desired product do not face any ethical problem as long as they trade their products on the market and obey the legal order. Joint production now implies that economic activity, in addition to the intended products, also results in unintended outputs, which often are unnoticed. These joint products are therefore outside the social and legal order. At the same time, they may be harmful, e.g., to other producers, consumers, or to the natural environment. As a consequence, both the producer and the wider society demanding the desired principal product now face an ethical problem. Inattention to joint production may therefore easily result in ethical negligence. An example is the inattention to waste in the nuclear industry. From the beginning of nuclear power, it was recognised that very dangerous and long-lived

waste materials would be produced as by-products. Nevertheless, for the first 30 years of commercial power generation, unconscionably little attention was paid to the disposal of this waste (Proops, 2000). Concerning the second issue, epistemology, the area to which we draw attention is that of surprise and ignorance. Even if one were to suppose that it were possible to produce only principal products, this could still give rise to unanticipated and unwanted environmental effects (e.g., CFCs are a principal product, not a by-product). However, we believe that unwanted waste by-products are likely to be a greater source of unpleasant environmental surprises because, as mentioned above, they are not the focus of attention of their producers. The story of waste chlorine in the nineteenth century is one of ignorance of, and inattention to, the effects of emitting this waste product, with damaging and unforeseen consequences for air and water quality. What lessons can we learn from this discussion? Considering the concept of joint production naturally leads one to address issues of ethics and epistemology, requiring one to discuss economic questions in a philosophical context. In particular, the concept creates an awareness of both: (i) the ethical dimension of economic action due to unintended joint outputs, and (ii) our potential ignorance, primarily of the effects of unwanted products.

The Concept of Joint Production and Environmental Policy

We have outlined the relationship of joint production to thermodynamics, economics and philosophy, and argued that joint production is also an eminently comprehensible notion. We should now indicate some issues where the notion of joint production is likely to be especially useful for the discussion of environmental policy. In particular, we want to show that the concept of joint production could naturally lead to these issues currently being discussed in ecological economics as part of a single framework of analysis. This further demonstrates, we believe, the power of the joint production approach.

The Universality of the Concept

The concept of joint production may be employed at several different levels. It can be used for the analysis of an individual production process, of a firm, of an economic sector, or of a whole economy. It is also suited to examine environment-economy interactions in which economic activities and resulting environmental effects are separated by long time intervals, e.g., with CO_2 emissions. In both cases, today's effects on the natural system are caused by stocks of these substances, which were accumulated mainly from emissions up to several decades ago.

Holistic Approach to Policy

Taking a joint production approach to economy-environment interactions stresses the necessary relationships between various sorts of inputs into production processes, and the corresponding sorts of outputs. As illustrated in Fig. 2.4, much, even most, production requires inputs of low-entropy fuels and high-entropy raw materials, and generates low-entropy desired goods and high-entropy wastes. Thus, this thermodynamically based joint production representation shows us that the two issues, of natural resource use and of pollution from waste, are necessarily and intimately related: the resource is the mother of the waste. So, it is conceptually incomplete to consider natural resources and pollution as separate issues. Seeking to understand either on its own misses this relationship with potentially profound implications for policy analysis. In summary, the theory of joint production tells us that sound environmental policies can come only from an integrated and holistic conceptualisation of the production and consumption processes.

Time Scales and Time Horizons

An aspect of the awareness of potential ignorance and responsibility towards which our analysis has led us is that of time scale. Desired principal products are generally produced and consumed over relatively short time scales, leading to relatively short time horizons of decision-makers with regard to such outputs. However, joint products which constitute waste are often emitted into the environment where they can accumulate over significantly longer time scales. Such accumulation may, and often does, lead to the unanticipated and unpleasant surprises discussed earlier. Clearly, the social management of such problems demands much longer time horizons than those applied to the principal products.

The Precautionary Principle

The discussion here concerning how the awareness of potential ignorance and responsibility follows from the perspective of joint production, gives additional support to applying the precautionary principle. Indeed, a frequently perceived weakness of this principle is its lack of apparent conceptual foundation. We consider it supportive of both the notions of joint production and precaution that an analysis of the former so directly gives rise to the latter.

External Effects

Within the environmental economics literature, with its roots in welfare economics, the usual analytical method for understanding environmental damage is through the notion of external effects. There is postulated a relationship between economic actors which is asymmetrical and not mediated by a market, e.g., if one smokes in a lift, it causes uncompensated offence to one's fellow passengers. In the usual externality approach, this relationship is conceptualised as an issue of welfare/utility loss of the person affected by the external effect, i.e., the description is based on the effect. One could, however, recast this relationship starting from the cause of the effect. Very often, one would observe that the starting point is an unintended joint product. In the example of smoking in the lift, the desired product of nicotine in the bloodstream has an unwanted joint product of smoke in the lift. Therefore, we observe that there exists a duality between an explanation based on the effect, i.e., the externality approach, and an explanation starting from the cause of the effect, i.e., the joint production approach.

We also note that welfare effects will only be taken account of once they have been experienced, i.e., external effects are matters of the ex-post. On the other hand, the concept of joint production can alert one to the potential of environmental harm, i.e., considering joint production ex-ante creates a motive for actively exploring as yet unknown potential welfare effects. We therefore argue that the concepts of joint production and external effects are complementary.

In this part, we have discussed the concept of joint production to be considered as a foundational notion for ecological economics. We have drawn upon thermodynamics, economics and philosophy in our exploration of joint production, and have shown that it is constitutive and supportive of such fundamental notions in environmental thought as the precautionary principle and external effects.

Within the ecological economics literature, it seems that the entropy concept already has foundational status. However, we believe that entropy has been less fruitful as a tool of analysis than was originally anticipated. This is partly due to the fact that concepts and methods of thermodynamics, e.g., the notion of entropy or the idea of adiabatic changes are fairly complicated and highly abstract. As a consequence, most economists who are not trained in that field find them strange, unfamiliar and probably not even plausible, let alone useful. In our opinion, it is therefore indispensable to provide some kind of translation of the insights from thermodynamics into a language that economists understand and with which they are familiar.

The concept of joint production provides such a translation. In contrast to the entropy notion, which is notoriously difficult, it can easily be explained and its relevance to environmental problems is usually obvious. At the same time, it allows ecologists to get in touch with mainstream economists and to make use of the large body of knowledge available in economics.

In summary, the notion of joint production might constitute a foundational concept for ecological economics since it is applicable to the natural systems with which humans interact it is descriptive of economic activity, it relates to the areas of responsibility and human knowledge, and it is transparent and comprehensible to practitioners, policy makers and the wider public.

Hence, the concept of joint production unifies thermodynamic-ecological, economic and philosophical principles. Viewing joint production in this way opens up directions for fruitful research drawing on various concepts and methods of economics and of the natural sciences. Hence, we believe that the concept of joint production has the potential to become an important conceptual element of ecological economics.

Joint Production and purchasing a Multi-Product Problem and determining the Optimal Order Value under Practical Constearints.

Mathematical Model, its objective is to reduce inventory cost.

$$\min TC = \sum_{i=1}^{N} \sum_{t=1}^{T-1} (C_{it} Q_{it} + C'_{it} \theta'_{it}) \frac{D_{it}}{Q_{it} + Q'_{it} + b_{it}} + (A_{it} + A'_{it})$$

$$\frac{D_{it}}{Q_{it} + Q'_{it} + b_{it}} + h_{it} \frac{D_{it}}{Q_{it} + Q'_{it} + b_{it}}$$

$$\left\{ \frac{(Q_{it} + R_{it}) \left(T_{it}^{Q} - \frac{b_{it}}{D_{it}} \right)}{2} + \frac{2R_{it} + T_{it}^{P}(P_{it} - D_{it})T_{it}^{P}}{2} + \frac{\left(R_{it} + T_{it}^{P}(P_{it} - D_{it}) \right) \left(T_{it}^{D} - \frac{b_{it}}{D_{it}} \right)}{2} y_{it} \right.$$

$$+ \frac{Q_{it} \left(T_{it}^{Q} + \frac{R_{it}}{D_{it}} \right)}{2} + \frac{W_{it}^{P} \left(T_{it}^{P}(P_{it} - D_{it}) + R_{it} \right) \left(T_{it}^{P} + \frac{R_{it}}{P_{it} - D_{it}} \right)}{2}$$

$$+ \frac{W_{it}^{D} \left((R_{it} + T_{it}^{P}(P_{it} - D_{it})) \left(T_{it}^{D} - \frac{b_{it}}{D_{it}} \right) \right)}{2},$$

$$(1 - y_{it}) \} + IT_{it} \frac{D_{it}}{Q_{it} + Q'_{it} + b_{it}} \left\{ \left(\frac{b_{it} \left(T_{it}^{B} - \frac{T_{it}^{P}(P_{it} - D_{it})}{D_{it}} + R_{it} \right)}{2} \right) y_{it} + \right.$$

$$\left(\frac{W_{it}^{P} R_{it}^{2}}{2D_{it}} + \frac{W_{it}^{P} R_{it}^{2}}{2(P_{it} - D_{it})} + \frac{\frac{b_{it}(T_{it}^{D} - D_{it})}{D_{it}}}{2} \right) (1 - y_{it}) \} + \sum_{i=1}^{M} \sum_{k=1}^{K} \sum_{t=1}^{T} W_{it}^{P} C_{k}'' M_{ikt}$$

Subject to:

- $T_{it}^Q = \dfrac{Q_{it} - R_{it}}{D_{it}}, \forall i \in I, t \in T$

- $T_{it}^P = \dfrac{Q'}{P_{it}}, \forall i \in I, t \in T$

- $T_{it}^D = \dfrac{R_{it} + T_{it}^P (P_{it} - D_{it}) + bit}{D_{it}}, \forall i \in I, t \in T$

- $R_{it} + \in \, \le y_{it}, \ \beta, \ \forall i \in I, t \in T$

- $R_{it} \ge - (1 - y_{it}). \ \beta, \ \forall I \in i, t \in T$

- $T_{it}^P \le W_{it}^P \text{-}\beta, \ \forall i \in I, t \in T$

- $T_{it}^D \le W_{it}^D .\beta, \ \forall i \in I, t \in T$

- $T_{it}^S = T_{it}^Q, + T_{it}^P + T_{it}^D, \ \forall i \in I, t \in T$

- $T_{it}^S = T_{total}', \ \forall i \in I, t \in T$

- $\displaystyle\sum_{i=1}^{N} M_{itk} \left(T_{it}^P + Si\right) \le T_{total}', t \in T, k \in K.$

- $\displaystyle\sum_{K=1}^{K} M_{itk} \le 1, i \in N, t \in T$

- $M_{itk} \le T_{it}^P .\beta, \ i \in N, k \in k, t \in T$

- $\displaystyle\sum_{i=1}^{N} w_i \lambda_{it} \le f, \ \forall t \in T$

- $\lambda_{it} = Z_{it} Q_{it} + (1 \text{-} Z_{it})(R_{it} + T_{it})^P (P_{it} - D_{it})), \ \forall i \in T, t \in T$

- $\lambda_{it} \ge Q_{it}, \ \forall i \in T, t \in T$

- $\lambda_{it} \ge R_{it} + T_{it}^P (P_{it} - D_{it}), \ \forall i \in T, t \in T$

- $Q_{it}, \ Q_{it}' \ge 0, y_{it}, Z_{it} \in \{0,1\}, \forall i \in I, t \in T, i = 1, 2 \ldots, N\}, t = [0, 1, 2, \ldots, T\}.$

Parameters

The Mathematical model parameters are:

D_{it} = Demand rate of 1st product in period t

F = Maximum storage capacity

L_{it} = Holding cost of 1st product in period t

A_{it}' = Set up cost of 1st product in period t

C_{it}' = Production cost of 1st product in period t

λ_{it} = Maximum inventory level of 1st product in period t

β = Infinite positive large number.

THC	=	Total inventory Cost
TOC	=	Total ordering and setup cost
TMC	=	Total Machinery Cost
P_{it}	=	Production rate of 1st product in period t
w_i	=	Coefficient of base product volume
A_{it}	=	Ordering cost of 1st product in period t
C_{it}	=	Purchasing cost of 1st product in period t
C_{itk}^m	=	Cost of using machine k for producing t
S_i	=	Set up time of production
TPC	=	Total purchasing and production cost
TBC	=	Total back order cost
T_{it}^Q	=	Interval between purchasing and using production i in time t until the level of inventory is in level P_{it}
T_{it}^P	=	Producing production i is time t until stop time of production
T_{it}^D	=	Interval between the stop time of producing production in time t until the end of period.1' = 1, 2, ..., N.

Example 1

If $u = x^3 + y^3 + z^3 - 3xyz$, prove that

$$x\frac{\delta u}{\delta x} + y\cdot\frac{\delta u}{\delta y} + z\cdot\frac{\delta u}{\delta z} = 3U$$

Solution:

$u = x^3 + y^3 + z^3 - 3xyz$

$\Rightarrow \dfrac{\delta u}{\delta x} = 3x^2 - 3yz$

$\dfrac{\delta u}{\delta y} = 3y^2 - 3xz$

$\dfrac{\delta u}{\delta z} = 3z^2 - 3xy$

$x\cdot\dfrac{\delta u}{\delta x} + y\cdot\dfrac{\delta u}{\delta y} + z\cdot\dfrac{\delta u}{\delta z} = (3x^2 - 3yz) + y\,(3y^2 - 3xz) + z\,(3z^2 - 3xy)$

$$= 3x^3 - 3xyz + 3y^3 - 3xyz + 3z^3 - 3xyz$$
$$= 3(x^3 - xyz + y^3 - xyz + z^3 - xyz)$$
$$= 3(x^3 - 3xyz + y^3 + z^3)$$
$$= 3(x^3 + y^3 + z^3 - 3xyz) \text{ (since, } u = x^3 + y^3 + z^3 - xyz)$$
$$= 3u$$

Notes

Example 2

Verify Euler's Theorem

$$x \cdot \frac{\delta f}{\delta x} + y \cdot \frac{\delta f}{\delta y} = 3f \text{ for the function}$$

$$f(x,y) = x^3 + 3y^3 - x^2y$$

Solution:

$$f(x,y) = x^3 + 3y^3 - x^2y$$

$$\frac{\delta f}{\delta x} = 3x^2 - 2xy$$

$$\frac{\delta f}{\delta y} = 9y^2 - x^2$$

$$
\begin{aligned}
x \cdot \frac{\delta f}{\delta x} + y \cdot \frac{\delta f}{\delta y} &= x(3x^2 - 2xy) + y(9y^2 - x^2) \\
&= 3x^3 - 2x^2y + 9y^3 - x^2y \\
&= 3x^3 + 9y^3 - 3x^2y \\
&= 3(x^3 + 3y^3 - x^2y) \\
&= 3f \text{ (since, } f = x^3 + 3y^3 - x^2y)
\end{aligned}
$$

Example 3

Find the marginal cost for the total cost function

$$C = 3x^4 - 4x^3 + 2x^2 - 9x$$

Solution:

$$C = 3x^4 - 4x^3 + 2x^2 - 9x$$

$$MC = \frac{dc}{dx} = 12x^3 - 12x^2 + 4x - 9$$

Example 4

What is marginal 4x at output v = 4 for average cost function $\left[\dfrac{18}{q} - 20 + 5q \right]$

Solution:

Average cost function $= \dfrac{18}{q} - 20 + 5q$

Total cost function $= Acxq = \left(\dfrac{18}{q} - 20 + 5q \right) q$

$$= 18 - 20q + 5q^2$$

$$\therefore \text{ Marginal cost} = \frac{d}{dx}(18 - 20q + 5q^2)$$

$$= -20 + 10q$$

If, q= 4, Marginal Cost = –20 + 10(4)

= –20 + 40 = 20

Example 5

Total Cost Function is

$C = \frac{1}{3}Q^3 + 6Q^2 + 12Q$, find AC and MC

Solution:

$C = \frac{1}{3}Q^3 + 6Q^2 + 12Q$

$AC = \frac{C}{Q} = \frac{\frac{1}{3}Q^3 + 6Q^2 + 12Q}{Q}$

$\therefore AC = \frac{1}{3}Q^2 + 6Q + 12$

$MC = \frac{d}{dQ}(C)$

$= \frac{d}{dQ}\left(\frac{1}{3}Q^3 + 6Q^2 + 12Q\right)$

$= Q^2 + 12Q + 12$

2.10 SUMMARY

1. Production is the result of co-operation of four factors of production, *viz.,* land, labour, capital and organisation.

2. The relationship between output and costs is expressed in terms of cost function.

3. Long-run cost functions provide useful information for planning the growth as well as the investment policies of a firm.

4. Have at different times examined many actual production functions and a famous production function is the Cobb-Douglas production function.

5. The CES function is homogenous of degree one.

2.11 SELF ASSESSMENT QUESTIONS

1. Explain in detail about the production function.

2. Explain in detail about the cost function.

3. Critically analyse Homogeneous Production Function.

4. Explain Euler's theorem.

5. Explain Cobb-Douglas Production Function.

❧❧❖❧❧

MARKET EQUILIBRIUM

Objectives

The objectives of this lesson are to learn:

- Market equilibrium
- Stability of equilibrium
- Imperfect competition

Structure:

3.1 MARKET EQUILIBRIUM: FACTOR MARKET EQUILIBRIUM

Market Equilibrium

Market equilibrium refers to a condition where a market price is established through competition such that the amount of goods or services sought by buyers is equal to the amount of goods or services produced by sellers. This price is often called the competitive price or market clearing price and will tend not to change unless demand or supply changes, and the quantity is called "competitive quantity" or market clearing quantity.

Factor market equilibrium (Equilibrium of a Firm in Factor Market: Perfect Competition and Imperfect Competition)

When an organisation decides to hire a factor of production, it makes comparison between MRP of the factor with that of its Marginal Factor Cost (MFC).

If the MRP is greater than the marginal cost of factor (MRP > MC), then the factor is employed because it would generate more marginal revenue.

On the other hand, when MRP is lesser than the marginal cost of factor (MRP < MC), then the organisation would not employ the factor as it would increase costs. In case, the MRP is equal to the marginal cost of factor (MRP = MC), then the organisation would attain equilibrium.

However, in modern times, the organisations determine the actual amount of factors that are required to achieve equilibrium. For determining the equilibrium point, it is necessary for an

organisation to analyse the factor market in different market structures, such as perfect competition and imperfect competition. Let us discuss the equilibrium of a firm in different market structures.

Question 1

Suppose there is a perfectly competitive industry where all the firms are identical with identical cost curves. Furthermore, suppose that a representative firm's total cost is given by the equation $TC = 100 + q^2 + q$ where q is the quantity of output produced by the firm. You also know that the market demand for this product is given by the equation $P = 1000 - 2Q$ where Q is the market quantity. In addition you are told that the market supply curve is given by the equation $P = 100 + Q$:

(a) What is the equilibrium quantity and price in this market given this information?

To find the equilibrium set market demand equal to market supply: $1000 - 2Q = 100 + Q$. Solving for Q, you get $Q = 300$. Plugging 300 back into either the market demand curve or the market supply curve you get $P = 400$.

(b) The firm's MC equation based upon its TC equation is $MC = 2q + 1$. Given this information and your answer in part (a), what is the firm's profit maximizing level of production, total revenue, total cost and profit at this market equilibrium? Is this a short-run or long-run equilibrium? Explain your answer.

From part (a) you know the equilibrium market price is $400. You also know that the firm profit maximizes by producing that level of output where MR = MC. Since the equilibrium market price is the firm's marginal revenue you know that $MR = \$400$. Setting MR = MC gives you $400 = 2q + 1$, or $q = 199.5$. Thus, the profit maximizing level of output for the firm is 199.5 units when the price is $400 per unit. Using this information it is easy to find total revenue as the price times the quantity: $TR = (\$400 \text{ per unit})(199.5 \text{ units}) = \$79,800$. Total cost is found by substituting $q = 199.5$ into the TC equation: $TC = \$40,099.75$. Profit is the difference between TR and TC: $\text{Profit} = TR - TC = 79,800 - 40,099.75 = \$39,700.25$. Since profit is not equal to zero this cannot be a long-run equilibrium situation: it must be a short-run equilibrium situation.

(c) Given your answer in part (b), what do you anticipate will happen in this market in the long-run?

Since there is a positive economic profit in the short run, there should be entry of firms in the long-run resulting in an increase in the market quantity, a decrease in the market price, and firms in the industry earning zero economic profit.

(d) In this market, what is the long-run equilibrium price and what is the long-run equilibrium quantity for a representative firm to produce? Explain your answer.

The long-run equilibrium price is that price that results in the representative firm earning zero economic profit. This will occur when MC = ATC for the representative firm. ATC is just the TC equation divided by q. Thus, $2q + 1 = (100 + q^2 + q)/q$. Solving for q, $q = 10$. Plugging 10 in for q into the ATC equation yields the following: $ATC = (100 + 10^2 + 10)/10 = 21$. So, when Price equals $MR = \min ATC = MC = \$21$, this firm will break even. To see this compute TR for the firm when it produces 10 units and sells each unit for $21: TR

$= \$210$. Notice that this is the same as the firm's TC: thus, the firm earns zero economic profit.

(e) Given the long-run equilibrium price you calculated in part (d), how many units of this good are produced in this market?

To find this quantity you need to substitute \$21 (the long-run equilibrium price) into the market demand curve to determine the quantity that the market must produce in order to be in long-run equilibrium. This quantity is equal to 489.5 units.

Question 2

The market for study desks is characterized by perfect competition. Firms and consumers are price takers and in the long run there is free entry and exit of firms in this industry. All firms are identical in terms of their technological capabilities. Thus the cost function as given below for a representative firm can be assumed to be the cost function faced by each firm in the industry. The total cost and marginal cost functions for the representative firm are given by the following equations:

$TC = 2q_s^2 + 5q_s + 50$

$MC = 4q_s + 5$

Suppose that the market demand is given by:

$P_D = 1025 - 2Q_D$

Note: Q represents market values and q represents firm values. The two are different:

(a) Determine the equation for average total cost for the firm.

ATC for the firm is TC/q, so dividing the total cost equation above by q gives us:

$ATC = 2q_s + 5 + 50/q_s$

(b) What is the long-run equilibrium price in this market? (**Hint:** Since the market supply is unknown at this point, it's better not to think of trying to solve this problem using demand and supply equations. Instead you should think about this problem from the perspective for a firm. Specifically, a long run equilibrium occurs where ATC = MC = Price)

In a long-run equilibrium, ATC equals Marginal Cost and profits equal zero. Setting the two equations equal:

$ATC = 2q_s + 5 + 50/q_s = 4q_s + 5 = MC$

$50/q_s = 2q_s$

$50 = 2q_s$

$25 = q_s^2$

Take the square root of both sides and find:

$5 = q_s$

However, the question wants us to find long run prices. We know that the firm produces were Price = MR = MC, so if we can determine the firm's MC, then we can determine the equilibrium price in the market.

We know that:

$MC = 4q_s + 5$

And solved for:

$5 = q_s$

Substituting:

MC = 4(5) + 5 = 25

The equilibrium price in the market is 25.

(c) What is the long-run output of each representative firm in this industry?

We solve for this in the previous part, $5 = q_s$

(d) When this industry is in long-run equilibrium, how many firms are in the industry? (**Hint:** Firms are identically sized).

Now we should determine the market quantity Q from the market demand curve, given that we know the market price is 25. Market demand is given as:

P_D = $1025 - 2Q_D$

And we know that market price = 25, so:

25 = $1025 - 2Q_D$

$1000 = 2Q_D$

500 = Q_D

Since each firm is making 5 units (as we found in parts b and c), there must be 100 firms, since they are all identically sized.

Now suppose that the number of students increases such that the market demand curve for study desks shifts out and is given by,

$P_D = 1525 - 2Q_D$

(e) In the short-run will a representative firm in this industry earn negative economic profits, positive economic profits, or zero economic profits? (**Hint:** You can solve this without calculation.)

The demand curve has shifted to the right. Given what we learned earlier in the semester, we should know that the market price will increase. If market prices are increasing, then firms are earning higher marginal revenues than they earn in a long-run equilibrium. This means that firms are earning positive economic profits.

(f) In the long-run will a representative firm in this industry earn negative economic profits, positive economic profits, or zero economic profits? (**Hint:** Again, no calculation required).

In the long-run economic profits are always zero since there is free entry/exit in a perfectly competitive market. Firms will either enter the industry until there are no possible profit opportunities. If there are economic losses, firms will leave the industry until profits hit zero.

(g) What will be the new long-run equilibrium price in this industry?

The same as it was before, P = 25, because that is where zero-profits occur for firms.

(h) At the new long-run equilibrium, what will be the output of each representative firm in the industry?

Firm output will still be 5 as this is the quantity where ATC = MC, and long-run profits are zero.

Notes

(i) At the new long-run equilibrium, how many firms will be in the industry?

This will be different since there is a new demand curve. Specifically, there is a new market demand. With the new market demand curve:

P_D = $1525 - 2Q_D$

We can substitute $P = 25$:

25 = $1525 - 2Q_D$ $1500 = 2QD$

750 = Q_D

We can see that the new market demand is 750. Since each firm produces 5 units and firms are all identical, there must be 750/5 or 150 firms.

Now, consider another scenario where technology advancement changes the cost functions of each representative firm. The market demand is still the original one (before the increase in the number of students). The new cost functions are:

TC = $q_s^2 + Sq_s + 36$

MC = $2q_s + S$

(j) What will be the new equilibrium price? Is it higher or lower than the original equilibrium price?

Similar to part (b), in a long-run equilibrium, ATC equals Marginal Cost and profits equal zero. Setting the two equations equal:

ATC = $q_s + 5 + 36/q_s = 2q_s + 5 = MC$

$36/q_s$ = q_s

36 = q_s

Take the squl1re root of both sides and find:

6 = q_s

However, the question wants us to find long run prices. We know that the firm produces were Price = MR = MC, so if we can determine the firm's MC, then we can determine the equilibrium price in the market.

We know that:

$MC = 2q_s + 5$

And solved for:

$6 = q_s$

Substituting:

$MC = 2(6) + 5 = 17$

The equilibrium price in the market is 17.

The price is lower than before, and this makes sense because the technological improvement has lowered the costs for the firm. With lower costs, the price is lower for firms to have zero profits.

(k) In the long-run given this technological advance, how many firms will there be in the industry?

Now we should determine the market quantity Q from the market demand curve, given that we know the market price is 17. Market demand is given as:

$$P_D = 1025 - 2Q_D$$

And we know that market price = 17, so:

$$17 = 1025 - 2Q_D$$

$$1008 = 2Q_D$$

$$504 = Q_D$$

Since each firm is making 6 units (as we found in parts b and c), there must be 84 firms, since they are all identically sized. (504/6 = 84)

Since each firm faces lower costs, more firms need to enter the industry to drive down prices so that there are zero profits in the long run. We see the number of firms increase, the price decrease, and the market quantity increase as a result of this competition in the long run.

Question 3 (Please put some thought into these)

(a) Describe the factors that drive profits to zero in perfectly competitive markets in the long run. Explain carefully the incentives that drive the market to a long run equilibrium.

The biggest factor driving this is the free entry/exit of firms in the long run, and that firms are selling identical products. With firms being able to enter and exit the market as they wish, profit opportunities cannot last. If I observe another firm making positive profits, there is an incentive for me to enter the industry (at no cost) and try to take advantage of some of these profits. Since there are many identical firms, there will be many firms entering the industry to take advantage of these profit opportunities. However, when many firms compete, the market price decreases and the profit opportunities disappear.

(b) Why would a firm choose to operate at a loss in the short run? Explain carefully.

If the firm can cover their variable costs in the short run, then they can start to pay down some of their fixed costs by producing. If they shut down they must pay all of their fixed costs. If the firm can cover the variable costs they can use any excess revenue towards paying their fixed costs, which is a better outcome than shutting down in the short run. In the long run a firm cannot constantly operate at losses and will eventually leave the industry unless costs change.

(c) When do firms decide to shut down production in the short run? Explain carefully.

If the firm cannot cover their variable costs, then the act of production is going to lead to larger losses than simply shutting down and paying the fixed costs. It doesn't make sense for a firm to lose more money by staying open than what they would lose if they simply chose not to produce.

(d) Draw a graph for a perfectly competitive market, specifically showing the short run supply curve. What is the relationship between the short run supply curve and what we talked about in parts (b) and (c)? Explain carefully.

The short-run supply curve is the marginal cost curve anywhere above the intersection with average variable costs. Specifically, the firm will only produce goods if the marginal revenue covers their variable costs (even if they operate at a loss). However, when marginal revenue falls below variable costs the firm will shut down in the short run.

Question 4

Consider a perfectly competitive market in the short run. Assume that market demand is P = 100 $4Q_D$ and market supply is P = Qs. Denoting firm level quantity by q, assume TC = 50 + 4q + $2q^2$ so that MC = 4 + 4q:

(a) What is the market equilibrium price and quantity?

Set demand equal to supply and find 100 – 4Q = Q, so Q = 20, P = 20.

(b) How many firms are in the industry in the short run?

Perfectly competitive firms will set P = MC, so 20 = 4 + 4q, so q = 4. If each perfectly competitive firm is producing 4, market output is 20, there will be 5 perfectly competitive firms in the industry.

(c) Do firms make a profit or loss in the short run, and how much are these profits/losses?

Firms will make losses in the short run. There are a variety of ways to see this. One is to calculate ATC = 50/q + 4 + 2q, set q = 4, and find ATC = 24.5, so price is less than ATC, by 4.5, and they are selling 4, so the losses are 18. Another way is to calculate this is to calculate total revenue (P*Q = 80) minus total cost 50 + 16 + 32 = 98 and see the difference is – 18 (or a loss of 18).

(d) What is the equilibrium price in the long run? What will be equilibrium profit in the long run? How many firms will there be in the long run? Hint, for the last part of the question, assume that there can be fractional firms, if necessary – if the numbers of firms are in units of 10,000, for example, the answer will be fine. Moreover, assume the entry or exit in the industry will cause the supply curve to shift, while the demand curve does not shift. Therefore, industry output can be found by taking the long-run price and plugging it into the demand curve.

There will obviously be exit from the industry if perfectly competitive firms arc making losses in the short run. Equilibrium long-run profits will be zero. In the long-run, firms will produce at the minimum of the average total cost curve. That occurs where MC = ATC. Setting these two equal, we have 4 + 4q = 50/q + 4 + 2q. Solving this for q, we get 2q = 50/q, or q = 5. Each of the remaining firms will be larger (before they produced 4, now they produce 5). The long-run equilibrium price will be equal to marginal cost (or ATC) when MC = ATC. So plug the quantity 5 into MC and find the long-run equilibrium price, P = 24. The exit of firms causes the supply curve to shift back (demand will stay constant). So, from the demand curve, the total output consistent with the market price is Q = 19 (this comes from 24 = 100 – 4Q). If total output in the market is 19 and each firm in the industry produces 5, there will be 3.8 firms in the industry.

Equilibrium in Factor Market: Perfect Competition

In the factor market, under perfect competition, an individual organisation cannot affect the prices of a factor of production by increasing or decreasing its consumption.

This is because the quantity demanded by an organisation of a particular factor is very small as compared to the market demand. In such a case, the organisation cannot affect the price of the factors. Thus, it has to purchase the factor at the prevailing market price. Even if the organisation increases the consumption of the factor, the price of the factor would remain same.

For example, in perfect competition, organisations need to pay wages to its employees according to the wage rates prevailing in the market. Similarly, if we look upon the supply side, a single supplier does not have ample amount of products to meet the demand of all the customers in the market. Therefore, in perfect competition, Marginal Product (MP) and Average Product (AP) are same and their curves would intersect each other. Thus, MP and AP would form a straight horizontal line. Here, we would again take the example of labour and wages to understand equilibrium in factor market under perfect competition.
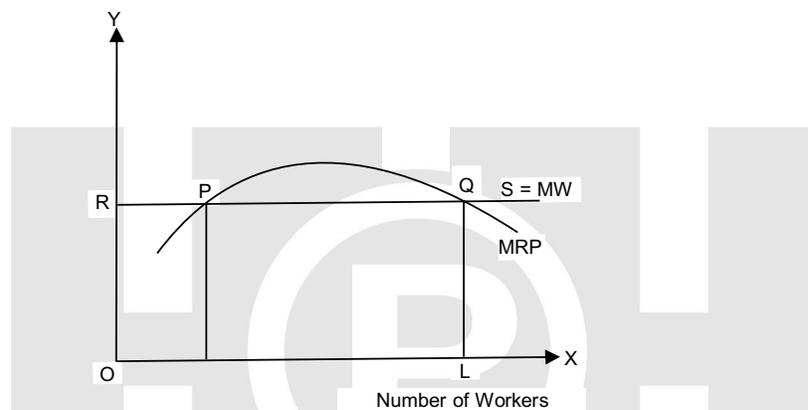
Fig. 3.1: Equilibrium in the Factor Market under Perfect Competition

In Fig. 3.1, we have assumed labour as a variable factor, while keeping the other factors at constant. The RS line shows the marginal wage rate. In the factor market, all organisations can hire any number of workers at the prevailing price OR. The MRP curve of labour intersects the line RS at two points P and Q.

An organisation cannot attain equilibrium at point P because at this point the number of workers employed is increased. Thus, in this case, the MRP of labour would be higher than the marginal wage OR. On the other hand, at point Q, when the organisation employs OL number of workers, the MRP of labour is equal to its marginal cost.

Therefore, the organisation would attain its equilibrium at point Q. Apart from this, if the organisation employs more than OL workers, the marginal cost of labour would exceed MRP. In such a case, the organisation would incur losses.

In summation, there are two conditions required for attaining equilibrium in the factor market under perfect competition, which are as follows:

(i)   MRP = MFC

(ii) MRP curve intersects marginal cost from above (as shown in Fig. 3.1)

However, from Fig. 3.1, we cannot determine whether the organisation would earn profit or incur loss.
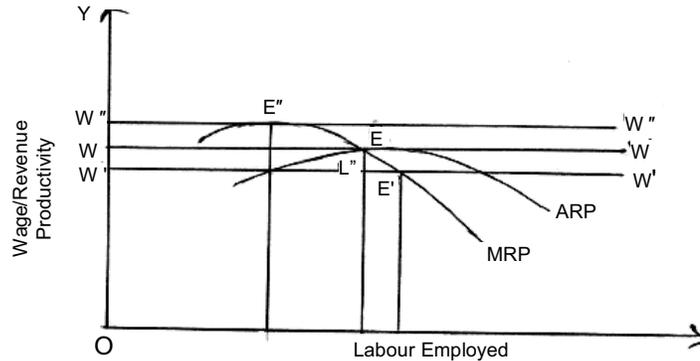
Fig. 3.2: Profit or Loss in Perfect Competition

In Fig. 3.2, MRP intersects Average Revenue Productivity (ARP) at point E. When the wages are at level OW, the equilibrium point is attained at E′. On the other hand, when the wages are at level OW″, equilibrium point is achieved at point E″. At point E′, extra profit is E′L′, which is in the short run only.

In the long run, supernormal profit attracts new organisations to enter the market. This increases the demand for labour. Therefore, the wage level of labour also increases and reaches OW. At OW wage level, the equilibrium shifts to E and supernormal profit disappears. This is because wages are equal to average revenue productivity.

However, at equilibrium point E″, the wages are more than the average revenue productivity. In such a case, the organisation would incur losses. In case of losses, many organisations would leave the market, which would result in the reduction of labour and wage rates. This again brought the wage level at OW and equilibrium point at E. At this point, MRP would become equal to ARP.

Equilibrium in Factor Market: Imperfect Competition

In the above, we have discussed the equilibrium of an organisation in the factor market under perfect competition. However, in the real world, the factor market is imperfect. Therefore, we would learn the equilibrium of an organisation in the factor market under imperfect competition.

For understanding the equilibrium in case of imperfect competition, we would take the case of monopsony. In monopsony, there is only one buyer of factors of production and a large number of sellers. In this case, there is no competitor in the market who wants to buy the factors of production.

Therefore, the single buyer has a control on the price of factors. This implies that he/she can bargain for the prices of factors as per his/her choice. For example, if the buyer wants to hire a factor say labour, then he/she can set wages according to him/her.
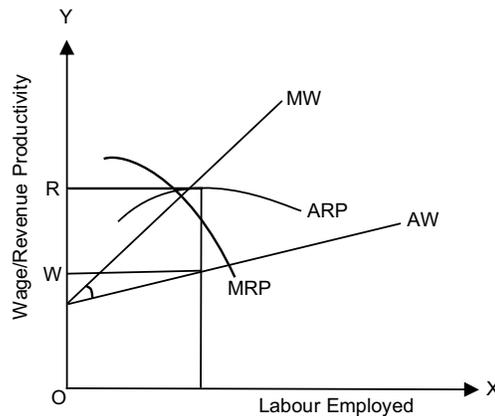
Fig. 3.3: Equilibrium in Imperfect Competition

In Fig. 3.3, Average Wage (AW) curve moves from left to right in upward direction and Marginal Wage (MW) curve is above the AW curve. In imperfect competition, equilibrium can be attained when MW is equal to MRP. In present case, equilibrium is at point E.

At point E, AW is NP, MRP is EN and NP is less than EN. This shows that services provided by labour is more than wages provided by the organisation to them. In other words, the labour is exploited by the organisation. This is also termed as monopolistic exploitation.

In imperfect competition, as the buyer has the power to decide the wages of labor, therefore, labour is exploited in this type of market structure. For example, in case of oligopoly or monopoly, the number of job opportunities is limited and the unemployment is high. In such cases, the labour is ready to work even at low wage rates.

The Model of Oligopolistic Competition

Suppose that the oligopolistic competition (of order n) occurs when several companies (firms), $F_1, \ldots, F_n$ appear on the market offering the same product X. It $q_1, \ldots, q_n$ denotes. Their demand and production volume of each company. Respectively, the total demand for to the sum of individual demand, i.e.,

$$q = q_1 + \ldots + q_n \qquad \ldots(3.1)$$

while the law of demand in general, inverse form is given by,

$$P = f(q) \qquad \ldots(3.2)$$

Where, P is these price of product X

Further, $T_1(q), T_2(q), \ldots T_n(q)$ denotes the total cost of product X in the companies $F_1, F_2, \ldots, F_n$ respectively and $C_1(q), C_2(q) \ldots, C_n(q_n)$. Their total income. Then total income is generally equal to the product of demand and price.

$$C_i(q) = q_iP = q_iP(q_1 + \ldots + q_n) \qquad \ldots(3.3)$$

While the difference between total revenue and total costs present the profit of the company, given by the functions.

$$P_i(q) = C_i(q) - T_i(q), i = 1, \ldots, n \qquad \ldots(3.4)$$

We can put the equation (3.4) is a developed form, taking the function of the total revenue (3.3). The profit function of company F_i, $i = 1, 2, ..., n$ these as follows.

$$P_i(q) = q_i f(q_1 + ... + q_n) - T_i(q) \qquad ...(3.5)$$

In a further analysis of the model of oligopolistic competition we will omit the possibility that in case of change in production in the i-th company other monopolistic are indifferent to the changes. Specifically, certainly all the other participants will short is a similar way i.e., changes is production volume is the company pulls F_i change in other businesses. Based on this, we conclude that for the demand q there is a corresponding functional dependency, which is displayed as the following

$$q_i = \begin{cases} \xi_{ij}(q_j) & i \neq j \\ q_i & i = j \end{cases} \qquad ...(3.6)$$

Therefore, a functional relationship between the various q_i and q_j exists only for $i \neq j$, which is consistent with the fact that this is the demand of different company. Now substituting (3.6) into (3.5) we obtain the following system of equations.

$$P_i(q) = q_i f\left(q_i + \sum_{j \neq 1}^{i} \xi_j^i(q_i)\right) = \frac{\delta T_i(q)}{\delta q_i} \qquad ...(3.7)$$

that completely describer the state of the market which a set of monopolistic firms. In economic terms, any solution (3.7) i.e., a finite sequence of number

$$(q_i, ..., q_n) \in R^n \qquad ...(3.8)$$

Which identically fulfills the system of equations, is the optional value of demand or optimal level of production for which each of monopolistic market. They are realized when the marginal revenues are equal to marginal costs i.e., if

$$C_g(q) = T_g(q) \qquad ...(3.9)$$

Where $C_g(q)$ and $T_g(q)$ are derivatives of function of total income $C(q)$ and total costs (Tq) expressed through the demand for X.

3.2 STABILITY OF EQUILIBRIUM (SIMPLE KEYNESIAN MODEL)

Let us make an in-depth study of the Stability of Equilibrium.

The assumption that $0 < b < 1$ is crucial for establishing stability in SKM.

Stability in this context refers to a stable equilibrium position in the commodity market.

The stability condition is that the slope of the $C + I + G$ schedule has to be less than unity.

For the sake of simplicity, we ignore government expenditure and taxes. So, we are now examining the SKM without government.

In a two-sector economy, the slope of $C + I$ schedule has to be less than unity. Here, the $C + I$ schedule is parallel to C schedule since I is autonomous. Hence, the slope of the $C + I$ schedule is the same as the slope of the C schedule (since the slope of I schedule is zero). Thus, for a given level of autonomous investment, the equilibrium value of Y is determined by the consumption function.

If the slope of the consumption function is less than 1, the slope of the $C + I$ schedule will also be less than 1. This will then be less than the slope of the income line $Y = C + S$ (ignoring taxes).

And equilibrium will be stable as shown in Fig. 3.4(a). Otherwise, it will be unstable as shown in Fig. 3.4(b). This point may now be proved.
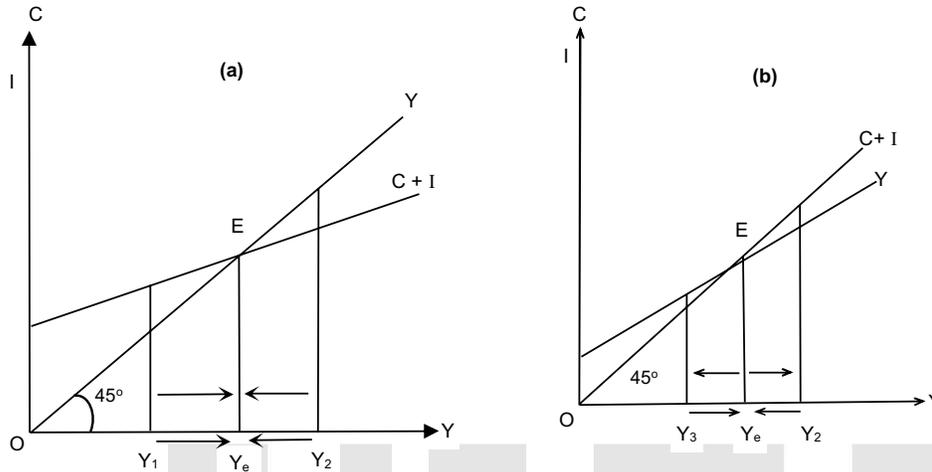
Fig. 3.4: Stable and Unstable Equilibria

In Fig. 3.4, AD is equal to C + I and AS is equal to Y. Suppose excess demand (E) is equal to AD – AS, i.e., E = C + I – Y. Here, E is a function of Y. According to the Walrasian stability condition, the commodity market will be stable if

$dE/dY < 0$

i.e., $dE/dY = dC/dY – 1 < 0$

or, $dC/dy < 1$

or, $b < 1$

Here, b is MPC.

If b > 1, the C + I schedule will intersect the income line from below and equilibrium will be unstable. Any deviation of Y_e in either direction will be cumulative in nature as shown in Fig. 3.4.

1. The Stability Condition

The stability condition in the SKM is that the MPC(b) should lie in-between zero and one. It has to be greater than zero and less than one. We may now rigorously demonstrate the income path in SKM with a lagged consumption function.

We know that the Keynesian consumption function is linear. If we assume that there is one period lag in the consumption function, then we can express the function as $C_t = a + bY_{t-1}$ where t-1 is the last period's income, a is the intercept of the function (showing autonomous or income-independent consumption) and b is the MPC($0 < b < 1$).

Since in SKM all investment is autonomous and thus remains constant at all levels of income, we write $I_t = \bar{I}$. Now, equilibrium in the goods (commodity) market requires that $Y_t = C_t + I_t$ or $Y_t, = a + bY_{t-1} + \bar{I}$ or $Y_t = bY_{t-1} + (a + \bar{I})$.

From the particular solution to this first order linear homogeneous difference equation, we arrive at the equilibrium value of national income (Y_e) in SKM. Assuming income to remain constant overtime, we put $Y_t = Y_{t-1} = Y_e$. Therefore, we get $Y_e = bY_e + a + \bar{I}$.

Here, Y_e is the equilibrium value of income. Since Y_e is assumed to remain unchanged period after period, if $Y_{t-1,} = Y_{e,}$ then Y_t will also be equal to Y_e. Thus, if Y was in equilibrium last year $(t - 1)$, then it will also be in equilibrium in the current year. Similarly, if Y is in equilibrium in the current year, it will also be in equilibrium in the next year.

2. The Derivation of the Stability Condition

The equilibrium level of income is said to be stable if any deviation from it tends to create forces that bring actual income back to the equilibrium level. Let us suppose, that in period $(t - 1)$, actual income exceeds its equilibrium level. Suppose $Y_{t-1} = Y_e + k$ where $k > 0$. So, we have

$Y_{t-1} = Y_e + k$ where $k > 0$.

So, we have

$$
\begin{aligned}
Y_t &= bY_{t-1} + (a + \bar{I}) \\
&= b(Y_e + k)\,(a + \bar{I}) \\
&= bY_e + bk + (a + \bar{I}) \\
&= \frac{b}{1-b}(a + \bar{I}, + bk + (a + \bar{I}) \\
&= (a + \bar{I})\left(\frac{b}{1-b} + 1\right) + bk \\
&= \left(\frac{1}{1-b}\right)(a + \bar{I}) + bk = Y_e + bk
\end{aligned}
$$

In the same way, we get

$$
\begin{aligned}
Y_{t+1} &= bY_t + (a + \bar{I}) \\
&= b(Y_e + bk) + (a + \bar{I}) \\
&= bY_e + b^2k + (a + \bar{I}) \\
&= \frac{b}{1-b}(a + \bar{I}) + b^2k + (a + \bar{I}) \\
&= \frac{1}{1-b}(a + \bar{I}) + b^2k \\
&= Y_e + b^2k
\end{aligned}
$$

If we repeat this process, we get,

$Y_{t+2} = Y_e + b^3k$

$Y_{t+3} = Y_e + b^4k, \ldots\ldots., $ etc.

This means that the income rises in different time periods in this dynamic SKM (where time enters the analysis as an important variable) and will be $(Y_e + k)$, $(Y_e + bk)$, $(Y_e + bk^2)$, $(Y_e + b^2k)$,..., etc. Now, if the MPC(b) is greater than 1, then we have $k < bk < b^2k < b^3k \ldots$

This simply implies that if in any period actual income goes above its equilibrium level, then the gap between the two (i.e., the excess of actual income over the equilibrium income will continue to increase with the passage of time. But if MPC(b) is less than 1, we have

$k > bk > b^2k > b^3k \ldots$

In this case, the gap between the two (i.e., the excess of actual income over equilibrium income) becomes gradually smaller and smaller. Thus, if $b > 1$, the equilibrium level of income in SKM is unstable, in the sense that any deviation of actual income (Y) from its equilibrium level (Y_e) does not bring Y back to Y_e. But if b is less than 1, Y_e is stable because in case of any deviation of Y from Y_e, Y moves toward Y_e in the next and subsequent periods. It is also quite obvious that if $k < 0$, i.e., $< Y_{t-1} Y_e$, then the gap between the two (deficiency of Y from Y_e) will increase $b < 1$ and the gap will become smaller and smaller, and ultimately disappear if $b < 1$.

If $b = 1$, we have $Y_t - Y_{t-1} + (a + \bar{I})$.

In this case if we assume that $(a + \bar{I}) = 0$, i.e., there is no autonomous (income independent) expenditure, then $Y_t = Y_{t-1}$ [from equation (1)]. If such a restrictive assumption is made, the level of income (Y) remains constant in all time periods. This means that the economy does not grow. It is in a stationary state where national income remains constant. Some economists even call this type of equilibrium behaviour by the name 'neutral equilibrium'.

However, if $(a + \bar{I})$ is positive or negative, equilibrium will not even exist. If $(a + \bar{I}) > 0$, then income in period T will be greater than Y_{t-1} by $(a + I)$, in which case Y will continue to grow without limit. It will explode. In contrast if $(a + \bar{I}) < 0$, Y will fall without limit, it will fall toward zero.

So, the basic point to note is that if and only if $b < 1$, the income path in SKM will be stable.

3. Dynamic Analysis

A. Consumption Lag of One Period

Although the SKM is static in nature, we can extend it to make a dynamic analysis of the income path by considering a lagged consumption function. This means that consumption in the current period (r) depends on the income of the last period $(t - 1)$. So, the Keynesian linear consumption function can be expressed as $C_t = a + bY_{t-1}$ where a is the intercept (a positive constant) and b is the slope (the MPC which is also a positive constant).

We continue to assume that all investment is autonomous and hence independent of income. So, $I_t = \bar{I}_t$.

Thus, the income equation in SKM is

$Y_t = a + bY_{t-1} + \bar{I}_t$

The solution of this first order linear homogeneous difference equation gives the time path of income. This approach follows from the assumption that there is a one period consumption lag.

B. Production Lag of One Period

Now, we can adopt an alternative approach to Keynesian dynamics by assuming a production lag of one period. But there is no consumption lag. So, we get,

$Y_t = C_{t-1} + I_{t-1}$ (where $C_{t-1} = a + bY_{t-1}$ since there is no consumption lag now)

$Y_t = a + bY_{t-1} + I_{t-1}$

$\quad = a + bY_{t-1} + A$

where $A = I_{t-1}$, i.e., output in the current period is equal to aggregate demand of the last period.

4. Two-part Solution to the Difference Equation

The particular solution to the difference equation is obtained by assuming stationary value of Y, i.e., $Y_t = Y_{t-1} = \overline{Y}$, and solving for $\overline{Y}$. So, we get,

$$\overline{Y} = a + b\,\overline{Y} + A$$

or $\overline{Y}(1 - b) = a + A$

or $\overline{Y} = \dfrac{a + A}{1 - b}$

This Particular Solution ($\overline{Y}$) gives us the equilibrium level of income. This reason is easy to find out.

If $Y_{t-1} = \overline{Y}$, then $Y_t = \overline{Y}$.

We way now consider the solution to the homogeneous part of the difference equation, i.e., $Y_t = bY_{t-1}$

Let, $Y_t = x^2$

Then, $x^1 = bx^{1-1}$

or $x = b$

$\therefore Y_t = b^1$

$\therefore Y_t = kb^t$ is also a solution where k is a positive arbitrary constant whose value has to be determined from the initial equilibrium condition of income. The complete solution to the difference equation is expressed as

$$Y_t = kb^t + \frac{a + A}{1 - b}$$

Suppose, $t = 0$, the we get $Y_0 = k + \dfrac{a + A}{1 - b}$

where Y_0 is the given initial level of income (which is a pre-determined variable). So, we get

$$k = Y_0 - \frac{a + A}{1 - b}$$

or $k = Y_0 - \overline{Y}$

If we incorporate this value of k is the complete solution, we get

$$Y_t = (Y_0 - \overline{Y})\,b^t + \overline{Y}$$

This equilibrium shows the time path of income in the sense that it indicate how Y change with t. From (2) we can also find out the stability condition of equilibrium.

$$Y_t = kb^1 + a + A/1 - b \qquad\qquad\qquad\qquad\qquad ...(3.10)$$

$$K = Y_0 - a + A/1 - b \qquad\qquad\qquad\qquad\qquad ...(3.11)$$

In this case also, $\overline{Y}$, b has to be less than 1 in which case as $t \to \alpha$, $b^t = 0$ in the limit and $Y_t \to Y$, even if $Y_0 \neq Y$. This means that even if the initial level of income is different from the equilibrium level of income, then actual income will tend towards equilibrium value over time. If, however, $b > 1$,

then as $t \propto b^t$ will also approach α, in which case the initial level of income will gradually diverge away from its equilibrium value. This means that the income path in SKM will be unstable.

Two possible income paths in Keynesian dynamic model are shown in Fig. 3.5. Now, we show time on the horizontal axis and income on the vertical axis.
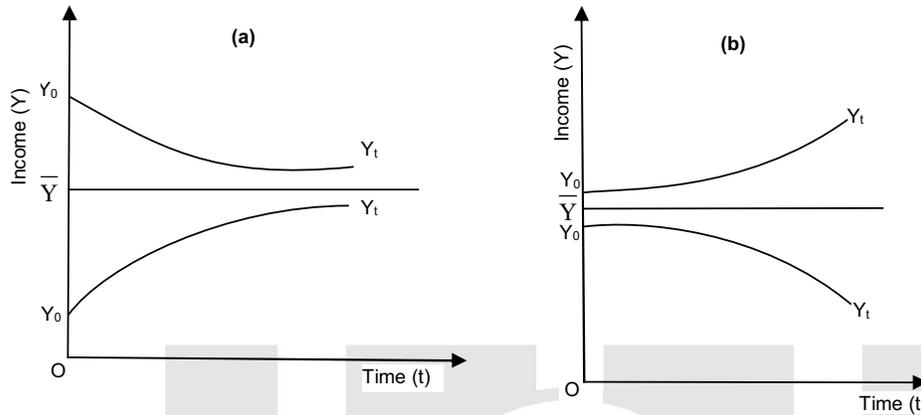
Fig. 3.5: Stable and Unstable Income Paths in Keynes' Dynamic Model

In Fig. 3.5(a), we assume that b < 1. Therefore, if $Y_0 > Y$, then Y_t comes closer and closer to Y over time. Conversely, if $Y_0 < Y$, then Y_t increases steadily and gradually moves toward Y with the passage of time. Thus, the equilibrium income is stable.

In Fig. 3.5(b), we assume that b > 1. Therefore, if $Y_0 > Y$, then income will continue to increase with the passage of time. This means that there will be no limit to the increase in income. Contrarily, if $Y_0 < Y$, actual income will continue to fall over time.

So, there will be no limit to the fall in income. In this case, actual income will move further and further away from its equilibrium value. In other words, the deviation from equilibrium becomes cumulative and equilibrium is unstable.

5. Induced Investment and Stability of Equilibrium

We now relax the assumption that all investment is autonomous. We now assume that investment is partly autonomous and partly induced. Thus, we can write

$I = I_a + I_p$

$\quad = I_a + iy$

where $i = dI_p/dY$ = the marginal propensity to invest.

In this case, the investment demand schedule, instead of being horizontal throughout will be upward sloping from left to right and its slope is the marginal propensity to invest (MPI) which is positive. The MPI is defined as the ratio of the change in investment to the change in national income which brings it about.

In this case, change in income leads to a change in investment while in the original investment has no relation to income. Now that investment has an induced component also, we have to modify the stability condition.

In Fig. 3.6(a), we see that equilibrium income is stable. If there is any deviation of point E, Y will fall since MPI > MPS and Y will come back to the original level. In Fig 3.6(b), we see that equilibrium income is unstable. If there is any deviation of point E, Y will continue to move further and further away from Y_e.

Thus, the condition of stability in this context is that MPI > MPS, i.e., the slope of the saving (S) schedule has to be less than that of the investment (I) schedule.
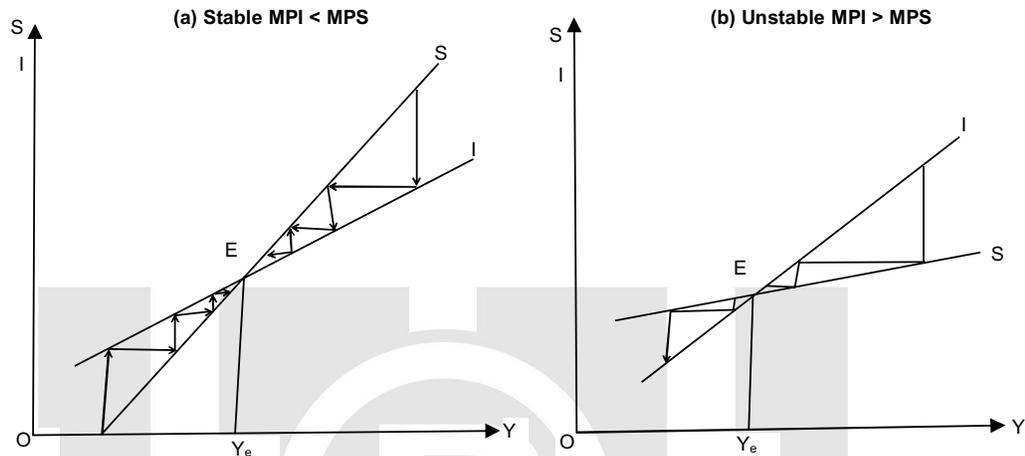
Fig. 3.6: Stable and Unstable Equilibria

6. Two Related Points

(i) Shift vs. Change in Slope:

If there is a change in any of the autonomous components of DE = C + I + G = a + bY – bT + I + G, the DE schedule will shift up or down. The autonomous components of E are a, –bT, I and G. There is a change in the slope of the DE schedule if Y changes. The slope of DE is b which is MPC (= $\Delta C/\Delta Y$) which indicates how C and hence DE will change when income changes. If b increases (falls), the E schedule becomes steeper (flatter).

(ii) Two Components of Equilibrium Income:

The essence of the process of income determination in the context of SKM is captured by the following equation:

Y = 1/1 – b(a – bT + I + G)

So, equilibrium income = (Autonomous expenditure multiplier) × (Autonomous expenditures).

Here, $1/(1 – b)$ is the autonomous expenditure multiplier. Here 'b' is the MPC and (1 – b) is the MPS. So, the multiplier is the reciprocal of the MPS. Since MPC < 1, the multiplier is a number which is greater than 1. If b = 0.5, m = 2; if b = 0.8, m = 5. Thus, if b increases, m also increases.

The term 'autonomous expenditure multiplier' is derived from the fact that every rupee of autonomous expenditure is multiplied by this number to find out its contribution to equilibrium income.

The second component of equation indicates the level of autonomous expenditures which is determined by factors other than current income. Here, I and G are fully autonomous. But C is partly autonomous and partly induced. The terms related to C but unrelated to Y are a and –bT.

These two terms measure the autonomous component of consumption expenditures (a) and the autonomous effect of tax collections on aggregate demand (–bT), which also works through consumption. Since these two terms affect the amount of consumption for a given level of income (Y) and are not themselves determined by income, they are treated as autonomous component of C.

Keynesian Model – Numerical Example

Given $Y = c + 1$

$$C = a + bY$$

Where, $a = 100$

$b = .75$

$I = 300$

Solving for Y_e

$$Y = 100 + .75Y + 300$$

$$\Rightarrow Y - .75Y = 100 + 300$$

$$\Rightarrow (1 - .75) = 100 + 300$$

$$\Rightarrow Y = (1/.25) \times 400$$

$$\Rightarrow Y = 4 \times 400$$

$$\Rightarrow Y_e = \$1600 \text{ billion}$$

Solving for equilibrium consumption and savings.

Once we have Y_e, we can find C_e and S_e:

$$C_e = a + bY_e$$

$$\Rightarrow C_e = 100 + .75(1600)$$

$$\Rightarrow C_e = 100 + 1200 = 1300$$

$$\Rightarrow S_e = -a + (1 - b)Y_e$$

$$= -100 + (1 - .75)1600$$

$$= -100 + .25(1600) = 300$$

$$Y_e = C_e + S_e$$

$$\Rightarrow Y_e - C_e = S_e$$

$$\Rightarrow 1600 - 1300 = 300$$

As, $S_e = I$ (Savings = Investment at Equilibrium)

$$\Rightarrow 300 = 300.$$

3.3 SUMMARY

1. Market equilibrium refers to a condition where a market price is established through competition such that the amount of goods or services sought by buyers is equal to the amount of goods or services produced by sellers.

2. In the factor market, under perfect competition, an individual organisation cannot affect the prices of a factor of production by increasing or decreasing its consumption.

3. The equilibrium level of income is said to be stable if any deviation from it tends to create forces that bring actual income back to the equilibrium level.

3.4 SELF ASSESSMENT QUESTIONS

1. Critically analyse Factor market equilibrium.

2. Critically explain Simple Keynesian Model.

❧❦✤❦❧

UNIT – III

Chapter	
4	**LINEAR PROGRAMMING**

Objectives

The objectives of this lesson are to learn:

- Linear programming
- Simplex method
- Duality theorem
- Complementary slackness theorem

Structure:

4.1 LINEAR PROGRAMMING: BASIC THEOREMS

Consider the Linear Program (P)

Minimize $c^T . x$

Subject to

$Ax = b$

$x \geq 0$

where A is an $m \times n$ matrix of rank m.

Definition 1: A feasible solution is an element $x \in R^n$ which satisfies the constraints $Ax = b$ and $x \geq 0$.

Among all solutions of the equation $Ax = b$, certain ones are called basic.

Definition 2: Let B be any $m \times n$ non-singular submatrix of A consisting of linearly independent columns of A. Then if the $n - m$ components of a solution x corresponding to the columns of A which do not appear in B are set equal to zero, the solution x_B, of the resulting set of equations is said to be a basic feasible solution with respect to the basic consisting of columns of B. The components of x_B associated with the columns of B are called the basic variables.

Note that equation Ax = b may not have any basic solutions in the general case. In order to ensure that basic solutions exist, it is usual to makes certain assumptions: (a) that n > m; (b) that the rows of A are linearly independent and (c) that the rank of A is m. These conditions are sufficient for the existence of at least one basic solution. It may well occur that some components of a basic solution are zero.

Definition 3: If one or more basic variables in a basic solution has the value zero, then that solution is said to be a degenerate basic solution.

We shall refer to a feasible solution which is also basic as a basic feasible solution. If it is also basic, then it is an optimal basic feasible solution.

Let us return to the linear programming problem P. The fundamental result is that we need only search among the basic feasible solutions for an optimal solution. Indeed, that is what the Simplex Method actually does.

Theorem (The Fundamental Theorem of Linear Programming) (Given the linear programming problem P, where A is an m× n matrix of rank m)

1.   If there is any feasible solution, then there is a basic feasible solution.

2.   If there is any optimal solution, then there is a basic optimal solution.

Proof: Suppose that a feasible solution exists. Choose any feasible solution among those with the fewest men-zero components. If there are no non-zero components, then x = 0 and x is a basic solution by definition. Otherwise, take the index set j: = { j_1, j_2, ..., j_r} with elements corresponding to those $x_{ji} > 0$. Then if we denote the corresponding columns by {$a^{(j1)}$, $a^{(j2)}$, ..., $a^{(jr)}$} there are two possibilities:

1. The set of columns is linearly independent. In this case, we certainly have r ≤ m. If r = m, the corresponding solution is basic and the proof is complete. If r < m, then since A has rank m, we choose m – r vectors from the remaining n – r columns of A so that the resulting set of m columns vectors is linearly independent. Assigning the value zero to the corresponding m – r variables yields a (degenerate basic feasible solution.)

2. The set of columns {$a^{(j1)}$, $a^{(j2)}$, ...· $a^{(jr)}$} is linearly dependent. Then there is a choice of scalars a_{ji}, not all zero, such that $a_{j1} a^{(j1)} + a_{j2} a^{(j2)} + ... + a_{jr} a^{(ty)} = 0$ ...(4.1)

Without loss of generality, we may take $α_{ji} \neq 0$ and indeed on $α_{ji} > 0$ (otherwise multiply (4.1) by – 1)

Now, since $x_{ji} > 0$, the corresponding feasible solution x_{ji} is just a linear combination of the columns $a^{(ji)}$. Hence,

$$A_x = \sum_{i=1}^{r} x_{ji} a^{(ji)} b$$

Now, multiplying the dependence relation (4.1) by a real number λ and subtracting, we have

$$A (x - \lambda \alpha) = \sum_{i=i}^{r} (x_{ji} - \lambda^\circ{}_{ji}) a^{(ji)} = b$$

and this equation holds for every λ although one or more components, $x_k - \lambda α_{jk}$ may violate the non-negativity condition. Now, let $\overline{\alpha} = (\overline{\alpha}_1, ..., \overline{\alpha}_n)^T$ which

$$\overline{\alpha}_K := \begin{cases} \alpha_{jk}, & k = ji \\ 0, & k \neq ji \end{cases}$$

Then, for each value of λ the vector $x - \lambda\,\overline{\alpha}$ is a solution of the constraint equations. Note that, for the special value $\lambda = 0$, this vector is the original feasible solution. Now, as λ increases from 0, the components of $x - x^{\overline{\alpha}}$ will individually change; each will either increase, decrease or stay constant, depending on whether the corresponding $\overline{\alpha}_i$ is positive, negative or zero.

Suppose, for some i_0, $\overline{a}_{i0} \leq 0$. Then $(x_{i0} - \lambda\,\overline{\alpha}_{i0}) \geq 0$ for all $\lambda > 0$. On the other hand, if for some i_0, $\overline{\alpha}_{io} > 0$, then $(x_{i0} - \lambda\,\overline{\alpha}_{i0})$ remains greater than 0 only for sufficiently small $\lambda > 0$. Now, we take

$$\widetilde{\lambda} := \min\left\{\frac{x_i}{\alpha_i}\Big|\, x_i > 0,\, \overline{\alpha}_i > 0\right\}$$

Then $\widetilde{\lambda}$ corresponds to the first value of λ for which one or more non-zero components of $x - \lambda\,\overline{\alpha}$ becomes 0. For this value of λ, the vector $x - \widetilde{\lambda}\overline{\alpha}$ is feasible and has at most $r - 1$ positive components.

Repeating this process if necessary, we can obtain a feasible solution with non-zero components corresponding to linearly independent columns. In this situation, the previous alternative applies.

This completes the proof of the first part of the theorem.

Now, assume that x^* is an optimal solution. There is no guarantee that this optimal solution is unique. In fact, we have seen cases where there is no uniqueness. Some of these solutions may have more positive components than others. Without loss of generality, we assume that x^* has a minimal number of positive components. If $x^* = 0$, then x^* is basic and the cost is zero. If $x^* \neq 0$ and if j is the corresponding index set, then there are two cases as before. The proof of the first case in which the corresponding columns are linearly independent is exactly as in the previous proof of this case.

In the second case, we proceed just as with the second case above. To do this, however, we must show that, for any λ, $x^* - \lambda\overline{\alpha}$ is optimal. To show this, we note that the associated cost is $(c, x^* - \lambda\overline{\alpha}) = (c, x^*) - \lambda(c, \overline{\alpha})$

Then, for sufficiently small $|\lambda|$, the vector $x^* - \lambda\overline{\alpha}$ is a feasible solution for positive or negative values of λ. Hence, we conclude that

$$(\lambda.\,\overline{\alpha}) = 0$$

Since, were it not, then we could determine a small λ of the proper sign, to make

$$(c, x^* - \lambda\overline{\alpha}) < (c, x^*)$$

which would violate the assumption of the optimality of x^*.

Production Planning Problem

Example: A firm manufacture 3 products A, B and C. The profit per unit sold of each product is ₹ 3, ₹ 2 and ₹ 4 respectively. The firm required to manufacture are unit of each of the there products and the daily capacity of the two machines P and Q is given in the table below:

Table 4.1

Machine	Time per Unit (Minutes) Product			Machine Capacity (Minutes/Day)
	A	**B**	**C**	
P	4	3	5	2000
Q	2	2	4	2500

It is required to determine the daily number of units to be manufacture for each product, so as to maximise the profit. However, the firm must manufacture at least 100 A's, 200 B's and 50 C's but no more than 150 A's. It is assumed that all the amount produced are consumed in the market.

Solution:

- **Step 1:** We study the situation to find the key decision to be mode and in this connection looking for variables help considerably.

- **Step 2:** Select symbols for variable quantitive identified in Step 1. Let the number of units of the products A, B and C manufacture daily be designated by x_1, x_2, x_3 respectively.

- **Step 3:** Express feasible alternative mathematically in terms of the variables. These feasible alternatives are those which are physically economically and financially possible. Since, it is not possible to manufacture any negative quantitive, it is quite obvious that in the present situation feasible alternatives are sets of values of x_1, x_2 and x_3 satisfying $x_1 \geq 0$, $n_2 \geq 0$, $x_3 \geq 0$.

- **Step 4:** Identity the objective quantitatively and express it as a linear function of variables. The objective here is to maximize the profit. In view of the assumption that all the units produced are consumed in the market, it is given by the linear function

  $z = 3x_1 + 2x_2 + 4x_3$

- **Step 5:** Express in words the influencing factors or constraints (or restrictions) which occur generally because of the constraints on availability (resources) or requirements (diamonds). Express then restrictions also as linear equatitive/inequalitive in terms of the variables. Here, in order to produce x_1 units of product A, x_2 units of product B and x_3 units of product C the total times needed an machines P and Q are given by

  $4x_1 + 3x_2 + 5x_3$ and $2x_1 + 2x_2 + 4x_3$ respectively. Since the manufacturer does not have more than 2000 minutes available an machine P and 2500 minutes available on machine Q, we must have

  $4x_1 + 3x_2 + 5x_3 \leq 2000$

  and $2x_1 + 2x_2 + 4x_3 \leq 2500$

  Also, the manufacturer has to satisfy the following given additional restrictions.

  $x_1 \leq 150$, $x_2 \geq 100$, $x \geq 200$ and $x_3 \geq 50$

Hence, the manufacture's problem can be put in the following mathematical form.

Determine there non-negative real member x_1, x_2 and x_3 such that:

(i) $4x_1 + 3x_2 + 5x_3 \leq 2000$

(ii) $2x_1 + 2x_2 + 4x_3 \leq 2500$

(iii) $100 \leq x \leq 150$, $x_2 \geq 250$, $x_3 \geq 50$

and for which the expression (objective function)

  $z = 3x_1 + 2x_2 + 4x_3$

May be a maximum (greatest).

The inequations (i), (ii), (iii) are called the constraint of the liners programming problem.

This problem is called a linear programming problem, since a liner function of there variability x_1, x_2 and x_3 defines by

$$z = 3x_1 + 2x_2 + 4x_3$$

is maximized subject to the liners constraints, (i), (ii), (iii).

4.2　THEORY OF THE SIMPLEX METHOD (NON-DEGENERACY EXCLUDED)

Simplex method is an iterative procedure that allows to improve the solution at each step. This procedure is finished when it is not possible to improve the solution.

Starting from a random vertex value of the objective function, Simplex method tries to find repeatedly another vertex value that improves the one you have before. The search is done through the side of the polygon (or the edges of the polyhedron, if the number of variables is higher). As the number of vertices (and edges) is finite, it will always be able to find the result.

Simplex method is based on the following property: "If objective function, F, does not take the maximum value in the A vertex, then there is an edge starting at A, along which the value of the function grows.

You should take care about Simplex method only works with "≤" type inequality and independent coefficients higher or equal to zero, and you will have to standardise the restrictions for the algorithm. In case after this procedure, "≥" or "=" type restrictions appear (or not modified), you should try other ways, being Two-phase Simplex method the best choice.

Preparing the model to adapt it to the Simplex method:

This is the standard way of the model.

Objective function: $c_1 . x_1 + c_2 . x_2 + ... + c_n . x_n$

Subject to:　　　　　　　$a_{11} . x_1 + a_{12} . x_2 + ... + a_{1n} . x_n = b_1$

　　　　　　　　　　　　$a_{21} . x_1 + a_{22} . x_2 + ... + a_{2n} . x_n = b$

　　　　　　　　　　　　$am_1 . x_1 + am_2 . x_2 + ... + a_{mn} . x_n = b_m$

　　　　　　　　　　　　$x_1, ..., x_n \geq 0$

To do this, you must follow these rules:

- The objective must be maximise or minimise the function.
- All restrictions must be equal.
- All variables are not negatives.
- The independent terms are not negatives.
- Changing the optimisation type.

If we want to minimise our model, we can keep it, but we must consider the new criteria for the halt condition (stop iterations when all coefficients in the value objective function row are less or equal to zero), and the leaving condition row. In order to not change criteria, we can convert the minimise objective function F to maximise objective F·(–1).

Advantages: We will not have to worry about halting criteria, or exit condition of rows, since they keep on.

Inconveniences: In the event if the function have all its basic variables positive, and further the restrictions are inequality "≤", they become negative when doing the change and plus signs remain in

the row of the value of the objective function, then Simplex method obeys the halting condition, and that optimal value obtained would be 0, by default.

Solution: In fact, this kind of problem does not exist, since the solution is greater than 0, any restriction should have the condition "$\geq$", and then we would go into a model for the Two-phase Simplex method.

Converting the independent term sign (constants to the right of restrictions):

We will have to arrange our model so that the independent terms of restrictions will be greater or equal to 0, if not, Simplex method cannot be used. The only thing that would be necessary to do is multiply by "–1" the restrictions where independent terms be less than 0.

Advantages: With this simple modification of signs in restriction, we can use Simplex method.

Inconveniences: It can work out in restrictions where we have to modify the signs of constants, the signs of inequalities be ("=", "$\leq$"), becoming ("=", "$\geq$") what in any event we will have to develop the Two-Phase Simplex method. This inconvenience is not controllable, although it would be able to benefit us only if terms of inequality exist ("$\leq$", "$\geq$"), and terms "$\geq$" coincide with restrictions where the independent term is negative.

Normalisation Restrictions

If an inequality of the type "$\geq$", appears in our model, we will have to add a new variable, called surplus variable s_i, with restriction $s_i \geq 0$. The new variable appears with coefficient equal to zero in the objective function, and subtracting in inequalities.

A problem appears to us, let us see how to solve inequalities that contains an inequality type "$\geq$":

$$a_{11} \cdot x_1 + a_{12} \cdot x_2 \geq b_1 \rightarrow a_{11} \cdot x_1 + a_{12} \cdot x_2 - 1 \cdot x_s = b_1$$

As all our models based on that all its variables are greater or equal than zero, when we do the first iteration in the Simplex's model, the basic variables will not be in the base and they will take value zero, and all others will maintain their values. In this case, our variable x_s, after doing zero to x_1 and x_2, will take the value $-b_1$. The condition of not negativeness will not come true. So it will be necessary to add a new variable, x_r, that will appear in the objective function with zero coefficient, and adding in the inequality of correspondent restriction. Would be left of the following way:

$$a_{11} \cdot x_1 + a_{12} \cdot x_2 \geq b_1 \rightarrow a_{11} \cdot x_1 + a_{12} \cdot x_2 - 1 \cdot x_s + 1 \cdot x_r = b_1$$

This type of variables are called artificial variables, and they will appear when there are inequalities with inequality ("=", "$\geq$"). This will take us compulsorily to accomplish the Two-phase Simplex method, that will explain later on.

If self mode, if inequality has "$\leq$" type, we will have to add a new variable, called slack variable s_i, with restriction s_i "$\geq$" 0. The new variable appears with zero coefficient in the objective function, and adding up in the inequalities.

To sum up, we can let this board, according to the inequality that appears, and with the value that the new variables must be with.

Types of Inequality	Types of Variable
$\geq$	– Surplus + Artificial
=	+ Artificial
$\leq$	+ slak

Starting the Simplex method:

Once we have standardised our model, it can happen to go into the Simplex method or Two-phase Simplex method.

Chart 4.1
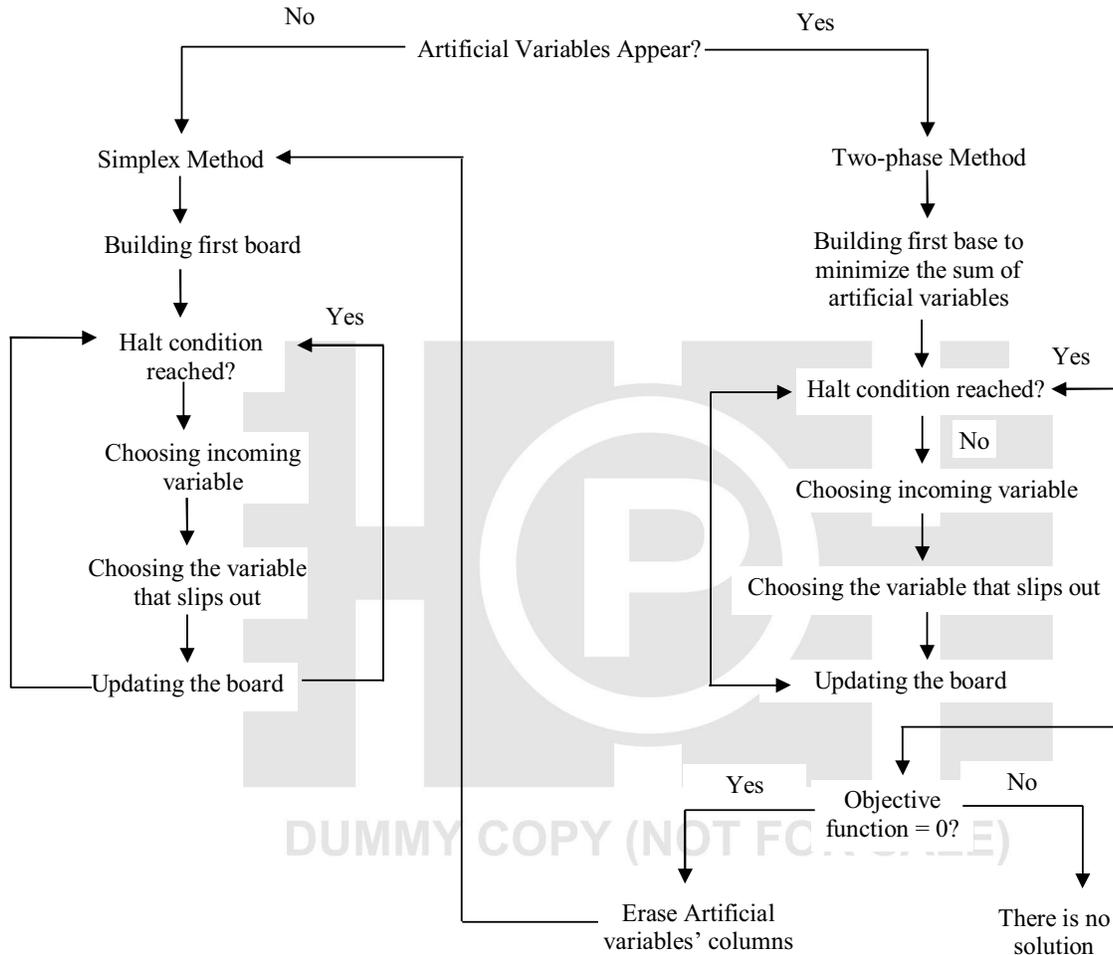
Let us explain each method step by step, concretising the aspects that are necessary to take into account.

Simplex Method

First Board Construction: In the board's first column will appear that we will call base; in the second one, the coefficient that each variable that appears at base has in the objective function (we will call this column C_b); in third column, the independent term of every restriction (P_0), and from this column will appear each variable of the objective function (P_i). In order to have a more obvious vision of the board, we will include a row that we will put each one of the names of columns in. On this board we have, we will include two new rows: one that will lead the board, where the constants of the coefficients of the objective function will appear, and another one that will be the last row, where the objective function will take value. Our final board will have such rows as restrictions.

Tableau

Base	C_b	P_0	C_1 P_1	C_2 P_2	....	C_n P_n
P_1	Cb_1	b_1	a_{11}	a_{12}	....	a_{1n}
P_2	Cb_2	b_2	a_{21}	a_{22}	....	a_{2n}
......	....	....	....	....…	....	......
P_m	Cb_m	b_m	am_1	am_2		a_{mn}
Z		Z_0	$Z_1 - C_1$	$Z_2 - C_2$	....	$Z_n - C_n$

Z row's values are obtained this way: Z_0 value will be the result of substituting C_{im} in the objective function (zero else appears in the base). The left columns are obtained by subtracting to this value the one belonging to the coefficient that appears in the board's front row.

It will be observed when realizing Simplex method, that slack variables will be in the base, in this first table:

- **Halt Condition:** We will check whether we must do a new iteration or not to do, that it will be known if in Z row appears any negative value. If this is not the case, it means we have reached the problem's optimal solution.

- **Choosing incoming variable:** If halt condition has not come true, we must choose one variable to enter the base in the next board. For it we look for strictly negative values of the Z row, and the minor will be which give us the incoming variable.

- **Choosing the variable that slips out:** Once we have obtained the incoming variable, the coming out variable will be reached, having nothing else to do that selects the row whose quotient P_0/P_j be the lowest among strictly positives (considering that only will be done when P_j be greater than 0). The intersection among incoming column and the coming out row will determine the pivot element.

- **Updating the board:** The correspondent rows to the objective function and titles will remain unaltered in the new board. Left rows will be calculated in two ways.

If we are trying with pivot row, each element will result from:

- New Pivot Row Element = Actually Pivot Row Element/Pivot

The left elements of rows will be reached so:

- New Row Element = Actually Pivot Row Element – (Pivot Column Element from Actually Row * New Row Element).

Two-phase Simplex Method

This method differs from Simplex method that first it is necessary to accomplish an auxiliary problem that has to minimise the sum of artificial variables. Once this first problem is resolved and reorganising the final board, we start with the second phase, that consists in making a normal Simplex.

First Phase

At this first phase, all can be done like in Simplex method, except the first board's construction, halt condition and preparing the board that will be used in the second phase.

First Board Construction: We proceed in same way as Simplex method, but with some differences. Objective Function row is different in the first phase, because the objective function

changes, thus it will appear every term with zero value, which are artificial variables of "–1" value because we are minimizing the sum of this variables (remember that minimize F is the same that maximize F·(–1)).

The other difference for this first board consists in the way of calculating the row Z. It will have to be calculated the following way: The $C_b \cdot P_j$ products will be added for all rows, and to this sum, we must subtract the value that appears (according to the column that we are doing) in the objective function row.

Tableau

Base	Cb	C_0	C_1	C_2	....	C_{n-k}	....	C_n
		P_0	P_1	P_2	....	P_{n-k}	....	P_n
P_1	Cb_1	b_1	a_{11}	a_{12}	....	a_{1n-k}	....	a_{2n}
P_2	Cb_2	b_2	a_{21}	a_{22}	....	a_{2n-k}	......	a_{2n}
....	....	....	....	....	....	....	......	....
P_m	Cb_m	b_m	am_1	am_2	....	a_{mn-k}	....	a_{mn}
Z	....	Z_0	Z_1	Z_2	....	Z_{n-k}	....	Z_n

Being $Z_j = \Sigma(C_b \cdot P_j) - C_j$ with $C_j = 0$ for all decision, slacks and surplus variables and $C_j = -1$ for artificial variables:

- **Halt condition:** The halt condition is the same that in Simplex method. The difference resides is that it can occur two cases when halt condition is reached: the function takes zero value, it means that the original problem has solution, or function takes a different value, suggesting that our model does not have solution.

- **Erasing artificial variables columns:** If we have reached the conclusion that the original problem has solution, we must prepare our board for the second phase. The artificial variables columns will be erased. Modify the objective function row instead original, and calculate Z row in the same way such as the 1st phase's first board.

Noticing Anomalous Cases and Solutions

(a) **Obtaining the solution:** When the halt condition is reached, you can see the values of the basic variables which are in the base and the optimal value that the function takes, looking at P_0 column. In case you are minimising,

(b) **Infinite solutions:** Once the halt condition is obeyed, if you notice that any variable that does not appear in the base, has a 0 value at row Z, it means there are other solutions that give you the same optimal value for the objective function. This is a problem which admits infinite solutions, all them among the segment (or plane portion, or space region, etc. depending on the number of variables) that defines $A_x + B_y = Z_0$. You could do more iterations using incoming variable as any of the variables in the Z row which have zero value, and you would have other solutions.

(c) **Unbounded solution:** When you are searching the outgoing variable you notice that every variable in the incoming variable column have all their elements negative or void, it is a problem which has an unbounded solution. So, there is no optimal concrete value. If the values of the variables grow, the objective function value also grows without violating any restriction.

(d) **Solution does not exist:** In case there seems to be no solution, then we will have to solve it using Two-phase Simplex method. So, at the end of the 1st phase, we will know if we are in such situation.

(e) **Tie of incoming variable:** You can choose any one of them, unless it affects the final solution. the inconvenience is that it presents informations and you will have to do more or less iterations.

(f) **Tie of coming out variable:** Again you can choose anyone of them. In order to avoid them as far as possible, we will have prejudice in favor of basic variables.

(g) **Curiosity in the 1st Phase:** When the first phase finalises, if the original problem has solution, all the artificial variables in the Z row must have value "1".

(h) **Can the pivot be 0?:** It cannot be 0, because quotients must be greater than 0.

Theorem Reduction of Feasible Solution to a Basic Feasible Solution

If an L.P.P. has a feasible solution, these it also has basic feasible solution.

Proof: Let the L.P.P. be to determine x so as to maximize $z = c^T x$ subject to the constraints. $Ax = b, x \geq 0$

Where, A in $m \times n$ real matrix and b, c are $m \times 1$, and $n \times 1$, real matrices respectively. Let $p(A) = m$.

Since, there does exist a feasible solution, we must have

$\xi P(A, b) = P(A)$ and $m < n$

Le, $x = (x_1, x_2 ...., x_n)$ be a feasible solution so that $x_j \geq 0$ for all j.

To be precise, let us suppose that x has p positive components and let the remaining $n - p$ components be all zero.

Let us so re-label our components that the positive components are the first p components and assume. That the column of A have re-labeled accordingly.

Then $\sum\limits_{j=1}^{P} a_j n_j = b$

Where, $a_1, a_2, ..., a_p$ are the first P column of A.

Two case new de arise:

(i) The vectors $a_1, a_2, ..., a_p$ form a linearly independent set. The $P \leq m$.

If $P = m$, the given solution is a non-degenerate basic feasible solution, with $x_1, x_2, ..., x_p$ as the basic variables. It pcm, then the set $(a_1, a_2 ...., a_p)$ can be extended to $(a_1, a_2 ...., a_p, a_{p+1}, ..., a_m)$ to form a basic for the columns of A.

Them, are have

$\sum\limits_{j=1}^{m} j_j a_j = b$

Where, $x_j = 0$ for $j = p + 1, p + 2, ..., m$.

Thus, me have, in this case, a degenerate basic feasible solution with m.p of the basic variables zero.

(ii) The set $(a_1, a_2 ...., a_p)$ is linearly dependent.

Let, $(\propto_1, \propto_2 ...., \propto_p)$ be a set of consonants (not all zero) such that,

$$\sum_{j=1}^{p} \propto_j a_j = 0$$

Suppose, that for any index r, $a_r \neq o$. Then

$$a_r = - \sum_{\substack{j=1 \\ j \neq r}}^{P} \frac{\propto_j}{\propto_r} a_j$$

Substituting in the relation $\sum_{j=1}^{p} a_j x_j = b$, we get

$$\sum_{\substack{j \pm r \\ j=1}}^{P} x_j a_j + \left(- \sum_{\substack{j \pm r \\ j=1}}^{P} \frac{\propto_j}{\propto_r} a_j \right) x_r = b$$

or,

$$- \sum_{\substack{j \neq r \\ j=1}}^{P} \left(x_j - x_r \frac{a_j}{a_r} \right) a_j = b$$

Thus, we have a solution with not more than $P - 1$ non-zero components. To ensure that then are non-negative, we shall choose a, in such a way that,

$$x_j - x_r \frac{\propto_j}{\propto_r} \geq 0 \text{ for all } j \neq r$$

This requires, that either aj = 0, or

$$\frac{x_j}{\propto_j} \geq \frac{x_r}{\propto_r} \text{ if } \propto_j > 0 \text{ and } \frac{x_j}{\propto_j} \leq \frac{x_r}{\propto_r} \text{ if } \propto_j < 0$$

Thus if we select a, such that,

$$\frac{x_r}{\propto_r} = \min_j \left\{ \frac{x_j}{\propto_j} \propto_j \geq 0 \right\}$$

There, for each of the p − 1 variables, $x_j - x. \dfrac{\propto_j}{\propto_r}$ is non-negative and so, we have a feasible solution with not more then p − 1 non-zero components.

Consider now this new feasible solution with not more than p − 1 non-zero components. If the corresponding not of p − 1 columns of A is linearly independent, case (i) applier and we have arrived at a basic feasible solution. If this set is again linearly dependent, we way repeat the process to arrive at a feasible solution with not more than p − 2 non-zero components. The argument can be repeated. Ultimately an get a feasible solution with associated set of column vectors of A, linearly independent. The discussion of case (i) then applier and we do get a basic feasible solution.

Examples: If $x_1 = 1$, $x_2 = 1$, $x_3 = 1$ is a feasible solution to the system of equations

$$x_1 + x_2 + 2x_3 = 4$$

$$2x_1 - x_2 + x_3 = 2$$

If this solution a basic feasible solution: it not reduce the given feasible solution to a basic feasible solution.

Solution: The given system of equations can be written as

$$\begin{pmatrix} 1 & 1 & 2 \\ 2 & -1 & 1 \end{pmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

or $A_x = b$.

Since $p(A) = 2$, therefore, there are two linearly independent columns vectors of A. Let the column vectors of A to be denoted by a_1, a_2, a_3 respectively.

Thus,

(i) $\alpha_1 a_1 + \alpha_2 a_2 + \alpha_3 a_3 = 0$, $\alpha_i^!$'s are scalars not all zero.

We must have,

$\alpha_1 + \alpha_2 + 2\alpha_3 = 0$ and $2\alpha_1 - \alpha_2 + \alpha_3 = 0$

These give $\alpha_1 = -\alpha_3$

Taking $\alpha_3 = 1$, we get $\alpha_1 = -1$ and $\alpha_2 = -1$, so we can write

$-a_1 - a_2 + a_3 = 0$

The given system of equations is $x_1 a_1 + x_2 a_2 + x_3 a_3 = b$,

where, $x_1 = x_2 = x_3 = 1$, is a solution.

In order to reduce the number of positive variables, the vectors to be removed is selected is accordance with above theorem.

New, $\dfrac{x_r}{\alpha_r} = \min_{j} \left\{ \dfrac{x_j}{\alpha_j} \alpha_j > 0 \right\} = \min_{j} \left\{ \dfrac{1}{1} \right\} = 1$

The vector, for which $\dfrac{x_r}{\alpha_r} = 1$, is a_3 and so are can remove the vector a_3 and obtain a new

solution with not more then non-zero variable. Thus, the values of the new variable are given by

$$\hat{x}_j = x_j - \dfrac{x_r}{\alpha_r} \alpha_j \quad j = 1, 2$$

This gives,

$$\hat{x}_1 = x_1 - \dfrac{x_r}{\alpha_r} \alpha_1 = 1 - 1(-1) = 2$$

$$\hat{x}_2 = x_2 - \dfrac{x_r}{\alpha_r} \alpha_2 = 1 - 1(-1) = 2$$

Thus, [2, 2, 0] is also a feasible solution and it has only two non-zero components and so $2a_1 + 2a_2 = b$.

Now, since, a_1, a_2 linearly independent, [2, 2, 0] is a basic feasible solution with $x_3 = 0$ as a non-basic variable.

Note: The choice of $\propto_1$'s in the relation.

$\propto_1 a_1 + \propto_2 a_2 + \propto_3 a_3 = 0$ has a bearing on the basic feasible solution obtained. For instance, if we take

$\propto_1 = 1$, $\propto_3 = -1$, and $\propto_2 = 1$, so that $a_1 + a_2 - a_3 = 0$

then,

$$\min_{j}\left\{\frac{n_j}{\propto_j} \propto_j > 0\right\} = \min\left\{\frac{1}{1},\frac{1}{1}\right\} = 1$$

Here,

$$\frac{x_1}{\propto_1} = \frac{x_2}{\propto_2} = 1 = \min_{j}\left\{\frac{n_j}{\propto_j} \propto_j > 0\right\}$$

Thus, a_1 and a_2 are equally eligible for removed

In case, a_1 is removed, we get

$$\hat{x}_2 = x_2 - \frac{x_r}{\propto_r} \propto_2 = 1 - 1 = 0$$

$$\hat{x}_3 = x_3 - \frac{x_r}{\propto_r} \propto_3 = 1 - (-1) = 2$$

and is case, a_2 is removed, we get,

$$\hat{x}_1 = x_1 - \frac{x_r}{\propto_r} \propto_1 = 1 - 1 = 0$$

$$\hat{x}_3 = x_3 - \frac{x_r}{\propto_r} \propto_3 = 1 - 1 = 2$$

If can be shown, as before, that the two solution are basic.

Hence, corresponding to the choice of $\propto_j$'s made here, we get two basic feasible solution (both degenerate):

(i) [u, u, 2] with $x_1 = 0$, non-basic, and (ii) [u, u, 2] with $x_2 = 0$ non basis'.

4.3 DUALITY THEOREMS

The dual to an LP in standard form

(P)	Maximise	$c^T x$
	Subject to	$Ax \le b$, $0 \le x$
	is the LP	

(D)   Maximise $b^T y$

Subject to $A^T y \geq c, \ 0 \leq y$

Since the problem D is a linear program, if also has a dual. The duality terminology suggests that the problems P and D came as a pair implying that the dual to D should be P. This is indeed the case as we we now show.

Minimise $b^T y$ – Maximise $(-b)^{T y}$

Subject to $A^T y \geq c$ = Subject to $(- A^T)y \leq (-C)$

$0 \leq y$ $0 \leq y$

The problem on the right is in standard form. So, we can take its dual to get the LP.

Maximise $(-c)^T{}_x$ $Ms|x|= -$ maximise $c^T{}_x$

Subject to $(- A^T)^T{}_x \geq (-b),$ subject to $Ax \leq b,$

$0 \leq x$ $0 \leq x$

The primal-dud pair of LPs P-D are related via the Weak Duality Theorem.

Weak Duality Theorem

If $x \in R^n$ is feasible for P and $y \in R^m$ is feasible for D, then

$c^T x \leq y^T Ax \leq b^T y$

Thus, if P is unbounded, then D is necessarily infeasible and if D is in unbounded, then P is necessarily infeasible. Moreover, if $c^T \bar{x} = b^T \bar{y}$ with $\bar{x}$ feasible for P and $\bar{y}$ feasible for D, then $\bar{x}$ must solve P and $\bar{y}$ must solve D.

Let us now use the Weak Duality Theorem in conjunction with the Fundamental Theorem of Linear Programming to prove the Strong Duality Theorem.

Strong Duality Theorem

If either P or D has a finite optimal value, then so does the other, the optimal values coincide, and optimal solutions to both P and D exit.

Remark: This result state that the finiteness of the optimal value implies the existence of a solution. This is not always the case for non-linear optimisation problems.

Min e^x
$x \in R$

This problem has a finite optimal value, namely zero. However, this value is not attained by any point $x \in R$, i.e., it has a finite optimal value, but a solution does not exist. The existence of solutions where the optional value is finite is one of the many special properties of linear programs.

Proof: Since the dual of the dual is the primal, we may assume that the primal has a finite optimal value. In this case, the Fundamental Theorem of Linear Programming says that an optimal basic feasible solution exists. By our formula for the general form, we know that there exists a non-singular record matrix $R \in R^{n \times n}$ and a vector $y \in R^m$ such that the optimal tableau has the form

$$\begin{bmatrix} R & 0 \\ -y^T & 1 \end{bmatrix} \begin{bmatrix} A & I & b \\ e^T & 0 & 0 \end{bmatrix} = \begin{bmatrix} RA & R & R_b \\ e^T - y^{TA} & -y^T & y^{Tb} \end{bmatrix}$$

Since this is an optimal tableau, we know that,

$$C - A^T y \leq 0, - y^T \leq 0$$

with $y^T b$ equal to optimal value in the primal problem. But, the $A^T y \geq 0$ and $0 \leq y$ so that y is feasible for the dual problem D. In addition, the Weak Duality Theorem implies that

$$b^T y = \text{Maximise } c^T x \leq b^T \hat{y}$$

$$\text{Subject to } Ax \leq b, 0 \leq x$$

for every vector $\hat{y}$ that is feasible for D. Therefore, y solves D.

In this part, it will be shown that the optimum dual solution can be obtained directly from the optimum solution of primal linear programming problem and *vice verse.*

Theorem Weak Duality Theorem

Let, x_0 be a feasible solution to the Primal Problem:

Maximize $(fx) = c^T x$, subject to $Ax \leq b$, $x \geq 0$ where, x and $c \in R^n$, $b \in R^m$ and A is an $m \times n$ real matrix. It w_0 be a feasible solution to the dual of the primal, namely minimize $g(w) = b^T w$ subject to $A^T w \geq c$, $w \geq 0$ where, $w \in R^m$, then $c^T x_0 \leq b^T w_0$

Prof: Since, x_0 and w_0 are the feasible solutions to the primal and its and dual respectively, we have,

$Ax_0 \leq b$, $x_0 \geq 0$ and $A^T w_0 \geq c$, $w_0 \geq 0$

New, $a^T w_0 \geq c \Rightarrow e^T \leq w_0^T A$

or, $c^T x_0 \leq w_0^T A x_0 \leq w_0^T b$ [since, $Ax_0 \leq b$]

or, $c^T x_0 \leq b^T w_0$ [since, $w^T b = b^T w_0$]

Theorem: Let, no. be a feasible solution to the primal problem: Maximize $g(w) = b^T w$ subject to $a^T w \geq c$, $w \geq 0$ where, x and $c \in R^n$, w and $b \in R^m$ and A is an $m \times n$ real matrix. If

$$c^T x_0 = b^T w_0$$

Then both x_0 and w_0 are optimal solutions to the primal and its dual respectively.

Proof: Let x_0^* be any other feasible solution to the primal problem. Then above theorem given

$$c^T x_0^* \leq b^T w_0$$

$$\leq c^T_0 \text{ (Since, } c^T x_0 = b^T w_0)$$

and hence, x_0 is an optimum solution to the primal problem, because primal is a maximization problem, because primal is a maximization problem.

Similarly if w_0^* is any there feasible solution to the and dual problem, then

$$b^T w_0 \leq b^T w_0^*$$

and thus, w_0 is an optimum solution to the dual problem.

Theorem Basic Duality Theorem

Let a primal problem be Maximize $f(x) = c^T_x$ subject to $Ax \leq b$, $x \geq 0$ x, $c \in R^n$ and the associated dual be minimize $g(w) = b^T w$ subject to $A^T w \geq c$, $w \geq 0$ and w, $b \in R^m$. If x_0 (w_0) is are

optimum solution to the primal (dual), then there exists a feasible solution $w_0(x_0)$ to the dual (primal), such that

$c^T x_0 = b^T w_0$

Proof: The standard primal can be written as maximize $z = c^T x$ subject to $Ax + ix_j = b$ where, $x_j \in R^m$ is the slack vector and I is the associated $m \times n$ identity matrix.

Let, $x_0 = [x_B, 0]$ be an optimum solution to the primal, where, $x_B \in R^m$ is the optimum basic solution given by $x_B = B^{-1} b$, B being the optimal basic of A. Then, the optimal primal objective function is

$z = c^T x_0 = c_B^T x_B$

Where, c_B is the cost vector associated with x_B.

Now, net evaluations with optimal a simplex table an given by

$$z_j - c_i = c_B^T y_j - c_j = \begin{cases} c_B^T B_{aj}^{-1} - c_j & \forall a_j \in A \\ c_B^T B_{\propto j}^{-1} - 0 & \forall \propto_j \in I \end{cases}$$

Since x_0 is optimal, we must have $z_j - c_j \geq 0$, for all j. This gives

$c_B^T B^{-1} a_j \geq c_j$ and $c_B^T B^{-1} \propto_j \geq 0$ for all j

or $c_B^T B^{-1} A \geq c^T$ and $c_B^T B^{-1} \geq 0$ in matrix form

or $A^T B^{-1} c_B \geq c$ and $B^{-1} c_B \geq 0$

New, if we let $B^{-1} c_B = w_0 \in R^m$, the above become

$A^T w_0 \geq c$ and $w_0 \geq 0$.

This means that w_0 is a feasible solution to the dual problem. Moreover, the corresponding dual objective function value is

$b^T w_0 = w_0^T b = c_B^T B^{-1} b = c_B^T \times B = c^T x_0$

Thus, given an optimal problem x_0 to the primal, there exists a feasible solution w_0 to the dual such that $c^T x_0 = b^T w_0$. Similarly, starting with w_0, the existence of x_0 can be proved.

Corollary: if x_0 is an optimal solution to the primal, an optimal solution to the dual is given by $w_0 = B^{-1} c_B$. Where B is the primal optimal basis.

Note: Observe that $c_B^T B^{-1}$ represent optimum $z_j - c_j$'s under primal slack columns.

4.4 COMPLEMENTARY SLACKNESS THEOREM

One of the major theorems in the theory of duality in Linear Programming is the Complementary Slackness Theorem. This theorem allows us to find the optimal solution of the dual problem when we know the optimal solution of the primal problem (and *vice versa*) by solving a system of equations formed by the decision variables (primal and dual) and constraints (primal and dual model).

The importance of this theorem is that it facilitates the resolution of the models of linear optimisation, allowing you to find the simplest model to address (from the algorithmic point of view) because either way you will get the results of the associated equivalence model (may it be a primal or dual model).

Let us consider the following Linear Programming model (here-in-after primal) in two variables whose optimal solution is $X = 14/5$ and $Y = 8/5$ with optimal value $V(P) = 20.8$.

$V(P) = 20.8.$

> Max $\quad 4X + 6Y$
> S.A $\quad 2X + 4Y \leq 12$
> $\qquad 4X + 3Y \leq 16$
> $\qquad X \geq 0, \quad Y \geq 0$

The dual model associated with the primal model is:

> Min $\quad 12A + 16B$
> S.a. $\quad 2A + 4B \geq 4$
> $\qquad 4A + 3B \geq 6$
> $\qquad AB \geq 0.$

Then the Complementary Slackness Theorem shows us the following relationships:

As we know $X = 14/5$ and $Y = 8/5$ (primal optimal solution), if we replace these values of X and Y in the third and fourth equation, we generate a 2×2 system of equations in terms of A and B whose solution corresponds to $A = 6/5$ and $B = 2/5$ (a feasible and optimal solution of the dual model). If we subsequently evaluate the objective function in the dual problem of this solution, we obtain: $V(D) = 12(6/5) + 16(2/5) = 20.8$ which is similar to the primal problem's optimal value (satisfies the Strong Duality Theorem).

Notes: Note that the 1 and 2 constraints of the primal problem are active at the optimum, i.e., equality is met.

Theorem: (Complementary Slackness)

Let x_0 and w_0 be the feasible solution to the primal {Max. $c^Tx \mid Ax \leq b, x \geq 0$} and its dual {Min. $b^Tw \mid A^Tw \geq c, w \geq 0$} respectively. Then, a necessary and sufficient condition for x_0 and w_0 to be optimal to, their respective problem is that,

$$w_0^T (b - Ax_0) = 0 \text{ and } x_0^T (A^Tw_0 - c) = 0$$

Proof: Necessity let $\propto = w_0^T (b - Ax_0)$ and $\beta = x_0^T (A^Tw_0 - c)$. Since, x_0 w_0 are feasible solutions, to the primal and dual, respectively, we have

$$\propto \geq 0, \beta \geq 0 \text{ and } \propto -1\beta = w_0^T\beta - x_0^Tc$$

Now, if x_0, w_0 are optimal, then $c^T_{x_0} = b^Tw_0$ so that $\propto + \beta = 0$. But, since, $\propto \geq 0$ and $\beta \geq 0$, this gives $\propto = 0$ and $\beta = 0$. Thus, the conditions are necessary.

Sufficiency: Let the given conditions hold for the feasible solutions x_0 and w_0. That is,

$$\propto = 0 \text{ and } \beta = 0$$

Thus, $0 = \propto + \beta = w_0^T b - x_0^T c$

$\Rightarrow c^T x_0 = b^T w_0$

$\Rightarrow x_0$ and w_0 are optimal.

Thus, the conditions are sufficient.

Corollary 1: If x^0 and w^0 be feasible solution to the primal and dual problems respectively, there the will be optimal if and only if

$$w_i^0 \left(b_1 - \sum_{j=1}^{n} a_i x_j \right) = 0, \, i = 1, 2, \ldots\ldots, m$$

and $x_i^0 \left(\sum_{i=1}^{m} a_{ji} w_i^0 - c_j \right) = 0, \, j = 1, 2, \ldots\ldots, n$

Proof: From the above theorem, x_0 and w_0 will be optimal if and only if

$$w_0^T (b - Ax_0) = 0 \text{ and } x_0^T (c - A^T w_0) = 0$$

Consider the first set of conditions. Since each term in the summation $w_0^T (b - Ax_0)$ is non-negative, if follow that

$$w_i \left(b_i - \sum_{j=1}^{n} a_{ij} x_j^0 \right) = 0, \, i = 1, 2, \ldots\ldots, m$$

Similarly, the second set of conditions is equivalent to

$$x_j \left(\sum_{i=1}^{m} a_{ij} w_i^0 - c_j \right) = 0, \, j = 1, 2, \ldots\ldots, n$$

Corollary 2: For optimal feasible solutions of the primal and dual systems, whenever, the i th variable is strictly positive in either system, the i th relation of its dual is an equality.

Proof: It follows from corollary 1, that

$$w_i^0 > 0 \Rightarrow \sum_{j=1}^{n} a_{ij} x_j^0 = b_i \text{ (i th primal relation)}$$

and $c_i^0 > 0 \Rightarrow \sum_{i=1}^{m} a_{ij} w_i^0 = c_{ij}$ (j th dual relation)

Corollary 3: For optimal feasible solutions of the primal and dual systems, whenever i th relation of either systems is satisfied as a strict inequality, then the i th variable of its dual vanishes.

Proof: It follows from corollary 1, that

$$\sum_{j=1}^{n} a_{ij} x_j^0 \leq b_i \Rightarrow w_i^0 = 0$$

and $\sum_{i=1}^{m} a_{ij} w_i^0 > c_j \Rightarrow x_j^0 = 0$

Remarks: The conditions of corollary 1 can also be written as

$$w_i^0 \cdot x_{n+i} = 0 \quad i = 1, 2, \ldots\ldots, m$$

and $x_j^0 \cdot w_{m+j} = 0 \quad j = 1, 2, \ldots\ldots, n$

Where , x_{n+i} is the i th slack variable in the primal problem and w_{m+j} is the j th surplices variable in the dual.

Thus, the theorem relates the variables are problem to the slack or surplices variable of the other.

The above relations are called complimentary slackness because they imply that whenever a constrains in are of the problems holds with strict inequality (so that there is a slack in the constraint) the complementary dual variable vanishes.

4.5 SUMMARY

1. Simplex method is an iterative procedure that allows to improve the solution at each step. This procedure is finished when it is not possible to improve the solution.

2. One of the major theorems in the theory of duality in Linear Programming is the Complementary Slackness Theorem. This theorem allows us to find the optimal solution of the dual problem when we know the optimal solution of the primal problem (and *vice versa*) by solving a system of equations formed by the decision variables (primal and dual) and constraints (primal and dual model).

4.6 SELF ASSESSMENT QUESTIONS

1. Critically explain Linear Programming. What is its importance?
2. Explain in detail 'Simplex Method'.
3. Critically analyse Duality Theorem.
4. Explain in detail 'Complementary Slackness Theorem'.

❧❦❖❦❧

NON-LINEAR PROGRAMMING

Objectives

The objectives of this lesson are to learn:

- Non-linear Programming
- Kuhn-Tucker Optimality Criteria
- Saddle Points
- Concave Programming

Structure:

5.1   NON-LINEAR PROGRAMMING

In mathematics, non-linear programming is the process of solving an optimisation problem where some of the constraints or the objective function are non-linear. An optimisation problem is one of calculation of the extrema (maxima, minima or stationary points) of an objective function over a set of unknown real variables and conditional to the satisfaction of a system of equalities and inequalities, collectively termed constraints. It is the sub-field of mathematical optimisation that deals with problems that are not linear.

Applicability

A typical non-convex problem is that of optimising transportation costs by selection from a set of transportation methods, one or more of which exhibit economies of scale, with various connectivities and capacity constraints. An example would be petroleum product transport given a selection or combination of pipeline, rail tanker, road tanker, river barge, or coastal tankship. Owing to economic batch size, the cost functions may have discontinuities in addition to smooth changes.

In experimental science, some simple data analysis (such as fitting a spectrum with a sum of peaks of known location and shape but unknown magnitude) can be done with linear methods, but in general, these problems also are non-linear. Typically, one has a theoretical model of the system under study with variable parameters in it and a model of the experiment or experiments, which may

also have unknown parameters. One tries to find a best fit numerically. In this case, one often wants a measure of the precision of the result as well as the best fit itself.

Definition

Let n, m, and p be positive integers. Let X be a subset of R^n. Let f, g_i and h_j be real-valued functions on X for each i in $\{1, …, m\}$ and each j in $\{1, …, p\}$, with at least one of f, g_i and h_j being non-linear.

A non-linear minimisation problem is an optimisation problem of the form:

Minimise f(x)

Subject to $g_i(x) \leq 0$ for each $i \in \{1, …, m\}$

$h_j(x) = 0$ for each $j \in \{1, …, p\}$

$x \in X$

A non-linear maximisation problem is defined in a similar way.

Possible Types of Constraint Set

There are several possibilities for the nature of the constraint set also known as the feasible set or feasible region.

An infeasible problem is one for which no set of values for the choice variables satisfies all the constraints, i.e., the constraints are mutually contradictory, and no solution exists; the feasible set is the empty set.

A feasible problem is one for which there exists at least one set of values for the choice variables satisfying all the constraints.

An unbounded problem is a feasible problem for which the objective function can be made to be better than any given finite value. Thus, there is no optimal solution, because there is always a feasible solution that gives a better objective function value than does any given proposed solution.

Methods for Solving the Problem

If the objective function f is linear and the constrained space is a polytope, then the problem is a linear programming problem, which may be solved using well-known linear programming techniques such as the Simplex method.

If the objective function is concave (maximisation problem) or convex (minimisation problem) and the constraint set is convex, then the program is called convex and general methods from convex optimisation can be used in most cases.

If the objective function is quadratic and the constraints are linear, quadratic programming techniques are used.

If the objective function is a ratio of a concave and a convex function (in the maximisation case) and the constraints are convex, then the problem can be transformed to a convex optimisation problem using fractional programming techniques.

Several methods are available for solving non-convex problems. One approach is to use special formulations of linear programming problems. Another method involves the use of branch and bound techniques, where the program is divided into sub-classes to be solved with convex (minimisation problem) or linear approximations that form a lower bound on the overall cost within the subdivision.

With subsequent divisions, at some point, an actual solution will be obtained whose cost is equal to the best lower bound obtained for any of the approximate solutions. This solution is optimal, although possibly not unique. The algorithm may also be stopped early, with the assurance that the best possible solution is within a tolerance from the best point found. Such points are called ε-optimal. Terminating to ε-optimal points is typically necessary to ensure finite termination. This is especially useful for large, difficult problems and problems with uncertain costs or values where the uncertainty can be estimated with an appropriate reliability estimation.

Under differentiability and constraint qualification, the Karush-Kuhn-Tucker (KKT) conditions provide necessary conditions for a solution to be optimal. Under convexity, these conditions are also sufficient. If some of the functions are non-differentiable, sub-differential versions of Karush-Kuhn-Tucker (KKT) conditions are available.

The linear programming problem which can be review as to

$$\text{Maximize } Z = \sum_{j=1}^{n} e_j x_j$$

$$\text{Subject to } \sum_{j=1}^{n} a_{ij} \ x_j \le b_i \text{ for } i = l, z, \ldots, m$$

and $xj \ge o$ for $j = l, L, \ldots, m$

The term 'non-linear programming' usually refer to the problem in which the objective function (1) become non-linear or one or more of the constraint inequalities (2) have non-linear or both.

Example: Consider the following problem

Maximize (Minimize) $Z = x_1^2 - x_2^2 + x_3^2$

Subject to $x_1 + x_2 + x_3 = 4$ and $x_1, \ x_2, \ x_3 \ge 0$

Multivariate Optimization without Constraints

A nonlinear multivariable optimization without constraints has the form:

Maximize $f(x_1, x_2, \ldots, x_n)$

with $x_1, x_2, \ldots, x_n \ge o$

Local and Global Maxima

Definition: An objective function $f(x)$ has a maximum at $\hat{x}$ there exist an ε - neighborhood around $\hat{x}$ ---- $f(x) \le f(\hat{x})$ for all x in this ∈ - neighborhood at which the function is defined. If the condition is must for every positive and them $f(x)$ has a global maximum at x.

Unconstrained optimization

We have to optimize $f(x_1, x_2, \ldots x_n)$

In unconstrained type of function we determine extreme points

$$\frac{\delta f}{\delta x_1} = 0$$

$$\frac{\delta f}{\delta x_2} = 0$$

$\vdots$

$$\frac{\delta f}{\delta x_n} = 0$$

For are variable

$$\frac{d^2 f}{\delta x^2} > 0 \quad \text{Then f is minimum.}$$

$$\frac{d^2 f}{\delta n^2} < 0 \quad \text{Then f is maximum.}$$

$$\frac{d^2 f}{\delta x^2} = 0 \quad \text{Then further investigation needed.}$$

For two variable

$rt - s^2 > 0$ Then the function is minimum.

$rt - s^2 < 0$ Then the function is maximum.

$rt - s^2 = 0$ Further investigation needed.

Where, $r = \dfrac{\delta^2 f}{\delta x_1^{\ 2}}, \quad s = \dfrac{\delta^2 f}{\delta x_1 x_2}, \quad t = \dfrac{\delta^2 f}{\delta x_2^{\ 2}}$

For 'n' variable

Hession Matrix

$$\begin{bmatrix} \dfrac{\delta^2 f}{\delta x_1^{\ 2}} & \dfrac{\delta^2 f}{\delta x_1 x_2} & \cdots & \dfrac{\delta^2 f}{\delta x_1 x_n} \\ \dfrac{\delta^2 f}{\delta x_1 x_2} & \dfrac{\delta^2 f}{\delta x_1^{\ 2}} & \cdots & \dfrac{\delta^2 f}{\delta x_2 xn} \\ \vdots & \vdots & & \vdots \\ \dfrac{\delta^2 f}{\delta x_1 x_2} & \dfrac{\delta^2 f}{\delta x_2 x_n} & \cdots & \dfrac{\delta^2 f}{\delta x_n^{\ 2}} \end{bmatrix}$$

$|H| > 0$ at p_1 , + attains minimum at P_1

$|H| < 0$ at p_1 , + attains maximum at P_1

Convex function: A function f(x) is said to be convex function over the region s if for any two points x_1, x_2 belongs S.

We have the function

$f(\lambda n_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2)$

Where, $0 \leq \lambda \leq 1$

S is strictly convex function, if

$f(\lambda x_1 + (1 - \lambda) x_2) < \lambda f(x_1) + (1 - \lambda) f(x_2)$

Concave function: A function f (x) is said to be concave function over the region s it for any two points x_1, x_2 belongs to S

We have the function

$f(xn_1 + (1 - \lambda)x_2) \geq \lambda\ f(x_1) + (1 - \lambda)\ f(x_2)$ where, $0 \leq \lambda \leq 1$

S is strictly concave function if

$f(xn_1 + (1 - \lambda)\ x_2) > \lambda f(x_1) + (1 - \lambda)\ f(x_2)$

Result:

1. Sum of two convex functions is also a convex function.

2. Let $f(x) = x^T\ Ax$ be positive semi definite quadratic form then f(x) is a convex function.

3. Let f(x) be a convex function over convex region S, then a local minima of f(x) is global minima of f(x) in the region S.

4. If f(x) is a strictly convex function over the convex set S them f(x) has unique global minima.

5.2 KUHN-TUCKER OPTIMALITY CRITERIA/KUHN-TUCKER RESULTS

In mathematical optimisation, the Karush-Kuhn-Tucker (KKT) conditions, also known as the Kuhn-Tucker conditions, are first derivative tests (sometimes called first-order) necessary conditions for a solution in non-linear programming to be optimal, provided that some regulatory conditions are satisfied. Allowing inequality constraints, the KKT approach to non-linear programming generalises the method of Lagrange multipliers, which allows only equality constraints. The system of equations and inequalities corresponding to the KKT conditions is usually not solved directly, except in the few special cases where a closed-form solution can be derived analytically. In general, many optimisation algorithms can be interpreted as methods for numerically solving the KKT system of equations and inequalities.

The KKT conditions were originally named after Harold W. Kuhn and Albert W. Tucker, who first published the conditions in 1951. Later, scholars discovered that the necessary conditions for this problem had been stated by William Karush in his Master's Thesis in 1939.

Non-linear Optimisation Problem

Consider the following non-linear minimisation or maximisation problem:

Optimise f(x)

Subject to

$g_i(x) \leq 0$

$h_j(x) = 0$

where x is the optimisation variable, f is the objective or utility function, $g_i(i = l, ..., m)$ are the inequality constraint functions and $h_j(j = 1, ..., l)$ are the equality constraint functions. The number of inequality and equality constraints are denoted m and *l* respectively.

Necessary Conditions

Suppose that the objective function $f : R^n \rightarrow R$ and the constant functions $g_i : R^n \rightarrow R$ and $h_j : R^n \rightarrow R$ are continuously differentiable at a point x*. If x* is a local optimum and the optimisation

problem satisfies some regularity conditions, then there exist constants $\mu_i(i = 1, ..., m)$ and $\lambda_j(j = 1, ..., l)$ *Notes*
called KKT multiplier, such that

Stationarity

For maximising f(x):

$$\nabla f(x^*) = \sum_{i=1}^{m} \mu_i \nabla g_i(x^*) + \sum_{j=1}^{l} \lambda_j \nabla h_j(x^*)$$

For minimising f(x):

$$-\nabla f(x^*) = \sum_{i=1}^{m} \mu_i \nabla g_i(x^*) + \sum_{j=1}^{l} \nabla h_j(x^*)$$

Primal Feasibility

$g_i(x^*) \leq 0$, for i = 1, ..., m

$h_j(x^*) = 0$, for j = 1, ..., l

Dual Feasibility

μ_i , ≥ 0, for i = 1, ..., m

Complementary Slackness

$\mu_i\, g_i(x^*) = 0$, for i = 1, ..., m

In the particular case, m = 0, i.e., when there are no inequality constraints, the KKT conditions turn into the Lagrange conditions, and the KKT multipliers are called Lagrange multipliers.

Sufficient Conditions

In some cases, the necessary conditions are also sufficient for optimality. In general, the necessary conditions are not efficient for optimality and additional information is necessary, such as Second-order Sufficient Conditions (SOSC). For smooth functions, SOSC involves the second derivatives, which explains its name.

The necessary conditions are sufficient for optimality if the objective function f of a maximisation problem is a concave function, the inequality constraints g_j are continuously differentiable convex functions and the equality constraints h_i are affine functions.

Second-order Sufficient Conditions

For smooth, non-linear optimisation problems, a second-order sufficient condition is given as follows:

The solution x^*, λ^* and μ^* found in the above section is a constrained local minimum if for the Lagrangian

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{m} \mu_i g_i(x) + \sum_{j=1}^{l} \lambda_j h_j(x)$$

Then, $s^T \nabla_{xx}^2 L(x^*, \lambda^*, \mu^*)\, s \geq 0$

where, $s \neq 0$ is a vector satisfying the following:

$$\left[\nabla_x g_i(x^*),\ \nabla_x h_j(x^*)\right]^T . s = 0$$

where only those active inequality constraints $g_i(x)$ corresponding to strict complementarity (i.e., where $\mu_i \geq 0$) are applied. The solution is a strict constrained local minimum in case the inequality is also strict.

The real applications of mathematical programming economics contain both types of constraints: inequalities as well as equation. Therefore, we define the general mathematical programming problem as follow:

Minimize $f_i(x, y)$

Subject to $f_i(x, y) \leq 0$ (i = 1, 2, ..., m)

   $g_L(x, y) = 0$ (h = m + 1 , .…., r)

   $x \gneqq 0,$

   $y \in R'$

Obviously, problem (s^a) with $\phi(x, y, u, v)$ $fo(x, y) + \sum_{i=1}^{m} u_i\, f_i(x,y) + \sum_{h=m+1}^{r} v_L \delta_L (x, y)$, we may verify that the Kuhn-Tucker condition take the symmetric form,

$$\frac{\delta\phi^0}{\delta x} \gneqq 0, \quad \frac{\delta\phi^0}{\delta y} = 0, \quad \frac{\delta\phi^0}{\delta u} \lneqq 0 \quad \frac{\delta\phi^0}{\delta v} = 0$$

$$x^0 \; \frac{\delta\phi^0}{\delta x} = 0, \quad u^0 \frac{\delta\phi}{\delta u} = 0, \quad x^0 \geq 0, \quad v^0 \geq 0$$

Where, $\phi^0 = \phi(x^0, y^0, u^0\, v^0)$, (x^0, y^0) denote the local minimum of the function fo (x, y) under the constraints of problem (5a) and (u^0, v^0) are the corresponding Lagrange multipliers. It is worth nothing that the Lagrange multiplier υ related to the equalities are not restricted to the non negation (as the classical Lagrange Theory).

A summary of the rules for the formulation of the Kuhn-Tucker condition for the general mathematical programming problem in an follows:

Rule 1: For a minimization (maximization) problem writ all inequality constraints in the form

$f_i (x) \lneqq 0 \; (f_i(x) \gneqq 0)$

Rule 2: Write the Lagrange function as mum of the objective function and the lighted constrains.

Rule 3: The partial derivatives of the Lagrange function:

(a) With respect to the non-negative variables are non negative (non-positive) for a minimization (maximization) problem and the complimentary slackness condition is fulfilled;

$x \; \frac{\delta\phi}{\delta x} = 0$

(b) With respect to the free variables are equal to zero

(c) With respect to the Lagrange multipliers corresponding to the inequality constraints are non-positive (non-negative) for a minimization (maximization) problem and the complementary slackness condition is fulfilled;

$u \; \frac{\delta\phi}{\delta u} = 0$

(d) with respect to the Lagrange multiplier, corresponding to the equality constraints are equal to zero for a numerical illustration we consider the following example:

minimize $f_0(x) = x_1^2 - 4x_1 + x_2^2 - 6x_2$

Subject to $\quad x_1 + x_2 \leqq 3$

$$- 2x_1 + x_2 \leqq 2$$

The lagrange function is

$$\phi(x, u) = x_1^2 - 4x_1 + x_2^2 - 6x_2 + u_1(x_1 + x_2 - 3) + u_2(-2x_1 + x_2 - 2)$$

Application of the Kuhn-Tucker conditions may be expressed as

$$\frac{\delta\phi}{\delta x_1} = 2x_1 - 4 + u_1 - 2u_2 = 0 \qquad\qquad \text{...(5.1)}$$

$$\frac{\delta\phi}{\delta x_2} = 2x_2 - 6 + u_1 + u_2 = 0 \qquad\qquad \text{...(5.2)}$$

$$\frac{\delta\phi}{\delta u_1} = x_1 + x_2 - 3 \leq 0 \qquad\qquad \text{...(5.3)}$$

$$u_1 \frac{\delta\phi}{\delta u_1} = u_1(x_1 + x_2 - 3) = 0 \qquad\qquad \text{...(5.4)}$$

$$\frac{\delta\phi}{\delta u_1} = -2x_1 + x_2 \leq 0 \qquad\qquad \text{...(5.5)}$$

$$u_2 \frac{\delta\phi}{\delta u_2} = u_2(-2x_1 + x_2 - 2) = 0 \qquad\qquad \text{...(5.6)}$$

$$u_1 \geq 0 \quad u_2 \geq 0 \qquad\qquad \text{...(5.7)}$$

This is no general no single computational procedure for the solution of thus conditions. In order to show how to use the Kuhn-Tucker condition, it is necessary to explore various case defined principally by reference to whether each u_i is f_e zero.

For the first case, suppose that $u_1 = 0$ and $u_2 = 0$. From conditions (5.1) and (5.2) in get $x_1 = 2$ and $x_2 = 3$. This vector can not be a solution of our problem because it violates the first constraint $x_1 + x_2 \leqq 3$

Second, suppose that $u_1 \neq 0$ and $u_2 = 0$. Then equations (5.1) and (5.2) are reduced to

$2x_1 + u_1 = 4$

$2x_2 + u_2 = 6$

$- 2x_1 + x_2 = 2$

Yields the solution $x_1 = \dfrac{4}{5}$, $x_2 = \dfrac{18}{5}$, $u_2 = -\dfrac{6}{5}$, which violates conditions (5.3) and (5.7).

The last possibility is $u_1 \neq 0$ and $u_2 \neq 0$, because of the complementary solution condition (5.4) and (5.6) must be satisfied an equalities:

$x_1 + x_2 = 3$

$- 2x_1 + x_2 = 2$

The solution is $x_1 = \dfrac{1}{3}$ and $x_2 = \dfrac{8}{3}$. Substituting these value is (5.1) – (5.2), we obtain a negative

value for the Lagrange multiplier $u_2 = -\dfrac{8}{9}$ which is a contradiction to condition (5.7).

Only the values $x_1 = 1$, $x_2 = 2$, $u_1 = 2$, and $u_2 = 0$ satisfy all Kuhn-Tucker conditions and a simple inspection of the feasible solutions illustrates that this is induced the optional solution of our examples.

In order to illustrate that the Kuhn-Tucker conditions are not sufficient conditions for a local minimum (maximum) of mathematical programming problems, we consider the following very single one-variable example.

Maximize $f_o(x) = (x-1)^3$ …(5.8)

Subject to $x \leqq 2$ …(5.9)

$x \geqq 0$ …(5.10)

According to rule 1 for the Kuhn-Tucker conditions, we rewrite the constraint as $2 - x \geqq 0$. Then the Lagrange function is

$\phi(x, u) = (x-1)^3 + u(2-x)$

Application of the Kuhn-Tucker conditions for the maximization problem gives

$\dfrac{\delta\phi}{\delta x} = 3(x-1)^2 - u \leqq 0$ …(5.11)

$x\,\dfrac{\delta\phi}{\delta x} = x[3(x-1)^2 - u] = 0$ …(5.12)

$\dfrac{\delta\phi}{\delta u} = 2 - x \geqq 0$ …(5.13)

$u\,\dfrac{\delta\phi}{\delta u} = u(2-x) = 0$ …(5.14)

$u \geqq 0$ (because of rule 1)   …(5.15)

We may verify that $x^0 = 1$ and $u^0 = 0$ satisfy the Kuhn-Tucker conditions (5.11) – (5.15). By simple observation, it can be shown that the maximum of function (5.11) under the constraint (5.12) – (5.10) is at the point $x = 2$ and not at the point $x^0 = 1$.

5.3 SADDLE POINTS WITH SPECIAL REFERENCE TO CONCAVE PROGRAMMING

In mathematics, a saddle point or minimax point is a point on the surface of the graph of a function where the slopes (derivatives) in orthogonal directions are all zero (a critical point), but which is not a local extremum of the function. An example of a saddle point shown on the right is when there is a critical point with a relative minimum along one axial direction (between peaks) and at a relative maximum along the crossing axis. However, a saddle point need not be in this form. For example, the function $f(x, y) = x^2 + y^3$ has a critical point at $(0, 0)$, i.e., a saddle point since it is neither a relative maximum nor relative minimum, but it does not have a relative maximum or relative minimum in the y-direction.

The name derives from the fact that the prototypical example in two dimensions is a surface that curves up in one direction, and curves down in a different direction, resembling a riding saddle or a mountain pass between two peaks forming a landform saddle. In terms of contour lines, a saddle point in two dimensions gives rise to a contour graph or trace in which the contour corresponding to the saddle point's value appears to intersect itself.

Mathematical Discussion

A simple criterion for checking, if a given stationary point of a real-valued function f(x, y) of two real variables is a saddle point, is to compute the function's Hessian matrix at that point: if the Hessian is indefinite, then that point is a saddle point. For example, the Hessian matrix of the function $z = x^2 - y^2$ at the stationary point (x, y, z) is the matrix

$$\begin{vmatrix} 2 & 0 \\ 0 & -2 \end{vmatrix}$$

which is indefinite. Therefore, this point is a saddle point. This criterion gives only a sufficient condition. For example, the point (0, 0, 0) is a saddle point for the function $z = x^4 - y^4$ but the Hessian matrix of this function at the origin is the null matrix, which is not indefinite.

In the most general terms, a saddle point for a smooth function (whose graph is a curve, surface or hypersurface) is a stationary point such that the curve, surface, etc. in the neighbourhood of that point is not entirely on any side of the tangent space at that point.

In a domain of one dimension, a saddle point is a point which is both a stationary point and a point of inflection. Since it is a point of inflection, it is not a local extremum.

Concave Programming

Concave programming is another special case of the general constrained optimisation problem.

$$\max_{x \in X} f(x)$$

subject to $g(x) \leq 0$

in which the objective function f is concave and the constraint functions g_j are convex. For such problems, an alternative derivation of the Kuhn-Tucker condition is possible, providing yet another perspective on the Lagrangian method.

Duration

Consider the family of optimisation problems

$$\max_{x \in X} f(x)$$

subject to $g(x) \leq c$

Where the parameter c measures the available resources. The value function

$$v(c) \max_{x \in X} \{f(x) : g(x) \leq c\}$$

Summarises what can be achieved with different amounts of resource c. The set of attainable outcomes is given by its hypograph

$$A = \{(c, z) \in y \times R : z \leq v(c)\}$$

Since v is concave, its hypograph A is convex.

Let $z^* = v(0)$

and define $B = \{(c, z) \in y \times R : c \leq 0,\ z \geq z^*\}$

B is convex and its interior is non-empty and disjoint from A.

The Karlin Uzawa Saddle Point – Theorem

Suppose that we want to maximize some function of M variables $g(x_1, ..., x_N)$, useful to M inequality constraints of the following form:

1. $f_1(x_1, ..., x_N) \geq 0.,$ $f_2(x_1, ..., x_N) \geq 0.,$ ...., $f_m(x_1, ..., x_N) \geq 0$

In addition, we restrict the column vector $x = [x_1, ..., x_N]$ to belong to a closed comma set of feasible n's., x. Here, we will assume that the objectives function, $g(x)$ and all of the constraint functions, $f_m(x)$, m = 1, 2, ..., M are concave functions defined over the convex set x. The notation can be simplified if we define $f(x)$ as the column vector $[f_1(x), f_2(x)........ f_M(x)]$.

There are basic concave programming problem can be written is a succinct manner as follows:

2. Max x $\{g(x) : f(x) \geq O_M; x \in x\}$

A large number of economic problems can be written as the concave programming problem (2), including linear programming problems were the functions g and f are all linear and the set x is defined by simple liequalities on the components of the x vector.

We start off by defining the Lagrangian, $L(u, u)$ that corresponds to the nonlinear programming problem that is defined by (2) above:

3. $L(x, u) = g(x)\ fu^T\ f(x) = g(x) - \sum_{m=1}^{M} u^m\ f^m(x)$

Where, $u^T = [u_1, u_2, ..., u_M]$ is a vector of variables. (Lagrange multipliers or dual prices). A saddle point of the Lagrangian is a $x^0\ \varepsilon\ x$ and $u^0 \geq o_M$. That satisfies the following is equalities:

4. $L(x, u^0) \leq L(x^0, u^0) \leq L(x^0, u)$ for all $x \in X$ and $u \geq o_M$.

The first inequality is (u) say that $L(x, u^0) = g(x) + u^{0T} f(x)$, regarded as a function of x, has an unconstrained maximum at $x = x^0$ over all x that belongs to the set x. The second inequality is (4) says that $L(x^0, u) = g(x^0) + u^T f(x^0)$, regarded as a function of u, has an unconstrained minimum at $u = u^0$ over the set of all non-negative u. Hence, we see that the Lagrangian has the curvature of a saddle at the point x^0, u^0, curving downward in the x direction (s) and curving upward in the u direction (s).

Now, question arises – what is the significance of a saddle point? Following theorem tells us that if use happen to find a saddle point of Lagrangian (x^0, u^0) say, then the x^0 part of the saddle point solves the nonlinear programming problem (2) above suppose $(x^0, u°)$ is a saddle point of the corresponding Lagrangian; i.e., $x^0\ \varepsilon\ x$ and $u° \geq o_M$ satisfy the inequalities is (u). Then x^0 solve the nonlinear programming problem (2)

Proof: Using the definition of L (x, u), the second set of inequalities is (u) is equivalent to

5. $g(n^0) + u^{0T} f(x^0) \leq g(x^0) + u^T f(x^0)$ for all $u \geq 0_M$.

Now, subtract $g(x°)$ from each side of (5) and we get following inequalities:

6. $u^T f(x^0) \geq u^{0T} f(x°)$ for all $u \geq 0_M$.

If any component of $F(x^\circ)$ were negative say $f_m(x^\circ)$ were negative thus by taking U_m large enough, we would contradict the inequality (6) and so we deduce that

7. $f(x^0) \geq 0_M$

Since, $u^0 \geq 0_M$ and $f(x^\circ) \geq 0_M$, we deduce that $u^{0T} f(x^\circ) \geq 0$. But, now set $u = 0_M$ and from (6), we find that $u^{0T} F(x^0) \leq 0$, Hence,

8. $u^{0T} f(x^\circ) = 0$

Now, look at the first set of inequalities is (4), which are equivalent to;

9. $g(x) + u^{0T} f(x) \leq g(x^\circ) + u^{0T} + f(x^0)$ for all $x \in x$ using (8), (9), becomes:

10. $g(n) + u^{0T} f(x) \leq g(x^\circ)$ for all $x \in x$

Since, $n^0 \geq 0_M$ for any $x \in x$ such that $f(x) \geq 0_M$, we have

11. $g(x) \leq g(x) + u^{0T} f(x)$ for all $x \in x$ such that

   $\leq g(x^\circ)$ $f(x) \geq 0_M$ using (10),

But, the inequalities is (11) show on that x° solve (2) note that we did not require the concavity of g, f_{11} ....., f_n or the convexity of the domain of definition set, is order to prove the above result.

The above theorem tells us that if we can find a saddle point of the theorem Lagrangian, then we have a solution to the nonlinear programming problem (2). The problem with this result is that it does not tell us it such sadden points exist. If general, They will not exist. Hence, to get the existence of a saddle points, we need to impose some regularity conditions on the functions g, f_i, ..., f_M and the domain of definition set x. We will impose concavity on the functions g, f_1 ,........., f_N and convexity on the domain of definition set x. However, these conditions are not quite sufficient to imply the existence of at least are saddle point for the Lagrangian that corresponds to the nonlinear programming problem (2) we require at least an additional institutions on (2); a constraint qualification condition.

Before, we list possible constraint qualification conditions, we need to explain in general terms why one is required. while we defined the nonlinear programming problem (2), we did not specify that the number of constraint functions M be less than the number of variables N. Hence, if M is bigger than N, These is the possibility that the constraints will be inconsistent; i.e., there is no $x \in x$ that satisfies all the constraint $f(x) \geq O_N$. To solve this problem of inconsistency, we introduce our first constraint qualification condition.

12. Feasible constraint qualification condition: there exists $x^* \in x$ such that $f(x^*) \geq 0_M$

Obviously, (12) is a minimal constraint qualification condition that must be satisfied in order for solutions to the nonlinear programming problem (2) to exist, however, it thus out to be not quite strong enough for our purpose. Consider the following condition:

13. Slater constraint qualification condition: There exists $x^* \in x$ such that $f(x^*) \gg O_M$

The meaning of (13) is that there exists an $x \in x$ such that each of the inequality constraints is strictly satisfied when evaluated at x, i.e., we have $f_1(x) > 0$; $f_2(x) > 0$, ..., $f_M(x) > 0$.

This constraint qualification condition is due to slater (1950). In most economic contexts, it will not be restrictive to assume that the slater constraint qualification condition holds,

14. Karlin constraint qualification condition: for each non negative vector a which is not zero (so that $u > O_M$), There exists $x^* \in x$ such that $u^T f(x) > 0$

Note: That x* can depend on u. Obviously (13) implies (14), i.e., if the slater constraint qualification condition holds then so does the Karlin.

5.4 SUMMARY

1. In mathematics, non-linear programming is the process of solving an optimisation problem where some of the constraints or the objective function are non-linear.

2. If the objective function f is linear and the constrained space is a polytope, the problem is a linear programming problem, which may be solved using well-known linear programming techniques such as the Simplex method.

3. If the objective function is a ratio of a concave and a convex function (in the maximisation case) and the constraints are convex, then the problem can be transformed to a convex optimisation problem using fractional programming techniques.

4. In mathematical optimisation, the Karush-Kuhn-Tucker (KKT) conditions, also known as the Kuhn-Tucker conditions, are first derivative tests (sometimes called first-order) necessary conditions for a solution in non-linear programming to be optimal, provided that some regulatory conditions are satisfied.

5. In mathematics, a saddle point or minimax point is a point on the surface of the graph of a function where the slopes (derivatives) in orthogonal directions are all zero (a critical point), but which is not a local extremum of the function.

5.5 SELF ASSESSMENT QUESTIONS

1. Explain in detail 'Non-linear Programming'.
2. Critically analyse 'Kuhn-Tucker Optimality Criteria'.
3. Explain in detail 'Saddle Points with special reference to Concave Programming'.

❧ ❧ ❖ ❧ ❧

UNIT – IV

Chapter	
6	**THEORY OF GAMES**

Objectives

The objectives of this lesson are to:

- Theory of games
- Two-person Zero-sum Game
- Saddle point
- Pure and mixed strategies
- Matrix game and linear programming
- Rectangular game theory

Structure:

6.1   THEORY OF GAMES

Game theory is the study of mathematical models of strategic interaction between rational decision-makers. It has applications in all fields of social science as well as in logic and computer science. Originally, it addressed zero-sum game, in which one person's gains result in losses for the other participants. Today, game theory applies to a wide range of behavioural relations, and is now an umbrella term for the science of logical decision-making in humans, animals and computers.

Modern game theory began with the idea regarding the existence of mixed-strategy equilibria in two-person zero-sum games and its proof by John von Neumann. Von Neumann's original proof used the Brouwer fixed-point theorem on continuous mappings into compact convex sets, which became a standard method in game theory and mathematical economics. His paper was followed by the 1944 book 'Theory of Games and Economic Behaviour', co-written with Oskar Morgenstern, which considered cooperative games of several players. The second edition of this book provided an

axiomatic theory of expected utility, which allowed mathematical statisticians and economists to treat decision-making under uncertainty.

Game Theory was developed extensively in the 1950s by many scholars. It was later explicitly applied to Biology in the 1970s, although similar developments go back at least as far as the 1930s. Game theory has been widely recognised as an important tool in many fields. As of 2014, with the Nobel Memorial Prize in Economic Sciences going to game theorist Jean Tirole, eleven game theorists have won the Nobel Prize for Economics. John Maynard Smith was awarded the Craford Prize for his application of Game Theory to Biology.

Game Types

Cooperative/Non-cooperative

A game is cooperative if the players are able to form binding commitments externally enforced (e.g., through contract law). A game is non-cooperative if players cannot form alliances or if all agreements need to be self-enforcing (e.g., through credible threats).

Cooperative games are often analysed through the framework of cooperative game theory, which focuses on predicting which coalitions will form, the joint actions that groups take and the resulting collective payoffs. It is opposed to the traditional non-cooperative game theory which focuses on predicting individual players' actions and payoffs and analysing Nash equilibria.

Cooperative game theory provides a high-level approach as it only describes the structure, strategies and payoffs of coalitions, whereas non-cooperative game theory also looks at how bargaining procedures will affect the distribution of payoffs within each coalition. As non-cooperative game theory is more general, cooperative games can be analyzed through the approach of non-cooperative game theory (the converse does not hold) provided that sufficient assumptions are made to encompass all the possible strategies available to players due to the possibility of external enforcement of cooperation. While it would thus be optimal to have all games expressed under a non-cooperative framework, in many instances, insufficient information is available to accurately model the formal procedures available to the players during the strategic bargaining process, or the resulting model would be of too high complexity to offer a practical tool in the real world. In such cases, cooperative game theory provides a simplified approach that allows analysis of the game at large without having to make any assumption about bargaining powers.

Symmetric/Asymmetric

A symmetric game is a game where the payoffs for playing a particular strategy depend only on the other strategies employed, not on who is playing them. If the identities of the players can be changed without changing the payoff to the strategies, then a game is symmetric. Many of the commonly studied 2×2 games are symmetric. The standard representations of chicken, the prisoner's dilemma, and the stag hunt are all symmetric games. Some scholars would consider certain asymmetric games as examples of these games as well. However, the most common payoffs for each of these games are symmetric.

Most commonly studied asymmetric games are games where there are not identical strategy sets for both players. For instance, the ultimatum game and similarly the dictator game have different strategies for each player. It is possible, however, for a game to have identical strategies for both players, yet be asymmetric. For example, the game pictured to the right is asymmetric despite having identical strategy sets for both players.

Zero-sum/Constant-sum

Zero-sum games are a special case of constant-sum games, in which choices by players can neither increase nor decrease the available resources. In zero-sum games, the total benefit to all players in the game, for every combination of strategies, always adds to zero (more informally, a player benefits only at the equal expense of others). Poker exemplifies a zero-sum game (ignoring the possibility of the house's cut), because one wins exactly the amount one's opponents lose. Other zero-sum games include matching pennies and most classical board games including Go and Chess.

Many games studied by game theorists (including the famed prisoner's dilemma) are non-zero-sum games, because the outcome has net results greater or less than zero. Informally, in non-zero-sum games, a gain by one player does not necessarily correspond with a loss by another.

Constant-sum games correspond to activities like theft and gambling, but not to the fundamental economic situation in which there are potential gains from trade. It is possible to transform any game into a (possibly asymmetric) zero-sum game by adding a dummy player (often called "the board") whose losses compensate the players' net winnings.

Simultaneous/Sequential

Simultaneous games are games where both players move simultaneously, or if they do not move simultaneously, the later players are unaware of the earlier players' actions (making them effectively simultaneous). Sequential games (or dynamic games) are games where later players have some knowledge about earlier actions. This need not be perfect information about every action of earlier players; it might be very little knowledge. For instance, a player may know that an earlier player did not perform one particular action, while s/he does not know which of the other available actions the first player actually performed.

The difference between simultaneous and sequential games is captured in the different representations discussed above. Often, normal form is used to represent simultaneous games, while extensive form is used to represent sequential ones. The transformation of extensive to normal form is one way, meaning that multiple extensive form games correspond to the same normal form. Consequently, notions of equilibrium for simultaneous games are insufficient for reasoning about sequential games (See '*sub-game perfection*').

Perfect Information and Imperfect Information

An important subset of sequential games consists of games of perfect information. A game is one of perfect information if all players know the moves previously made by all other players. Most games studied in game theory are imperfect-information games. Examples of perfect-information games include Tic-tac-toe, Checkers, Infinite Chess and Go.

Many card games are games of imperfect information, such as Poker and Bridge. Perfect information is often confused with complete information, which is a similar concept. Complete information requires that every player know the strategies and payoffs available to the other players but not necessarily the actions taken. Games of incomplete information can be reduced, however, to games of imperfect information by introducing "moves by nature".

Combinatorial Games

Games in which the difficulty of finding an optimal strategy stems from the multiplicity of possible moves are called combinatorial games. Examples include Chess and Go. Games that involve imperfect information may also have a strong combinatorial character, e.g., Backgammon. There is

no unified theory addressing combinatorial elements in games. There are, however, mathematical tools that can solve particular problems and answer general questions.

Games of perfect information have been studied in combinational game theory, which has developed novel representations, e.g., surreal number as well as combinatorial and algebraic (and sometimes non-constructive) proof methods to solve games of certain types, including "loopy" games that may result in infinitely long sequences of moves. These methods address games with higher combinatorial complexity than those usually considered in traditional (or "economic") game theory. A typical game that has been solved this way is hex. A related field of study, drawing from computational complexity theory is game complexity, which is concerned with estimating the computational difficulty of finding optimal strategies.

Research in artificial intelligence has addressed both perfect and imperfect information games that have very complex combinatorial structures (like Chess, Go or Backgammon) for which no provable optimal strategies have been found. The practical solutions involve computational heuristics, like alpha-beta pruning or use of artificial neural networks trained by reinforcement learning, which make games more tractable in computing practice.

Infinitely Long Games

Games, as studied by economists and real-world game players, are generally finished in finitely many moves. Pure mathematicians are not so constrained, and set theorists in particular study games that last for infinitely many moves, with the winner (or other payoff) not known until after all those moves are completed.

The focus of attention is usually not so much on the best way to play such a game, but whether one player has a winning strategy. (It can be proven, using the axiom of choice, that there are games – even with perfect information and where the only outcomes are "win" or "lose" – for which neither player has a winning strategy.) The existence of such strategies, for cleverly designed games, has important consequences in descriptive set theory.

Discrete and Continuous Games

Much of game theory is concerned with finite, discrete games, that have a finite number of players, moves, events, outcomes, etc. However, Many concepts can be extended. Continuous games allow players to choose a strategy from a continuous strategy set. For instance, Cournot competition is typically modeled with players' strategies being any non-negative quantities, including fractional quantities.

Differential Games

Differential games such as the continuous pursuit and evasion game are continuous games where the evolution of the players' state variables is governed by differential equations. The problem of finding an optimal strategy in a differential game is closely related to the optimal control theory. In particular, there are two types of strategies: the open-loop strategies are found using the Pontryagin maximum principle while the closed-loop strategies are found using the Bellman's Dynamic Programming method.

A particular case of differential games are the games with a random time horizon. In such games, the terminal time is a random variable with a given probability distribution function. Therefore, the players maximise the mathematical expectation of the cost function. It was shown that the modified

optimisation problem can be reformulated as a discounted differential game over an infinite time interval.

Evolutionary Game Theory

Evolutionary game theory studies players who adjust their strategies over time according to rules that are not necessarily rational or far-sighted. In general, the evolution of strategies over time according to such rules is modeled as a Markov chain with a state variable such as the current strategy profile or how the game has been played in the recent past. Such rules may feature imitation, optimisation or survival of the fittest.

In biology, such models can represent (biological) evolution, in which offspring adopt their parents' strategies and parents who play more successful strategies (i.e., corresponding to higher payoffs) have a greater number of offspring. In the social sciences, such models typically represent strategic adjustment by players who play a game many times within their lifetime, and consciously or unconsciously, occasionally adjust their strategies.

We have two player. Alice (abbreviated as A and referred the plancus 'she') and Bob (B, 'he') each of which has the choice between two actions. For the choice of a_1' of A and b_j of B. The (i, j) entry in the table lists the pay-off of A factor and the one of B after the comma.

Table 6.1

	b_1	b_2
a_1	3, 2	1, 3
a_2	2, 1	0, 0

So, how should the two players reason and act to maximize their pay-off, assuming that both know the structure of the game and the pay-off matrix we assume at their point that the two players play simultaneously.

The first observation is that for player A, given an action of B, the first row is always better than the second. One says that action a_i dominates a_2 and play a_i in any case. When B realize this, he should play b_2. This, A will get the pay-off. Whereas B gets 3. This, A will get the pay-off, whereas B gets 3. This represents a so called Nash equilibrium indicating that neither player can unilaterally change from a_1 to a_2, but B keeps b_2, her/his pay-off. If A changes from a_1 to a_2, but B keeps b_2, her pay-off would be reduced from 1to 0. If B switched from b_2 to b_1, which A continues to play a_1, his pay-off would be reduced from 3 to 2.

Obviously, this equilibrium leaves B better off than A. If the game wear played sequentially instead of simultaneously, write A playing first, she should choose a_2 in place of Q_1, even though that action is dominated in the simultaneous game, as this would force B to play b_1, giving A the pay-off 2 which is higher than aQ_1 achieved, in the Nash equilibrium.
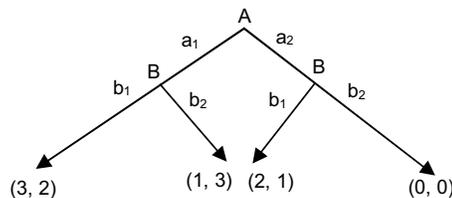
Fig. 6.1

This observation several a problem with the dominance arrangements that A should disregard a_2 in the simultaneous game. After all, if she can make it plausible to B that she will play a_2, this would force him to play b_1. The dominance argument compares the outcome of each action against the same action of the opponent. The opponent, however will react differently to the different actions of A_0 the will play b_2 against a_1, but b_1 against a_2, and the latter is better for A.

If, however, B can move first, he should play b_2, forcing A to play a_1 which is the above trash equilibrium which is optimal for B.
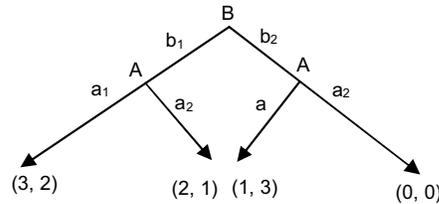
Fig. 6.2

We how, consider the situation when A can move first, but B cannot observe A's wave, as indicated by the dashed line in the following figure:
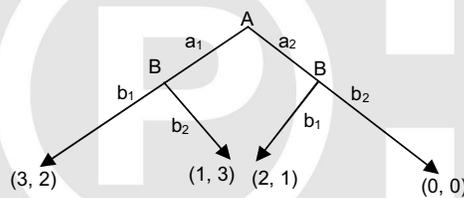
Fig. 6.3

B might reason that A had player a_2, in order to force him to play b_1, and he should correspondingly do so. Now, however, A could think that because of this reasoning, B will play b_1 anyway because of this reasoning, B will play b_1 anyway and he could therefore decide to play a_2 to maximise her pay-off. In term, B anticipation that reasoning, he could play b_2. However, when A believes, B to play b_2, she should play a_1, and lower the consistent beliefs repeat themselves. When, B believes A to play a_1, he should play b_2 and whenever A experts that B will play b_2. She should play a_1. Since, only chain of consistent higher order belief, will eventually arrive at the point, it seems that (a, b_2) will be the only equilibrium, we need a formal definition of an equilibrium her, however, in order to substantiate that claim. Incidentally, here, it helps B to be ignorant about A's more, as otherwise, A could have forced him to play b_1, which leads to a worse pay-off for him. Thus, in such games, it can be disadvantageous to acquire more information about the opponent. On should be careful with the interpretation of this finding, however, what harms B is not that he has that information, but the fact than A knows that he has that information, but in fact that A knows that he has that information. It indicates, A also possesses some additional information. If A did not know that B knows her move, Then B would not be at disadvantage. Conversely A should try to transmit the information we when again in a situation to which the preceding analysis applies.

We can gain further insight by who difying the pay-off matrix. For instance, when are consider.

Table 6.2

	b₁	b₂
a₁	1, 2	1, 3
a₂	2, 1	0, 0

That is, we lower the pay-off for A when playing a₁ against b₁, then it is perfectly reasonable for A to play a₂ to which B should react with b₁. Thus, when we remove the temptation for A to play a₁ instead of a₂ against b₁, we improve her position.

Examples:

1. We consider the following complimentarity game. Each player can contribute 0, 1, 2 or 3 (units). When the sum of the two contributes is at last 3, each player receives 3 minus her own contribution, else both receives 0. The pay-off matrix then is

Table 6.3

0, 0	0,0	0,0	3, 0
0, 0	0, 0	2, 1	2, 0
0, 0	1, 2	1, 1	1, 0
0, 3	0, 2	0, 1	0, 0

We can already make a simple, but important, observation her: Then is no point for player 1 to play the last strategy, that is contribute 3, because her pay-off for any other strategy is always at least as high as there for 3, and is several case higher. The same applies for player 2, of course, as this game is symmetric. Therefore, player i can assume that player i will never play 3. Therefore, we only used to consider the reduced pay-off matrix.

Table 6.4

0, 0	0, 0	0, 0
0, 0	0, 0	2, 1
0, 0	1, 2	1, 1

Applying the same reasoning to this new pay-off matrix, it there will pay for neither player to play the first strategy, that is contribute 0. Thus, we can reduce the pay-off matrix once more to arrive at

Table 6.5

0, 0	2, 1
1, 2	1, 1

Thus, each player would contribute 1 or 2. For each player, the best situation is if she herself contributes 1 while the other one contributes 2. When she knows however, that the other one will contribute only then she has no choice but to contribute if she is to maximise her pay-off. That is, this represents a situation where neither player can change her strategy unilatural without decreasing her pay-off .

2. We consider a game, the so called matching firm is game whose each player has only two strategies, and player they play different strategies, for player 2, the situation is the opposite, that is zero-sum game. Meaning that for every strategy profile, the pay-offs of the two player sum to 0. Thus, the pay-off matrix is

Table 6.6

1, – 1	– 1, 1
– 1, 1	1, – 1

In this game, whatever the actual strategy profile it is always advantageous for precisely are of the two players to change her strategy. In particular, if the game were to be played repeated, it would be disadvantageous for either player to always play the same strategy because the other one could then choose a winning strategy more generally, it would be disadvantageous for either player to play is a manner that is predictable for her opponent. Thus, in this game, the best option for either player would be to play the two strategies randomly with probability 1/2 each. At least, if both of them played that way, then for neither of them it would carry an advantage to unitaturally change her strategy because the opponent would respond correspondingly.

3. This game is sometimes called the bottle of the sixth game a girl (player 1) and her friend (player 2) would like to play this time together, but each of them prefer, a different activity. The girl, likes to attend their time together but each of the no prefers a different activity. The girl likes to attend a heavy wright boxing fight (action 1 whereas the boy prefer to go to the fashion show (action 2). The pay-off matrix is

Table 6.7

4, 2	1, 1
50	2, 4

For this game, either way of attending the same activity is an equilibrium where neither of them would benefit from charging her/his action unilaturally. As represented, this game is non-symmetric, but there is an equivalent symmetric form. Either player could insist on his/her preferred activity or yield to the performance of the partner. When insisting is action 1, yielding action 2, the pay-off matrix became,

Table 6.8

1, 1	4, 2
2, 4	0, 0

Which is symmetric. In fact, the situation is row the same which is symmetric. In fact the situation is similar as in Example 1.

4. The next game has the structure of the so called prisoner's dilemma

Table 6.9

4, 4	2, 6
6, 2	3, 3

Even though it is better for both of them if they play their first strategy (cooperates) than when they both play the second one (defect) each player has an incentive to switch to defecting when the other player cooperates. The cooperating player would then be put up at disadvantage and should also

switch to defecting to avoid that. Therefore, the only equilibrium is where they both defect and get only 3 each. This outcome looks some what paradoxial because it sums perfectly possible that they agreed to cooperate and received the pay-off of 4.

After having analyzed these examples in detail, let us once more emphasize the key point. Player i commonly choose her own strategy, but her pay-off also depends on the strategy of her opponent s.i. She should therefore select her won strategy si leaves i writs the highest pay-off among all her possible strategy choices. In other words, i wants to be the best off under under. The assumption that her opponent choose, her - the opponent's – best response. And the opponent applies the seam reasoning.

Keeping these examples and observations in mind, we now develop some general corrupts. In particular, the examples may have created the impression that even though a 2×2 game is completely described by 8 numbers, there is a bewildering multitude of possible phenomena. The question therefore arise whether any kind of classification is possible.

First of all, an observed that, so far, only pay-off differences were relevant writs their absolute values did not matter. Also, in our analysis of Example 3 we have seen that relabeling the strategies simply interchanges some rows or columns, but does not change the game. In that way, many (but not all) game can be transforms into a symmetric firm.

Classification of symmetric 2×2 games: we consider a symmetric 2×2 game with pay-off matrix.

$$A = \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix}$$

We observe that we, obtain an equivalent game, is the sense of ranking of strategic when we substract fixed numbers from each column. More precisely, it does not make difference for the choice of strategies of the player. When we pay-off to a particular strategy of her opponent is changed by a fixed amount that does not depend her own strategy. Thus, will may substract π_2, from first and π_{12} from the second column, to obtain the diagonal pay-off matrix.

$$A = \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix}$$

With $a_1 = \pi_{11} - \pi_{21}$; $a_2 = \pi_{22} - \pi_{12}$, Thus, any symmetric 2×2 game is represented by the pair $(a_1, a_2) \in R^2$. The classification is then in terms of the signs a_1 and a_2. Category consists of these games with $a_1 < 0$, $a_2 < 0$.

Here, strategy 2 strictly dominates strategy 1. Thus, both player will play 2. This category includes the prisoner's dilemma a the reader will easily check. Of course, after this rearrange must, the paradox discussed above is gone. Category IV where $a_1 > 0$, $a_2 < 0$ is of course equivalent to this one by rebelling the two strategy. Category it means $a_1 > 0$, $a_2 > 0$. Here, for neither player it is advantageous to unilaterally deviate from strategy α if her apparent plays that strategy. Finally, category III comprises the games with $a_1 < 0$, $a_2 < 0$. Here playing different strategies is the best opinion from while no unilatural deviation pays off.

6.2 TWO-PERSON ZERO-SUM GAME

Game Theory provides a mathematical framework for analysing the decision-making processes and strategies of adversaries (or players) in different types of competitive situations. The simplest

type of competitive situations are two-person zero-sum games. These games involve only two players. They are called zero-sum games because one player wins whatever the other player loses.

Zero sum games with only two players or competitors are called two-person zero. Sum or rectangular games. In this case the loss (gain, of our player is exactly equal to the gain (loss) of the opponent and each player knows the out can for all possible strategies that he and his opponent may use during a play of the game. The resulting out came, representing gain (or loss) to a particular player can be displayed in the form of payoff matrix $A = (a_{ij})$, where a_{ij} is the payoff to player I (say), when he employs his ith move which player II (the opponent) employs his jth move. Thus, if player I has strategies (move) available to him and player II her in moves available to him, then the payoffs for various strategy combinations may be represented by an $m \times n$ payoff matrix (a_{ij}). For this reason, the two-person zero-sum games are also called matrix games.

Example: Odds and Evens

Consider the simple game called odds and evens. Suppose that Player 1 takes evens and Player 2 takes odds. Then, each player simultaneously shows either one finger or two fingers. If the number of fingers matches, then the result is even, and Player 1 wins the bet ($2). If the number of fingers does not match, then the result is odd, and Player 2 wins the bet ($2). Each player has two possible strategies: show one finger or show two fingers. The payoff matrix shown below represents the payoff to Player 1.

Payoff Matrix

Table 6.10

Strategy		Player	
		1	**2**
Player	1	2	–2
	2	–2	2

Basic Concepts of Two-person Zero-sum Games

This game of odds and evens illustrates important concepts of simple games:

- A two-person game is characterised by the strategies of each player and the payoff matrix.

- The payoff matrix shows the gain (positive or negative) for Player 1 that would result from each combination of strategies for the two players. Note that the matrix for Player 2 is the negative of the matrix for Player 1 in a zero-sum game.

- The entries in the payoff matrix can be in any units as long as they represent the utility (or value) to the player.

- There are two key assumptions about the behaviour of the players. The first is that both players are rational. The second is that both players are greedy meaning that they choose their strategies in their own interest (to promote their own wealth).

Example: Suppose a two-person zero-sum game has two players, R and c with their respective available streets gives R_1, R_2 an C_1 C_2, C_3 respectively. Let the pay offs to the player R be expressed in terms of gains to him. Let the pay offs to player R be given by the following 2×3 pay off matrix.

$$A = \begin{array}{c} R_1 \\ R_2 \end{array} \begin{pmatrix} C_1 & C_2 & C_3 \\ -1 & 3 & 0 \\ 2 & -4 & 1 \end{pmatrix}$$

Then the following explanation may be given for the various payoffs:

Table 6.11

Strategy of R	Strategy of C		
	C_1	C_2	C_3
R_1	R_1 loser 1 unit	R gain 3 units	More gains
R_2	R gains 2 units	R loses 4 units	R gains 1 unit.

Since, the game is zero-sum, every gain of player R is an equal loss of player C and *vice versa.* Thus, the payoff matrix for the player C will be just the negative of the pay off matrix of R, so that the sum of the payoff matrices of the two players is ultimately a null matrix.

For a given payoff matrix, an shall adopt the following:

(a) Row designation are the courses of action available to player R who will be called the new player.

(b) column designations are the course of action available to player C who will be called the column player.

(c) The various pay offs are the payoffs to the row player.

6.3 SADDLE POINT

Saddle Point and Value of the Game

Definition (Equilibrium as Saddle Point)

A saddle point or equilibrium point of a payoff matrix is that position in the payoff matrix where the maximum of low minimise coincides with the minimum of the column maxima. The payoff at the saddle point is called the value of the game and is obviously equals to the maximin are minimax values of the game.

Thus, (k, r) the position of the payoff matrix (a_{ij}) will be a saddle point if and only if

$$a_{kr} = \begin{array}{c} \max \\ i \end{array} \left[\begin{array}{c} \min \\ j \end{array} , \{a_{ij}\} \right] = \begin{array}{c} \min \\ j \end{array} \left[\begin{array}{c} \max \\ i \end{array} \{a_{ij}\} \right]$$

The saddle point and hence the value of the game, need not be unique we shall find out Q value of the game by 2. The importance of the saddle point arises from the fact that, in general, the optimum play consist its sticking to the strategies which correspond to the saddle point. To solve a game all therefore merely need to look for the saddle point of the payoff matrix. If it exists, the game is solved. But, unfortunately, most payoff matrices do not posses any saddle point. In general following theorem (6.1), the value of the game is satisfies.

Maximin value $\leq \upsilon \leq$ minimax value he shall do not the maximin value of the game by υ and the minimax value of the game by $\overline{v}$. These valrus all also called the lower value and upper value of the game, respectively. A game is said to be a fair game if $\underline{v} = 0 = \overline{v}$. A game is said to be strictly determinable if $\underline{v} = v = \overline{v}$.

Theorem (6.1)

Let (a_{ij}) be $m \times n$ payoff matrix for a two person, zero-sum game. If $\underline{v}$ denotes the maximin value and $\bar{v}$ the minimax value of the game, $\bar{v} \geq \underline{v}$, i.e.,

$$\min_{\leq j \leq n} \left[\max_{1 \leq i \leq m} \cdot \{a, i\} \right] \geq \max_{1 \leq i \leq m} \left[\min. \{a_{ij}\} \right]$$

Proof: We have

$$\max_{1 \leq i \leq m} [a_{ij}] \geq a_{ij} \text{ for all } j = 1, 2, ..., i \text{ and } \min_{1 \leq j \leq n} [a_{ij}] \leq a_{ij} \text{ for all } i = 1, 2, ..., m$$

Let, the above maximum be attained at $i = i'$ and the minimum be attained at $j = j'$, i.e.,

$$\max_{1 \leq i \leq m} [a_{ij}] = a_i'j \text{ and } \min_{1 \leq j \leq n} \{a_{ij}\} = a_i'_j$$

Then, we must have

$$a_i'_j \geq a_{ij} \geq a i'_j \text{ for all } i = 1, 2, ..., m; j = 1, 2, ..., n \text{ from this, we get}$$

$$\min_{1 \leq j \leq n} \left\{ a'_{i\,j} \right\} \geq a_{ij} \geq \max_{1 \leq i \leq m} \{a_{ij'}\} \text{ for all } i = 1, 2, ..., m; j = 1, 2, .., n$$

$$\therefore \min_{1 \leq j \leq n} \left[\max_{1 \leq i \leq m} \{a_{ij}\} \right] \geq \max_{i \leq i \leq m} \left[\min_{1 \leq j \leq n} \{ai'_j\} \right] q \bar{v} \geq \underline{v}$$

Please refer Chapter 5, Point No. 5.3, Page No. 175.

6.4 PURE AND MIXED STRATEGIES

In game theory, a player's strategy is any of the options which he or she can choose in a setting where the outcome depends not only on their own actions but on the actions of others. A player's strategy will determine the action which the player will take at any stage of the game.

The strategy concept is sometimes (wrongly) confused with that of a move. A move is an action taken by a player at some point during the play of a game (e.g., in chess, moving white's Bishop a_2 to b_3). A strategy, on the other hand, is a complete algorithm for playing the game, telling a player what to do for every possible situation throughout the game.

A strategy profile (sometimes called a strategy combination) is a set of strategies for all players which fully specifies all actions in a game. A strategy profile must include one and only one strategy for every player.

Strategy Set

A player's strategy set defines what strategies are available for them to play.

A player has a finite strategy set if they have a number of discrete strategies available to them. For instance, a game of rock-paper-scissors comprises a single move by each player and each player's move is made without knowledge of the other's, not as a response. So, each player has the finite strategy set {rock, paper, scissors}.

A strategy set is infinite otherwise. For instance, the cake cutting game has a bounded continuum of strategies in the strategy set {Cut anywhere between 0% and 100% of the cake}.

In a dynamic game, the strategy set consists of the possible rules a player could give to a robot or agent on how to play the game. For instance, in the Ultimatum game, the strategy set for the second player would consist of every possible rule for which offers to accept and which to reject.

In a Bayesian game, the strategy set is similar to that in a dynamic game. It consists of rules for what action to take for any possible private information.

Choosing a Strategy Set

In applied game theory, the definition of the strategy sets is an important part of the art of making a game simultaneously solvable and meaningful. The game theorist can use knowledge of the overall problem to limit the strategy spaces and ease the solution.

For instance, strictly speaking in the Ultimatum game, a player can have strategies such as: of ($1, $3, $5, ..., $19), accept offers of ($0, $2, $4, ..., $20). Including all such strategies makes for a very large strategy space and a somewhat difficult problem. A game theorist might instead believe they can limit the strategy set to: {Reject any offer ≤ x, accept any offer > x; for x in ($0, $1, $2, ..., $20)}.

Pure and Mixed Strategies

A pure strategy provides a complete definition of how a player will play a game. In particular, it determines the move a player will make for any situation he or she could face. A player's strategy set is the set of pure strategies available to that player.

Pure strategy

Interestingly pure strategy can be explained as a special case of mixed strategy in which a particular move is assigned probably one and all others probably zero. For example, $e_1 = [1, 0, 0, ...] \in S_m$ and $e_i = [0, 0, ...1, ...0] \in S_m$ are pure strategies of the row player.

So, the mixed strategy $e_i \in S_m$ whose all components except the ith are zero, is called ith pure strategy for the row player. Similarly, $ej \in S_n J = 1, 2, ..., n$ are called the pure strategies for the column player. It will be observed that the probability that player A choose his ith. Choice is same as the probability of his choosing the pure strategy $eJ \in S_m$. Thus, for a game it is sufficient to specify the payoff matrix and the pure strategies for both the players.

Suppose, a_{ij} is the payoff to raw player when he choose the pure strategy $ei \in S_m$ and the column player choose the pure strategy $ej \in S_n$. Then the expected payoff to the pure strategy $ej \in S_n$. Then the expected pay off to row player, given that the column player uses his pure strategy $ej \in S_n$, is

$$E(P, ej) = \sum_{i=1}^{r} a_{ij}Pi'$$

The expected payoff to the row player when the column player uses mixed strategy $q = (q_1, q_2, ..., q_n) \in S_n$, is

$$E(P, q) = \sum_{J=1}^{n} q_j \in (P, l_j) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}, P_i q_j = pTAq$$

We shall call $E(P, q)$ as the expectation function of the metangular game, where $P \in S_m$, $q \in S_n$. Observe that if P is kept fixed at some value and q is varied, then $E(p, q)$ will be a minimum for some value of q. Let this minimum value be

$$\phi = \min_{q \in S_n} E(p,q)$$

When, p is given some other fixed value, a different value of ϕ. Obtained. Thus, by assigning different values to p, a set of values of ϕ is obtained, assuming that ϕ arises is every ease. This implies that ϕ is a function of p and therefore, we can write

$$\phi(p) = \min_{q \in S_n}, E(p,q)$$

Now, if we assume that $\phi(p)$ is a maximum for some value of p, than we can write

$$\max_{p \in S_m} \phi(p) = \max_{p \in Sm} \min_{q \in Sm} E(p,q)$$ by first finding the maximum value of E(p, q, with respect to p keeping q fixed and then finding the minimum of the function so obtained with respect to q)

Theorem (6.2)

Let, E(p, q) be such that both

$$\max_{q \in S_m} \min_{q \in S_n}, E(p, q) \text{ and } \min_{q \in S_n} \max_{p \in S_n}, E(p, q) \text{ exist, then}$$

$$\min_{q \in S_n} \max_{p \in S_m}, E(p, q) \geq \max_{p \in S_m}. \min_{q \in S_n} \in (p, q)$$

Proof: Lets $p°$ and $q°$ be some availability chosen points is S_m and S_n respectively. Thus, we must have $\max\limits_{p \in S_m} \in (p, q°) \geq \in (p°, q°)$ and $\min\limits_{q \in S_n} E(p°, q)$ But, $q°$ is arbitarily chosen and could have been any point n S_n, and for every one of them the inequality holds. Even, if we had chosen $q°$ to be that point for which $\max\limits_{p \in S_m} E(p, q)$ has the value, the inequality remains true.

$$\therefore \min_{q \in S_n} \max_{p \in S_m} E(p,q) \geq \min_{q \in S_n} E(p,q)$$

Again, since $p°$ is any point is S_m, the inequality holds even if we choose that $p°$ which gives the maximum value of $\min\limits_{q \in S_n} E(p, q)$. Hence $\min\limits_{q \in S_n} \max\limits_{p \in S_m} E(p, q) \geq \max\limits_{p \in S_m} \min\limits_{q \in S_n} E(p, q)$

Mixed Strategy

A mixed strategy is an assignment of a probability to each pure strategy. This allows for a player to randomly select a pure strategy. (See the following section for an illustration.) Since probabilities are continuous, there are infinitely many mixed strategies available to a player.

Let $A = (a_{ij'})$ be the payoff matrix of an m × n matrix game between two player, call them now player and column player. The row-player has available to him a set of m strategies that be can employ with same specified probability.

Let $P = [P_1, P_2, ..., P_m]$ determine there probabilities, where $P_i \geq 0$ for i = 1, 2, ..., m, $\sum P_i = 1, P_i$ will being the probability that the row player employ his i the strategy.

Let $S_m = \{P = P_i \geq 0 \forall 1 \text{ and } \sum P_i = 1\}$

Then, $P \in Sm$ is called a mixed strategy of row player S_m is called the set of mixed strategies of the row player.

Likewise, for the column player, we define an n-tuple $q = [q_1, q_2..., q_n] \in S_n$, ($q_i \geq 0$ for $j = 1, 2, ...,$ n, $\sum q_j = 1$) as the mixed strategy for the column player S_m is called the set of mixed strategies for the column player.

Of course, one can regard a pure strategy as a degenerate case of a mixed strategy, in which that particular pure strategy is selected with probability and every other strategy with probability 0.

A is a mixed strategy in which the player assigns a strictly positive probability to every pure strategy. (Totally mixed strategies are important for equilibrium refinement such as trembling hand perfect equilibrium.)

Illustration

Consider the payoff matrix pictured to the right (known as a coordination game). Here, one player chooses the row and the other chooses a column. The row player receives the first payoff, the column player the second. If row opts to play A with probability 1 (i.e., play A for sure), then he is said to be playing a pure strategy. If column opts to flip a coin and play A if the coin lands heads and B if the coin lands tails, then he is said to be playing a mixed strategy, and not a pure strategy.

Table 6.12

	A	B
A	1,1	0,0
B	0,0	1,1
Pure	Coordination Game	

Significance

In his famous paper, John Forbes Nash proved that there is an equilibrium for every finite game. One can divide Nash equilibria into two types. Pure strategy Nash equilibria are Nash equilibria where all players are playing pure strategies. Mixed strategy Nash equilibria are equilibria where at least one player is playing a mixed strategy. While Nash proved that every finite game has a Nash equilibrium, not all have pure strategy Nash equilibria, e.g., a game that does not have a Nash equilibrium in pure strategies. However, many games do have pure strategy Nash equilibria (e.g.. The Coordination game, the Prisoner's dilemma and the Stag hunt). Further, games can have both pure strategy and mixed strategy equilibria. An easy example is the pure coordination game, where in addition to the pure strategies (A, A) and (B, B), a mixed equilibrium exists in which both players play either strategy with probability 1/2.

6.5 EQUIVALENCE OF MATRIX GAME AND LINEAR PROGRAMMING

Matrix Game

A matrix game, which is short for finite two-person zero-sum game, allows a game to be represented in matrix form as its name implies. This is a direct consequence of the fact that two opponents with exactly opposite interests play a game under a finite number of strategies, independently of his or her opponent's action. Once both players each make an action, their decisions are disclosed. A payment is made from one player to the other based on the outcome, such that the gain of one player equals the loss of the other, resulting in a net payoff summing to zero.

Objective

The objective of game theory is to analyse the relationship between decision-making situations in order to achieve a desirable outcome. The theory can be applied to a wide range of applications, including, but not limited to, economics, politics and even the biological sciences. In essence, game theory serves as s means to create a model to represent certain scenarios that have a variety of variables and potential outcomes. With these models developed from game theory, one can determine if assumptions made for a certain scenario are valid or whether additional models should be created that could more accurately assess the current problem. Game theory can be broken into a variety of different "games," each analysing different situations in which a decision is to be made by one player with other players potentially affecting the process.

Game Types

As aforementioned, there are many types of "games" that have been developed due to game theory. These games model various scenarios and differ from each other depending on how the players in the game cooperate with each other:

1. **Strategic Games:** A strategic game that models a set of players, each with a set of actions with preferences for performing each of their corresponding actions.

2. **Extensive Games:** A game that consists of players that choose a terminal history that results in a player function depending on the chosen terminal history. Much like the strategic game, players have preferences over their set of terminal histories.

3. **Coalitional Game:** A game that consists of a set of players, with each player being part of a group. These groups of players are called coalitions and have a set of actions that can be taken based on player preferences.

Payoff Matrix

From the definition of the finite two-person zero-sum game, all payments (P_{ij}) can be represented as a matrix $P = [P_{ij}]$ where i is an action in the finite set of chains that one player makes and j is that of the other for all $i \in \{1, 2, ..., m\}$ and of $j \in \{1, 2, ..., n\}$. Here, matrix P is called the payoff matrix. Note that the payments are made in one direction, i.e., P_{ij} represents payments made from the first player to the second so that the elements of P can be positive, 0, or negative (i.e., $P \in R^{m \times n}$).

Example 1: Consider the game rock-paper-scissors as a simple example. Let the choice of rock, paper and scissors be denoted as 1, 2 and 3, respectively. This means that the first row of the payoff matrix indicates the first person playing a rock, while the columns represent the second player's choices. Further, let a "win," "draw" and "loss" be respectively assigned the values +1, 0 and –1. Based on the definition of the payoff matrix, a game of rock-paper-scissors then has the payoff matrix.

$$P = \begin{bmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix}$$

It is well known that a clear winning strategy does not exist for rock-paper-scissors, as can be seen in the payoff matrix. In other words, each player equally has a chance of winning.

Example 2: Now, consider a non-trivial game of rock-paper-scissors where the payoff matrix is

$$P = \begin{bmatrix} 0 & 2 & -3 \\ -1 & 0 & 4 \\ 3 & -4 & 0 \end{bmatrix}$$

In this game, an equal chance of 1/3 is no longer optimal, and the first player has a total payoff of 9 while the second player has that of 8. It appears that the first player has an advantage, but this may not necessarily be true. For example, it may be possible for the second player to make a profit if for each game played, the second player were to make a payment to the first player that is less than the expected gain.

6.6 RECTANGULAR GAME THEORY AND ITS SOLUTION

Rectangular game theory is applied to the response of organisms to random environmental changes: A hypothetical example of the response of a facultative algae to random changes in illumination is analysed. Two strategies are discussed: Bayes' risk and maximin. It is shown how to detect such strategies in populations. The analysis does not require assumptions about the form of the relationship between metabolic activities and selective advantage, but assumptions about evolutionary optimisation are required. The maximin strategy is shown to be related to metabolic homeostasis.

Fundamental Theorem of Rectangular Games

If (a_{ij}) be any $m \times n$ payoff matrix of a game, than $\min\limits_{q \in Sn} \max\limits_{p \in S_m} ECP$, and $\max\limits_{p \in S_m} \min\limits_{p \in S_n} E(p,q)$ both exist and are equal.

Proof: The expectation function $E(p, q)$ is a continuous linear function of P defined over the closed and bounded subset S_m of R_m for each $q \in S_n$. Then $\max\limits_{q \in S_m} E(p,q)$ exist and is a continuous function of q

Similarly, $\max\limits_{p \in S_m} \min\limits_{q \in S_n} E(p,q)$ exists.

Now, if $\sum\limits_{i=1}^{m} p_i a_{ij} > 0$ for all j and multiplying by the component q_j of q and them taking summation was j. $\in (p,q) = \sum\limits_{j=1}^{n} \sum\limits_{i=1}^{m} P_i a_{ij} q_j > 0$ for all $q \in S_n$

Therefore, $\min\limits_{q \in S_j} E(p, q) > 0$ and consequently $\max\limits_{p \in S_n} \min\limits_{q \in S_n} E(p, q) > 0$

On the other hand, if assume,

$$\sum\limits_{j=1}^{n} q_j a_{ij} \le 0$$

Then by a similar argument, we can write

$$\min_{q \in S_n} \max_{p \in S_m} E(p, q) \leq 0$$

Thus, we have shown that it is not possible to simultaneously have $\displaystyle\min_{q \in S_n} \max_{p \in S_m} E(p, q) \leq 0$ and

$\displaystyle\max_{p \in S_m} \min_{q \in S_n} E(p, q) > 0$ Consider, now the payoff matrix $(a_{ij} - k)$ formed by subtracting k constant, positive or negative. If the expectation function of this reduced matrix is denoted by $E_k(p, q)$ then

$$E_k(p, q) \quad = \sum_{i=1}^{m} \sum_{j=1}^{n} p_i (a_{ij} - p) q_j$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{n} p_i \, a_{ij} \, q_j - k \sum_{i=1}^{m} \sum_{j=1}^{n} p_i q_j$$

$$= E(p - q) - k.$$

Since, $\displaystyle\sum_{i=1}^{m} p_i = 1 = \sum_{j=1}^{n} p_j$

Because, (a_{ij}) 0 any real matrix, what is true for (a_{ij}) is also true for $(a_{ij} - k)$. Therefore, we find that,

$$\max_{p \in S_m} \min_{q \in S_n} Ek(p, q) > 0 \text{ and } \min_{q \in S_n} \max_{p \in S_m} Ek(p, q) \leq 0$$

Or, $\displaystyle\max_{p \in S_m} \min_{q \in S_n} E_k(p, q) > k$ and $\displaystyle\min_{q \in S_n} \max_{p \in S_m} E(p, q) < k$ does not hold for any value of k.

We therefore, conclude that

$$\max_{p \in S_m} \min_{q \in Sn} E(p, q) < \min_{q \in S_n} \max_{p \in S_m} E(p, q)$$

Does not hold, Hence,

$$\max_{p \in S_m} \min_{q \in S_n} E(p, q) \geq \min_{q \in S_n} \max_{p \in S_m} \in (p, q)$$

Also,

$$\min_{q \in S_n} \min_{q \in S_m} E(p, q) \geq \max_{p \in S_m} \min_{q \in S_n} \in (p, q)$$

Thus,

$$\max_{p \in S_m} \min_{q \in S_n} E(p, q) = \min_{q \in S_n} \max_{p \in S_m} \in (p, q)$$

6.7 SUMMARY

1. Game theory is the study of mathematical models of strategic interaction between rational decision-makers.

2. Game theory was developed extensively in the 1950s by many scholars. It was later explicitly applied to Biology in the 1970s, although similar developments go back at least as far as the 1930s.

3. Game theory provides a mathematical framework for analysing the decision-making processes and strategies of adversaries (or players) in different types of competitive situations. The simplest type of competitive situations are two-person zero-sum games.

4. In game theory, a player's strategy is any of the options which he or she can choose in a setting where the outcome depends not only on their own actions but on the actions of others. A player's strategy will determine the action which the player will take at any stage of the game.

5. A matrix game, which is short for finite two-person zero-sum game, allows a game to be represented in matrix form as its name implies. This is a direct consequence of the fact that two opponents with exactly opposite interests play a game under a finite number of strategies, independently of his or her opponent's action.

6.8 SELF ASSESSMENT QUESTIONS

1. Explain in detail the Theory of Games.

2. Explain in detail about the Saddle Point and its importance.

3. Critically analyse 'Pure and Mixed Strategies'.

4. Explain in detail the Rectangular Game Theory.

Chapter	LEONTIEF SYSTEM
7	

Objectives

The objectives of this lesson are to learn:

- Leontief system
- Hawkins-Simon theorem
- Theorem of non-substitution

Structure:

7.1   THE LEONTIEF SYSTEM: STATIC INPUT-OUTPUT ANALYSIS

Input-output is a novel technique invented by Professor Wassily W. Leontief in 1951. It is used to analyse inter-industry relationship in order to understand the interdependencies and complexities of the economy, and thus the conditions for maintaining equilibrium between supply and demand.

Thus, it is a technique to explain the general equilibrium of the economy. It is also known as "inter-industry analysis". Before analysing the input-output method, let us understand the meaning of the terms "input" and "output". According to Professor J.R. Hicks, an input is "something which is bought for the enterprise" while an output is "something which is sold by it."

An input is obtained but an output is produced. Thus, input represents the expenditure of the firm, and output its receipts. The sum of the money values of inputs is the total cost of a firm and the sum of the money values of the output is its total revenue.

The input-output analysis tells us that there are industrial interrelationships and interdependencies in the economic system as a whole. The inputs of one industry are the outputs of another industry and *vice versa*, so that ultimately their mutual relationships lead to equilibrium between supply and demand in the economy as a whole.

Coal is an input for steel industry and steel is an input for coal industry, though both are the outputs of their respective industries. A major part of economic activity consists in producing intermediate goods (inputs) for further use in producing final goods (outputs).

There are flows of goods in "whirlpools and cross currents" between different industries. The supply side consists of large inter-industry flows of intermediate products and the demand side of the final goods. In essence, the input-output analysis implies that in equilibrium, the money value of aggregate output of the whole economy must equal the sum of the money values of inter-industry inputs and the sum of the money values of inter-industry outputs.

1. Main Features

The input-output analysis is the finest variant of general equilibrium. As such, it has three main elements; Firstly, the input-output analysis concentrates on an economy which is in equilibrium. Secondly, it does not concern itself with the demand analysis. It deals exclusively with technical problems of production. Lastly, it is based on empirical investigation. The input-output analysis consists of two parts: the construction of the input-output table and the use of input-output model.

2. The Static Input-output Model

The input-output model relates to the economy as a whole in a particular year. It shows the values of the flows of goods and services between different productive sectors especially inter-industry flows.

Assumptions

This analysis is based on the following assumptions:

 (i) The whole economy is divided into two sectors—"inter-industry sectors" and "final-demand sectors," both being capable of sub-sectoral division.

 (ii) The total output of any inter-industry sector is generally capable of being used as inputs by other inter-industry sectors, by itself and by final demand sectors.

 (iii) No two products are produced jointly. Each industry produces only one homogeneous product.

 (iv) Prices, consumer demands and factor supplies are given.

 (v) There are constant returns to scale.

 (vi) There are no external economies and diseconomies of production.

 (vii) The combinations of inputs are employed in rigidly fixed proportions. The inputs remain in constant proportion to the level of output. It implies that there is no substitution between different materials and no technological progress. There are fixed input coefficients of production.

Explanation

For understanding, a three-sector economy is taken in which there are two inter-industry sectors, agriculture and industry, and one final demand sector.

Table 7.1 provides a simplified picture of such economy in which the total output of the industrial, agricultural and household sectors is set in rows (to be read horizontally) and has been divided into the agricultural, industrial and final demand sectors. The inputs of these sectors are set in columns. The first row total shows that altogether the agricultural output is valued at ₹ 300 crores per year.

Table 7.1: Input-Output Table

(In value terms) (₹ crores)

Sectors		Purchasing Sectors			
		1 **Input to** **Agriculture**	**2** **Inputs to** **Industry**	**3** **Final** **Demand**	**Total Output** **or** **Total Revenue**
Selling Sectors	Agriculture	50	150	100	300
	Industry	100	250	150	500
	Value added*	150	100	0	250
	Total input **or** **Total cost**	**300**	**500**	**250**	**1050**

Of this total, ₹ 100 crores go directly to final consumption (demand), i.e., household and government, as shown in the third column of the first row. The remaining output from agriculture goes as inputs: 50 to itself and 150 to industry. Similarly, the second row shows the distribution of total output of the industrial sector valued at ₹ 500 crores per year. Columns 1, 2 and 3 show that 100 units of manufactured goods go as inputs to agriculture, 250 to industry itself and 150 for final consumption to the household sector.

Let us take the columns (to be read downwards). The first column describes the input or cost structure of the agricultural industry. Agricultural output valued at ₹ 300 crores is produced with the use of agricultural goods worth ₹ 50, manufactured goods worth ₹ 100 and labour and/or management services valued at ₹ 150. To put it differently, it costs ₹ 300 crores to get revenue of ₹ 300 crores from the agricultural sector. Similarly, the second column explains the input structure of the industrial sector (i.e., $150 + 250 + 100 = 500$).

Thus, "a column gives one point on the production function of the corresponding industry." The 'final demand' column shows what is available for consumption and government expenditure. The third row corresponding to this column has been shown as zero. This means that the household sector is simply a spending (consuming) sector that does not sell anything to itself. In other words, labour is not directly consumed.

There are two types of relationships which indicate and determine the manner in which an economy behaves and assumes a certain pattern of flows of resources. They are:

(a)   The internal stability or balance of each sector of the economy, and

(b)   The external stability of each sector or inter-sectoral relationships.

Professor Leontief calls them the "fundamental relationships of balance and structure." When expressed mathematically, they are known as the "balance equations"' and the "structural equations".

If the total output of say X of the 'ith' industry is divided into various numbers of industries 1, 2, 3, ..., n, then we have the balance equation:

$$X_1 = x_{i1} + x_{i2} + x_{i3} + x_{in} + ... + D_1 \qquad \qquad ...(7.1)$$

and if the amount say Y absorbed by the "outside sector" is also taken into consideration, the balance equation of the ith industry becomes

$$X_i = x_{11} + x_{12} + x_{13} + ... + x_{in} + D_i + Y_i \qquad \qquad ...(7.2)$$

or

$$\sum_{i=1}^{n} X_{ij} + Y_i = X_i$$

It is to be noted that Y_i stands for the sum of the flows of the products of the ith industry to consumption, investment and exports net of imports, etc. It is also called the "final bill of goods" which is the function of the output to bill. The balance equation shows the conditions of equilibrium between demand and supply. It shows the flows of outputs and inputs to and from one industry to other industries and *vice versa.*

Since x_{12} stands for the amount absorbed by industry 2 of the ith industry, it follows that X_{ij} stands for the amount absorbed by the ith industry of jth industry.

The "technical coefficient" or "input coefficient" of the ith industry is denoted by:

$a_{ij} = X_{ij}/x_j$ ...(7.3)

where X_{ij} is the flow from industry i to industry j, x_j is the total output of industry a_{ij} and a_{ij}, as already noted above, is a constant, called "technical coefficient" or "flow coefficient" in the ith industry. The technical coefficient shows the number of units of one industry's output that are required to produce one unit to another industry's output.

Equation (7.3) is called a "structural equation." The structural equation tells us that the output of one industry is absorbed by all industries so that the flow structure of the entire economy is revealed. A number of structural equations give a summary description of the economy's existing technological conditions.

Using equation (7.3) to calculate the a_{ij} for our example of the two-sector input-output Table 1, we get the following technology matrix.

Table 7.2: Technology Coefficient Matrix A

	Agriculture	Industry
Agriculture	50/300 = .17	150/500 = .30
Industry	100/300 = .33	250/500 = .50

These input coefficients have been arrived at by dividing each item in the first column of Table 7.2 by first row total, and each item in the second column by the second row, and so on. Each column of the technological matrix reveals how much agricultural and industrial sectors require from each other to produce a rupee's worth of output. The first column shows that a rupee's worth of agricultural output requires inputs worth 33 paise from industries and worth 17 paise from agriculture itself.

The Leontief Solution

The table can be utilised to measure the direct and indirect effects on the entire economy of any sectoral change in the total output of final demand.

Again using equation (7.3),

$a_{ij} = x_{ij}/x_j$ ...(7.4)

Cross multiplying, $x_{ij} = a_{ij} . x_j$

By substituting the value of x_{ij} into equation (7.2) and transposing terms, we obtain the basic input-output system of equations.

$$X_i - \sum_{i=1}^{n} a_{ij} x_j = Y_i$$

In terms of our two-sector economy, there would be two linear equations that could be written symbolically as follows:

(a) $x_1 - a_{11} x_1 \dots a_{12} x_2 = y_1$

(b) $x_2 - a_{21} x_2 \dots a_{22} x_2 = y_2$

The above symbolic relationship can be shown in matrix form:

$X - [A]X = Y$

$X [I - A] = Y$

where matrix $(I - A)$ is known as the Leontief Matrix.

$$(1 - A)^{-1} (1 - A)X = (1 - A)^{-1}Y$$

$$X = (1 - A)^{-1}Y \qquad [\because (1 - A)^{-1} (1 - A)]$$

and I, the identity matrix $= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

Hence, $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - [A] \right\}^{-1} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$

Numerical Solution

Our technology matrix as per Table 7.2 is

$A = \begin{bmatrix} .1 & .3 \\ .3 & .5 \end{bmatrix}$ and $Y = \begin{bmatrix} 100 \\ 150 \end{bmatrix}$

$(1 - A) = \begin{bmatrix} .9 & -.3 \\ .3 & -.5 \end{bmatrix}$

The value of inverse $= \dfrac{\text{Adjoint}}{\text{Determinant}} = \dfrac{A_{ij}}{|A|}$

$$|A_{ij}| = \begin{bmatrix} .5 & .3 \\ .3 & .9 \end{bmatrix}$$

By transposing, $A_{ij} = \begin{bmatrix} .5 & .3 \\ .3 & .9 \end{bmatrix}$

The value of determinant $= .9(.5) - (-.3)(-.3)$

$$= .45 - .09 = .36$$

Hence, $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \dfrac{1}{.36} \begin{bmatrix} .5 & .3 \\ .3 & .9 \end{bmatrix} \begin{bmatrix} 100 \\ 150 \end{bmatrix}$

The total output of agriculture sector $(x_1) = \dfrac{.5 \times 100 + .3 \times 150}{.36} = 264$

The total output of industrial sector $(x_2) = \dfrac{.3 \times 100 + .9 \times 150}{.36} = 458$

3. The Dynamic Input-Output Model

So far, we have studied an open static model. "The model becomes dynamic when it is closed by the linking of the investment part of the final bill of goods to output. The dynamic input-output model extends the concept of inter-sectoral balancing at a given point of time to that of inter-sectoral balancing over time.

This necessarily involves the concept of durable capital. The Leontief dynamic input-output model is a generalisation of the static model and is based on the same assumptions. In a dynamic model, the output of a given period is supposed to go into stocks, i.e., capital goods, and the stocks, in turn, are distributed among industries.

The balance equation is:

$$X_i(t) = x_{i1}(t) + x_{i2}(t) + x_{i3}(t) + ... + x_{in}(t) + (s'_{i1} + s'_{i2} + s'_{i3} + ... + s'_{in}) + D_i(t) + Y_i(t) \qquad ...(7.5)$$

Here, $X_i(t)$ represents the total flow of output of ith industry in period t, which is used for three purposes:

(i) For production in the economy's n industries, $x_{11}(t)$, $x_{12}(t)$, etc. in that period;

(ii) As net addition to the stock of capital goods in n industries, i.e., S'_t which can also be written as $\Delta S_1(t) = S_1(t + 1) - S_1(t)$, where $S_1(t)$ indicates the accumulated stock of capital in the current period (t), and $S_1(t + l)$ is next year's stock; and

(iii) As consumption demand for the next period $D.(t + 1)$. If we ignore depreciation and wear-tear, then $S.(t + 1) - S_1(t)$ is the net addition to capital stock out of current production. Equation (7.4) can, therefore, be written as:

$$X_i(t) \, t \, x_{1i1} + x_{i2} + x_{i3} + x_{in} + S.(t + 1) - S_1(t) + D_2(t) + Y_i(t) \qquad ...(7.6)$$

where $Y_i(t)$ stands for the amount absorbed by the outside sector in period t.

Just as the technical coefficient was derived in the case of the static model, the capital coefficient can be found out in a similar manner. Capital coefficient of the ith product used by the jth industry is denoted by

$$b_{ij} = S_{ij}/X_j$$

Cross multiplying, we have $S_{ij} = b_{ij}.X$

where, S_{ij} represents the amount of capital stock of the ith product used by the jth industry. X_j is the total output of industry j, and b_{ij} is a constant called capital coefficient or stock coefficient. Equation (7.5) is known as the structural equation in a dynamic model.

If the b_{ij} coefficient is zero, it means that no stock is required by an industry and the dynamic model becomes a static model. Moreover, b_{ij} can neither be negative nor infinite. If the capital coefficient is negative, the input is, in fact, an output of an industry.

Leontief Model

The open Leontief Model

Let the n industries denoted by $S_1, S_2, ... S_n$. Here, a_{ij} denote the number of units produced by industry S_i necessary to produce one unit by industry S_j and b_j with number of externally demanded units of industry S_i

The following equations are satisfied:

Production of	Total output	=	Internal consumption	+ External Demand
Farming industry (intons):	x	=	$0.05x + 0.5y$	$+ 8000$
Horse industry:	y	=	$0.01x$	$+ 2000$

(in 100 km horse rides)

In general, let $x_1, x_2, ..., x_n$ be the total output of industry $S_i, S_2, ... S_n$ respectively.

Then

$x_1 = a_{11} x_L + a_{12} x_1 + ...a_{1n} \; x_1 + b_1$

$x_2 = a_{21}, x_1 + a_{22} x_1 + ...+ a_{2n} x_2 + b_2$

…

$x_n = a_{n1} x_1 + a_{n2} x_1 + ...+ a_{nn} x_n + b_n$

Since, a_{ij}, x_j is the number of units produced by industry S_i and consumed by industry S_j. The total consumption equals the total production for the product of each industry S_i

Let,

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}, \; B = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}, \; x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

A is called the input output matrix, B the external demand vector and x the production level vector. The alone system of linear equations is equivalent to the matrix equation

$x = Ax + B$

In the open Leontief model, A and $B \neq \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$ are given and the problem is to determine x form

this matrix equation.

We can transform this equation as follows:

$I_n Y - Ax = B$

$(I_n - A)x = B$

$x = (I_n - A) B$

If the inverse of the matrix $I_n - A$ exist, $(I_n - A)$ is there called the Leantief inverse. For a given realistic economy, a solution obviously must exist.

For our example, we have

$$A = \begin{pmatrix} 0.05 & 0.5 \\ 0.1 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 8,000 \\ 2,000 \end{pmatrix}, \qquad X = \begin{pmatrix} x \\ y \end{pmatrix}$$

We obtain therefore the solution

$$x = (I_L - A)_B$$

$$= \left(\begin{pmatrix} 10 \\ 01 \end{pmatrix} - \begin{pmatrix} 0.05 & 0.5 \\ 0.1 & 0 \end{pmatrix} \right) \begin{pmatrix} 8000 \\ 2000 \end{pmatrix}$$

$$= \begin{pmatrix} 0.95 & -0.5 \\ -0.1 \end{pmatrix} \begin{pmatrix} 8000 \\ 2000 \end{pmatrix}$$

$$= \frac{1}{9} \begin{pmatrix} 10 & 5 \\ 1 & 9.5 \end{pmatrix} \begin{pmatrix} 8000 \\ 2000 \end{pmatrix}$$

$$= \begin{pmatrix} 10,000 \\ 3,000 \end{pmatrix}$$

i.e., x = 10,000 tons wheat and y = 3 million kn horse side. If the external demand changes, to $B°$

$$= \begin{pmatrix} 7300 \\ 2500 \end{pmatrix}, \text{ we get } \begin{pmatrix} x \\ y \end{pmatrix}° = (I_2 - A) \ B° = \frac{1}{a} \begin{pmatrix} 10 & 5 \\ 1 & 9.5 \end{pmatrix} \begin{pmatrix} 7300 \\ 2500 \end{pmatrix} = \begin{pmatrix} 9500 \\ 3450 \end{pmatrix}$$

i.e., are does not need to recompute $(I_2 - A)$

One difficulty with the model: how to determine the matrix A from a given economy? Typically,

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \text{ is known}$$

$$B = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \text{ is known and } (a_{ij}x_j)_{i,j} = 1, ..., n \text{ is known}$$

One tanks therefore the matrix $(a_{ij}x_j)_{i,j}$ 1, ..., n and divides the 7th column by xy for j = 1, ..., n to get A

Example: An economy has the two industries R and S, the current consumption is given by the table.

Table 7.3: Consumption

	R	S	External
Industry R production	50	50	20
Industry S production	60	40	100

Assume the new external demand is 100 units of R and 100 units of S. Determine the new production levels.

Solution: The total production is 120 units for R and 200 units for S. We obtain

$$X = \begin{pmatrix} 120 \\ 100 \end{pmatrix} \quad B = \begin{pmatrix} 120 \\ 100 \end{pmatrix} \quad A = \begin{pmatrix} \dfrac{50}{120} & \dfrac{50}{200} \\[2mm] \dfrac{60}{120} & \dfrac{40}{200} \end{pmatrix} \text{ and }$$

$$B° = \begin{pmatrix} 100 \\ 100 \end{pmatrix}.$$

The solution is $x° = (I_2 - A)B° = \dfrac{1}{41} \begin{pmatrix} 96 & 30 \\ 60 & 70 \end{pmatrix} \begin{pmatrix} 100 \\ 100 \end{pmatrix} = \begin{pmatrix} 30 & 7 & 3 \\ 31 & 7 & 0 \end{pmatrix}$

The new production levels are 307.3 and 317.0 for R and S respectively.

7.2 HAWKINS-SIMON THEOREM

Hawkins-Simon Theorem

The Hawkins-Simon condition refers to a result in mathematical economics, attributed to David Hawkins and Herbert A. Simon, that guarantees the existence of a non-negative output vector that solves the equilibrium relation in the input-output model when demand equals supply. More precisely, it states a condition for $|I - A|$ under which the input-output system

$$|I - A|.x = a$$

has a solution $\hat{x} \geq 0$ for any $d \geq 0$. Hence, I is the identity matrix and A is called the input-output matrix or Leontief matrix after Wassily Leontief, who empirically estimated it in the 1940s. Together, they describe a system in which

$$\sum_{i=1}^{n} a_{ij} x_j + d_i = x_i, \qquad i = 1, 2, \ldots, n$$

Where a_{ij} is the amount of the ith good used to produce one unit of the jth good, x_j is the amount of the jth good produced, and d_i is the amount of final demand for good i. Rearranged and written in vector notation, this gives the first equation.

Define $[I - A] = B$, where $B = [b_{ij}]$ is an $n \times n$ matrix with $b_{ij} \leq 0$, $i \neq j$. Then the Hawkins-Simon theorem states that the following two conditions are equivalent:

(i)   There exists an $x > 0$ such that $B . x > 0$.

(ii) All the successive principle minors of B are positive, i.e.,

$$b_{11} > 0, \quad \begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix} > 0, \quad \ldots, \quad \begin{vmatrix} b_{11} & b_{12} & \ldots & b_{1n} \\ b_{21} & b_{22} & \ldots & b_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ b_{n1} & b_{n2} & \ldots & b_{nn} \end{vmatrix}$$

Let A be a real square non matrix. A is said to be inverse (semi) positive, if it is not singular and A^{-1} is (semi) positive. A tilde on a real vector x or a real square matrix denote transposition. Notations $x \geqq 0$ (a $x \in R_x^n$), $x \geq 0$, $x > 0$ (a $x \in R_{xx}^n$) mean respectively that vector x is non-negative, semipositives or positive. A bar on a vector or a matrix either denotes function of the last components or suggests a vocation to further extension; a denote bar denotes function of the first components.

An LU factorization of A is a decomposition A = LU, where L is a lower triangular matrix with unit diagonal matrix, and U is an upper triangular matrix. If is well known (and this results from the ensuring calculations) that much a factorization exist when all the leading principal minor ('leading minors' for short are nonzero and then the factorization is unique.

We shall consider a classical transform of the system of equations

$A_x = Y$

$a_{11} x_1 + a_{12} x_2 + ...+ a_{1n} x_n = y_1$

$a_{21} x_1 + a_{22} x_2 + ...+ a_{2n} x_n = y_2$

$a_{x1} x_1 + a n_2 x_2 + ... + a_{nn} x_n = y_n$

If $a_{11} \neq 0$, the first equality can be used to eliminate x_1 from the other equations. Here, we especially denote to the properties of the transformed system and its associated $(n - 1) \times (n - 1)$ matrix S_1, more generally to these of the $(n - k) \times (n - k)$ matrix S_k obtained after the successive eliminations of $x_1, ..., xk$.

A fruitful interpretation of the elimination of x_1 is to consider that we have premultiplied both number of the equality $A_x = y$ by the lower triangular matrix

$$L_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -a_{21}/a_{11} & 1 & 0 & 0 \\ \cdots & 0 & \ddots & 0 \\ -a_{n1}/a_{11} & 0 & 0 & 1 \end{bmatrix}$$

In $L_1 A$, the first row coincides with that of A and the entire 2 to n of the first column are zero. Let us denote $A_{1i\,1j} = a_{11} a_{1j} - a_{11} a_{1j}$ the 2×2 minor extracted from rows 1 and i and columns 1 and j of A. The $(n - 1) \times (n - 1)$ sub-matrix S_1 made of rows and column 2 to x of L, A is written as

$$S_1 = \begin{bmatrix} \Delta_{1212}/a_{11} & \cdots & \Delta_{1212}/a_{11} \\ \cdots & \cdots & \cdots \\ \Delta_{1n12}/a_{11} & \cdots & \Delta_{1n\,1n}/a_{11} \end{bmatrix}$$

$S_1 = S_1 (A)$ is called the sehur compliment of a 11. The initial system of equations is transformed into the equivalent systems.

$a_{11} x_1 + a_{12} + x_2 + ... + a_{1n} x_n = y_1$

$$S_1 \begin{pmatrix} x_2 \\ x_n \end{pmatrix} = \begin{pmatrix} y_L \\ \overline{y_n} \end{pmatrix} + y_1 \begin{pmatrix} -a_{21}/a_{11} \\ -a_{n1}/a_{11} \end{pmatrix}$$

The n – 1 of above equation are written more compactly as

$$s_1 \overline{\overline{x}}_1 = y(1) + y_1 \overline{\overline{L}}_{(1)}$$

When, $\overline{\overline{x}}$ (1) (respectively $\overline{\overline{y}}$ (1)) denotes the vector x (respectively y) truncated of its component, and $\overline{\overline{l}}_{(1)}$ the column vector made of the last n-1 components of the first column of L_1.

7.3   THEOREM ON NON-SUBSTITUTION

The non-substitution theorem asserts that the choice of technique is independent of patterns of final demand when efficiency prevails as to the use of a single primary factor, say labour, while the asserted constancy of the input-output table no longer holds when more than one kind of primary factor is involved.

The aim of the Non-substitution Theorem lies in the justification of constancy of an input-output table, and it asserts:

(i)   That the choice of technique is independent of patterns of final demand when efficiency prevails as to the use of a single primary factor, say labour; and

(ii)   That the commodity price vector is uniquely determined by the production structure, independently of final demand patterns.

These two findings have a solid foundation in the static context, but the dynamic non-substitution theorems allow various degrees of freedom (Mirrlees (1969), Burmeister and Kuga (1970), Stiglitz (1970), Samuelson (1991)) in that the input-output table including the capital input coefficients is stable in the long-run along the wage-profit frontier. It is well known that the asserted constancy of the input-output table no longer holds when more than one kind of primary factor is involved. Yet, no question has been posed so far as to whether the commodity price vector is determined independently of final demand patterns when there are multiple primary factors of production.

The Model

Every industry indexed by $j \in N \overset{dif}{=} \{1, 2, \ldots, n\}$ has a no-joint, concave, constant returns to scale technology. The input consists of intermediate outputs from industries is N and primary factors indexed by $k \in M \overset{dif}{=} \{1, 2, \ldots, m\}$. We begin with the description of the unit cost functions. The domain of each cost function is $P \times W$, where $P \overset{dif}{=} R_+^n$ consists of prices of n commodities.

$p = (p_1, \ldots, p_n)$ and $W \overset{dif}{=} R_+^m \setminus \{0\}$ consists of factor prices $w = (w_1, \ldots, w_n)$ of m primary factors. Our first concern is to examine the preliminary question of whether or not to the set of unit cost functions $c^j : p \times w \to R_+^n$ of industry $j \in N$, there exists a certain price sector $p = (p_1, p_2, \ldots, p_n) \in P$ given $w = (w_1, w_2, \ldots, w_n) \in W$ satisfying.

$$P_j = c^j(p, w), \ j \in N \hspace{3cm} \ldots(7.7)$$

We may write (7.7) as $P = c(p, w)$ and call p that satisfies (7.7) the unit cost equating price relative to w. To ensure the existence of such a price $p \in P$ to (7.7), we assume

(A.1)   Given $w \in W$, there is a certain $\overline{p} = (\overline{p}_1, \ldots, \overline{p}_n) \in P$, depending upon w, satisfying $\overline{p}_j \geq c^j(\overline{p}, w)$ for all $j \in N^3$;

(A.2)   Given $w \in W, c^j$ is non-decreasing, i.e., $c^j(p, w) \leq c^j(p', w)$ for $p \leq p'(p, p' \in P)$ and $j \in N$.

Theorem 1: Suppose an economy satisfies (A.1) and (A.2). Then for each $w \in W$, there is a certain $p \in P$, satisfying $p_j = c^j(p, w)$, j.

A subset $P \in s \times s$ is a partial order is S_i f the following there conditions are met: (i) $(x, x) \in P$ for $x \in s$, (ii) $[(x, y) \in P \wedge (y, x) \in P] \Rightarrow x = y$, (iii) $[(x, y) \in P \wedge (y, 2) \in P] \Rightarrow (x, z) \Rightarrow P$. A function f: $P \rightarrow P$ is called non-decreasing if $u \leq v \Rightarrow f(u) \leq f(v)$.

Exploiting these concepts, we use the well-known Tarski's fixed point theorem:

Tarski's Fixed Point Theorem

Let $(P, \leq)$ be a conditionally complete partially ordered set and f a non-decreasing function from P into P. If P has two elements a end to b such that $a \leq f(a) \leq f(b) \leq b$ then there exists are elements e $\in P$ c = f(c).

A partially ordered set is which every non empty subset that is bounded above has a least upper bound is conditionally complete. $P = R^n +$ is conditionally complete with respect to the usual partial order $\leq$ we show that the required conditions are met is our case. Our (A.2) is the non-decreasingness is Tarski's Fixed Point Theorem. We use (A.1) for the existence of an upper bound P satisfying $c(P, w) \leq P$. The lower bound is our case may be taken to be zero. There, by (A.2) is follows that $o \leq e(o, w) \leq e(P, w)$ in sum, we have $0 \leq c(o, w) \leq c(p, w) \leq P$.

Theorem 2: Suppose in an economy for each w in the interior of w, there exists a unique positive cost equating price $p \in P$ satisfying $P = c(p, w)$.

Remark: Theorem 2 can be regarded as an extension of Theorem 1. He assumes that each cost function c does not contain P_i as its arguments and that there is only one kind of primary factor, labour. Theorem 2 is more general is that c does not contain P_i as its arguments, is addition to the structure is which more than one kind of primary factor is involved.

7.4 SUMMARY

1. Input-output is a novel technique invented by Professor Wassily W. Leontief in 1951.

2. It is used to analyse inter-industry relationship in order to understand the interdependencies and complexities of the economy, and thus the conditions for maintaining equilibrium between supply and demand.

3. The non-substitution theorem asserts that the choice of technique is independent of patterns of final demand when efficiency prevails as to the use of a single primary factor, say labour, while the asserted constancy of the input-output table no longer holds when more than one kind of primary factor is involved.

7.5 SELF ASSESSMENT QUESTIONS

1. Critically explain the Leontief System: Static Input-Output Analysis.

2. Critically explain Hawkins-Simon Theorem.

3. Critically explain Theorem on Non-substitution.

❧❦❖❧❧

UNIT – V

<table>
<tr><td>Chapter

8</td><td>THEORY OF ECONOMIC GROWTH</td></tr>
</table>

Objectives

The objectives of this lesson are to:

- Harrod-Domar Model
- Neo-classical model of Solow
- Neo-Keynesian model of Passineti
- Two-sector model of Uzawa
- Optimal Economic Growth – Ramsey Problem

Structure:

8.1   THEORY OF ECONOMIC GROWTH: HARROD-DOMAR MODEL

Ever since the end of Second World War, interest in the problems of economic growth has led economists to formulate growth models of different types. These models deal with and lay emphasis on the various aspects of growth of the developed economies. They constitute in a way alternative stylised pictures of an expanding economy. A feature common to all of them is that they are based on the Keynesian saving-investment analysis. The first and the simplest model of growth—the Harrod-Domar Model—is the direct outcome of projection of the short-run Keynesian analysis into the long-run.

This model is based on the capital factor as the crucial factor of economic growth. It concentrates on the possibility of steady growth through adjustment of supply of demand for capital. Then there is Mrs. Joan Robinson's model which considers technical progress also, along with capital formation, as a source of economic growth. The third type of growth model is that built on neo-classical lines. It assumes substitution between capital and labour, and a neutral technical progress in the sense that technical progress is neither saving nor absorbing of labour or capital. Both the factors

are used in the same proportion even when neutral technical takes place. We deal with the prominent growth models here.

Although Harrod and Domar models differ in details, they are similar in substance. One may call Harrod's model as the English version of Domar's model. Both these models stress the essential conditions of achieving and maintaining steady growth. Harrod and Domar assigned a crucial role to capital accumulation in the process of growth. In fact, they emphasised the dual role of capital accumulation.

On one hand, new investment generates income (through multiplier effect); on the other hand, it increases productive capacity (through productivity effect) of the economy by expanding its capital stock. It is pertinent to note here that classical economists emphasised the productivity aspect of the investment and took for granted the income aspect. Keynes had given due attention to the problem of income generation but neglected the problem of productive capacity creation. Harrod and Domar took special care to deal with both the problems generated by investment in their models.

General Assumptions

The main assumptions of the Harrod-Domar models are as follows:

(a) A full-employment level of income already exists.

(b) There is no government interference in the functioning of the economy.

(c) The model is based on the assumption of "closed economy." In other words, government restrictions on trade and the complications caused by international trade are ruled out.

(d) There are no lags in adjustment of variables, i.e., the economic variables such as savings, investment, income and expenditure adjust themselves completely within the same period of time.

(e) The Average Propensity to Save (APS) and Marginal Propensity to Save (MPS) are equal to each other (APS = MPS) or written in symbols,

$S/Y = \Delta S/\Delta Y$

(f) Both propensity to save and "capital coefficient" (i.e., capital-output ratio) are given constant. This amounts to assuming that the law of constant returns operates in the economy because of fixity of the capital-output ratio.

(g) Income, investment and savings are all defined in the net sense, i.e., they are considered over and above the depreciation. Thus, depreciation rates are not included in these variables.

(h) Saving and investment are equal in *ex-ante* as well as in *ex-post* sense, i.e., there is accounting as well as functional equality between saving and investment.

These assumptions were meant to simplify the task of growth analysis; these could be relaxed later.

Harrod's growth model raised three issues:

(i) How can steady growth be achieved for an economy with a fixed (capital-output ratio) (capital-coefficient) and a fixed saving-income ratio?

(ii) How can the steady growth rate be maintained? Or what are the conditions for maintaining steady uninterrupted growth?

(iii) How do the natural factors put a ceiling on the growth rate of the economy?

In order to discuss these issues, Harrod had adopted three different concepts of growth rates:

- The actual growth rate, G,

- The warranted growth rate, G_w

- The natural growth rate, G_n

The Actual Growth Rate is the growth rate determined by the actual rate of savings and investment in the country. In other words, it can be defined as the ratio of change in income (ΔT) to the total income (Y) in the given period. If actual growth rate is denoted by G, then

$$G = \Delta Y/Y$$

The actual growth rate (G) is determined by saving-income ratio and capital-output ratio. Both the factors have been taken as fixed in the given period. The relationship between the actual growth rate and its determinants was expressed as:

$$GC = s \qquad\qquad\qquad\qquad \ldots(8.1)$$

where G is the actual rate of growth, C represents the capital-output ratio $\Delta K/\Delta Y$ and s refers to the saving-income ratio $\Delta S/\Delta Y$. This relation states the simple truism that saving and investment (in the *ex-post* sense) are equal in equilibrium. This is clear from the following derivation.

Since $G = \dfrac{\Delta Y}{Y}$

$C = \dfrac{\Delta K}{\Delta Y} \cdot \dfrac{1}{\Delta Y}$ $\left[\because \Delta K = I \right]$

As $s = \dfrac{S}{Y}$

Substituting the value of G, C and s in equation (8.1), we get

$$\frac{\Delta Y}{Y} \times \frac{I}{\Delta Y} = \frac{S}{Y}$$

or $\dfrac{I}{Y} = \dfrac{S}{Y}$

or $I = S$

This relation explains that the condition for achieving the steady state growth is that *ex-post* savings must be equal to *ex-post* investment. "Warranted growth" refers to that growth rate of the economy when it is working at full capacity. It is also known as full-capacity growth rate. This growth rate denoted by G_w is interpreted as the rate of income growth required for full utilisation of a growing stock of capital, so that entrepreneurs would be satisfied with the amount of investment actually made.

Warranted growth rate (G_w) is determined by capital-output ratio and saving-income ratio. The relationship between the warranted growth rate and its determinants can be expressed as

$$G_w C_r = s$$

where C_r shows the needed C to maintain the warranted growth rate and s is the saving-income ratio.

Let us now discuss the issue: how to achieve steady growth? According to Harrod, the economy can achieve steady growth when

$G = G_w$ and $C = C_r$

This condition states, firstly, that actual growth rate must be equal to the warranted growth rate. Secondly, the capital-output ratio needed to achieve G must be equal to the required capital-output ratio in order to maintain G_w, given the saving coefficient (s). This amounts to saying that actual investment must be equal to the expected investment at the given saving rate.

Instability of Growth

We have stated above that the steady-state growth of the economy requires an equality between G and G_w on one hand, and C and C_r on the other. In a free-enterprise economy, these equilibrium conditions would be satisfied only rarely, if at all. Therefore, Harrod analysed the situations when these conditions are not satisfied:

(i)   If $G > G_w$* then $C < C_r$

(ii)   If $G < G_W$ then $C > C_r$

We analyse the situation where G is greater than G_w. Under this situation, the growth rate of income being greater than the growth rate of output, the demand for output (because of the higher level of income) would exceed the supply of output (because of the lower level of output) and the economy would experience inflation. This can be explained in another way too when $C < C_r$. Under this situation, the actual amount of capital falls short of the required amount of capital.

This would lead to deficiency of capital, which would, in turn, adversely affect the volume of goods to be produced. Fall in the level of output would result in scarcity of goods and hence inflation. This, under this situation, the economy will find itself in the quagmire of inflation.

On the other hand, when G is less than G_w, the growth rate of income would be less than the growth rate of output. In this situation, there would be excessive goods for sale, but the income would not be sufficient to purchase those goods. In Keynesian terminology, there would be deficiency of demand, and consequently, the economy would face the problem of deflation. This situation can also be explained when C is greater than C_r.

Her, the actual amount of capital would be larger than the required amount of capital for investment. The larger amount of capital available for investment would dampen the marginal efficiency of capital in the long period. Secular decline in the marginal efficiency of capital would lead to chronic depression and unemployment. This is the state of secular stagnation.

From the above analysis, it can be concluded that steady growth implies a balance between G and G_w. In a free-enterprise economy, it is difficult to strike a balance between G and G_w as the two are determined by altogether different sets of factors. Since a slight deviation of G from G_w leads the economy away and further away from the steady-state growth path, it is called 'knife-edge' equilibrium.

G_n, the natural growth rate, is determined by natural conditions such as labour force, natural resources, capital equipment, technical knowledge, etc. These factors place a limit beyond which expansion of output is not feasible. This limit is called Full-employment Ceiling. This upper limit may change as the production factors grow or as technological progress takes place. Thus, the natural growth rate is the maximum growth rate which an economy can achieve with its available natural

resources. The third fundamental relation in Harrod's model showing the determinants of natural growth rate is

$G_n C_r$ is either = or ≠ s

Interaction of G, G_w and G_n
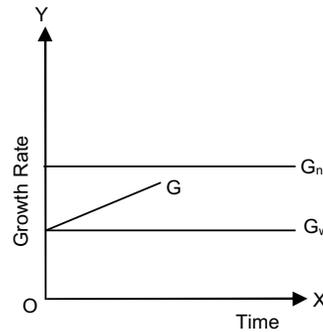
Fig. 8.1: When $G_n > G$ Employment Increases so as to Achieve full Employment

Comparing the second and the third relations about the warranted growth rate and the natural growth rate which have been given above, we may conclude that G_n may or may not be equal to G_w. In case G_n happens to be equal to G_w, the conditions of steady growth with full employment would be satisfied. But such a possibility is remote because of the variety of hindrances are likely to intervene and make the balance among all these factors difficult. As such, there is a definite possibility of inequality between G_n and G_w. If G_n exceeds G_w, G would also exceed G_w for most of the time as is shown in Fig. 8.1, and there would be a tendency in the economy for cumulative boom and full employment.

Such a situation will create an inflationary trend. To check this trend, savings become desirable because these would enable the economy to have a high level of employment without inflationary pressures. If on the other hand, G_w exceeds G_n, G must be below G_n for most of the time and there would be a tendency for cumulative recession resulting in unemployment (Fig. 8.1).

The Domar Model

The main growth model of Domar bears a certain resemblance to the model of Harrod. In fact, Harrod regarded Domar's formulation as a rediscovery of his own version after a gap of seven years.

Domar's theory was just an extension of Keynes' General Theory, particularly on two counts:

1. Investment has two effects:

   (a) An income-generating effect and

   (b) Productivity effect by creating capacity.

   The short-run analysis governed by Keynes ignored the second effect.

2. Unemployment of labour generally attracts attention and one feels sympathy for the jobless, but unemployment of capital attracts little attention. It should be understood that unemployment of capital inhibits investment and hence reduces income. Reduction of income brings about deficiency in demand and hence unemployment.

Thus, the Keynesian concept of unemployment misses the root cause of the problem. Domar wanted to analyse the genesis of unemployment in a wider sense.

To understand the implications of Domar model, one should get familiar with the relations listed below:

1. Income is determined by investment through multiplier. For simplicity, saving-income ratio (s) is assumed constant. This implies that

$Y_{(t)} = I_{(t)}/s$

where Y is the output, I is the actual investment and s is saving-income ratio (saving propensity), and (t) shows the time period.

2. Productive capacity is created by investment to the extent of the potential (social) average productivity of investment denoted by α. For simplicity, this is also assumed to be constant. In notation form, the relation can be written as

$Y_{(t)} - Y_{(t-1)} = I_{(t)/\alpha}$

where Y shows the productive capacity for output and α is the actual marginal capital-output ratio which is the reciprocal of "potential social average investment productivity" ($\alpha = 1/\sigma$). Therefore, Equation (2) can also be expressed as $\Delta Y_t = \sigma I_t$. This equation shows that the change in productive capacity is the product of capital productivity (σ) and investment. As such, it reveals the productivity effect.

3. Investment is induced by output growth along with entrepreneurial confidence. The latter is adversely affected by "Junking" which means the untimely loss of capital value due to the unprofitable operation of older facilities. This may be due to the shortage of labour or invention of new products or labour-saving inventions. This assumption can be shown by the relation.

$$\frac{I_{(t)}}{Y_{(t-1)}} = G\left[\frac{Y_{(t-1)}Y_{(t-2)}}{Y_{(t-2)}Y_{(t-3)}} \cdot d_{(t)}\right]$$

where G is an increasing function of the rate of output acceleration, but a decreasing function of the "Junking ratio" $d_{(t)}$.

If junking ratio is zero, then investment increases at the same rate as output.

4. Employment depends upon the 'utilisation ratio' expressed as the ratio between actual output and productive capacity. It may be expressed as

$$\frac{N_{(t)}}{L_{(t)}} \cdot H \cdot \left[\frac{Y_{(t)}^{d}}{Y_{(t)}^{s}} \dots\right], H > 0$$

Here, N refers to employment and L to the labour force. H is the employment coefficient, Y^d the actual output and the productive capacity, (t) being the time period. This equation explains that the ratio of employment to labour force is determined by employment coefficient (H) and the ratio of output to productivity. The dots are meant to indicate the existence of other determinants of the employment ratio. If we assume that the employment coefficient takes the maximum value of unity (i.e., H = I), then $Y^d_{(t)} = Y^s_{(t)}$.

5. Past as well as present investment can generate productive capacity at a given ratio. But due to managerial miscalculation, the new investment projects will cause untimely demise of old project and plants. If "junking" exists, it would dampen the productivity of investment. This assumption is considered the central theme of Domar's model. In the form of notations, it can be expressed as

$$K_{(t)} - K_{(t-1)} = I_{(t)} - [d_{(t)} K_{(t)}]$$

where K is capital, I shows investment, $d_{(t)}$, $K_{(t)}$ is the amount of capital junked, and $d_{(t)}$ is the junking ratio.

Domar viewed growth from the demand as well as the supply side. Investment on the one side increases productive capacity and on the other generates income. Balancing of the two sides provides the solution for steady growth. The following symbols are used in Domar's model:

-   Y_d = level of net national income or level of effective demand at full employment (demand side)

-   Y_s = level of productive capacity or supply at full-employment level (supply side)

-   K = real capital

-   I = net investment which results in the increase of real capital, i.e., $\Delta K \alpha$ = marginal propensity to save, which is the reciprocal of multiplier. σ = (sigma) is productivity of capital or of net investment.

The demand side of the long-term effect of investment can be summarised through the following relation. This relation is a simple application of Keynes' investment multiplier.

$$Y_d = 1/\sigma \,.\, I$$

This relation tells us that: (i) the level of effective demand (Y_d) is directly related to the level of investment through the multiplier whose value is given by $1/\alpha$. Any increase in the level of investment will directly increase the level of effective demand and *vice versa*. (ii) The effective demand is inversely related to the marginal propensity to save (α). Any increase in marginal propensity to save (σ) will decrease the level of effective demand and *vice versa*.

The supply side of the economy in the Domar model is shown through the relation

$$Y_s = \sigma K$$

This relation explains that the supply of output (Y_s) at full employment depends upon two factors: productive capacity of capital (σ) and the amount of real capital (K). Any increase or decrease in one of these two factors will change the supply of output. If the productivity of capital (σ) increases, that would favourably affect the economy's supply. Similar is the effect of the change in the real capital K on the supply of output.

For the economy's long-term equilibrium, the demand Y_d and supply Y_s sides should be equal. Therefore, we can write:

$$Y_d = Y_s$$

$$\text{or} \quad \frac{I}{\alpha} = \sigma K$$

$$\Rightarrow I = \alpha \sigma K$$

This relation tells us that steady growth is possible when investment over a period of time equals the product of saving-income ratio, capital productivity and capital stock.

The demand and the supply equation in the incremental form can be written as follows. The demand side is

$$\Delta Y_d = \Delta I/\alpha \qquad \qquad …(8.2)$$

But the increment has not been shown in α because it is a constant in terms of the assumptions. Since $1/\alpha$ is nothing, but α and ΔI leads to ΔK, we can write the supply relation as follows:

$$\Delta Y_s = \sigma \Delta K$$

This equation shows that a change in the supply of output (ΔY_s) can be expressed as the product of the change in real capital (ΔK) and the productivity of capital (σ). Substituting the value of ΔK as I in the above equation, we get the supply side of the economy as

$$\Delta Y_s = \sigma I \qquad \qquad …(8.3)$$

From equations (8.2) and (8.3), we can derive the condition for steady growth. Using equations (8.2) and (8.3), we get

$$\Delta Y_d = \Delta Y_s$$

$$\Rightarrow \frac{\Delta I}{\alpha} = s.I$$

and by cross multiplying, we get

$$\frac{\Delta I}{I} = \alpha \, . \, \sigma$$

$$\text{or} \quad \frac{\Delta Y}{Y} = \alpha \, . \, \sigma \qquad \qquad …(8.4)$$

Equation (8.4) explains that if steady growth is to be maintained, the income growth rate $\Delta Y/Y$ should be equal to the product of marginal propensity to save (α) and the productivity of capital (σ). In the words of K.K. Kurihara, "It is an increase in productive capacity (ΔY_s) due to increment of real capital (ΔC) which must be matched by an equal increase in effective demand (ΔY_d) due to an increment of investment (ΔI), if a growing economy with an expanding stock of capital is to maintain continuous full-employment."

Thus, income and investment must grow at the annual rate of 3% if steady growth rate is to be maintained.

Doman's equation of steady growth can be explained with the help of a numerical example

1. Suppose $\sigma = 5\%$ and $\sigma = 6\%$ then

$$\frac{\Delta z}{I} = \frac{50}{100} \times \frac{6}{100} = \frac{3}{100} \text{ or } 3\%$$

2. Let $\sigma = 25$ percent per year, $\propto = 12\%$ and $y = 150$ billion dollars per year. If full employment is to be maintained an amount equal to $150 \times 12/100 = 18$ billion dollars should be invested. This will raise productive capacity by the amount invested or times i.e., by $150 \times 12/100 = 4.5$ billion dollars, and the national income will have to rise by the same amount but the relative rise in income will equal the absolute increase divided by the income itself, i.e.,

$$150 \times \frac{\dfrac{12}{100} \times \dfrac{25}{100}}{150} = \frac{12}{100} \times \frac{25}{100} = \alpha\,\sigma = 3\%$$

Analysis of Disequilibrium

Disequilibrium (non-steady state) prevails:

(i)   When $\dfrac{\Delta I}{I}$ or $\dfrac{\Delta Y}{Y}$ $> \alpha \,.\, \sigma$

(ii) When $\dfrac{\Delta I}{I}$ or $\dfrac{\Delta Y}{Y} < \alpha \,.\, \sigma$

Under the first situation, long-term inflation would appear in the economy because the higher growth rate of income will provide greater purchasing power to the people and the productive capacity ($\sigma\alpha$) would not be able to cope with the increased level of income. The first situation of disequilibrium will, therefore, create inflation in the economy.

The second situation, under which growth rate of income or investment is lagging behind the productive capacity, will result in overproduction. The reduced growth rate of income will put a constraint on the purchasing power of the people, thereby reducing the level of demand and resulting in overproduction. This is the situation in which there would be secular stagnation. We have thus arrived at the same conclusion of instability of steady growth which we had derived from the Harrod model.

Summary of Main Points

The main points of the Harrod-Domar analysis are summarised below:

1. Investment is the central variable of stable growth and it plays a double role; on one hand, it generates income and on the other, it creates productive capacity.

2. The increased capacity arising from investment can result in greater output or greater unemployment depending on the behaviour of income.

3. Conditions concerning the behaviour of income can be expressed in terms of growth rates, i.e., G, G_w and G_n, and equality between the three growth rates can ensure full employment of labour and full-utilisation of capital stock.

4. These conditions, however, specify only a steady-state growth. The actual growth rate may differ from the warranted growth rate. If the actual growth rate is greater than the warranted rate of growth, the economy will experience cumulative inflation. If the actual growth rate is less than the warranted growth rate, the economy will slide towards cumulative inflation. If the actual growth rate is less than the warranted growth rate, the economy will slide towards cumulative deflation.

5. Business cycles are viewed as deviations from the path of steady growth. These deviations cannot go on working indefinitely. These are constrained by upper and lower limits, the 'full employment ceiling' acts as an upper limit and effective demand composed of autonomous investment and consumption acts as the lower limit. The actual growth rate fluctuates between these two limits.

Diagrammatic Representation

Refer to Fig. 8.2 where Income is shown on the horizontal axis, Saving and Investment on vertical axis. The line S(Y) drawn through the origin shows the levels of saving corresponding to different levels of income. The slope of this line (tangent α) measures the average and marginal propensity to save. The slopes of lines Y_0I_0, Y_1I_1 and Y_2I_2 measures the acceleration coefficient v which remains constant at each income level of Y_0, Y_1, and Y_2.

At the initial income level of Y_0, the saving is S_0Y_0. When this saving is invested, income rises from Y_0 to Y_1. This higher level of income increases saving to S_1Y_1. When this amount of saving is reinvested, it will further raise the level of income to Y_2. The higher level of income will again raise saving to S_2Y_2. This process of rise in income, saving and investment shows the acceleration effect on the growth of output.
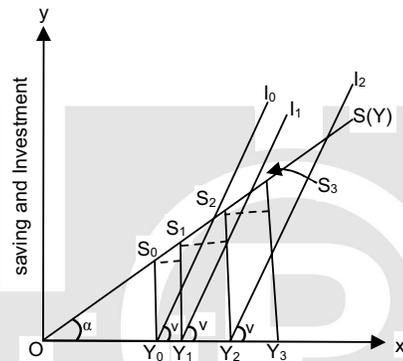
Fig. 8.2: The Harrod Model: Acceleration and Growth
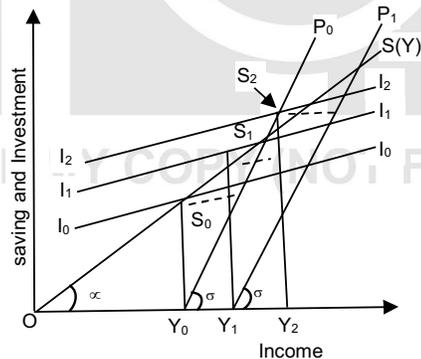
Fig. 8.3: Domar Model: Productivity of capital and Growth

Now, we give the diagrammatic exposition of the Harrod's model with the help of Fig. 8.3.

In this figure, income is shown on horizontal axis, saving and investment on vertical axis. The line S(Y) passing through the origin indicates the level of saving corresponding to different levels of income. I_0I_0, I_1I_1 and I_2I_2 are the various levels of investment. Y_0P_0 and Y_1P_1 measure the productivity of capital corresponding to different levels of investment.

The lines Y_0P_0 and Y_1P_1 are drawn parallel so as to show that productivity of capital remains unchanged. This diagram shows that the level of income is determined by the forces of saving and

investment. The level of income Y_0 is determined by the intersection of saving line S(Y) and the investment line I_0I_0.

At the level of income Y_0, the saving is Y_0S_0. When the saving Y_0S_0 is invested, it will increase the income level from OY_0 to OY_1. The productive capacity will also rise correspondingly. The extent of the income increase depends upon the productivity of capital, which is measured by the slope of the line $Y_0P_0(\alpha)$.

Higher the level of income, higher is the productive capacity. Similarly, when the level of income is OY_1, the level of saving is S_1Y_1. With investment of S_1Y_1, income will further rise to the level Y_2. This increase in income means expansion of purchasing power of the economy. But the coefficient of capital productivity would remain constant, this being an important assumption of Domar's model.

Example:

1. China's GDP increased from \$ 997.5 billion in 1999 to \$ 1,076.9 billion in 2000. Calculate the growth rate of china's GDP in 2000?

Ans.: The expression for the growth rate is year t + 1, is $g = \dfrac{Y_{t+1} - Y_t}{Y_t}$, where g represents the growth rate of GDP, it represents GDP in year t and Y_{t+1} represents GDP is year t + 1. In this example, \$ 997.5 billion corresponds to Y_t and \$ 1,076.9 corresponds to y_{+1}. Thus,

$g = \dfrac{\$1076.9 - \$997.5}{\$997.5} = 0.0796 = 8.0\%$. During the year 2000, China's GDP grew at a rate of about 8% per year.

2. India's real GDP per capita (PPP) grow at an average annual rate of 2.00% from 1960 through 1996, increasing from \$ 769 to \$ 1,546. Assuming India's GDP per capita continues growing at this average rate from 1996 through 2046, what will India's real GDP per capita equal in 2046?

Ans.: If the average growth rate over the fifty year period 1996 through 2046 equals 2.00%, then India's GDP per capita in 2046 can be expressed as $Y_{2046} = Y_{1996}(1+ g)^{50}$, where, Y_{2046} represents real GDP per capita in 2046, Y_{1996} represents real GDP per capita in 1996 and g represents the average annual growth rate. Substituting in the appropriate number.

$Y_{2046} = (\$ 1,546) (1 + 0.0200)^{50} = \$ 4, 161$, so real GDP per capita in 2046 (Y_{2046}) is predicted to be \$ 4, 161.

3. China;'s real GDP per capita (PPP, 1985 constant prices increased from \$ 564 in 1960 to \$ 2,374 in 1996. Calculate the average annual growth rate of china's real GDP per capita over the period 1960 -1996.

Ans.: Real GDP per capita in 1960, \$ 564 grew over a 36 year period to \$ 2,374. This can be expressed as \$ 2,374 = \$564 $(1 + g)^{36}$, where g is the average annual growth rate. To solve for g, divide both sides by \$ 564 : $\dfrac{\$2374}{\$564} = (1+g)^{36}$. Now, raise each side of the equation to 1/36:

$$\dfrac{(\$2374)^{1/36}}{(\$564)} = \left((1+g)^{36}\right)^{\frac{1}{36}} = 1 + g$$

So, g = $\dfrac{(\$2374)^{\frac{1}{36}}}{(\$564)} - 1 = 0.0407$.

Thus, china's real GDP per capita grew an average of about 4.1% annually.

4. From 1980 – to 1990 real GDP in India grew by 5.8% per annum, while investment averaged 23.1 percent of GDP. What was the ICOR for India between 1980 and 1990:

Ans.: The annual growth rate (g) and investment rate are given an 5.8% and 23.1% respectively. The Harrod-Domar model assumes that the investment rate equals the saving rate (s), so that saving rate equals 23.1%. To calculate the ICOR v use the Harrod-Domar equation, g $= \dfrac{s}{v}$ which can be rewritten as $v = \dfrac{s}{g}$ so, $v = \dfrac{23.1\%}{5.8\%} = \dfrac{0.231}{0.058} = 3.98 \approx 4.0$

Thus, the Incremental Capital Output Ratio (ICOR) was about 4.0

5. In Indonesia during the 1970s the Incremental Capital Output Ratio (ICOR) averaged 2.50:

   (a) Using the Harrod-Domar growth equation, what saving rate would have been required for Indonesia to achieve an aggregate growth rate of 8 percent per annum?

   (b) With the same IWRI what growth target could be achieved with a saving rate of 27 percent?

   (c) If there is a large increase in the saving rate and therefore a large increase in the amount of new capital formation, is the ICOR likely to rise, tall or remain the same? Explain.

Ans.: (a) Given that ICOR (v) equals 2.50 and the targeted annual growth rate is 8% (g), the Harrod-Domar equation can be a rewritten s = gv is order to solve for the required saving rates (s). Thus, s = (0.08) (2.50) = 0.20 = 20%. In order to achieve an 8%. growth rate. Indonesia would have to have a saving rate of 20%.

   (b) Given the ICOR (v) equals 2.5 and the saving rate (s) equals 27%, the Harrod-Domar equation can be used to determine the expected annual growth rate (g): $g = \dfrac{s}{v}$. Substituting in the given values for s and v : g = $\dfrac{0.27}{2.5}$ = 0.108 = 10.8%. If the saving rate were 27% the growth rate would be 10.8% per year.

   (c) The Herrod-Domar model assume the ICOR(v) remains constant. Thus, according to the Harrod-Domar model, an increase with saving rate has no effect on the ICOR. One might ----- however that the model's assumption is wrong and that a large increase is savings might result is some decline is the productivity of capital, hence a higher ICOR.

6. The government of a pour developing country from that a political cupheaval will occur unless the growth rate is at least 4 percent per annum. The ICOR and the saving rate are projected to be v = 50 and s = 14 percent respectively:

   (a) Show that 4 percent growth cannot be achieved under these circumstances.

   (b) With the saving sale an given, what ICOR would be required to achieve the 4 percent growth target?

Ans.: (a) If the saving rate is 14% and the ICOR is 5.0 then the growth rate can be expected, to be $g = \dfrac{s}{\upsilon} = \dfrac{0.14}{5.0} = 0.028$

This falls short of the desired 4% growth rate.

(b) If the targeted growth rate is 4% and the saving rate is 14% there the ICOR must equal to $v = \dfrac{s}{g} = \dfrac{14\%}{4\%} = \dfrac{0.14}{0.04} = 3.5$

Now that if the ICOR were lower than 3.5, growth would be greater than 4% per year (assuming the saving rate remains 14%)

7. Consider an economy is which the labour force grows be 2.7% per annum, while the capital stock grew by 4% per annum. Suppose 55% of national income goes to labour and 45% to capital:

   (a) If the residual were a = *v*, what rate of growth would the economy achieve?

   (b) The country's actual rate of growth has been 4.5% per annum, which is faster than the growth rate generated by the accumulation of capital and labour stocks.

   Calculate the value of the residual (a)

Ans.: (a) If the residual (a) equals the annual labour force growth rate (g_L) equals 2.7%, the annual capital stock growth rates equals 4% (g_K) and labour's share of national income (w_L) is 55%, which capital's show of national income (w_K) is 45%, then the growth rate of national income (gr) is $g_r = a + (w_K \times g_k) + (w_L \times g_L)$

$$= a + (0.45 \times 0.04) + (0.5.5 \times 0.027) = 0.03285$$

$$= 3.3\%$$

Thus, the economy will grow at a 3.3% annual rate.

(b) In this case, the growth of the inputs (capital and labour) insufficient to explain the growth of the economy. The difference between the growth rate and the sum of the weighted factor growth rates is the residual (a). This residual can be calculated using the same equation as is part a: $4.5\% = 0.045 = g_r = a + (w_k \times g_L) + (w_L \times g_L)$

$$= a (0.45 \times 0.04) + (0.55 \times 0.027)$$

Simplifying and solving this equation for a: a = 0.01215. Thus, the residual equals 0.01215.

8. From 1970 to 1989 Singapore's growth rate averaged 8.4 percent per year. A recent growth accounting study showed that the residual accounted for only 1.2% per year of Singapore's outstanding growth performances. This growth accounting analysis used weights of 0.33 for labour and 0.67 for capital including human capital.

   (a) Singapore's labour force grew by 2.6 percent per year during this period. What can you conclude about the annual growth rate of Singapore's capital stock (including both human and physical capital)?

Ans.: (a) Given that the growth rate of income (g_r) equal 8.4% per year the residual (a) equal 1.2% the labour force growth rate (g_L) is 2.6% and the weights on the growth rates of labour (w_L) and capital (w_k) and 0.33 and 0.67, respectively, the growth accounting

equation can be solved for the growth rate of capital (g_k). The growth accounting equation is

$$g_r = a + (w_k \times g_k) + (w_L \times g_L) \, S_u, \quad 8.4\%$$

$$= 1.2\% + (0.67 \times g_k) + (0.33 \times 2.6\%)$$

Simplifying and solving g_k = 0.0947 = 9.5%. Thus, the annual growth rate of Singapore's capital stock annual growth rate of Singapore's capital stock averaged 9.5%.

(b) We are given that income grows by 8.4% annually during this period. Growth of the capital stock is responsible for $\dfrac{6.37\%}{8.4\%}$ = 0.76 = 76% of the growth rate of income.

8.2 NEO-CLASSICAL MODEL OF SOLOW

Prof. Robert M. Solow made his model an alternative to Harrod-Domar model of growth.

It ensures steady growth in the long-run period without any pitfalls. Prof. Solow assumed that Harrod-Domar's model was based on some unrealistic assumptions like fixed factor proportions, constant capital-output ratio, etc.

Solow has dropped these assumptions while formulating its model of long-run growth. Prof. Solow shows that by the introduction of the factors influencing economic growth, Harrod-Domar's Model can be rationalised and instability can be reduced to some extent.

He has shown that if technical coefficients of production are assumed to be variable, then the capital-labour ratio may adjust itself to equilibrium ratio in the course of time.

In Harrod-Domar's model of steady growth, the economic system attains a knife-edge balance of equilibrium in growth in the long-run period.

This balance is established as a result of pulls and counter-pulls exerted by natural growth rate (G_n) (which depends on the increase in labour force in the absence of technical changes) and warranted growth rate (G_w) (which depends on the saving and investment habits of household and firms).

However, the key parameter of Solow's model is the substitutability between capital and labour. Prof. Solow demonstrates in his model that, "this fundamental opposition of warranted and natural rates turns out in the end to flow from the crucial assumption that production takes place under conditions of fixed proportions."

The knife-edge balance established under Harrodian steady growth path can be destroyed by a slight change in key parameters.

Prof. Solow retains the assumptions of constant rate of reproduction and constant saving ratio, etc., and shows that substitutability between capital and labour can bring equality between warranted growth rate (G_w) and natural growth rate (G_n), and economy moves on the equilibrium path of growth.

In other words, according to Prof. Solow, the delicate balance between G_w and G_n depends upon the crucial assumption of fixed proportions in production. The knife-edge equilibrium between G_w and G_n will disappear if this assumption is removed. Solow has provided solution to twin problems of disequilibrium between G_w and G_n and the instability of capitalist system.

In short, Prof. Solow has tried to build a model of economic growth by removing the basic assumptions of fixed proportions of the Harrod-Domar model. By removing this assumption, according to Prof. Solow, Harrodian path of steady growth can be freed from instability. In this way, this model admits the possibility of factor substitution.

Assumptions

Solow's model of long-run growth is based on the following assumptions:

1. The production takes place according to the linear homogeneous production function of first degree of the form:

   Y = F(K, L)

   where Y = Output

      K = Capital stock

      L = Supply of labour force

   The above function is neo-classic in nature. There is constant returns to scale based on capital and labour substitutability and diminishing marginal productivities. The constant returns to scale means if all inputs are changed proportionately, then the output will also change proportionately. The production function can be given as

   aY = F(aK, aL)

2. The relationship between the behaviour of savings and investment in relation to changes in output. It implies that saving is the constant fraction of the level of output. In this way, Solow adopts the Harrodian assumption that investment is in direct and rigid proportion to income.

   In symbolic terms, it can be expressed as follows:

   I = dk/dt = SY

   where S = Propensity to save

      K = Capital stock, so that investment I is equal

3. The growth rate of labour force is exogenously determined. It grows at an exponential rate given by

   $L = L0e^{nt}$

   where L = Total available supply of labour

      n = Constant relative rate at which labour force grows

4. There is full employment in the economy.

5. The two factors of production are capital and labour, and they are paid according to their physical productivities.

6. Labour and capital are substitutable for each other.

7. Investment is not of depreciation and replacement charges.

8. Technical progress does not influence the productivity and efficiency of labour.

9. There is flexible system of price-wage interest.

10. Available capital stock is fully utilised.

Following these above assumptions, Prof. Solow tries to show that with variable technical coefficient, capital-labour ratio will tend to adjust itself through time towards the direction of equilibrium ratio. If the initial ratio of capital-labour ratio is more, then capital and output will grow more slowly than labour force and *vice versa.*

To achieve sustained growth, it is necessary that the investment should increase at such a rate that capital and labour grow proportionately, i.e., capital-labour ratio is maintained.

Solow's model of long-run growth can be explained in two ways:

1. Non-mathematical Explanation
2. Mathematical Explanation

1. Non-mathematical Explanation

According to Prof. Solow, for attaining long-run growth, let us assume that capital and labour both increase but capital increases at a faster rate than labour so that the capital-labour ratio is high. As the capital-labour ratio increases, the output per worker declines, and as a result, national income falls.

The savings of the community decline, and in turn investment and capital also decrease. The process of decline continues till the growth of capital becomes equal to the growth rate of labour. Consequently, capital-labour ratio and capital-output ratio remain constant, and this ratio is popularly known as "Equilibrium Ratio".

Prof. Solow has assumed technical coefficients of production to be variable, so that the capital-labour ratio may adjust itself to equilibrium ratio. If the capital-labour ratio is larger than equilibrium ratio, then that of the growth of capital and output capital would be lesser than labour force. At some time, the two ratios would be equal to each other.

In other words, this is the steady growth. According to Prof. Solow as there is the steady growth, there is a tendency to the equilibrium path. It must be noted here that the capital-labour ratio may be either higher or lower.

Like other economies, Prof. Solow also considers that the most important feature of an underdeveloped economy is dual economy. This economy consists of two sectors—capital sector or industrial sector, and labour sector or agricultural sector. In industrial sector, the rate of accumulation of capital is more than the rate of absorption of labour.

With the help of variable technical coefficients, many employment opportunities can be created. In agricultural sector, real wages and productivity per worker is low. To achieve sustained growth, the capital-labour ratio must be high and underdeveloped economies must follow Prof. Solow to attain the steady growth.

This model also exhibits the possibility of multiple equilibrium positions. The position of unstable equilibrium will arise when the rate of growth is not equal to the capital-labour ratio. There are other two stable equilibrium points with high capital-labour ratio and the other with low capital-labour ratio.

If the growth process starts with high capital-labour ratio, then the development variables will move in forward direction with faster speed and the entire system will grow with high rate of growth. On the other hand, if the growth process starts with low capital-labour ratio, then the development variables will move in forward direction with lesser speed.

To conclude the discussion, it is said that high capital-labour ratio or capital-intensive technique is very beneficial for the development and growth of capitalist sector and on the contrary, low capital-labour ratio or labour-intensive technique is beneficial for the growth of labour sector.

2. Mathematical Explanation

This model assumes the production of a single composite commodity in the economy. Its rate of production is $Y(t)$ which represents the real income of the community. A part of the output is consumed, and the rest is saved and invested somewhere.

The proportion of output saved is denoted by s. Therefore, the rate of saving would be $sY(t)$. The capital stock of the community is denoted by $K(t)$. The rate of increase in capital stock is given by dk/dt and it gives net investment.

Since investment is equal to saving, we have the following identity:

$$K = sY \qquad \qquad \qquad \qquad …(8.5)$$

Since output is produced by capital and labour, the production function is given by

$$Y = F(K, L) \qquad \qquad \qquad …(8.6)$$

Putting the value of Y from (8.6) in (8.5), we get

$$S = sF(K, L) \qquad \qquad \qquad …(8.7)$$

where L is total employment and F is functional relationship.

Equation (8.7) represents the supply side of the system. Now, we are to include demand side too. As a result of exogenous population growth, the labour force is assumed to grow at a constant rate relative to n. Thus,

$$L(t) = L_0 e^{nt} \qquad \qquad \qquad …(8.8)$$

Where L = Available supply of labour

Putting the value of L in equation (8.7), we get

$$K = sF(K, L_0 e^{nt}) \qquad \qquad …(8.9)$$

The right hand of the equation (8.8) shows the rate of growth of labour force from period o to t or it can be regarded as supply curve for labour.

"It says that the exponentially growing labour force is offered for employment completely in elastically. The labour supply curve is a vertical line, which shifts to the right in time as the labour force grows. Then the real wage rate adjusts so that all available labour is employed and the marginal productivity equation determines the wage rate which will actually rule."

If the time path of capital stock and of labour force is known, the corresponding time path of real output can be computed from the production function. Thus, the time path of real wage rate is calculated by marginal productivity equation.

The process of growth has been explained by Prof. Solow as, "At any moment of time, the available labour supply is given by (8.8) and available stock of capital is also a datum. Since the real return to factors will adjust to bring about full employment of labour and capital, we can use the production function (8.6) to find the current rate of output. Then the propensity to save tells us how much net output will be saved and invested. Hence, we know the net accumulation of capital during the current period. Added to the already accumulated stock, this gives us the capital available for the next period and the whole process can be repeated."

Possible Growth Patterns

To find out whether there is always a capital accumulation path consistent with any rate of growth of labour force, we should know the accurate shape of production function. Otherwise, we cannot find the exact solution.

For this, Solow has introduced a new variable

$$r = \frac{K}{L}$$

Where K/L = Capital-labour ratio

K = rL

But, $L = L_0 e^{nt}$

$K = r L_0 e^{nt}$

Differentiating with output to t, we get

$$\frac{dk}{dt} = nr L_0 e^{nt} + L_0 e^{nt} \frac{dr}{dt}$$

$$\Rightarrow \quad \frac{dk}{dt} = \left(nr + \frac{dr}{dt} \right) L_0 e^{nt} \qquad \qquad ...(8.10)$$

$$\Rightarrow \quad \left(nr + \frac{dr}{dt} \right) L_0 e^{nt} = sF(K, L_0 e^{nt})$$

$$\Rightarrow \quad \left(nr + \frac{dr}{dt} \right) L_0 e^{nt} = sF L_0 e^{nt} \frac{K}{L_0 e^{nt}}, 1$$

$$\text{or,} \quad nr + \frac{dr}{dt} = sF \left(\frac{K}{L_0 e^{nt}}, 1 \right)$$

Since $\dfrac{K}{L_0 e^{nt}} = r$

$$nr + \frac{dr}{dt} = sF(r, 1)$$

$$\Rightarrow \quad \frac{dr}{dt} = sF(r, 1) - nr$$

$$\text{or, } r = sF(r, 1) - nr \qquad \qquad ...(8.11)$$

where r = K/L

n = Relative share of choice of labour force (i/1)

The function F(r, 1) gives output per worker or it is the total product curve as varying amounts 'r' of capital are employed with one unit of labour. The equation (8.11) states that, "the rate of change of the capital-labour ratio as the difference of two terms, one representing the increment of capital and one the increment of labour."

The diagrammatic representation of the above growth pattern is as under:

Notes

In diagram 8.4, the line passing through origin is nr. The total productivity curve is the function of sF(r, 1) and this curve is convex to upward. The implication is that to make the output positive, it must be necessary that input must also be positive, i.e., diminishing marginal productivity of capital. At the point, of intersection, i.e., nr = sF(r, 1) and r′ = 0 when r′ = 0, then the capital-labour ratio corresponding to point r* is established.

Capital-Labour Ratio
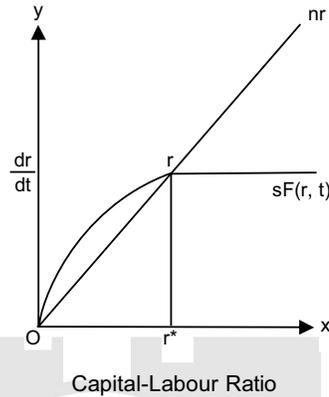
Fig. 8.4: Growth Pattern

Now, capital and labour will grow proportionately. Since Prof. Solow considers constant returns to scale, real output will grow at the same rate of n and output per head of labour, force will remain constant.

In mathematical terms, it can be explained as:

$$\frac{dr/dt}{t} = \frac{dK/dt}{K} = \frac{dL/dt}{L}$$

Now, $\dfrac{dL/dt}{K} = n$

$$\frac{dK}{dt} = sF(K, L)$$

$$\frac{dr}{dt} = r\frac{sF(K,L)}{K} - nr$$

Since it has assumed to have constant returns to scale,

$$\frac{dr}{dt} = rsF\left(1\frac{L}{K}\right) - nr$$

$$= \frac{rsF(K/L,1)}{K/L} - nr$$

$$= \frac{rsF(K/r,i)}{K/L} - nr$$

Thus, $\dfrac{dr}{dt} = sF(r, 1) - nr$

Path of Divergence

Here, we are to discuss the behaviour of capital-labour ratio, if there is divergence between r and r*. There are two cases:

(i)   When $r > r^*$

(ii)   When $r < r^*$

If $r > r^*$, then we are towards the right of intersection point. Now, $nr > sF(r, 1)$ and from equation (8.11), it is easily shown that r will decrease to r*. On the other hand, if we move towards left of the intersection point where $nr < sF(r, 1)$, $r > 0$ and r will increase towards r*. Thus, equilibrium will be established at point E and sustained growth will be achieved. Thus, the equilibrium value of r* is stable.

According to Prof. Solow, "Whatever the initial value of the capital-labour ratio, the system will develop towards a state of balanced growth at a natural rate. If the initial capital stock is below the equilibrium ratio, capital and output will grow at a faster rate than the labour force until the equilibrium ratio is approached. If the initial ratio is above the equilibrium value, capital and output will grow more slowly than the labour force. The growth of output is always intermediate between those of labour and capital."

The stability depends upon the shape of the productivity curve $sF(r, 1)$ and it is explained with the help of a diagram given below:
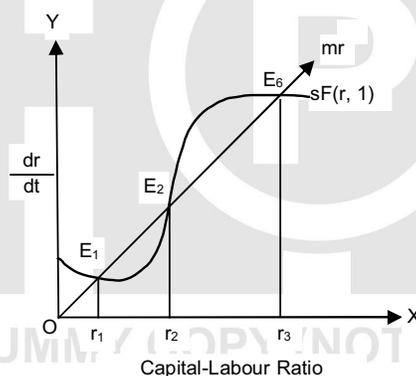

Fig. 8.5: Path of Divergence

In Fig. 8.5, the productivity curve $sF(r, 1)$ intersects the ray nr at three different points E_1, E_2, E_3. The corresponding capital-labour ratio is r_1, r_2 and r_3. The points are r_3 stable, but r_2 is not stable. Taking point r_1 first, if we move slightly towards right, $nr > sF(r, 1)$ and r is negative implying that r decreases.

Thus, it has a tendency to slip back to r_1. If we move slightly towards its left, $nr < sF(r, 1)$ and r is positive which shows that r increases and there is a tendency to move upto point r_1. Therefore, a slight movement away from r_1 creates conditions that forces a movement towards showing that r_1 is a point of stable equilibrium.

Likewise, we can show that r_3 is also a point of stable equilibrium. If we move slightly towards right of r_2, $nr > sF(r, 1)$, nr and r is positive and there is a tendency to move away from r_2.

On the other hand, if we move slightly towards left of r_2, $nr > sf(r, 1)$ so that r is negative and it has a tendency to slip downwards towards r_1. Therefore, depending upon initial capital-labour ratio,

the system will develop to balanced growth at capital-labour ratio r_1 and r_3. If the initial ratio is between 0 and r_2, the equilibrium is at r_1 and if the ratio is higher than r_2, then equilibrium is at r_3.

To conclude Solow puts, "When production takes place under neo-classical conditions of variable proportions and constant returns to scale, no simple opposition between natural and warranted rates of growth is possible. There may not be any knife-edge. The system can adjust to any given rate of growth of labour force and eventually approach a state of steady proportional expansion", i.e.,

$$\Delta K/K = \Delta L/L = \Delta Y/Y$$

Applicability to Underdeveloped Countries

Unlike Harrodian model, Solow's model also does not apply to development problem of underdeveloped countries. Most of the underdeveloped countries are either in 'pre-take-off' or 'take-off' condition and this model does not analyse any policy formulation to meet the problems of underdeveloped countries.

But certain elements from the Solow model are still valid and can be used to chalk out the problem of underdevelopment. The remarkable feature of Solow's model is that it provides deep insight into the nature and type of expansion experienced by the two sectors of underdeveloped countries.

The interpretation of underdevelopment is explained with the help of Fig. 8.6 given below.
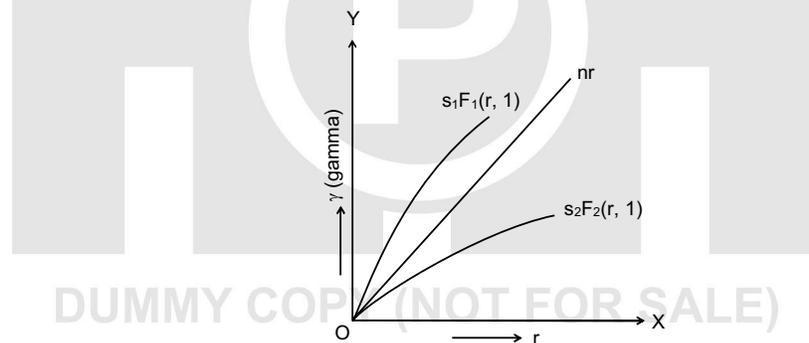
Fig. 8.6: Interpretation of Development

The line nr represents the balanced requirement line. When the warranted growth rate and natural growth rate are equal, then steady growth is achieved.

Along this path, there is full employment and unchanging capital-labour ratio. The curve represented by $s_1F_1(r, 1)$ gives productive system in terms of both output and savings. On the other hand, $s_2F_2(r, 1)$ gives unproductive system and the per capita income and savings would decline. Both the systems have low marginal productivity.

The first system can be identified by industrial sector of underdeveloped countries which tends to grow with ever-increasing intakes of capital in relation to labour. The second system conforms to the agrarian sector of underdeveloped countries. There is more labour supply due to rapid population growth. Investment is also positive.

Once the initial growth of population has occurred and land has become scarce, the real wage rate tends to be fixed at certain level, though the marginal productivity declines. The result of this is disguised unemployment.

In nutshell, we can conclude the discussion of validity of Solow's model is that there are certain elements which could be gainfully utilised for analysing the problem of underdevelopment. The phenomenon of technological dualism which is commonly prevalent in these economies can be better explained in terms of Solow's model.

Though Solow's model is basically embedded in a different setting, yet its concept of technical coefficient provides elegant and simple theoretical apparatus to solve the problems of under development.

Merits of the Model

Solow's growth model is a unique and splendid contribution to economic growth theory. It establishes the stability of the steady-state growth through a very simple and elementary adjustment mechanism.

Certainly, the analysis is definitely an improvement over Harrod-Domar model, as he succeeded in demonstrating the stability of the balanced equilibrium growth by implying neo-classical ideas. In fact, Solow' growth model marks a breakthrough in the history of economic growth.

The merits of Prof. Solow's model are undermentioned:

1. Being a pioneer of neo-classical model, Solow retains the main features of Harrod-Domar model like homogeneous capital, a proportional saving function and a given growth rate in the labour forces.

2. By introducing the possibility of substitution between labour and capital, he gives the growth process and adjustability, and gives more realistic touch.

3. He considers a continuous production function in analysing the process of growth.

4. Prof. Solow demonstrates the steady-state growth paths.

5. He successfully shunted aside all the difficulties and rigidities of modern Keynesian income analysis.

6. The long-run rate of growth is determined by an expanding labour force and technical process.

Shortcomings of the Model

1. **No study of the problem of balance between G and G_w:** Solow takes up only the problem of balance between warranted growth (G_w) and natural growth (G_n), but it does not take into account the problem of balance between warranted growth and the actual growth (G and G_w).

2. **Absence of investment function:** There is a absence of investment function in Solow's model and once it is introduced, the problem of instability will immediately reappear in the model as in the case of Harrodian model of growth.

3. **Flexibility of factor price may bring certain problems:** Prof. Solow assumed the flexibility of factor prices, but it may bring certain difficulties in the path of steady growth. For example, the rate of interest may be prevented from falling below a certain minimum level and this may in turn, prevent the capital-output ratio from rising to a level necessary for sustained growth.

4. **Unrealistic assumptions:** Solow's model is based on the unrealistic assumption that capital is homogeneous and malleable. But capital goods are highly heterogeneous and may create the problem of aggregation. In short, it is not easy to arrive at the path of steady growth when there are varieties of capital goods in the market.

5. **No study of technical progress:** This model has left the study of technological progress. He has merely treated it as an exogenous factor in the growth process. He neglects the problem of inducing technical progress through the process of learning, investment and capital accumulation.

6. **Ignores the composition of capital stock:** Another defect of Prof. Solow's model is that it totally ignores the problem of composition of capital stock and assumes capital as a homogeneous factor which is unrealistic in the dynamic world of today. Prof. Kaldor has forged a link between the two by making learning a function of investment.

8.3   NEO-KEYNESIAN MODEL OF PASSINETI

A Pasinetti Model of Savings and Growth

This model develops a two-sector growth model in which institutional investors play a significant role. A necessary and sufficient condition is established under which these investors own the entire capital stock in the long run. The dependence of the long-run growth rate on the behaviour of such investors, and the effects of a productivity increase are analysed.

In the Keynes-Kaldor-Pasinetti post-Keynesian growth model, two classes of agent, workers and capitalists, save constant proportions of their income. On a balanced growth path, the rate of profit is independent of the workers' savings propensity. Meade (1963), and Samuelson and Modigliani (1966) prove an "anti-Pasinetti" theorem which establishes the existence of an alternative balanced growth path on which pure capitalists cease to exist and all capital is owned by workers. Kaldor (1966) proposed an alternative institutional setting for post-Keynesian growth theory in which large capitalist corporations play an important role in savings and investment decisions. This kind of corporate economy is described in Marris (1964), Marris and Wood (1971), Wood (1975) and Eichner (1976, 1985). Moss (1978) extends Pasinetti's analysis to a corporate economy by dividing Kaldor's (1966) household sector into workers and financial capitalists whose income arises only from financial capital. O'Connell (1985, 1995) develops an alternative approach to the corporate economy, showing that the "anti-Pasinetti" theorem does not hold when firms reinvest a proportion of their profits. Commendatore (1999) extends the Post-Keynesian Growth Model to a corporate economy, analysing the effects of firm and shareholder behaviour. Feldíman (1928) and Mahalanobis (1953/4) analyse the effects of investment allocation on economic growth. This analysis is extended to a multi-sector model by Araujo and Teixeira (2011).

In this part, we consider a two-sector model in which institutional investors such as pension funds, unit trusts and insurance companies have an important role. It reflects Pasinetti's idea that workers must own the capital to which their savings have given rise, but also acknowledges that, in a modern capitalist economy, these savings are typically mediated by institutions such as pension funds. Dinenis and Scott (1993) argue that pension funds are a major vehicle for personal long-run saving in the UK economy. They report that such funds controlled over £ 250 billion of funds in 1989, their total net assets constituting 38% of personal sector net financial wealth. These funds owned 23% of UK equity, 21% of British government securities and 18% of British holdings of foreign equity. Apilado (1972) investigates whether pension savings in the US economy between 1955 and 1970 are

a substitute for other forms of saving. He concludes that they were in fact an addition to other forms of saving and that, via an increase in total saving, generated an increase in the growth rate. Pension funds have obligations to pay pensions, and in many jurisdictions (e.g., the UK), pensioners are allowed to withdraw a proportion of their pension pot prior to retirement. Workers' savings/consumption decisions are not explicitly modelled here, rather institutional investors are assumed to invest a proportion s < 1 of their income, where s is treated as exogenous. Van Groezen *et al*. (2007) developed a two-sector growth model with a capital-intensive commodity sector (with endogenous growth) and a labour-intensive services sector. They analyse the effects on economic growth of a switch to a more funded pension scheme. In this model, increased savings resulting from the pension reform generate higher growth in a closed economy provided capital and labour are not strong substitutes. However, the opposite is true for a small open economy. Hachon (2010) analyses the effect of the structure of pension systems on the growth rate. He contrasts "purely Beveridgian" pension systems, where every agent receives the same pension, with "purely Bismarckian" systems, where pensions depend on agents' wages. Hachon's focus is on the redistributional effects of pensions, in similar vein to a paper of Docquier and Paddison (2003).

Pasinetti was concerned to provide a normative description of the economic system, focussing on the physical requirements for reproduction. But his insights can be reinterpreted as providing a positive analysis of modern capitalism. In such an economy, savings and investment are mediated by institutional investors. So, two questions arise naturally: Will the long-run growth rate in such an economy be determined by the behaviour of institutional investors and, if so, how? Will institutional investors own the entire capital stock in the long run? What are the implications for long-run growth and capital ownership, of a one-shot productivity increase? All three questions are analysed below. A capitalist economy with institutional investors works in a complicated way, but adopting and developing Pasinetti's insights allows an analysis of these questions which is simple enough to be tractable.

Structure of the Model

In an economy with institutional investors, investment and hence growth are likely to be influenced by the decisions of such investors. But under modern capitalism, there are many high technology firms (e.g., IT, software, etc.) which present institutional investors with substantially greater problems of risk and asymmetric information than firms with less dynamic technologies (e.g., consumer durables). It is, therefore, reasonable to assume a correlation between technological level and the degree to which accumulation is financed from retained profits. We refer to capital accumulated from retained profits as "corporate capital", and that accumulated through institutional investment as "institutional capital". To capture this distinction in a two-sector model, we assume two different production sectors at opposite ends of this "technology spectrum". Sector 1 consists of high technology, capital-intensive firms which invest all their profits, and also obtain investment from outside institutional investors. It produces an output Q_1 using labour L_1 and capital K_1. Sector 2 consists of medium technology, less capital-intensive firms whose investment expenditure comes exclusively from outside institutional investors. It produces an output Q_2 using labour L_2 and capital K_2:

Total output of the economy will be denoted by

$Q = Q_1 + Q_2$

Total labour employed in the economy will be denoted by

$L = L_1 + L_2$

Total capital employed in the economy will be denoted by

$K = K_1 + K_2$

Both factors are assumed perfectly mobile, equalising wage and profit rates between the two sectors. Capital is assumed fully employed, but there may be unemployed labour in the economy. Outside institutional investors receive income based on wages (e.g., pension contributions) and from profits earned on their portion of the capital stock. They invest a proportion, s, of their income, of which a share, $1 - \theta$, goes to sector 1 and θ, goes to Sector 2. We establish conditions under which the growth rate of the economy is independent of the institutional investors' behaviour. In this case, the share of the capital stock funded from retained profits remains strictly positive. There is also a balanced growth path along which the growth rate depends on the behaviour of institutional investors. In this case, the share of the capital stock funded from retained profits disappears in the long run and the entire capital stock is owned by institutions.

Wage and Profit Rate

Sector 1 consists of high technology firms with capital-output ratio $k_1 = K_1/Q_1$ and output-labour ratio $q_1 = Q_1/L_1$. Sector 2 consists of medium technology firms with capital-output ratio $k_2 = K_2/Q_2$ and output-labour ratio $q_2 = Q_2/L_2$. It will be assumed that:

$q_1 > q_2$ and $k_1 > k_2$ ...8.12)

Together these inequalities imply that:

$K_1/L_1 > K_2/L_2$ ...(8.13)

Wage-profit frontiers can readily be derived for the two sectors. Let w denote the wage rate and r the profit rate. Then:

$$Q_1 = wL_1 + rK_1 \qquad ...(8.14)$$
$$1 = wq_1 + rk_1 \qquad ...(8.15)$$
$$\text{and } Q_2 = wL_2 + rK_2 \qquad ...(8.16)$$
$$1 = wq_2 + rk_2 \qquad ...(8.17)$$

The two wage-profit frontiers are illustrated in Fig. 8.7.
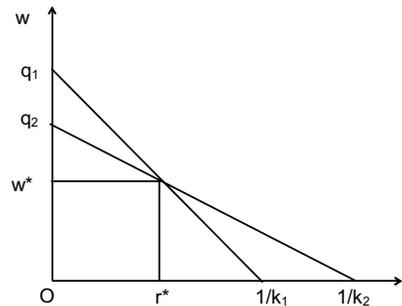

Fig. 8.7: Wage-profit Frontiers for Sectors 1 and 2

Inter-sectoral mobility of the two factors ensures that wage and profit rates are determined at the intersection of the two frontiers yielding.

$w^* = k_2 - k_1/k_2/q_1 - k_1/q_2$ ...(8.18)

and $r^* = q_2 - q_1/k_2q_2 - k_1q_1$ ...(8.19)

Note that the wage-profit frontiers do not assume full employment of labour; the availability of labour is never a constraint on growth. Capital is assumed fully employed, and both factors are assumed instantaneously and costlessly mobile between the two sectors.

Capital Accumulation

Sector 1 (high-technology) firms will be assumed to reinvest all their profits and also to receive a share θ of institutional investment. So, let $K_1 = X + Y$ where X = "corporate capital" (i.e., that portion of Sector 1 capital funded from retained profits) and Y = "institutional capital" (i.e., that portion of Sector 1 capital funded by outside institutions). Sector 2 (medium-technology) firms will be assumed to fund their capital accumulation entirely from outside institutional sources. For notational consistency, let $Z = K_2$. Institutional investors own a portion $Y + Z$ of the capital stock. Assume all capital depreciates at a rate ∂. X, Y and Z are governed by the linear dynamical system given by equations (8.20, 8.21, 8.22) below.

$$X = (r^* - \delta)X \qquad \text{...(8.20)}$$

$$Y = \left[(1-\theta)s\left[\frac{tw^*}{k_1q_1} + r^*\right] - \delta\right]Y + (1-\theta)s\left[\frac{tw^*}{k_2q_2} + r^*\right]Z + \frac{(1-\theta)stw^*}{k_1q_1}X \qquad \text{...(8.21)}$$

$$Z = \theta s\left[\frac{tw^*}{k_1q_1} + r^*\right]Y + \left[\theta s\left(\frac{tw^*}{k_2q_2} + r^*\right) - \delta\right]Z + \frac{\theta stw^*}{k_1q_1}X \qquad \text{...(8.22)}$$

It is further established that this dynamical system has two different types of steady state, depending on whether or not the condition

$$\frac{stw^*}{(1-s)}\left[\frac{1-\theta}{k_1q_1} + \frac{\theta}{k_2q_2}\right] \leq r^* \qquad \text{...(8.23)}$$

is satisfied. Now, define shares in the total capital stock:

$$x = \frac{X}{K}, \quad y = bv \; \frac{Y}{K}, \quad z = \frac{Z}{K} \qquad \text{...(8.24)}$$

We focus on a condition necessary and sufficient for the institutional investors to own the whole economy in the steady state (that is $x = 0$ or $Y + Z/K = 1$ in the steady state).

Steady State 1

It is shown in the that, the dynamical system consisting of equations (8.20, 8.21 and 8.22) converges to a steady state in which $x > 0$ and the growth rate is given by

$$g = r^* - \delta$$

if and only if Condition (8.24) is satisfied. In this steady state, the long-run growth rate does not depend on the savings behaviour of institutional investors, and they do not own the whole economy in the long run.

Steady State 2

The dynamical system consisting of equations (8.20, 8.21 and 8.22) converges to a steady state in which $x = 0$ and the growth rate is given by

$$\lambda_2 = stw * \left(\frac{1-\theta}{k_1 q_1} + \frac{\theta}{k_2 q_2} \right) + sr * -\delta \qquad \text{...(8.25)}$$

if and only if Condition (8.25) is violated. In this steady state, the long-run growth rate does depend on the savings behaviour of institutional investors. In particular, it increases with t (proportion of the wage bill received by institutional investors), s (the invested proportion of institutional investors. income) and θ (the proportion of that investment that goes to Sector 1 (high technology) firms). Moreover, in this steady state, institutional investors own the whole economy in the long run.

The Passineti Paradox

1. Passineti showed that when worker save, invest and earn a return on their investment, their saving propensity does not affect the rate of profit is a Golden Age model. In the golden age, the rate of profit is determined by the saving propensity and rate of growth of the economy alone.

Passineti adds two assumption to Kaldor's theory to derive this result. The first assumption is that workers out a return on their investment, and their rates of return on capital is the same as that capitalists. Algebraically,

$$r = \frac{P_c}{K_c} = \frac{P_w}{K_w} \qquad \text{...(8.26)}$$

The second assumption is called a "condition of equilibrium growth" by Passineti is his correspondence with the author. This holds that the share of capitalists is total capital remains constant in the golden age *viz.*,

$$\frac{S_c}{S} = \frac{K_c}{K} \qquad \text{...(8.27)}$$

In his book "Growth and Income Distribution" (1974) Passineti shows that if I/Y stay constant for a sufficiently long period of time, condition (8.27) results is a steady state model given S_c and S_w.

Given assumption (8.26) and condition (8.27), Passineti (1960-61) derives his well known distribution relation g = s, r after a large number of substitution.

Proof: Condition (8.27) gives:

$$\frac{S_e P_l}{S} = \frac{K_c}{K} \qquad \text{...((8.28)}$$

Substituting S with I, and shifting I to the RHS and K_c to LHS, we get

$$\frac{S_c P_c}{k_c} = \frac{1}{k} \qquad \text{...(8.29)}$$

Which by eq (8.12) become (since $P_c/K_c = r$ and $I/K = g$)

$$S_c^{\ r} = g \qquad \text{...(8.30)}$$

Which is Passineti for fundamental equation

The Passineti Model of Profit and Growth

Passineti builds his model on the following assumption:

1. The national income (Y) consists of wages (w) and profits

2. Wages an distribution to workers is proportion to the amount labour they contribute and profit are distributed to capitalists is proportion to the capital they own.

3. There is full employment.

4. Each class saves a fixed proportion of its income and the capitalist propensity to save (Sc) is greater than of worker.

Following these assumption, the national income identify is

$$y = W + P$$

$$P = P_1 + P_w \qquad \qquad ...(8.31)$$

$$y = w + P_w + P_c$$

P_w, P_c – profit occurring to capitalist and workers, respectively.

The saving functions of workers and capitalists are

$$S_w = S_w(w + P_w) \qquad \qquad ...(8.32)$$

or $\qquad S_c = S_c \, P_c$

So that $\qquad S_c = S_w(w + P_w) + S_c \, P_c \qquad \qquad ...(8.33)$

We know that $\quad I = S$

$$I = S_w(w + P) + S_c \, P_c$$

$$Y = w + P_w + P_c$$

$W + P_w = Y - P_c$

$$\begin{aligned} I \; &= S_w \, (y - P_c) + S_c \, P_c \qquad \qquad [\therefore (W + Pw) = (y - PU)] \\ &= S_w Y - S_w \, P_c + S_c \, P_c \\ &= S_w \, Y + (S_c - S_w) \, P_c \qquad \qquad ...(8.34) \end{aligned}$$

The ratio of investment to national income:

$$\frac{I}{Y} = \frac{Sw^y + (S_c - S_w) P_c}{y} \qquad \qquad ...(8.35)$$

$$\frac{g}{Y} = sw + \frac{P_c}{Y}(S_c - S_w)$$

$$\frac{P_c}{Y}(S_c - S_w) = \frac{1}{Y} - S_w$$

$$\frac{P_c}{Y} = \frac{S}{S_c - S_w} = \frac{1}{Y} - \frac{S_w}{S_c - S_w} \qquad \qquad ...(8.36)$$

The above expression explain the distribution of income between capitalists and workers: similarly, we can derive the ratio of investment total capital

$$\frac{1}{k} = \frac{Sw^y + (S_c - S_w)P_c}{k}$$

Or, $\qquad \dfrac{1}{k} = S_w \dfrac{Y}{K} = \dfrac{P_c}{K}(S_c - S_w)$

Or, $\dfrac{P_c}{k}(S_c - S_w) = \dfrac{1}{k} - sw\dfrac{y}{k}$

Or, $\dfrac{P_c}{k} = \dfrac{1}{S_c - S_w} \times \dfrac{1}{k} - \dfrac{S_w}{S_c - S_w} \times \dfrac{y}{k}$ ...(8.37)

The expression (8.36) and (8.37) refer to the part of profit which due to capitalists alone. The distribution of income between profits and wages is obtained by adding the share of workers profit into income P_w/Y to both sides of equation. This equation simply represents the share of capitalist profit is national income. The distribution of income between profits and wages can be expressed

$\dfrac{P}{Y} = \dfrac{P_c}{Y} + \dfrac{P_w}{Y}$ ...(8.38)

The equation (8.37) represents the ratio of capitalists profit to total capital and not the ratio of total profits to total capital. The rate of profit is calculated by adding worker's profit into capital P_w/k to both sides of equation (8.37)

$\dfrac{P}{k} = \dfrac{P_c}{k} + \dfrac{P_w}{k}$ ...(8.39)

According to Passineti, then is a fundamental relation profits and saving. The profits are distributed is proportion to the savings is the long run. They are same for both the workers and the capitalists. Thus,

$\dfrac{P_w}{S_w} = \dfrac{P_c}{S_c}$ ...(8.40)

This is based on the principle that profits are distributed is proportion to ownership of capital. To calculate the real value of the ration of profits to savings, are substitute the value of saving function is (8.40).

$\dfrac{P_w}{S_w(w + P_w)} = \dfrac{P_c}{S_c P_c}$

$S_w = S_w(W + P_w)$ and $S = S_c P_c$

$S_w = (w + P_w) = S_c P_c$

In the long run, when workers save, they receive an amount of profits (P_w) which make their total savings exactly equal to the amount that capitalist would have saved out of worker's profit these profits remained to them. This means that the rates of profit is indeterminate on the part of woman.

There is a close relation between savings and profit in case of capitalist because their saving come out of profits. Thus, for any given S_c, there is only one proportional relation between profit and savings which make $\dfrac{P_c}{P_c} = \dfrac{P_c}{s}$. Thus proportional relation can be nothing but Sc, which will determine the ratio of profits to savings for all the saving groups and a result, the income distribution between profit and wages for the whole system of the economy.

Further, the amount of investment is determined by technical progress and population growth to maintain full employment over time. In this case, there is only are equilibrium rate of profit which is determined by the natural rate of growth divided by capitalist propensity to save this can be written as:

$$\frac{P}{k} = \frac{n}{s} \qquad \qquad ...(8.41)$$

It is only this rate of profit (P/k) that keeps the system on the dynamic path of full employment. This condition for stability is such a system, where full employment investments are carried out and prices are flexible w.r.t. wages is $S_c > o^2$.

Reformulation of the Model:

Which reformulating the model equations (8.31), (8.32) and (8.33) should be considered and further derive an identify that explains total profit are divided between profits accruing to capitalists and worker.

The new equation depicting profit function then becomes $P = P_c + P_w$ ...(8.42)

and new equation depicting saving function then become $S_w = S_w (w + P_v)$ ...(8.43)

Units equations (8.42) and (8.43) a new set of questions can be derived an formulated below:

$$\frac{P}{k} = \frac{1}{S_c - S_w} \frac{1}{k} - \frac{S_w}{S_c - S_w} \frac{y}{k} + \left[\frac{S_w S_c}{S_c S_w} \frac{Y}{1} \frac{S_w}{S_c - S_w} \right] \qquad ...(8.44)$$

$$\text{and} \quad \frac{P}{Y} = \frac{1}{S_c - S_w} - \frac{1}{Y} - \frac{Sw}{S_c - S_w} + \left[\frac{S_w S_c}{S_c S_w} \cdot \frac{k}{1} - \frac{S_w}{S_c S_w} \cdot \frac{k}{y} \right] \qquad ...(8.45)$$

8.4 TWO-SECTOR MODEL OF UZAWA

Two-sector extensions of the Solow-Swan growth model were introduced by Hirofumi Uzawa (1961, 1963), James E. Meade (1961) and Mordecai Kurz (1963). This led to an explosion of research in the 1960s, conducted primarily in the Review of Economic Studies, on the two-sector growth model.

Equations of the Model

Before we explain the model, following are its basic equations when the subscripts 1 and 2 demote the capital goods sector and the consumption goods sector respectively.

$L = L_o e^{nt}$ ...(8.46)

Equation (8.46) express the assumption of a constant proportionate growth rate of population.

$K = Y_1 - \mu^k$ ...(8.47)

It defines the set increase is the total capital stock (k) at only moment of time, given by the output of the capital sector (Y_t) minus depreciation (μ) which is assumed to be proportional to the existing stock of capital.

$Y_2 = f_2 (K_2, L_2)$ ...(8.48)

Equation (8.48) express the production function of the consumption sector where its output depends on quantities of capital and labour employed

$Y_1 = f_1 (k_1, L_1)$ ...(8.49)

Similarly, equation (8.49) express the production function of the capital sector.

The production function in equation (8.48) and (8.49) are assumed to be well-behaved because they show constant returns to scale with positive and decreasing marginal productivities.

$$Y = Y_2 + P^Y 1 \qquad \qquad ...(8.50)$$

Equation (8.50) defines GNP (y) measured is terms of the consumption goods.

Y_2 and P as the price of the capital goods, Y_1, is terms of the consumption goods.

$$K = K_1 + K_L \qquad \qquad ...(8.51)$$

$$L = L_1 + L_2 \qquad \qquad ...(8.52)$$

Equations (8.51) and (8.52) express full employment condition of capital and labour is sectors 1 and 2 respectively.

$$w = \frac{\delta F_2}{\delta L_2} = P \frac{\delta F_1}{\delta L_1} \qquad \qquad ...(8.53)$$

Equations (8.53) determines the wages rate (w) which is equal to the value of the marginal product of labour of labour in both sectors under perfect competition.

$$r = \frac{\delta F_2}{\delta k_2} = P \frac{\delta F_1}{\delta k_2} \qquad \qquad ...(8.54)$$

Similarly, equation (8.54) determine the ------ (r) which is equal to the value of the marginal product of capital is both sectors under perfect competition.

$$P^Y = S^Y \qquad \qquad ...(8.55)$$

Equation (8.55) shows investment saving ex-ante equality on the assumption that a constant fraction is saved out of GVP (Y) which is automatically invested.

At any moment of time, the labour force is given exogeneousty and the capital stock is given as an outcome of post accumulation so that equations (8.48) to (8.55) determine the short-run equilibrium and equations (8.46) and (8.47) determine the long-run equilibrium or the path of growth equilibrium concerned with the short-run equilibrium.

To explain the workers of the model, we first define the following derived variables:

$$k = \frac{k}{L}, \qquad y = \frac{Y}{L}, \qquad w = \frac{w}{r},$$

$$k_1 = \frac{k_1}{L_1}, \qquad y_1 = \frac{Y_1}{L}, \qquad L_1 = \frac{L_1}{L},$$

and $n = \dfrac{\Delta L}{L}$ i.e, Labour supply growing geometrically at rate n.

On the assumption of the homogeneous production functions of the first degree, equations (8.48) and (8.49) can be expressed as

$$Y_1 = L_1 f_1 \left(\frac{k_1}{L_1}, 1 \right) = L, f, \left(\frac{k_1}{L_1} \right)$$

So that $\dfrac{Y_1}{L} = \dfrac{L_1}{L} f_1 \left(\dfrac{k_1}{L_1} \right)$

Substituting the derived variables y_1, l and k, is the above equation.

$Y_1 = f_1(k_1) l_1$

$$\therefore \quad \frac{\delta F_1}{\delta K_1} = f_1(k_1)$$

From the assumption made on f_1, we have

$f_1(k_1) > 0, \quad f_1(k_1) > 0, \quad f^{11}(k_1) > 0$ and for both $k_1 > 0$

Basic Set-up

Hirofumi Uzawa's (1961, 1963) two-sector growth model considers a Solow-Swan type of Growth Model with two produced commodities, a consumer good and an investment good. Both these goods are produced with capital and labour. So, we have two outputs and two inputs, of which the most interesting feature is that one of the outputs is also an input. To use the old Hicksian analogy, in the Uzawa two-sector model, we are using labour and tractors to make corn and tractors. For the following exposition, we have benefited particularly from Burmeister and Dobell (1970), and Siglitz and Uzawa (1970).

Let us follow the basic set-up of the Uzawa two-sector model. We begin with the following definitions:

Y_c = output of consumer good

L_c = labour used in consumer good sector

K_c = capital used in consumer good sector

Y_i = output of investment good

p = price of investment good (in terms of consumer good)

L_i = labour used in investment good sector

K_i = capital used in investment good sector

Y = total output of economy

L = total supply of labour

K = total supply of capital

w = return to labour (wages)

r = return to capital (profit/interest)

The principal equations of the two-sector model can thus be set out as follows:

$Y_c = F_c(K_c, L_c)$ – consumer sector production function ...(8.56)

$Y_i = F_i(K_i, L_i)$ – investment sector production function ...(8.57)

$Y = Y_c + pY_i$ – aggregate output ...(8.58)

$L_c + L_i = L$ – labour market equilibrium ...(8.59)

$K_c + K_i = K$ – capital market equilibrium ...(8.60)

$w = dY_c/dL_c = p_i^{1/2} \cdot (dY_i/dL_i)$ – labour market prices ...(8.61)

$r = dY_c/dK_c = p_i^{1/2} \cdot (dY_i/dK_i)$ – capital market prices ...(8.62)

$g_L = n$ – labour supply growth ...(8.63)

$g_K = Y_i/K$ – capital supply growth ...(8.64)

These equations should be self-evident. The consumer goods sector and the investment goods sector each use both capital and labour to produce their output. We capture this with equations (8.56) and (8.57), where $F_c(\ddot{\imath}^{1/2}\cdot)$ is the consumer goods industry production function and $F_i(\ddot{\imath}^{1/2}\cdot)$ the investment goods industry production function. Both production functions $F_c(\ddot{\imath}^{1/2}\cdot)$ and $F_i(\ddot{\imath}^{1/2}\cdot)$ are nicely Neo-classical, in the sense of exhibiting constant returns to scale, continuous technical substitution, diminishing marginal productivities to the factors, etc.

Equation (3) is merely the definition of aggregate output, expressed in terms of the consumer good. Equations (4) and (5) are also self-evident: the market demand for labour is $L_c + L_i$ and the market demand for capital is $K_c + K_i$. As L and K are the respective supplies, then equations (4) and (5) are merely the factor markets equilibrium conditions so that demand equals supply in each market.

Now, we assume no barriers competition in the factor markets, so that there is free movement of labour and capital across sectors. This implies that the wage rate (w) and the profit rate (r) must be the same in both the consumer goods and investment goods industry. Neo-classical economic theory tells us that the marginal productivity schedules for each factor in each industry form those industries' demand functions for the factors. As such, in labour market equilibrium, the return to labour (w) must be equal to the marginal product of labour in the consumer goods sector (dY_c/dL_i) and the marginal product of labour in the investment goods sector $p_i^{1/2}(dY_i/dL_i)$. This is equation (6). Equation (7) asserts the analogous condition in capital market equilibrium, i.e., the rate of return on capital (r) is equal to the marginal product of capital in both sectors.

Finally, as the investment goods industry produces all the new capital goods in the economy, then, ignoring depreciation, we can define the change in the total stock of capital as that sector's output, i.e., $dK/dt = Y_i$, so the growth rate of capital is $g_K = (dK/dt)/K = Y_i/K$, which is equation (8). Labour supply is assumed to grow exogenously at the exponential rate n, thus the growth rate of labour is $g_L = (dL/dt)/L = n$, which is equation (9).

We would now like to express everything in intensive form, i.e., in *per capita* or per labour unit terms. This gets a bit tricky. But defining:

1. $y_c = Y_c/L$

2. $\lambda_c = L_c/L$

3. $k_c = K_c/L_c$

4. $||f_c(k_c) = f_c(K_c, L_c)/L_c$

5. $y_i = Y_i/L$

6. $\lambda_i = L_i/L$

7. $k_i = K_i/L_i$

8. $||f_i(k_i) = f_i(K_i, L_i)/L_i$

9. $y = Y/L$

10. $k = K/L$

Then equations (1) to (9) above can be converted to the following:

- $y_c = \lambda_c ||f_c(k_c)$ – consumer sector intensive production function   ...(1||′)

- $_i = \lambda_i ||f_i(k_i)$ – investment sector intensive production function   ...(2||′)

- $y = y_c + py_i$ – aggregate output per capita   ...(3||′)

- $\lambda_c + \lambda_i = 1$ – labour market equilibrium   ...(4| |′)
- $\lambda_c k_c + \lambda_i k_i = k$ – capital market equilibrium   ...(5| |′)
- $w = |\,|f_c - k_c|\,|f_c|\,|' = p_i^{\frac{1}{2}}(\,|\,|f_i - k_i|\,|f_i|\,|'$ – labour market prices   ...(6| |′)
- $r = |\,|f_c|\,|' = p_i^{\frac{1}{2}}|f_i|\,|'$ – capital market prices   ...(7| |′)
- $g_L = n$ – labour supply growth   ...(8| |′)
- $g_K = y_i/k$ – capital supply growth   ...(9| |′)

Equations (1| |′) and (2| |′) are the intensive production functions. These are derived as follows. Recall from (1) that

$Y_c = f_c(K_c, L_c),$

then dividing through by L_c, we obtain

$Y_c/L_c = f_c(K_c/L_c, 1) = |\,|f_c(k_c)$

But $Y_c/L_c = (Y_c/L)(L/L_c) = y_c/\lambda_c$

Thus, $Y_c = \lambda_c|\,|f_c(K_c)$, which is (1| |′).

The transformation from (2) to (2| |′) follows a similar procedure.

Each of these intensive production functions have simple properties. For instance, their first derivatives are the marginal product of capital, i.e.,

$|\,|\partial F_c/dK_i = |\,|f_c|\,|'(k_c)$ and $|\,|\partial F_i/dK_i = |\,|f_i|\,|'(k_i)$

So, diminishing marginal productivity implies

$|\,|f_c|\,|'|\,|'(k_c) < 0$ and $|\,|f_i|\,|(k_i) < 0.$

The production functions also fulfill the famous "Inada conditions", formulated by Ken-Ichi Inada (1963). Specifically:

$|\,|f_c(0) = 0, |\,|f_c(|\,|\infty) = |\,|\infty$

$|\,|f_c|\,|'(0) = |\,|\infty, |\,|f_c|\,|'\,(|\,|\infty) = 0$

for the intensive production function for the consumption good. The equivalent Inada conditions apply to the intensive production function for the investment good:

$|\,|f_i(0) = 0, f_i(|\,|\infty) = |\,|\infty$

$|\,|f_i|\,|'(0) = |\,|\infty|\,|, |\,|f_i|\,|'(|\,|\infty) = 0$

Equations (3| |′) and (4| |′) are obtained merely by dividing (3) and (4) by L. Equation (5| |′) is obtained by dividing (5) by L, which yields

$K_c/L + K_i/L = K/L = k$

But $K_c/L = (K_c/L_c)(L_c/L) = k_c\lambda_c$ and $K_i/L = (K_i/L_i)(L_i/L) = k_i\lambda_i$

So, $\lambda_c k_c + \lambda_i k_i = k$, as we have in (5| |′).

Equations (6| |′) and (7| |′) use Euler's theorem. Now, it is a simple matter to show that

$dY_c/dK_c = |\,|f_c|\,|'(k_c)$ and $dY_i/dK_i = |\,|f_i|\,|'(k_i)$

So, the competitive condition in (7) is converted to

$r = |\,|f_c|\,|' = p_i^{\frac{1}{2}}|\,|f_i|\,|'$

Notes

By constant returns to scale, we know from Euler's theorem that

$Y_c = (dY_c/dK)\cdot K + (dY_c/dL_c)\cdot L_c$

Thus, dividing through by L and rearranging

$dY_c/dL_c = (Y_c/L) - (K/L)\cdot(dY_c/dK) = y_c - k\cdot|f_c|'$

The corresponding transformation can be done for dY_i/dL_i. This is how we convert (6) to (6|′). Finally, equation (9|′) is obtained simply by multiplying (9) through by

$1 = L/L$

So, $g_K = (Y_i/L)/(K/L) = y_i/k$

Now, following Uzawa's notation, let us define ω (omega) as the wage-profit ratio, i.e., ω = w/r. Thus, combining equations (6|′) and (7|′):

$\omega = w/r = [|f_c(k_c) - k_c\cdot|f_c|']/|f_c|' = [|f_i(k_i) - k_c\cdot|f_i|']/|f_i|'$

or simply:

$\omega = (|f_c(k_c)/|f_c|') - k_c = (|f_i(k_i)/|f_i|') - k_i$

Now, notice that:

$d\omega/k_c = -|f_c|'|'\frac{1}{2}\cdot|f_c(k_c)\}/(|f_c|'(k_c))^2 > 0$

$d\omega/k_i = -|f_i|'|'\frac{1}{2}\cdot|f_i(k_i)\}/(|f_i|'(k_i))^2 > 0$

Thus, ω is positively related to k_c and k_i. It is not difficult to see that these are monotonic relationships. Consequently, we can define the functions:

$k_c = k_c(\omega)$ where $k_c|' = (|f_c|')^2/(|f_c|'|'\frac{1}{2}\cdot|f_c) > 0$

$k_i = k_i(\omega)$ where $k_i|' = (|f_i|')^2/(|f_i|'|'\frac{1}{2}\cdot|f_i) > 0$

which will be used extensively as they will form the boundaries of our equilibrium path.

The growth story can be quickly told. At steady-state, the capital-labour ratio k must be constant.

As $k = K/L$

then $g_k = g_K - g_L$

so, using our expression for g_K and g_L

$(dk/dt)/k = y_i/k - n$

so, $dk/dt = y_i - nk$

which is our fundamental differential equation. So, we have a steady-state where dk/dt = 0.

Of course, this is not the end, for we have yet to consider the question of macroeconomic equilibrium. Specifically, note that while we have laid out the supply of consumer and investment goods, we have said nothing so far about the demand for these outputs. As it turns out, this will depend crucially on the consumption-savings behaviour of households. Specifically, the demand for consumer goods will depend on the amount of income households consume, while the demand for investment goods will depend on the amount of savings. Now, we can follow the "Classical" economists and presume that all wages are consumed and all profits are saved (as Uzawa (1961) did); or we allow for some saving out of both wages and profits (as Uzawa (1963) allows) and we can even impose that the propensity to save out of these two categories of income is different (as Drandakis (1963) presumes).

Whatever the case, the model will not be closed until we consider the demands for outputs explicitly. This is, after all, a Neo-classical model, which means that the imputation theory should hold: output demands will determine output supplies and consequently factor market equilibrium. Causality thus runs from preferences of households to factor market equilibrium.

Working of the Model

In the Uzawa model, there are two perfectly productive factors producing capital goods is sector 1 and consumer goods is sector 2. These goods are produced under constant returns to seals. They have homogeneous production functions. There being full employment of labour and capital, a given value of (w/r) determine the division of labour force between the two sectors and their outputs (Y_2 and Y_1). Both industries make optional adjustments and there yield unit costs. Competition then sets the price ratio P (= P_2/P_1) for the two goods equal to the ration of unit costs. Thus, any a (= w/r) determines an equilibrium price ratio p write only one set of prices of goods such that no producer makes a profit or loss.

The technique of production is sector 1 is shown by the capital labour ratio k_1 since, there are constant returns to reach, the least-cost technique depends on σ (w/r). Thus, there can be only are k_1 associated with each σ and to each σ there corresponds unique price ratio, p, of the two goods.

Assuming that the producer is the two sectors plan to supply as much of each goods as is demanded at a given σ (w/r) and associated price-ratio, p, there are two types of incomes i.e., WL (wages of labour) and rk (rented on capital). A given proportion of each types of income is spent an each types of good (Y_2 and Y_1). Since, the initial capital labour ratio, K is known and the price ratio, P, is uniquely determined by σ (w/r), the ratio is which the two goods (Y_2/Y_1) an demanded is a function of a which is a Singh valued and continuous function. It follows that the ration is which capital and labour an demanded is also a function of σ (w/r). Thus σ is a unique wages rental ratio for which both equilibrium conditions are satisfied.

Monetary Equilibrium

The may normalize w and r by taking then to be non-negative and that they add up to one. This can be done as we have sum above because only their ratio σ (= w/r) matters. Normalized w and r refer to equilibrium price of wages and rentals is relation to the demands for labour (L) and capital (k) normalized w and r are called a momentary equilibrium if at these values the excess demands for labour and capital are each negatives. This is because on the assumptions of the model, the workers law take the form:

$$w (L' - L) + r (K' - K) = 0$$

Where, w and r represent wage and rental. L' and K' are demands for labour and capital and L and K their given supplier. The above equations shows that labour or capital can be is equilibrium excess supply only if its price is zero.

Since the excess demand functions of labour and capital are continuous over the normalized price space, an equilibrium exists.

Thus, w > 0, $v = 0$ is a momentary equilibrium then it is a uniquer equilibrium only if for all

w ≠ w, r ≠ r, w + r = 1, then

$$(w - w) Y_L (w, r) + (v - r) Y_k (w, r) > 0$$

Where Y_L (w, r) and Y_k (w, r) are the excess demand functions for labour and capital respectively.

Stability Proposition

The stability proposition status that with the labour supply glowing at the rate n (= $\Delta L/L$), the system will eventually approach a situation the which k = (k/L) is constant so that the capital stock is also growing at the rate n and the whole economy changes only in seals. By assuming n = $\Delta L/L$ and all rentals are stock an gross investment.

$$\Delta = rk/P, - \mu K$$

The product rental of capital is sector 1 is equal to the marginal product of capital is sector 1, fi (k_1), since under constant returns to such it depends on $k_1(= k_1/L_1)$ combining all these, we have

$$\Delta = k/k = \Delta k/k - \Delta L/L = f_i (k_l) - n - \mu$$

Where μ is depreciation of capital.

This is Uzawa's stability proposition which assets that both sides of this equation tend to zero.

To prove it, we have to prove that where k is very high, $\Delta k/k$ becomes negative and where k is very low, $\Delta k/k$, becomes positive. But the marginal product of capital $f(k_s)$ is a decreasing function of K_1. Now, we have to show that K and k_1 always make is the same direction and $f_1(k_1)$ decrease and ultimately becomes equal to or less than (n + μ). Then is a possibility that k and k_1 always more in the same direction. K_1 increase where as (= w/v) increase and so does k_2. It means that the capital labour ratio increase is each sector whenever the wage rental ratio rises. This stability proposition leads to the balanced growth path.

Condition for stability of the balanced growth equilibrium.

The stability of the equilibrium growth path is also unique an the assumption that sector -2 (consumer goods sector) is more capital intensive than sector -1 (capital goods sector). To prove this, suppose that at any σ (= w/r) ratio, $k_2/L_2 > k_1 /L_1$; so is $rK_2/wL_2 > rk_1 /wL_1$.

Under constant returns to scale and perfect competition, when or rises, the price ratio, $P(= P_2/P_1)$ increase or decrease according as the relative share of wages in factor -2 is greater or smaller than is sector -1. Thus, p falls where σ rises and *vice versa.*

From the assumption that all wages are subject on consumer goods and all rentals are saved, we have

$$w_L/r_k = P_2Y_2/P_1Y_1 \text{ and } \quad \frac{K}{L} = \frac{w}{r} = \frac{P_1Y_1}{P_2Y_2}$$

If w/r rise, P_2/P_2 will also rise. So, K/L must rise unless Y_1/Y_2 falls. But, if Y_1/Y_2 falls, there is a shift in favour of consumption goods which are more capital intensive. As proved above, with the capital labour ratio increasing is both sectors and the capital intensive consumer goods sector -2 gaining at the expense of sector -1, k(= K/L) must rise. Thus, k_1 and k must p more together and the stability condition for equilibrium growth holds.

But the condition that sector -2 is more capital intensive than sector -1 is a sufficient condition for stability and not a necessary condition. How this condition is violated but stability occurs is explained is the following example.

Suppose both sectors have cobb-Douglas production functions with elasticities of $\propto_1$ and $1 - \propto_1$ for k_1 and L_1 and $\propto_2$ and $1 - \propto_2$ for K_2 and L_2

Then,

$$rk = \propto_1 P_1 Y_1 + \propto_2 P_2 Y_2$$

and $\quad rk = P_1 Y_1$

Then, $\quad P_1 Y_1 = \dfrac{\alpha_2}{1-\alpha_1} P_2 Y_2$

and $\quad \dfrac{P_1 Y_1}{P_2 Y_2} = \dfrac{\alpha_2}{1-\alpha_1}$

But $\quad \dfrac{P_1 Y_1}{P_2 Y_2} = \dfrac{rk}{wL}$

So, $\quad \dfrac{rk}{wL} = \dfrac{\alpha_L}{1-\alpha_1}$

In the above equation, the right hand side is a constant. Hence, whenever, r/w falls, k, rise and K/L (= k) must also rise. So, we find that k_1 and k more together and stability occur. It does not matter whether sector -2 is more capital intensive than sector -1.

8.5 OPTIMAL ECONOMIC GROWTH – RAMSEY PROBLEM

Ramsey Problem

The Ramsey problem, or Ramsey-Boiteux pricing, is a second best policy problem concerning what price a public monopolist or a firm faced with an irremovable revenue constraint should set, in order to maximise social welfare. A closely related problem arises in relation to optimal taxation of commodities.

In a first best world, the optimal solution would be to use prices equal to marginal cost and charge an optimal lump-sum charge that would cover the fixed cost or revenue requirement. Nevertheless, this is usually impossible to implement, thus price distortion is inevitable.

This principle is applicable to pricing of goods that the government is the sole supplier of (public utilities) or regulation of natural monopolies, such as telecommunications firms. It is also applicable to situations where there is perfect competition in the private sector, but the government needs to distort the prices of the goods it provides in order to break even, or to earn a profit. In this case, the "constraint" is that the revenue requirement cannot be covered by a lump-sum tax. So, prices must be distorted

Description

For any monopoly, the price mark-up should be inverse to the price elasticity of demand. The more elastic demand for the product, the smaller the price markup. Frank P. Ramsey found such a result in 1927 in the context of taxation. The rule was later applied by Marcel Boiteux (1956) to natural monopolies (decreasing mean cost): a natural monopoly experiences profits/losses if it is forced to fix its output price at the marginal cost, subject to Economies of Scale being exhausted. Hence, the Ramsey-Boiteux pricing consists of maximising the total welfare under the condition of non-negative profit, i.e., zero profit. In the Ramsey-Boiteux pricing, the mark-up of each commodity is also inversely proportional to the elasticities of demand, but it is smaller as the inverse elasticity of demand is multiplied by a constant lower than 1.

Ramsey pricing is sometimes consistent with a government's objectives because Ramsey pricing is economically efficient in the sense that it can maximise welfare under certain circumstances. There are, however, problems with Ramsey pricing. A profit-maximising operator will choose Ramsey prices only if all markets are equally monopolistic or equally competitive. If markets are not equally monopolistic or competitive, then the regulator has an interest in ensuring that the extent to which the operator can use Ramsey pricing is limited to groups of services that are subject to similar degrees of competition. Regulators typically do this by forming groups of services that are subject to similar degrees of competition and allowing the operator price flexibility within each service group.

Even though Ramsey pricing can be economically efficient, it may not be consistent with the government's goal of providing affordable service to the poor and the rate by which prices change to achieve Ramsey-efficient prices may not be consistent with political sustainability. As a result of these two concerns, the regulator sometimes limits the operator's ability to pursue Ramsey pricing within a service group. In the case of services to the poor, the regulator may place upper limits on the prices. In the case of services where traditional prices were different from Ramsey prices, there are equity issues in changing from the traditional pricing structure to a new structure, even if the new structure would be more efficient in an aggregate sense. In such situations, the regulator may impose pricing restrictions that prevent Ramsey pricing or that impose a slower transition to Ramsey pricing than the operator would choose left to its own devices.

Lastly, regulators often note that Ramsey pricing is a form of price discrimination—although not necessarily a bad form of price discrimination—and customers sometimes object to it on that basis. The public sometimes believes that it is unfair to cause one type of customer to pay a greater mark-up above marginal cost than another type of customer. In such situations, regulators may further limit an operator's ability to adopt Ramsey prices.

Practical issues exist with attempts to use Ramsey pricing for setting utility prices. It may be difficult to obtain data on different price elasticities for different customer groups. Also, some customers with relatively inelastic demands may acquire a strong incentive to seek alternatives if charged higher mark-ups, thus undermining the method. Politically, customers with relatively inelastic demands may also be considered as those for whom the service is more necessary or vital; charging them greater mark-ups can be challenged as unfair. Crucially, many economists deny this, considering less vital services as unnecessary depending on its price elasticity of demand.

Formal Presentation and Solution

A formal presentation was given by Ramsey in a journal article titled: "A Contribution to the Theory of Taxation". The mathematical derivation follows:

Consider the problem of a regulator seeking to set prices $(P_1, ..., P_n)$ for a multi-product monopolist with costs $C(q_1, q_2, ..., q_n) = C(q)$ where, q_n is the output of good n and P_n is the price. Suppose that the products are sold is separate markets. So, demands are independent and demand for good n is $q_n(P_n)$, with inverse demand function $P_n(q)$.

Total revenue is given by

$$R(p, q) = \sum_n p_n q_n(P_n).$$

Total welfare is given by

$$w(p, q) = \sum_n \left(\int_o^{q_n} P_n(q)d_q \right) - c(q)$$

The problem is to maximise w(p, q) subject to the requirement that profit n = R – C should be equal to same field value π^*. Typically, the fixed value is zero to guarantee that the profits/losses are eliminated.

$$R(p, q) - C(q) = \pi^*$$

This problem may be solved using the Lagrange multiplier technique to yield the optimal output values and backing out the optimal prices. The first order conditions on q are

$$P_n - C_n(q) = -\lambda \left(\frac{\partial R}{\partial q_n} - C_n(q) \right)$$

$$= -\lambda \left(P_n \left(1 - \frac{1}{\varepsilon_n} \right) - C_n(q) \right)$$

where λ is a Lagrange multiplier, $C_n(q)$ is the partial derivative of C(q) with respect to q_n evaluated at q and $\varepsilon_n = -\dfrac{\partial q_n}{\partial P_n} \dfrac{P_n}{q_n}$ is the elasticity of demand for good n.

Dividing by P_n and rearranging yield,

$$\frac{P_n - C_n(q)}{P_n} = \frac{K}{\varepsilon_n}$$

Where $K = \dfrac{\lambda}{1 + \lambda} < 1$, i.e., the price margin compared to marginal cost for good n is again inversely proportional to the elasticity of demand. The Ramsey mark-up is smaller than the ordinary monopoly where K = 1, since λ = 1 (the fixed profit requirement, $\pi^* = R - C$ is non-binding). The Ramsey price setting monopoly is in second best equilibrium between ordinary monopoly and perfect competition.

Theorem

Given any positive integers p and q, these exists a smallest integer n = R (P, q) such that every 2 colouring of the edges of K_n contains either complete subgraph on p vertices, all of whose edge an in colour or a complite subgraph on q vertices, all of whose edgee are is colour 2.

Proof: We will proceed b_o induction on P + r

First we consider the box case is which P + q = 2. The only way this can be thus is if P + V = 1, and it is learn that R(1, 1) = 1

Now, we assume that the theorem holds whenever P + σ < x, for some positive integer N. Let p and Q be integer such that P + Q = N. Then, P + Q – 1 < N, so by our assumption we know that R (P – 1, Q) and R (P, Q – 1) exist. Consider,only colouring of the edges of K_v in two colour c_1 and c_2 where $v \geq$ R (P – 1, Q) + R (P, Q – 1)

Let, x be a vertex of k_v. By the pigeonhole principle and because $v \geq$ R (P – 1, Q) + R (P, Q –1), we know that of the v– 1edgn that x is incident to, either R (P – 1, Q) edge an colour c_1 or R (P, Q –1) edge an is colour C_2.

If x is incident to R (P − 1, Q) edge of colour c_1, consider the K_R (P − 1,Q) whose vertion an the vertion joined to x by edge of colour c_1, that is the subgraph induced by the neighbourhood of x. Because we know that R (P − 1, Q) exists, there are two possible case to consider. One is that this graph contains a $K_{P−1}$ with all edge is colour c1, is which case this $K_{P−1}$ together with *x* forms a monochromatic K_P is colour c_1. The other possibility is that k_R (P − 1, Q) contains a K_Q with all edge is colour C_2. In either case, we can that R(P, Q) exists.

A parallel argument holds if x is incident to R (P, Q − 1) edge of colour C_2 and K_x again contains one of the required monochromatic complete graph.

Thus, R(P, Q) exists, and is facts, because we close *v* such that $v \geq$ R(P − 1, Q) + R(P, Q − 1), we know that R(P, Q) ≤ R(P,Q − 1) + R(P −1, Q).

Ramsay's theorem guarantee that this smallest integer R_1 P, q exists but does little to helps us determine. That its value is given some positive integers, pond *v*. In general, this is actually are exceedingly difficult problem.

8.6 SUMMARY

1. The Harrod-Domar Model is the direct outcome of projection of the short-run Keynesian analysis into the long-run.

2. Harrod's model is the English version of Domar's model. Both these models stress the essential conditions of achieving and maintaining steady growth.

3. Prof. Robert M. Solow made his model an alternative to Harrod-Domar model of growth. It ensures steady growth in the long-run period without any pitfalls. Prof. Solow assumed that Harrod-Domar's model was based on some unrealistic assumptions like fixed factor proportions, constant capital-output ratio, etc.

4. In the Keynes-Kaldor-Pasinetti post-Keynesian growth model, two classes of agent, workers and capitalists, save constant proportions of their income.

5. In an economy with institutional investors, investment and hence growth are likely to be influenced by the decisions of such investors. But under modern capitalism, there are many high technology firms (e.g., IT, software, etc.) which present institutional investors with substantially greater problems of risk and asymmetric information than firms with less dynamic technologies (e.g., consumer durables).

8.7 SELF ASSESSMENT QUESTIONS

1. Critically explain Harrod-Domar Model.

2. Explain Neo-Keynesian model of Passineti.

3. Explain Optimal Growth – Ramsey Problem.

❧❦❖❧❧

MULTISECTOR GROWTH MODELS

Objectives

The objectives of this lesson are to learn:

- Von Neumann growth model
- Turnpike theorems

Structure:

9.1 VON NEUMANN GROWTH MODEL AND CONCEPT OF EFFICIENCY AND OPTIMISATION FOR VON NEUMANN MODEL

In the immediate post-war years, Mark Blaug identified the emergence of a new paradigm in economics, the so-called "formalist paradigm", which marked the arrival of the pre-eminence of (mathematical) form over (theoretical) content, and which is mostly characterised by the crucial importance economists give to a specific (non-constructive) kind of demonstration of existence of equilibrium. This revolution took shape in the 1950s and 1960s around the works of Arrow, Debreu, Patinkin, Solow, Dorfmann, Samuelson and Koopmans.

The objective is to interpret John von Neumann's growth model (1937) as a decisive step of this formalist revolution, and by doing so, contribute to the definition of the formalist paradigm in economics. In this model, it will be argued, is the manifestation of von Neumann's involvement in the formalist programme of mathematician David Hilbert, and provides economists with the new mathematical tools and methodology that will characterise the emerging paradigm in economics.

The 1937 model gave rise to an impressive variety of contrasting comments as far as the filiations (classical versus neoclassical) of the growth model are concerned, and constitutes one of those enigmas which historians of economic thought are so fond of. However, the identification of an economic formalist paradigm allows one to go beyond the traditional demarcation line between classical and neo-classical economics and challenges the legitimacy of such a criterion. The issue of the nature of the assumptions upon which the 1937 model is based becomes much less relevant than that of the extent of the methodological innovation introduced by von Neumann, namely, the introduction of the modern axiomatic approach in economics.

The aim is to elucidate this interpretation through a rational reconstruction of the epistemological approach adopted by von Neumann in the 1937. The result of this reconstruction may be summarised in this way: von Neumann gives here an economic interpretation to a specific formal system which he initially elaborated in his previous work of 1928 on game theory. Each term here has a precise meaning: a "formal system" is composed of: (a) a set of symbols, (b) a set of rules for transforming these symbols into formulae, (c) a set of rules for transforming the formulae, and (d) a reduced number of formulae representing the axioms of the system to be observed. By construction, a formal system has no semantic content and may take on different interpretations. A "model" is an interpretation that is given to a formal system. The clear-cut separation between syntax and semantics – between the formal aspects of the system and its various interpretations – is one of the most salient characteristics of modern axiomatics.

In order to prove that the scope of the 1937 model may be correctly grasped by understanding von Neumann's global epistemological approach, we will proceed as follows. It is first necessary to offer a brief overview of the growth model and of the controversy over the filiations; the variety of the comments is by itself an invitation to consider an alternative interpretation. We found such an alternative in von Neumann's involvement in the formalist Hilbertian programme so that the classical/neo-classical demarcation line may well be replaced by the formalist/non-formalist criterion, as Blaug (2003) and Nicola Giocoli (2003) suggest. The term "formalism" is ambiguous and requires further elucidation. In particular, the question of the impact of Godel's discoveries on the formalist programme is of primary interest to us to the extent that, it will be argued that it is a manifestation of the pragmatic turn that Godel lays on formalist mathematicians. We will then have all the elements to show that von Neumann's main achievement that has been to propose to economists the substitution of the mechanical analogy with the mathematical analogy, as a result of his participation in the post-Godelian mathematical formalist programme.

Its Various Interpretations

von Neumann characterises the equilibrium configuration of an economy expanding at a uniform rate. In equilibrium, prices are constant, as are the quantity ratios between different goods. Several simplifying assumptions are introduced by von Neumann to make equilibrium possible: constant returns to scale; pure and perfect competition; unlimited quantities of goods available through the productive process (this applies to land and labour, no primary factors existing in the model); no savings from workers who are depicted as draft animals; and no consumption from producers who save the totality of their income.

Production is considered a temporal process (of length of one period) of transforming one set of goods into another; for reasons of simplicity and for ensuring the unity of the solution. von Neumann also had to make the assumption that each good entered the productive process of all goods, be it as input or output, and also in an arbitrarily small proportion. The cost of production of one good depends on the value of the goods necessary for its production, plus the interest rate; the prices of goods correspond to their production costs, whatever the preferences of workers or producers are.

Solving this model allows identifying the following:

- Which, among the set of goods in the economy, are the free goods whose price must be fixed equal to zero, and what the prices are of the other non-free goods; free goods are goods whose produced quantity exceeds the quantity used in the production process in a proportion higher than the rate of growth of the economy. Introducing the free goods rule allowed von Neumann to avoid the occurrence of negative prices at equilibrium, and, from

a mathematical point of view, transform the representation of the economy by introducing linear inequalities into the model;

- Which are the profitable production processes and which ones are non-profitable and will, therefore, not be implemented (a profitability rule which, like the free goods rule, leads to the use of linear inequalities in the model); the model allows the determination of the maximum intensity with which each profitable process will be implemented, i.e., the produced quantities of each good, and, thus, given the constant returns to scale assumption, the growth rate of the economy;

- The dual symmetry of the model is one of its essential properties and manifests as follows:

  — Solving the model may be interpreted on the one hand as a problem of technological choice: given the price vector, it is possible to determine the vector of the maximum possible produced quantities and the optimal growth rate, under the constraint of the free goods rule and given the impossibility of consuming more than is produced;

  — Solving the model may also be interpreted on the other hand as a problem of economic expansion, which turns out to be the mirror image of the previous problem. It consists of determining the optimal price vector and interest rate which prevail, given the intensities of production processes, the efficiency rule, and the competitive constraint according to which no extra profits are allowed.

Von Neumann showed that an equilibrium solution exists, that it is unique, and that the interest rate of this configuration is equal to the growth rate. The proof of existence breaks with the traditional attempts of demonstrating the existence of a general equilibrium configuration consisting of counting the numbers of equations and unknowns. Such an approach did not constitute sufficient proof of existence, and, furthermore, the model was formalised in terms of inequalities (the free goods rule and the profitability rule) and thus required specific mathematical tools. The demonstration of existence provided by the author consisted in an extension of Brouwer's Fix Point Theorem and represented the first introduction of topological tools in economic analysis: Von Neumann introduced a new function, $\phi(X, Y)$, which represents the ratio between the total incomes and the total costs, and demonstrates the existence of a solution of the growth model, amounting to demonstrating the existence of a saddle point for function ϕ. Now, the existence of this saddle point is itself the consequence of von Neumann's demonstration of a fix point lemma. This demonstration is non-constructive in the sense that no method is provided for the determination of the fix point. With this kind of demonstration, equilibrium thus becomes a purely logical concept. Existence is demonstrated by showing that non-existence would involve a logical contradiction. As emphasised by Giocoli and also Blaug, this kind of non-constructive proof (or "negative proof") allows a direct jump from the axioms of the model to its final outcome and accounts for the neglect of mainstream economists in the analysis of the economic process that leads to equilibrium.

With the notable exception of Harold W. Kuhn and Albert W. Tucker (1958) who provide an analysis of the mathematics of von Neumann's proof, economists in the 1950s and 1960s mainly concentrated their comments on the economic filiations of this model. In 1959, the Kaldor-Solow debate that unfolded during the conference on Capital was the starting point of a long controversy over the interpretation of the 1937 model. Kaldor insisted upon the classical underpinnings of von Neumann's growth model, whereas Solow emphasised the possibility of integrating this model into the neo-classical framework. The arguments advanced by the two economists set the tone of future debate.

- Supporters of a classical interpretation insist on the heterodox nature of the assumptions on which the model is built. Kaldor, for instance, essentially based his position on von Neumann's assumption of infinite expansion of primary factors for, according to him, one of the defining features of mainstream economics is precisely the existence of a physical constraint on the available quantity of these resources. In the same way, Luigi Pasinetti (1977) stressed the circular character of the production process, whereas Heinz Kurz and Neri Salvadori (1993) insisted on its temporal dimension and on the proximity of certain of the model's characteristics with past contributions of classical authors, from Petty to Remak and von Bortkiewitcz. It is worth remarking that according to this line of interpretation, and contrary to what is defended below, the nature of the mathematical techniques used in demonstrations does not constrain the theoretical nature of the model. Accordingly, von Neumann's model would offer proof that optimisation tools do not constitute a selective feature of neoclassical economics;

- Supporters of a neo-classical interpretation put to the fore more technical arguments to show that the model may be understood as a special case of the more general neo-classical framework. Such generalisations entail, among others, the introduction into the model of the intertemporal preferences of consumers (Edmont Malinvaud, 1953), the consideration of labour as a primary factor constrained by an exogenous growth rate (Michio Morishima, 1964), a relaxation of the assumption of circularity according to which each production process uses or produces a given quantity of each good produced in the preceding period, etc. This interpretation consists ultimately in presenting here 1937 model as a crucial step in the construction of the neo-classical paradigm, starting from Leon Walras (through the formulation given by Gustav Cassel) and extending to the modern demonstration of existence by Kenneth Arrow, and Gerard Debreu.

It is possible to appraise the relevance of the controversy over the filiations of von Neumann's model from different perspectives. If it were simply a question of situating the model either in the classical or the neo-classical camp, then the extent of the confrontation would be rather narrow and the relevance of the debate questionable. However, from an analytical viewpoint, the implications of this confrontation have turned out to be very significant for both sides: in the orthodox camp, von Neumann's growth model is at the roots of linear programming, the turnpike theorem of Dorfman, Samuelson and Solow, and of modern proofs of existence of general equilibrium; in the heterodox camp, the growth model is certainly an important source of the classical revival of the 1960s that followed the publication of Sraffa's book. For instance, Goodwin's limit cycle model formalises short-term economic fluctuations along the quasi-stationary long-term equilibrium trend of von Neumann; Andras Brody (1970) starts from a simplified version (with no joint production) in matrix form of von Neumann's model in order to propose a mathematical rehabilitation of the labour theory of value.

The variety of the interpretations ultimately shows that von Neumann's growth model hardly fits into the traditional classical/neo-classical classification system. It is a characteristic of path-breaking contributions to upset the prevailing schemes. Interpreting the growth model in the light of the forthcoming formalist revolution of the 1950s means focusing on the nature of the mathematical innovations introduced by von Neumann in economics. These innovations may be appraised from different perspectives.

From a strictly technical viewpoint, von Neumann's contribution is easy to identify: it consists in the generalisation of Brouwer's Fix Point Theorem. This idea is endorsed by the fact that the

Minimax Theorem is an unnecessarily heavy tool to demonstrate the existence of an equilibrium solution of this economy. Nicholas Georgescu-Roegen (1951) provides a demonstration exclusively based on the properties of convexity and separation of hyper-plans, supporting the idea that the growth model represented to von Neumann only a specific support which allowed him to back up his mathematical results.

From a methodological perspective, the contribution of the 1937 model is much more complex to identify. It is the objective of this rational reconstruction to show that von Neumann's path-breaking contribution consisted of extending the standards of rigour of mathematical formalism to the community of economists. Discussion about the nature of the model's theoretical foundations is relegated to the background.

It is worth noting that the majority of the protagonists to the filiations debate make a point of mentioning the limitations of their comments, recognising to a certain extent that the field of economics does not represent the privileged field of investigation of the author: Tjalling C. Koopmans declared along this line that despite the unquestionable theoretical advance provided by the 1937 growth model. In the same way, David G. Champernowne conceded that the author approached the question of existence as a mathematician, putting the emphasis on aspects of the problems distinct from those upon which an economist would have insisted. Notice also the comment of Sukhamoy Chakravarty who, before introducing the Kaldor-Solow debate, asserted that it was possible ultimately that von Neumann himself considered as essentially technical in nature. Paul Samuelson declared with reference to this part, explaining further on that the genius of von Neumann's contribution fitted any capital model. Von Neumann himself cleared the question of the filiations in a lapidary style: "It is obvious to what kind of theoretical models the above assumptions correspond", as if this was not the issue at stake, drawing attention once more to the technical aspects and the nature of the mathematical approach itself.

Von Neumann and the Formalist Programme of Hilbert: Before and After Godel

From the start, a significant problem seems to threaten interpretation. It is of chronological order. The article of 1937 was designed, then published after Von Neumann was informed of the famous theorem of impossibility of Godel, devastator of the mathematical formalist programme and unanimously recognised as an element of rupture in the evolution of modern mathematics. von Neumann is also one of the first mathematicians to seize the range of Godel's theorem and to take into consideration its methodological consequences. It is necessary at this level to reconsider the definition of the formalist Hilbertian programme in order to understand more precisely what the impact of Godel's discoveries was, and to what extent it modified mathematical practices.

The term "formalism" itself is ambiguous because it bears a double significance. In its commonly accepted sense, formalism indicates nothing other than the mere use of symbols and unspecified mathematical techniques to express an idea. It is not acceptance that this term implies when it is associated with Hilbert. By formalism, one then understands a particular philosophy of mathematics which reduces it to a formal language, and is opposed to intuitionism and logicism on the question of the foundations of mathematics.

The debate on the foundations emerges among mathematicians at the end of the nineteenth century, while attempts to extend the traditional axiomatic (Euclidean) method to branches of mathematics other than geometry are multiplying. This method consists in accepting without demonstration a reduced set of postulates, the axioms, and deducing by logical inference a set of theorems. For a long time, the empirical obviousness of axioms seemed to guarantee the veracity of

the theorems which it was possible to deduce. But the growing abstraction of the mathematical practice and the discovery by Cantor and Russell of logical antinomies bring to the foreground the question of the consistency of formal systems. "Consistency" refers to a precise property: a formal system is consistent when it is impossible to deduce from its axioms two contradictory theorems. Three types of answer were advanced to give back to mathematicians their confidence in the rigour of mathematical practices.

Logicists try to found the consistency of mathematics by defining it as a branch of logic. The Principia Mathematica of Whitehead and Russell, published in 1910, falls under this head. There, the authors proposed a formalisation of arithmetic, whose goal is to clarify and make explicit all the logical inferences used in the reasoning and to show that all the concepts of arithmetic can be brought back to concepts of pure logic. However, this step did not gain much support from mathematicians as this solution did nothing but move the problem: the consistency of arithmetic depended on that of logic, and the consistency of logic was then itself under discussion.

Intuitionists, headed by Poincare and Brouwer, placed the authority of the perception and of the intuition of the mathematician above that of the logical principles and inference rules whose historical and cultural relativity were underlined. To be consistent, a system of calculation must thus be built from obvious and unimpeachable axioms and from rules of inference subjectively considered as reliable by the mathematician. Luitzen Brouwer, the fundamental dissension which exists between intuitionalism and formalism is that, a different answer is given to the question of knowing where the mathematical accuracy exists: to the intuitionalist, in human intellect; to the formalist. Thus, the consistency of a mathematical theory does not require a demonstration for intuitionalists insofar as it results from the construction itself of the theory, following the principles and the procedures acceptable to the majority of mathematicians.

On the contrary, the response of formalists to the uncertainty on foundations consisted of trying to establish rigorous evidence of consistency of the various branches of mathematics. Demonstrations of consistency initially take the form of relative proofs. Thus, Hilbert showed that the consistency of Euclidean geometry depends on that of algebra. Thereafter, he tried, with the assistance of his disciples (the first of whom was von Neumann, but also Ackermann and Bernays) to provide an absolute demonstration of consistency of arithmetic. It is at this level that the famous impossibility theorem of Gödel intervenes. In 1931, Gödel arrived at a devastating result on the question of the foundations of mathematics. He, in fact, showed that it was impossible to provide a demonstration of absolute consistency of arithmetic. Gödel did not prove the inconsistency of arithmetic, rather, the impossibility of showing that it was consistent, leaving the door open to the potential occurrence of new logical antinomies. In his book, Morris Kline (1980) presented in a provocative way the debate on the foundations of mathematics as a major intellectual rout, liquidating the hitherto-dominant design of mathematics like point of organ of rigour and scientific exactitude. The title of his work 'The Loss of Certainty' returned precisely to this radical reconsideration: mathematics cannot be unanimously regarded any more as a set of firmly established eternal truths.

This result certainly cooled down the enthusiasm of formalists but did not put an end to the programme of Hilbert whereof the work on foundations constitutes only one part. Formalists gave up the hope to be able to show that mathematics were consistent, but they did not give up their confidence in the power of modern axiomatics as an engine for discovering new scientific knowledge. As Giorgio Israel and Ana Gasca (1995) note indeed, the formalism of Hilbert was founded on the belief in a pre-established harmony between mathematics and physical reality, a harmony which makes it possible to conceive mathematics like the base of all exact scientific knowledge of nature.

The normative aspect of Hilbert's programme can consequently be interpreted as follows: the mathematical analogy, understood as the systematic adoption of the modern axiomatic approach represents the good scientific practice and this, whatever the scientific field considered. I believe: anything at all that can be the object of scientific thought becomes dependent on the axiomatic method, and thereby indirectly on mathematics, as soon as it is ripe for the formation of a theory. By pushing ahead to ever deeper layers of axioms, we also win ever-deeper insights into the essence of scientific thought itself, and we become ever more conscious of the unity of our knowledge. In the sign of the axiomatic method, mathematics is summoned to a leading role in science.

The association between axiomatic method and scientific rigour thus justifies the second side of the formalist programme of Hilbert consisting concretely of trying to extend this approach to other scientific disciplines, physics initially, but also economics. Thus, Hilbert's formalism has a double finality: to solve the problem of the foundations of mathematics (and, at this level, the results of Gödel are without call); and to extend modern axiomatics to all the scientific disciplines. This second aspect of the programme, the aspect that can be described as the imperialist or normative side, survived to Gödel. Weintraub identified these two aspects of the formalist programme. He distinguished between the Finitist Programme for the Foundations of Arithmetic (FPFA) whose objective was to found the consistency of arithmetic and the axiomatic approach, the only aspect of the formalist programme which has actually influenced the process of mathematisation of economics through the contributions of von Neumann for the strictly Hilbertian version of the AA programme, and Debreu for the Bourbakist version.

Until 1931, von Neumann was strongly implicated in the two aspects of Hilbert's formalist programme. As far as the work on foundations is concerned, he contributed to the axiomatisation of Cantor's set theory. This theory, known as the "naive" theory of sets because it was then not yet in axiomatic form, leads to logical inconsistencies discovered around 1900 by Cantor himself and by Russell. Since his doctorate thesis, von Neumann contributed to looking further into the axiomatisation of set theory proposed by Zermelo, Fraenkel and Skolem through the introduction of new axioms and methods, making it possible to avoid the occurrence of these contradictions. The axiomatic method is used in order to allow a rigorous representation of the theory within which the origin of contradictions can be easily found and possibly eliminated.

Regarding the normative aspect of the formalist programme, since 1926 von Neumann tackled the question of the mathematical axiomatisation of quantum physics, then defined around the two competing presentations of Heisenberg and Schrödinger. This work led to the publication in 1932 of the Mathematical Foundations of Quantum Mechanics in which the author managed to unify these two visions within a single formal system. Game theory is another field where the project of exporting modern axiomatics to new fields of scientific knowledge appears: von Neumann followed at the beginning the developments of Zermelo on the axiomatisation of chess, a question much debated in discussions in mathematical circles of the inter-war period. It was a question of showing that a formal system could receive an interpretation in terms of social phenomena rather than in strictly natural terms. von Neumann generalised the application of Zermelo to the context of any type of zero-sum games. This work led him to the determination of the Minimax Theorem in 1928. From thereon, Hilbertian formalism could penetrate the field of individual interactions and be used for the analysis of social phenomena.

The Pragmatic Turn

Gödel's discoveries affected von Neumann deeply. They contributed to immediately putting a term to his work on the foundations of mathematics and signalled the beginning of what many commentators describe as a pragmatic turn in the scientist's method. Hilbert's programme on the foundations conveyed the hope of justifying the axiomatic method, to carry mathematical results to the statute of eternal truth. Gödel destroyed this hope, but the majority of mathematicians (von Neumann among them) decided to use this method all the same because it remained, in spite of the loss of certainty, a rigorous way of producing scientific knowledge. The second side of Hilbert's programme was unharmed.

It was accepted that it was impossible to found mathematics absolutely. However, indirect ways existed to comfort scientists and to relativise the loss of certainty they suffered in full measure. First of all, should a contradiction emerge, formalisation makes it easier to search for its origins and eventually to eliminate it thanks to the baring of all of the concepts and reasoning intervening in the theory. The position of the Bourbakist programme is for this reason evocative: the objective of this radical version of formalism is not to found mathematics any more, rather, to clarify, through the linking of formal systems with one another, the architecture and unity of mathematics. The mathematician must face contradictions, if they emerge, on a case-by-case basis.

There is a second means of reassuring the scientist about the consistency of his formal system. It consists of putting back to the foreground considerations of a semantic nature. This assertion requires further elaboration. A prominent characteristic of Hilbertian formalism is without any doubt the strict separation between syntax and semantics. To formalise a theory in the sense of Hilbert means indeed emptying it from all of its semantic content and giving an abstract representation of it – the formal system – in the form of symbols, formulae (among them axioms) and sequences of formulae having no more obvious bond with the theory of departure. The formal system thus formed is like an abstract box, deprived of any significance, on which the mathematician works in order to draw theorems. At this stage, the question of the realism of the axioms is completely irrelevant. But it would be erroneous to say that, in axiomatics, reality does not matter at all, for in the next stage of the axiomatisation process, the objective is precisely to assign models to each formal system, i.e., to find an interpretation in terms of real phenomena for the formal system. A model consists of an interpretation of the formal system, each symbol receiving a meaning, and the same abstract box being able to receive various interpretations. The initial theory which inspired the formal system constitutes one model, among others. Formalism as a philosophy of mathematics is attached at this level with Plato's realism consisting of supporting the thesis that mathematics does not create anything, does not invent objects, rather, discovers pre-existent objects in the intellect. The power of axiomatisation is due precisely to the fact that the "discovery" of an abstract box makes it possible to explain several real phenomena, and rests on the belief of a preset adequacy between the structure of mathematics and reality.

This vision of the world is opposed to constructivism, of which intuitionism is a specific form, and which considers that a mathematical object exists only through its elaboration. To formalists, on the contrary, the very existence of any mathematical concept refers to a precise property that it is free from any contradiction.

Before paradoxes and logical antinomies were discovered and encouraged mathematicians to work out absolute demonstrations of consistency, it was sufficient, in order to found a formal system, to find a model in which its axioms were valid. For a long time, the obviousness of the Euclidian axioms was sufficient to ensure the consistency of Euclidian geometry. If axioms were valid, then it

was also the case for the theorems that one could derive from them. The so-called method of the models consisting of finding an interpretation to an abstract system in which its postulates are valid was largely used to give relative demonstrations of consistency to formal systems less intuitive than the Euclidean one. Gödel's discoveries led mathematicians to reconsider the value of this method. One cannot found the consistency of a formal system absolutely, but the discovery of a new and adequate model for this system reinforces its heuristic validity and comforts the mathematician regarding its consistency. The 1937 contribution of von Neumann may be interpreted in that way a new semantic correspondence is associated with a formal system elaborated beforehand. In particular, von Neumann gave an economic interpretation to a formal structure which he previously discovered in game theory (1928). This idea was expressed explicitly by the author himself when he declared that "the question whether our problem has a solution is oddly connected with that of a problem occurring in the Theory of Games dealt with elsewhere".

The formal similitude between the 1928 and 1937 models is, however, not immediate. In 1928, von Neumann demonstrated the existence of a solution for a two-person zero-sum game without ever defining a system of linear inequalities and equations. As Tinne H. Kjeldsen (2001) states, the Minimax Theorem was developed in 1928 with no explicit connection with the theory of linear inequalities, and there are no elements which shows that von Neumann would be aware at that time of this connection. However, the fact that this connection does exist is sufficient to corroborate this rational reconstruction. Kuhn and Tucker (1958) explicitly link the solutions of the minimax problem with a system of linear inequalities and equations which corresponds to the problem raised in 1937. They state explicitly that if the intensity and price vectors are both normalised, they form probability vectors which may be regarded as mixed strategies for the players of a zero-sum two-person game. Dore (1989b) also studied the connection between the system of inequalities and equations of the model and the two-person zero-sum game of 1928, the strategies of player I are represented by the set of vectors of production intensities, those of player II by the set of price vectors. Payoff functions depend on the strategies chosen by each player: player I chooses the vector of the intensities of production which maximises his payoff function, given the choice of player II, supposed for his part to choose the least satisfactory solution for the first player. A symmetrical reasoning relates to the choices of player II. The Minimax Theorem ensures the existence of a saddle point which corresponds to the situation where the rate of growth is equal to the interest rate.

Here, it will be explained the separation and hierarchy between syntax and semantics, typical of the axiomatic approach it.

The same formal system, the same box, indeed receives different interpretations, i.e., different models: one in game theory, one in economics, and even one in thermodynamics. Thanks to Gödel, we know that the consistency of this formal system is impossible to prove. However, the fact that this system fits different interpretations is a reassuring symptom of its consistency. The economic interpretation is, in this connection, the manifestation of the pragmatic turn of the mathematical formalist programme which consisted in considering not only the syntax aspect, but also the semantic step of the axiomatisation process through the identification of adequate new models. Further, with a new domain of application, economics, opened itself up to formalist mathematics, and, more generally, to mathematical analogy.

From the Mechanical to the Mathematical Analogy

The growth model was elaborated in 1931 in the United States and first presented to a mathematical seminar at Princeton, but it has definitely been arousing interest and enthusiasm since

its discussion in the Karl Menger seminar in Vienna in 1934. One reason for the particular interest of Viennese scholars in the growth model lies in the total adequacy between von Neumann's epistemological approach in this paper and the specific philosophical context of the Vienna Circle, marked by analytical philosophy, logical positivism, and a project of unification of sciences.

One finds a definite parallelism between the concerns of formalist mathematicians on one side and of logical positivist philosophers on the other. The major concern of mathematicians is to eliminate the possibility of contradictory theorems; the major concern of philosophers is to eliminate from their discourse all metaphysical proposition, i.e., any pseudo-scientific assertion whose intrusion in the reasoning may lead to logical inconsistencies. In both cases, discussions are directed towards the research of certainty in scientific reasoning.

The principal theses of logical positivism are presented by Otto Neurath, Rudolf Carnap and Moritz Hahn in an article of 1929, "The Scientific Conception of the World: The Vienna Circle", better known as the "Manifesto of 29". Logical positivism falls under the continuation of the positivist programme of Auguste Comte, Hume and Mach, whose objective was to base knowledge directly on experience. To this end, members of the Vienna Circle used the latest developments of modern logic from Frege, Peano and Russell. More precisely, logical positivism was born from the introduction of logical analysis into the positivist framework. Logical analysis consists in reducing scientific concepts and propositions to experience, to direct observation, from which all the remainder logically arises. In the same way, that axiomatisation makes it possible to uncover the source of possible contradictions easily, logical analysis tracks pseudo-propositions and contributes to eliminating them from philosophical discourse. The project of Carnap is even more ambitious. The philosopher has been working on a project to work out a formal logico-mathematical language used to guard scientists against the surreptitious intervention of pseudo-propositions in their reasoning. Philosophy thus becomes analytical. It is finalised with the revelation of the significance of propositions and the elimination of meaningless propositions. This "turning point of philosophy" is an indicator of the ambition of logical positivism to aim at unitary science. With analytical philosophy, it will not be necessary any more to speak about philosophical problems, because all problems will be discussed philosophically, i.e., clearly and meaningfully. The call for the unity of science, explicit in the Manifesto, claims to be epistemological. It is a means for scientists of working out a way of making science, whatever be the field of production of knowledge, which ensures rigorous reasoning, free from metaphysics. This is logical analysis for Russell, the universal formal language for Carnap, and modern axiomatics for Hilbert.

The unifying ambition of formalism asserts itself gradually. Initially, it was a question of unifying, through the development of modern axiomatics, all the branches of mathematics. Formalists, rather, their predecessors, analyticals, were then opposed to the purist vision of mathematics dominant by the end of the nineteenth century. According to purists, mathematics was to remain split in various branches, each defined by its own method of investigation. For example, purists refused geometric demonstrations based on Cartesian algebra. Analyticals, on the contrary (with Hilbert in the forefront), believed in the interaction of the various branches and shared an ideal of unification of mathematics, conceived as a unified system of knowledge. In a second step, this strong optimism exceeded the borders of the discipline; building from the success of the axiomatisation of quantum physics, formalists then invested the field of social phenomena.

Economics is implied in the philosophical programme of the Vienna Circle through the active interaction of the members of Hans Mayer's Economic Seminar with those of the Mathematical Colloquium run by Menger, son of the founder of the Austrian economic tradition. Collaboration

between mathematicians and economists crystallised in the resolution of the problem of imputation as defined by Menger in 1871. It consists of deducing the prices of factors of production starting from the value of the consumption goods which they contributed to produce. The solution suggested in 1889 by Wieser encounters a problem of surdetermination. Schlesinger, asked by Mayer to harness himself with the question, radically modified the nature of the problem. He endogenised the prices of consumption goods that Menger and Wieser took as data and posed the equations of a system of generalised interdependence. The question of imputation thus becomes that of demonstrating the existence of a general equilibrium configuration. Schlesinger, however, did not start from the Walrassian model, but from the very similar one of Gustav Cassel (1923), in which he integrated the free goods rule in order to avoid obtaining negative prices in equilibrium. The adoption of this rule has important consequences on the formal structure of the model: inequalities are introduced into the model; inequalities are relations of exclusion which constrain the prices of goods and which have the statute of axioms in the formulations offered by the mathematicians (Abraham Wald and later von Neumann) called to the rescue to solve the new system thus defined. The introduction of inequations is typical of formalist mathematics. According to Israel and Gasca, the motto "less differential equations, more inequalities" perfectly describes the tendency of the new mathematics.

From his collaboration with Schlesinger, Wald produced three articles, presented at the Mathematical Colloquium between 1934 and 1936. Over the course of the various articles, the mathematician refined the mathematical conditions necessary for the demonstration of existence (the syntax aspect) and concentrated himself more particularly on the question of their economic significance (the semantic aspect). Von Neumann became aware of Wald's demonstrations thanks to Menger in 1934 and announced the proximity with a model of general equilibrium which he had presented a few times earlier at Princeton. Menger then made an offer to von Neumann to publish his article in Ergebnisse (1937). According to Arrow, it is extremely probable that the models of Schlesinger and Wald on one hand and of von Neumann on the other were independently inspired by Cassel. Whereas Schlesinger introduced inequalities in the static model of Cassel, and Wald showed the existence of an equilibrium solution, von Neumann's model axiomatises the verbal developments Cassel made of an economy of generalised interdependence in a situation of uniform growth. Nicholas Kaldor said, from his conversations with von Neumann, that the dissatisfaction of the mathematician with regard to the Walrassian model had a double origin: the possibility of negative prices at equilibrium and the disinterest in dynamic forces. The 1937 model answered these two criticisms appropriately by proposing a model of expansion in which the free goods rule, with the statute of axiom of the formal system, eliminated the possibility of negative prices in equilibrium. This model, however, also addressed a more general criticism to economists.

According to Leonard, the fundamental criticism of von Neumann here related to the kind of mathematical instruments used since Walras in economic formalisation. However, if one replaced the 1937 contribution within the second part of the formalist programme of Hilbert (the imperialist aspect of the programme, with its project of extension of modern axiomatics to various fields), then, more than the type of tools used, it is the concept itself of scientific rigour which seems to be at the heart of von Neumann's criticisms on the state of the discipline. Walras used the mechanical analogy with the stated aim of giving economics the scientific rigour which was lacking till that point. Walrassian economics, like the other sciences based on the mechanical analogy, adopts as scientific criterion of rigour confrontation with reality. Accordingly, a model is an economy in miniature which is sufficiently simplified to allow mathematical treatment. The adoption of the mathematical analogy radically modifies this perception. Scientific rigour is defined according to internal criteria, mainly aesthetical (von Neumann 1947); rigour becomes synonymous with purity, abstraction, and

consistency of the formal system. Certainly, scientific rigour is a relative and changing concept. Thanks to Gödel, von Neumann paid the price. In the ultimate analysis, Gödel's discoveries resounds like a bulwark against possible drift towards abstraction, of which Hilbertian formalism could be the thin end of the wedge.

Of course, these critics are not concerned specifically with economics but with the most abstract practices of mathematicians, as, for instance, in the Bourbakist programme, the radical extension of formalism. But by substituting the term "mathematical" by "economical" in the preceding quotation, the criticism remains valid to some extent, testifying to the success of the imperialist incursion of formalism in economics.

The thesis of this paper is that the 1937 article is a contribution to the mathematical formalist programme. We defined this programme around two finalities: (i) the search for certainty and (ii) the project of unifying sciences. After Gödel's discoveries, the first part of the programme has faded deeply, whereas the second aspect remains intact. At the end of our reflection, it seems to us that the 1937 article fully fits the second aspect of this programme and reflects to a certain extent its new pragmatic dimension. We indeed tried to show that, a posteriori, von Neumann's contribution fulfils a twofold motivation:

1. To find a new model of a formal system insofar as, if it is not possible to prove the consistency of a system, it is nevertheless possible to consolidate the certainty of scientists through the exhibition of a new adequate interpretation;

2. To replace the use in economics of the mechanical analogy by the mathematical analogy.

Admittedly, much has already been written on the "most important paper done in mathematical economics". It was disguised with the most various interpretations. Ours is a contribution to the more restricted set of comments which concentrate less on possible filiations of the model than on the range of the original methodological approach of the author, positioning the 1937 contribution in the formalist revolution in economics.

The Model

The Neumann Model examines the possibility of balanced growth a constant rate $\propto$ in an expending economy. The model their to find out whether. There is a maximal value of $\propto$ and a maximum uniform rate of expansion of the economy and characteristics of this highest attainable growth rate. The model assumes that the economy is engaged in the production of n can commodities. ($j = 1, 2, 3, …., n$) with m process ($i = 1, 2, 3, ….., m$). Each of the processes terms out same inputs into outputs in fixed proportions.

Suppose, the process i, is carried out at the unit level of intensity. Further, aij and bij denote the input and output respectively of the jth commodity in the ith process. Thus the economy will use aij of j commodity as an input and will produce bij of j commodity as an output. It process i does not use j commodity a an input, there aij = 0. similarly, j process i does not producing commodity, there bij = 0. But in the economy aij $\geq$ 0 big $\geq$ 0.

Now, Newmann uses a set of process intensities μ_1, μ_2, …., μ_m corresponding to m process so an to equalise the growth rate of commodities in the economy. Suppose μ, units of process i an used by the economy during a period. So the total use of j commodity or an input to process i will be aij μi and the total output of j commodity by this process will be bij μi. Thus, the total amount of j commodity used by all process will be $\sum_{i=1}^{m} bij\mu i$ and its total output will be $\sum_{i=1}^{m} bij\mu i$.

If inputs and outputs are to grow at the growth rates $\propto$ any input was $\sum\limits_{i=1}^{m} bij\mu i$ is period t will

expend to $\propto \sum\limits_{i=1}^{m} bij\mu i$ one period later. But, that input must come from the preceding period's total

supply of output $\sum\limits_{i=1}^{m} bij\mu i$. Therefore, the first condition for the Neumann model is that the inputs used

tomorrow must not exceed the supply made available today. This is expressed as

$$\propto \sum_{i=1}^{m} aij\mu i \leq \sum_{i=1}^{m} bij\mu i \qquad \qquad \dots(9.1)$$

(for all j = 1 to n)

This inequality shows that same goods grow faster with given process intensities. This leads to maximising the growth rate of the economy. According to Newmann, for this there is nothing as a negative level of operation of a process so that

$$\sum \mu i > 0$$

Neumann carrier his model further by taking a set of output price P_1, ...P_2, ..., P_n and an interest factor, $\beta = 1 + 2/100$, (when z with percentage interest rate) in a system of perfect competition. In this case, prices will be so set that any process will yield zero profits.

To explain this no-profit condition, suppose a unit of process i uses aij unity of j input and each

unit cost P_j dollar. So, the total input cost of the unit of process i will by $\sum\limits_{i=1}^{n} bij\, P_j$. But where the

inputs an used after one period, their cost, including their interest cost, will increase from, $\sum\limits_{j=1}^{n} aij\, P_j$ to

$$\beta \sum_{j=1}^{n} aij\, P_j.$$

In that period, the process will produces bij quantitive of outputs having total market value of

$\sum\limits_{j=1}^{n} bij\, P_j$. Therefore, with no profits, the total cost of the unit of process $\beta \sum\limits_{j=1}^{n} aij\, P_j$ must equal its total

money yield Newmann $\sum\limits_{j=1}^{n} bij\, P_j.$

Since, under perfect competitive, profits must be non-positives, therefore for the i process, its total cost must be grater than or equal to the sum of all revenue in that period. That leads to the second condition of the Neumann model.

$$\beta \sum_{j=1}^{n} aij\, P_j \geq \sum_{j=1}^{n} bij\, P_j \qquad \qquad \dots(9.2)$$

If inequality holder for some processes which yield only looses. (i.e., total cost is greater those total revenue) they get subjected as inefficient processes. For only such inefficient process i′, we will have $\mu_i' = 0$ which means that the process will remain unutilised. It may be expressed as

If $\beta \sum\limits_{j=1}^{n} a_{ij} \, P_j > \sum\limits_{j=1}^{n} b_{ij} \, P_j$ …(9.3)

then $\mu_i' = 0$

Leaving this exceptional can of losses, we assume that at least one process is not inefficient where one equality sign hold is equation (9.2) of the system.

The last requirement of the system what any particular commodity j will be a frie good for which its price $P_j i = 0$, if as exist quantity of this commodity is available. In other words, if its available supply from all processes exceed the quantity of that commodity which will be needed as an input in the next period it will be frie good and hence its price will be zero. This condition can be expressed as

If $\propto \sum\limits_{i=1}^{n} a_{ij} \, \mu_i < \propto \sum\limits_{i=1}^{n} b_{ij}' \mu_i$

Thus, $P_j 1 = 0$

The central point in Neumann's analysis is that there exists a unique solution to his model. The proof given by him is beyond the mathematical level of one understanding. He shows that whatever the values of the parameters of his model so long it satisfies the conditional described above there will always be set of values for μ_1 ….., μ_2, P_1, …….P_n and $\propto$ and β which satisfy the requirement of maximum growth in the model.

The model determine, the value of $\propto$ and β uniquely in the system. Neumann process that is an optimal solution, a single maximum growth rate $\propto$ of the system must equal a single minimum interest rate β of the system to caver the cost of investment in inputs. This is in keeping with the no profit condition following from the assumption of perfect competition. If some output exceeds inputs by more than $\propto$ percent per period while β, the interest cost of much process is less than $\propto$, the inter process will be profitable.

In the final analysis, there will be no sustainable growth rate, say $\propto'$, greater than the equilibrium growth rate, $\propto$. For if there were available alternative processes capable of yielding a higher growth rate, $\propto'$, then the $\propto$ growth rate would not be consistent with equilibrium. Entrepreneurs would suitable to these alternative processes because with interest rate at the old rate, β, they would make a profit. With the interest rate then raised to $\beta' = \propto'$, to eliminate this profit, the old new maximal growth rate process would only be operational at a loss.

Prof. Koopmans gives a geometrical representative of the Nenmann growth path on a two-dimentional diagram. According to him, Nenmann limited his explanation to a proportional growth path where all goods grow at the same rate and at the fastest rate.
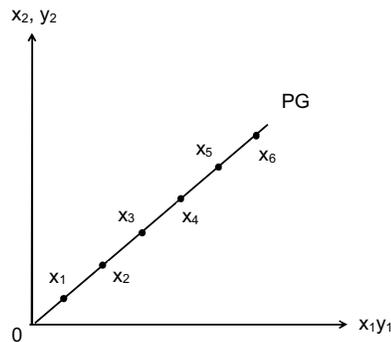
Suppose, there are two activities x_1, x_2 and y_1, y_2 where proportions are constant are time and they grow at the same rate. They are measured an x axis and y axis respectively. In fig (9.1), their proportional growth path is represented by a rising straight line PG where the output pair of each period's activity equal the input pair for that of the next period. PG is a unique maximal growth path of the system which is called the proportional growth path.

9.2 TURNPIKE THEOREMS – SAMUELSON AND TURNPIKE RESULT

In the current study, we investigate efficient capital accumulation in a stochastic neo-classical aggregate growth model. The underlying uncertainty is driven by Brownian-motion shocks and the major results do not rely on the specification of production functions. The stochastic balanced path of the capital-labour ratio is naturally derived by a Martingale, and the corresponding modified Golden Rule path of capital accumulation is well-defined. The powerful Martingale Theory is thus employed, and a stochastic Turnpike theorem involving the modified Golden Rule is proved, i.e., the underlying path of capital accumulation is asymptotically efficient in the sense of consumption maximisation. We focus on asymptotic Turnpike theorems and our Turnpike theorem only relies on the requirement that the modified Golden Rule Path of Capital Accumulation is reachable in any almost surely finite Markov time. Finally, it is asserted that the modified Golden Rule Path of Capital Accumulation indeed provides us with a robust Turnpike under very weak assumptions.

Merton (1975) extends the one-sector neo-classical growth model of Solow type to stochastic cases where the dynamics of capital-labour ratio is driven by a diffusion process, thereby providing us with an asymptotic theory of economic growth under uncertainty. Later on, Chang and Malliaris (1987) prove a theorem that confirms the existence and uniqueness of the stochastic growth path derived by Merton under certain assumptions. Therefore, noting the important and interesting properties reflected by Merton's model, the motivation of present exploration is to derive a well-defined modified Golden Rule Path of Capital Accumulation and establish corresponding Turnpike theorem based upon Merton's framework and also the theorem demonstrated by Chang and Malliaris. In other words, the current study enriches Merton's model and conclusion by uncovering a robust Turnpike theorem involving the modified Golden Rule implicitly implied by the basic model.

In deterministic neo-classical models, Golden Rule or modified Golden Rule is usually derived through the balanced path of capital-labour ratio. Similarly, the present modified Golden Rule is established via letting the drift term of the diffusion process of capital-labour ratio be equal to zero, thereby producing a Martingale Path of Capital Accumulation. That is, we define the stochastic balanced path of capital-labour ratio of the current stochastic neo-classical model by the Martingale Path of Capital-Labour Ratio. As a matter of fact, there is a natural one-to-one correspondence

between the modified Golden Rule and the martingale-path of capital-labour ratio. Consequently, it is ensured that the modified Golden Rule derived through the stochastic balanced path, i.e., the martingale path, of capital-labor ratio is well defined. And this creates a natural opportunity such that the powerful martingale theory can be appropriately employed to demonstrate the turnpike theorem. Rather, the present turnpike theorem shows that the martingale-path of capital-labour ratio will converge to the modified Golden Rule almost surely and in the sense of uniform topology (Dai, 2012) as long as the modified Golden Rule is reachable in any almost surely finite time. And one can easily notice the differences between the present result and those proved by Cass (1966) and Samuelson (1965) in deterministic neoclassical aggregate growth models.

When we define the concept of capital in a very broad sense, i.e., including human capital, health capital, environmental capital, and so on, then capital accumulation indeed plays a crucial role in modern economic growth. The major contribution can be summarized as follows:

1. Since Brownian motion shocks are widely used in continuous-time stochastic growth models, we provide an appropriate definition of the modified Golden Rule Path of Capital Accumulation in such kind of economies and this definition does not rely on the explicit specification of the Brownian shocks.

2. We develop a systematic mathematical method for providing robust Turnpike theorems in such kind of circumstances involving the above modified Golden Rule Path of Capital Accumulation, and we believe that our method is general enough to be used in other related environments. In other words, our contribution mainly focuses on theoretical issues of macroeconomic growth theory.

There are some related literatures. As is argued by Yano (1985), existing Turnpike theorems in optimal growth theory can be summarised as the following two types:

1. Neighbourhood Turnpike theorem which asserts that an optimal path in a growth model converges to a small neighbourhood of a stationary path.

2. Asymptotic Turnpike theorem which means that an optimal path converges to a stationary path. It focuses on asymptotic Turnpike theorem and we have confirmed the corresponding robustness in a continuous-time stochastic growth model.

On the other hand, it is well-known that the Golden Rule Path has been playing a very important role in neo-classical theory of capital accumulation starting from the pioneering papers of Phelps. Recently, Schenk-Hoppe (2002) also studied the Golden Rule in stochastic Solow growth model. Schenk-Hoppe employs dynamical systems theory, especially the concept of a random fixed point, to prove the existence of a Golden Rule Savings Rate for the stochastic Solow model. Methodologically, in studying the Golden Rule Path of Capital Accumulation, Brock and Mirman (1972) used the classical stochastic stability theory of Markov chains while Bayer and Walde (2011) expanded their distributional analysis by using the stability theory for Markov processes in continuous time. We heavily employ Martingale theory, which depends on continuous-time Markov processes driven by Brownian motions in the present economy, to demonstrate the corresponding Turnpike theorem involving the modified Golden Rule Path of Capital Accumulation. This method can be regarded as a useful complement to existing literatures involving the issue of efficient capital accumulation under uncertainty.

The Model

The major goal of the model is to introduce the stochastic path of capital accumulation in a one-sector neo-classical growth model with the uncertainty coming from the population size L(t), i.e., following Merton,

$$dL(t) = nL(t)dt + \sigma L(t)dB(t) \qquad \qquad …(9.4)$$

which is based upon the underlying filtered probability space (Ω, F, $\{F_t\}_{0 \le t \le T}$, P) with E denoting the expectation operator depending on

$$f_o \triangleq \{W\!\!\text{Æ}\}.$$

As usual, the neo-classical production function Y(t) = F(K(t), L(t)) is assumed to be strictly concave, homogeneous of first degree and also exhibit constant returns to scale with the law of motion of capital accumulation expressed as follows:

$$\frac{dK(t)}{dt} = F(K(t), L(t)) - \delta K(t) - C(t) \qquad \qquad …(9.5)$$

in which δ, an exogenously given constant, denotes the depreciation rate and C(t) represents aggregate consumption at time t.

Now, combining (9.4) with (9.5) and applying the classical rule yields the following SDE of capital-labour ratio,

$$dk(t) = [f(k(t)) - (\delta + n - \sigma^2) k(t) - c(t)]dt - \sigma k(t)dB(t) \qquad \qquad …(9.6)$$

Subject to $k(0) \equiv k_0 > 0$, a deterministic constant, and $f(k(t)) \triangleq Y(t)/L(t)$, $c(t) \triangleq C(t)/L(t)$ stand for per capita output and per capita consumption, respectively at time t. Specifically, for the SDE of capital-labour ratio given by (9.6), Chang and Malliaris (1987) proved the following result:

Proposition 1:

If the production function f is strictly concave, continuously differentiable as [0, ∞], f(0) = 0 and

$$\lim_{k(t) \to \infty} f'(k(t)) \triangleq \lim_{k(t) \to \infty} \frac{df(k(t))}{dk(t)} = 0$$

In order to make things much easier, we need,

Assumption 1: The assumptions or conditions given by Proposition 1 are assumed to be fulfilled throughout this section.

Turnpike Theorem

Here, we will derive a modified Golden Rule and establish the corresponding Turnpike theorem under relatively weak conditions. For the SDE of capital-labour ratio given by (9.6), we denote the drift term by

$$b(t) \triangleq f(k(t) - (\delta + n - \sigma^2) \, k(t) - c(t) \qquad \qquad \text{…(9.7)}$$

Which implies that the capital-labour ratio $k(t)$ tends to increase if $b(t) > 0$ and the capital-labour ratio tends to decrease if $b(t) < 0$. Golden Rule or modified Golden Rule is usually derived via the balanced path of capital-labour ratio in the deterministic case. Similarly, we derive the modified Golden Rule by $b(t) = 0$, which corresponds to the stochastic balanced path of capital-labour ratio and gives rise to

$$c(t) = f((k)(t)) - (\delta + n - \sigma^2) \, k(t) \qquad \qquad \text{…(9.8)}$$

Hence, the corresponding stochastic Golden Rule k^* is determined by

$$f'(k^*) = \delta + n - \sigma^2 \qquad \qquad \text{…(9.9)}$$

Meanwhile, substituting (9.8) into (9.6) leads us to

$$dk(t) = -\sigma k(t) dB(t) \qquad \qquad \text{…(9.10)}$$

Which hence defines a Martingale path of capital-labour ratio. Now, we can establish:

Theorem 1 (Turnpike theorem): If the following Markov time,

$$\tau(w) \triangleq \inf\{t \geq 0; \, k(t) = k^*\} < \infty \text{ a.s.}$$

then we get that the Martingale path of capital-labour ratio given by (9.10) will strongly converge to the stochastic Golden Rule k^* given by (9.9) a.s. and in the sense of uniform topology.

Proof: By the Doob's Martingale Inequality,

$$P\left(\sup_{0 \leq t \leq T} |k(t)| \geq \lambda\right) \leq \frac{1}{\lambda} E\left[|k(T)|\right] \frac{k_0}{\lambda}, \, \forall \lambda > 0, \, \forall T > 0$$

Without loss of generality, we put $\lambda = 2^m$ for $m \in N$, then,

$$P\left(\sup_{0 \leq t \leq T} |k(t)| \geq 2^m\right) \leq \frac{1}{2^m} k_0, \, \forall m \in N, \, \forall T > 0$$

Using the well-known Borel-Cantelli Lemma, we arrive at,

$$P\left(\sup_{0 \leq t \leq T} |k(t)| \geq 2^m \text{ for infinitely many m}\right) = 0, \, \forall T > 0$$

So far a.a. $w \in \Omega$, there exists in $\overline{m}(\omega) \in N$ such that,

$$\sup_{0 \leq t \leq T} |k(t)| < 2^m \text{ a.s. for } \overline{m} \geq m(\omega), \, \forall T > 0 \qquad \qquad \text{…(9.11)}$$

Thus, $k(t) = k(t, \omega)$ is uniformly bounded for $t \in [0, T]$, $\forall T > 0$ and for a.a. $\omega \in \Omega$.

Define $B_{2-m}(\tau^*(\omega)) \triangleq \{\tau(\omega) \geq 0; |\tau(\omega) - \tau^*(\omega)| < 2^{-m}\} \, \forall m \in N$

Thus, for $\forall \tau^m \in B_{2-m}(\tau^*)$, and based on the assumption that $\tau^*(\omega) < \infty$ a.s., applying Doob's optimal sampling theorem and Doob's Martingale inequality lead us to,

$$P\left(\sup_{0 \le t \le \tau} \left| k(t) - k* \right| \ge \epsilon\right) \le \frac{1}{\epsilon} E\left[\left| k(\tau''') - k* \right|\right], \forall \epsilon > 0$$

According to (8) and the continuity of Martingale with respect to time t, an application of Lebesgue bounded convergence theorem shows,

$$\limsup_{m \to \infty} P\left(\sup_{0 \le t \le \tau'''} \left| k(t) - k* \right| \ge \epsilon\right) \le \frac{1}{\epsilon} \limsup_{m \to \infty} E\left[\left| k(\tau''') - k* \right|\right] = 0, \forall \epsilon > 0$$

which yields

$$\limsup_{m \to \infty} P\left(\sup_{0 \le t \le \tau'''} \left| k(t) - k* \right| < \epsilon\right) = 1, \forall \epsilon > 0$$

It follows, from Fatou's Lemma that,

$$P\left(\sup_{0 \le t \le \tau*} \left| k(t) - k* \right| < \epsilon\right) = 1, \forall \epsilon > 0$$

Thus, we get,

$$\sup_{0 \le t \le \tau*} \left| k(t) - k* \right| < \epsilon, \text{a.s. for } \forall \epsilon > 0$$

i.e., $\limsup_{\tau* \to \infty} \sup_{0 \le t \le \tau*} \left| k(t) - k* \right| < \epsilon, \text{a.s. for } \forall \epsilon > 0$

Noting the arbitrariness of ϵ, the required assertion follows.

Next, the proceed to analyse the robustness of the Turnpike theorem given by Theorem 1, i.e., we show that the modified Golden Rule k* indeed provides in with a robust Turnpike under relatively wave assumptions. Based upon the Martingale path given by (9.7), we set

$$d\tilde{k}(t) = -\tilde{\sigma}\tilde{k}(t)dB(t) \qquad \qquad \dots(9.12)$$

subject to $\tilde{k}(0) \overset{\Delta}{=} k_0 > 0,$ a deterministic constant, such that,

$$\left| \sigma - \tilde{\sigma} \right| \le \xi$$

for any non-zero constant σ and $\tilde{s}$ with $\left| \sigma \right| \vee \left| \tilde{\sigma} \right| < \infty.$

As preparations, we need the following two lemmas,

Lemma 1: There exist constants $e(k_0, p, T) < \infty$ and $\tilde{e}(k_0, p, T) < \infty$

(i) $E\left(\sup_{0 \le t \le T} \left| k(t) \right|^p\right) \le e(k_0, p, T)$

and

(ii) $E\left(\sup_{0 \le t \le T} \left| \tilde{k}(t) \right|^p\right) \le \tilde{e}(k_0, p, T)$

for $\forall_0 < T < \infty, \forall_p \in N, p \ge 2$

Lemma 2: For k(t), defined in (9.7) and $\tilde{k}(f)$ defined in (9.9), one can get that,

(i) $E\left(\sup_{0 \le t \le T} |k(t)|^p\right) \le \left(\frac{p}{p-1}\right)^p \lambda^{p-1} k_0$

and $E\left(\sup_{0 \le t \le T} |\widetilde{k}(t)|^p\right) \le \left(\frac{p}{p-1}\right)^p \widetilde{\lambda}^{p-1} k_0$

for $\forall_0 < \lambda < \infty$, $\forall_0 < \widetilde{\lambda} < \infty$, $\forall_0 < T < \infty$ and $\forall_p \in N$, $p \ge \infty - 2$

Thus, the following proposition is derived.

Proposition 2: Based upon the above assumptions, and Lemma 1 or 2, then we have,

$E\left[\lim_{\tau > \infty} \sup_{0 < t \le \tau} |k(t) - \widetilde{k}(t)^2|\right] \to 0 \text{ as } \xi \to 0$

Both Turnpike theorems and the Phelps-Koopmans Theorem play very important roles in macroeconomics. Turnpike theory characterises the mathematical properties of the equilibrium or optimal path of resource allocation while the classical Phelps-Koopmans Theorem clearly uncovers that the efficient path of capital accumulation will definitely converge to the Golden Rule in the long run. Otherwise, dynamically inefficient accumulation happens. Samuelson (1965) proved a neighbourhood Turnpike theorem involving the Golden Rule in the classical Ramsey (1928) model, while the present paper demonstrates an asymptotic Turnpike theorem involving the modified Golden Rule in a stochastic neo-classical growth model,which implies that the path of capital accumulation is dynamically efficient.

Finally, it is also confirmed that the modified Golden Rule Path of Capital Accumulation is indeed a robust Turnpike. Some interesting extensions can be taken into account in future research. For example, since we only proves an asymptotic Turnpike theorem in the underlying economy, one interesting and possible extension is to find out conditions supporting a neighbourhood Turnpike theorem for neighbourhood efficiency characterisation of stochastic capital accumulation. Moreover, notice that the (modified) Ramsey rule also plays a crucial role in savings behaviour and macroeconomic growth. It is interesting to investigate the corresponding Turnpike theorems involving Ramsey rules by effectively employing the mathematical method developed in the present study. As a final point, if we are motivated to investigate the effect of stochastic TFP imposed on the efficient path of capital accumulation, geometric Brownian motion can be naturally employed with the purpose of introducing technology fluctuation into the underlying economy.

The Turnpike Theorems in then Deterministic Setting

It is important to distinguish the asymptotic and neighbourhood turn pike theorems for solutions to discounted finite-horizon optimal control. Problems from the classical. Turnpike theorem for solutions to undiscounted finite-horizon problem.

The classical turnpike theorems considers a deterministic finite-horizon (T > 0) optimal control problem on a given state space x. In a discrete time case, the problem is

$$\begin{cases} \text{Maximise} & \sum_{t-0}^{T-1} L(x_t, x_{t+1}) \\ \text{Subject to} & x_0, x_T \text{ EX given, } \{xt\}_{t}^{T} = 0^{CX} \end{cases} \qquad \dots(9.13)$$

with a cost (utility) function $L : X^2 \to R$, $T \in N$. In a continuous time case, the problem is

$$
\begin{cases}
\text{Maximise} & \int_0^T L(x_t, u_t)dt \\
\\
\text{Subject to} & xt = t(xt, ut),\ ut \in U \qquad\qquad \text{...(9.14)} \\
& x_0,\ x_t \in X\ \text{given},\ \{x_t\}_t \in (0, T)^{CX}
\end{cases}
$$

with a cost (utility) function $L : X \times xU \to R$, $T \in N$. Where $h : X \times U \to R$ is a continuous mapping with respect to both arguments, U is the set of admissible controls and $T \in Rt$. The initial value x_0 and the terminal value x_T are hold fixed. Let x^* denote a solution of the maximization problem.

Maximise $L(x, x)$, $x \in X$

Then path $x^* := \{x_t : x_t = x^*,\ \forall\, t \in (0,\ T)\}$ is called a stationery optimal path. Often, the uniqueness of the stationary optimal path is assumed.

In the case, where the terminal value x_T is not given, an additional term, inter-prated an the terminal pay off function, should be added to the above problems. That is, the optional problem is in a discrete time setting.

$$
\begin{cases}
\text{Maximise} & \sum_{t=0}^{T-1} L(x_t, x_{t+1}) + S(x_T) \qquad\qquad \text{...(9.15)} \\
\\
\text{Subject to} & x_0\ \text{given},\ \{xt\}_t^T = 0^{CX}
\end{cases}
$$

and, is a continuous time case

$$
\begin{cases}
\text{Maximise} & \int_0^T L(x_t, u_t)dt + S(x_T) \\
\\
\text{Subject to} & x_t = L(x_t, u_t),\ u_t \in U \qquad\qquad \text{...(9.16)} \\
& x_0\ \text{given},\ \{x_t\}_t \in [0,\ T]^{CX}
\end{cases}
$$

Where, the terminal payoff function $S : X \to R$.

The original formulation of the continuous time problem (9.14) is ascribed to Samuelson – slow and the discrete time problem to Gala and Mekanzid. It was further investigated by Raduer, Morishima etc. and they discussed is detail about the development of the turnpike theory in both the Von Neumann growth model and the Ramsuj's growth model and specified a classification of samuelson turnpike and Ramsay turnpike. The classical techniques which were developed to prove the classical turnpike theorems are critically dependent on the convexity of the control set u and technology function u and the preference function L. These classical techniques fall into two classes: value loss method and monotonicity method.

The classical turnpike theorem states that if T is large enough, the optimal path $\left\{x_t^T\right\}$. That transfers this system from x_0 to x_y approaches the unique optimal stationary level x^* and stays close to it for a large fraction of T and mover a way toward the terminal states only is the final periods, that is, the for each $\in > 0$. there exist an integer to such that for each $T \geq 2T_0$

$$\| x_t^* - x^* \| \leq \in, \forall t \in [T_0, T - T_0] \qquad\qquad \text{...(9.17)}$$

This is called the middle Turnpike theorems. However, if this result holds for any initial value $x_0 \in x$, it is the global version; if the result holds for some initial value x_0 which is near enough to x, it is the local version further, if for each $\in > 0$, There exists an integer T_0 and a number $d > 0$ such that

$\left\| x_0 - x \right\| \leq d$ implies (2.5) for all $t \in 0\,[0,\, T - T_0]$ with $T \geq 2\,T_0$, this was called the early turnpike theorems.

A version of the classical turnpike theorems for games was also presented. For instance, two-player zero sum differential game were studied by zaslavski, showing that there exists a pair of optional stationery paths x and y for the two players and presenting a version of classical middle turnpike theorem and by Alvariz and Bardi proving a turnpike theorem for the value of the game.

In the classical setting, the state space x was often a compact, convex bounded and closed subset of a finite dimensional space; the cost (utility) function L is often continuous smooth and strictly concave (or convey) hence the turnpike theorems was proved by the concavity analysis. Later under upper (lower) semicontinuity, some extensions to general Banach state space were obtained by e.g; kolokoltsov, Yakovenko etc.

When the future is discounted the deterministic optimal control problem with infinite horizon is described, is a disctete time setting as

maximise $\sum_{t=0}^{\alpha}\ \delta_L^e\ (x_t,\, x_{t+1})$

subject to $x_0 \in x,\ \{x_t\}_{t=0}^{\alpha}\ C_x$ …(9.18)

and is a continuous time setting as

maximise $\int_0^{\alpha - Pt_L} (xt,\, ut)dt$

Subject to $x_t = L\,(x_t,\, u_t),\ u_t \in U$ …(9.19)

$x_0 \in x,\ \{x_t\}_t \geq 0\ C\ x$

With a discounts factor $0 < \delta \leq 1$ and a discount rate $P \geq 0$. The case where $\delta = 1$ and $P = 0$ is reforred to a the undiscounted version of the problem (9.18) and (9.19). The asymptotic behaviour of the optimal path in the discounted framework depends critically on the magnitude of the initial stock and the discount factor.

Two classes of turnpike theorems should be distinguished where the future is discounted. It status that there exists a discount factor $\delta^1(0,\, 1)$ such that for x_0 x and any $\delta[\delta',\, 1]$ the optimal path $\{x_t\}$ starting at x_0 converge to the stationary path x

The asymptotic turnpike theorems was proved under strictly concavity assumption on L by e.g; -Schin kman, Bewly etc. interpreted the function as the consumption and set $L\,(x_t,\, x_t+1) := f(x_t) - x_t + 1 > 0$ for all t and considered a non convex (convex - concave) technology function f, i.e., f is assumed to be strictly increasing and twice continuously differentiable such that f statisfies $f'(\alpha) < 1 < f^1(0) < \alpha$ and there is a real number $k_1 > 0$ such that $f^1(x) = 0$ for $x = k_1$, $f^1(x) > 0$ for $0 \leq x < k$, and $f^{11}(x) < 0$ for $x > k_1$.

Mastrucchio assumed certain curvature restriction, i.e., The function L is continuous and Arandy $(\alpha,\, \beta)$ concave i.e.; $L\,(x,\, y) + \dfrac{1}{2} \propto\ |x|^2 + \dfrac{1}{2}P|y|^2$ is concave with $\alpha + \beta > o$ and that the value function

$w\,(x_0) = \max \sum_{t=0}^{\alpha}\ \delta^t\ L(x_t\ x_{t+1})$ is concave -r for all $\delta\ (\delta_0,\, 1)$, i.e., $w\,(x) + \dfrac{1}{2}\,r\ |x|^2$ is convex for all $\delta\ (\delta_0,\, 1)$. He proved that there exists $\phi\ f' = 1 - \dfrac{\alpha + \beta}{r}$ such that for a discount factor $\delta\ (\max\ [\delta_0,\, 1 - \dfrac{\alpha + \beta}{r}],\, 1)$ the local asymptotic turnpike theorem holds, where $\propto + \beta$ i.e., measure of the lower curvature of L

and r is a measure of its upper curvature. Carlso and Hawris proved version of the asymptotic turnpike theorem for infinite horizon open-loop differential games with $\delta = 1$ under a strict diagonal concavity condition on L an uniform asymptotic turnpike theorem was proved by Cuang Levan and Lisa Morhasim under similar assumptions, i.e; if the function $L(x, y)$ is a strictly concave function, increasing in x and decreasing in y, them for any initial point $x_0 \in x$, there exist $\beta \in (0, 1)$ such that for any $P \in [\beta', 1]$ the asymptotic turnpike properly holds.

A Aranyo applied the implicit function theorem to prove the asymptotic turnpike theorem by assuming a dominant diagonal blocks condition for an infinite matrix B formed by n × n blocks Dij i, j $\in \{1, 2, ...\}$ D_{ii} is invertible $\sup_i \left| D_{ii}^{-1} \right| < \propto$ and $\sup_{ii} \sum_{j=1, i \neq y}^{\alpha} \left| D_{ii} D_{ij} \right| < 1$ where the norm

$$\left| D_{ij} \right| = \sup_{|z|=1} \left| D_{ij} z \right|$$

The second class is the neighborhood turnpike theorem status that for any $\in > 0$, there exists $T > 0$ and $\delta'(0, 1)$ such that, an optimal paths stating them x_0 at the discrete factor $\delta(\delta', 1)$ eventually stage with is the $\in$ neighbourhood of a stationary path x, i.e,

$$\left\| x_\varepsilon - x \right\| \leq \varepsilon, \forall t \geq T$$

If a smaller neighborhood $\in$ is chosen, the closer the discount factor δ is to 1.

9.3 SUMMARY

1. In the immediate post-war years, Mark Blaug identified the emergence of a new paradigm in economics, the so-called "formalist paradigm", which marked the arrival of the pre-eminence of (mathematical) form over (theoretical) content, and which is mostly characterised by the crucial importance economists give to a specific (non-constructive) kind of demonstration of existence of equilibrium.

2. In the current study, we investigate efficient capital accumulation in a stochastic neo-classical aggregate growth model.

3. Merton (1975) extends the one-sector neo-classical growth model of Solow-type to stochastic cases where the dynamics of capital-labour ratio is driven by a diffusion process, thereby providing us with an asymptotic theory of economic growth under uncertainty.

9.4 SELF ASSESSMENT QUESTIONS

1. Critically explain 'Multisector Growth Models: Von-Neumann growth model and concept of efficiency and optimisation for Von-Neumann model'.

2. Critically explain 'Turnpike Theorems – Samuelson and Turnpike Result'.

❧❦❖❧❧