MICRO ECONOMICS I

Study material

CORE COURSE (I)

For

I SEMESTER BA ECONOMICS

(2011 Admission)

UNIVERSITY OF CALICUT

SCHOOL OF DISTANCE EDUCATION

CALICUT UNIVERSITY P.O. MALAPPURAM, KERALA, INDIA - 673 635

261

UNIVERSITY OF CALICUT

SCHOOL OF DISTANCE EDUCATION

Study material

CORE COURSE (I)

For I SEMESTER BA ECONOMICS

MICRO ECONOMICS I

Prepared by

1. **Module I and V** : **Dr. P. CHACKO JOSE**
   Associate professor,
   Department of Economics,
   Secred Heart College Chalakkudy

2. **Module II** : **Dr. K. RAJAN**
   Associate Professor,
   M.D. College, Pazhanji.,
   Thrissur-680 542.

3. **Module III** : **Sri. K.P. SHABEER**
   Assistant Professor,
   Government College, Kodanchery.

4. **Module IV** : **Dr. M.P. ABDULLA**
   Associate Professor,
   E.M.E.A. College, Kondotty.

Edited & Scrutinised by : **Dr.C.KRISHNAN**
Associate Professor,
Department of Economics,
Govt. College, Kodanchery

CONTENTS

Module -I
Introduction to Social Sciences

What are social sciences?

1.1. Introduction

The social sciences are one of three divisions of science, along with natural science and formal science. Social science is, in its broadest sense, the study of society and the manner in which people behave and influence the world around us. The social sciences are a group of academic disciplines that study the human aspects of the world. They diverge from the arts and humanities in that the social sciences emphasize the use of the scientific method and rigorous standards of evidence in the study of humanity, including quantitative and qualitative methods. The social sciences are also known pejoratively as the soft sciences in contrast to the hard sciences. Social science theories typically deal with aggregated, not individual, behaviour. Social science investigations examine an individual's relationship with and interaction in society.

Social science -cultural definition

- The study of people living together in groups, families, etc., and their customs, activities, etc.
- The study of how groups of people behave, often in an effort to predict how they will behave in the future.
- A branch of science that deals with the institutions and functioning of human society and with the interpersonal relationships of individuals as members of society
- A science (as economics or political science) dealing with a particular phase or aspect of human society

Academic Definition

The social sciences are a group of academic disciplines that study human aspects of the world. They diverge from the arts and humanities in that the social sciences tend to emphasize the use of the scientific method in the study of humanity, including quantitative and qualitative methods. The social sciences, in studying subjective, inter-subjective and objective or structural aspects of society, are sometimes referred to as soft sciences. This is in contrast to hard sciences, which may focus exclusively on objective aspects of nature. Social scientists engage in research and theorize about both group and individual behaviours.

1.2. History of Social Sciences

Let us see a chronological time line of the history of social sciences.

Ancient Greece

In ancient philosophy, there was no difference between mathematics and the study of history, poetry or politics. Only with the development of mathematical proof did there gradually arise a perceived difference between "scientific" disciplines and others, the "humanities" or the liberal arts. Thus, Aristotle studies planetary motion and poetry with the same methods, and Plato mixes geometrical proofs with his demonstration on the state of intrinsic knowledge.

The Enlightenment

This unity of science as descriptive remains, for example, in the time of Thomas Hobbes who argued that deductive reasoning from axioms created a scientific framework, and hence his *Leviathan* was a scientific description of a political commonwealth. What would happen within decades of his work was a revolution in what constituted "science", particularly the work of Isaac Newton in physics. Newton, by revolutionizing what was then called "natural philosophy", changed the basic framework by which individuals understood what was "scientific". While he was merely the archetype of an accelerating trend, the important distinction is that for Newton, the mathematical flowed from a presumed reality independent of the observer, and working by its own rules. For philosophers of the same period, mathematical expression of philosophical ideals was taken to be symbolic of natural human relationships as well: the same laws moved physical and spiritual reality. For examples see Blaise Pascal, Gottfried Leibniz and Johannes Kepler, each of whom took mathematical examples as models for human behavior directly. In Pascal's case, the famous wager; for Leibniz', the invention of binary computation; and for Kepler, the intervention of angels to guide the planets. In the realm of other disciplines, this created a pressure to express ideas in the form of mathematical relationships. Such relationships, called "Laws" after the usage of the time (see philosophy of science) became the model which other disciplines would emulate.

19th Century

Auguste Comte (1797-1857) argued that ideas pass through three rising stages, Theological, Philosophical and Scientific. He defined the difference as the first being rooted in assumption, the second in critical thinking, and the third in positive observation. This framework, still rejected by many, encapsulates the thinking which was to push economic study from being a descriptive to a mathematically based discipline. Karl Marx was one of the first writers to claim that his methods of research represented a scientific view of history in this model. With the late 19th century, attempts to apply equations to statements about human behaviour became increasingly common. Among the first were the "Laws" of philology, which attempted to map the change over time of sounds in a language.

It was with the work of Darwin that the descriptive version of social theory received another shock. Biology had, seemingly, resisted mathematical study, and yet the Theory of Natural Selection and the implied idea of Genetic inheritance - later found to have been enunciated by Gregor Mendel, seemed to point in the direction of a scientific biology based, like physics and chemistry, on mathematical relationships.

20th Century

In the first half of the twentieth century, statistics became a free-standing discipline of applied mathematics. Statistical methods were used confidently, for example in an increasingly statistical view of biology. The first thinkers to attempt to combine inquiry of the type they saw in Darwin with exploration of human relationships, which, evolutionary theory implied, would be based on selective forces, were Freud in Austria and William James in the United States. Freud's theory of the functioning of the mind, and James' work on experimental psychology would have enormous impact on those that followed. Freud, in particular, created a framework which would appeal not only to those studying psychology, but artists and writers as well.

One of the most persuasive advocates for the view of scientific treatment of philosophy would be John Dewey (1859-1952). He began, as Marx did, in an attempt to weld Hegelian idealism and logic to experimental science, for example in his "Psychology" of 1887. However, it is when he abandoned Hegelian constructs, and joined the movement in America called Pragmatism, possibly under the influence of William James' "Principles of Psychology" that he began to formulate his basic doctrine, enunciated in essays such as "The Influence of Darwin on Philosophy" (1910).

This idea, based on his theory of how organisms respond, states that there are three phases to the process of inquiry:

1. Problematic Situation, where the typical response is inadequate.
2. Isolation of Data or subject matter.
3. Reflective, which is tested empirically.

With the rise of the idea of quantitative measurement in the physical sciences, for example Lord Rutherford's famous maxim that any knowledge that one cannot measure numerically "is a poor sort of knowledge", the stage was set for the conception of the humanities as being precursors to "social science." This change was not, and is not, without its detractors, both inside of academia and outside. The range of critiques begin from those who believe that the physical sciences are qualitatively different from social sciences, through those who do not believe in statistical science of any kind, through those who disagree with the methodology and kinds of conclusion of social science, to those who believe the entire framework of scientificizing these disciplines is solely, or mostly, from a desire for prestige and to alienate the public.

Theodore Porter argued in "The Rise of Statistical Thinking" that the effort to provide a synthetic social science is a matter of both administration and discovery combined, and that the rise of social science was, therefore, marked by both pragmatic needs as much as by theoretical purity. An example of this is the rise of the concept of Intelligence Quotient, or IQ, a test which produces a number which it is not clear what, precisely, is being measured, except that it has pragmatic utility in predicting success in certain tasks.

The rise of industrialism had created a series of social, economic, and political problems, particularly in managing supply and demand in their political economy, the management of resources for military and developmental use, the creation of mass education systems to train individuals in symbolic reasoning and problems in managing the effects of industrialization itself. The perceived senselessness of the "Great War" as it was then called, of 1914-1918, now called World War I, based in what were perceived to be "emotional" and "irrational" decisions, provided an immediate impetus for a form of decision making that was more "scientific" and easier to manage. Simply put, to manage the new multi-national enterprises, private and governmental, required more data. More data required a means of reducing it to information upon which to make decisions. Numbers and charts could be interpreted more quickly and moved more efficiently than long texts.

In the 1930s this new model of managing decision making became cemented with the New Deal in the US, and in Europe with the increasing need to manage industrial production and governmental affairs. Institutions such as The New School for Social Research, International Institute of Social History, and departments of "social research" at prestigious universities were meant to fill the growing demand for individuals who could quantify human interactions and produce models for decision making on this basis. Coupled with this pragmatic need was the belief that the clarity and simplicity of mathematical expression avoided systematic errors of holistic thinking and logic rooted in traditional argument. This trend, part of the larger movement known as Modernism provided the rhetorical edge for the expansion of social sciences.

Present state

There continues to be little movement toward consensus on what methodology might have the power and refinement to connect a proposed "grand theory" with the various midrange theories which, with considerable success, continue to provide usable frameworks for massive, growing data banks.

1.3. Evolution of Social sciences

The social sciences have existed at least since Ancient Greece, where philosophers such as Plato and Aristotle studied numerous aspects of the world and passed them down via texts. To these thinkers, there was no fundamental distinction between social and natural science the way there is today. Disciplines such as geometry and psychology were intermixed and practiced by the same communities. During the 18th century there was a distinction made between the different types of sciences studied. Natural sciences were defined as sciences that are experimental and applied, whereas the social sciences are those that grew from moral philosophy. The generally accepted branches of social science include anthropology, economics, history, political science, psychology and sociology. Some additions to these have been made since their development to include education, geography, law, linguistics, criminology and archaeology. These all examine and study man's interactions with his fellow man and with his society.

The history of the social sciences has origin in the Western philosophy and shares various forerunners, but began most intentionally in the early 19th century with the positivist philosophy of science. Since the mid-20th century the term "social science" has come to refer more generally, not just to sociology, but to all those disciplines which analyse society and culture; from anthropology to linguistics to media studies.

The idea that society may be studied in a standardized and objective manner, with scholarly rules and methodology, is comparatively recent. Whilst there is evidence of early sociology in medieval Islam, and whilst philosophers such as Confucius had long since theorised on topics such as social roles, the scientific analysis of "Man" is peculiar to the intellectual break away from the Age of Enlightenment and toward the discourses of Modernity. Social sciences came forth from the moral philosophy of the time and were influenced by the Age of Revolutions, such as the Industrial revolution and the French revolution. The beginnings of the social sciences in the 18th century are reflected in various grand encyclopaedia of Diderot, with articles from Rousseau and other pioneers. The growth of the social sciences is also reflected in other specialized encyclopaedias. In the modern period, the term "*social science*" first used as a distinct conceptual field.

Around the turn of the 20th century, Enlightenment philosophy was challenged in various quarters. After the use of classical theories since the end of the scientific revolution, various fields substituted mathematics studies for experimental studies and examining equations to build a theoretical structure. The development of social science subfields became very quantitative in methodology. Conversely, the interdisciplinary and cross-disciplinary nature of scientific inquiry into human behaviour and social and environmental factors affecting it made many of the natural sciences interested in some aspects of social science methodology. Examples of boundary blurring include emerging disciplines like social studies of medicine, socio-biology, neuro-psychology, bio-economics and the history and sociology of science. Increasingly, quantitative and qualitative methods are being integrated in the study of human action and its implications and consequences. In the first half of the 20th century, statistics became a free-standing discipline of applied mathematics. Statistical methods were used confidently.

In the contemporary period, there continues to be little movement toward consensus on what methodology might have the power and refinement to connect a proposed "grand theory" with the various midrange theories which, with considerable success, continue to provide usable frameworks for massive, growing data banks.

The idea of social science is distinctively modern. Four developments set the stage for its emergence between the 17th and 19th centuries. First, the 17th century revolution in science was pivotal. It generated the notion of science as a cumulative empirical project, and complemented this with an ethos favouring the public sharing of knowledge and the foundation of social institutions to further both inquiry and publication. Science, in this new sense, combined inductive inquiry with explicit testing of propositions and formulation of theories based on empirical evidence. Second, the rise of the modern state (in both its domestic and colonial forms) gave social science both a topic and a client. States sought knowledge as the basis for policy. And the state itself could be an important object of science, as scholars sought to understand which policies worked and which did not, what factors made for better rule, and what organization of the state advanced human liberty. Closely related, the notion of nation as a pre-political definition of the people who rightly belonged in a given state, helped frame "society" as bounded, integrated, and developing through history. Third, the dramatic expansion of trade, division of labour, industry, and capital accumulation that marked the modern era provided both an impetus to study society and a basis for differentiating directly societal sources of change and self-organization from the effects of political rule. If the idea of nation suggested seeing society as a culturally unified entity with its own history, the modern idea of economy added the notion that society could develop on its own through material transformations in its productive capacity as well as through knowledge.

Fourth, Europeans in the early modern era undertook projects of exploration and eventually empire on a scale the world had never seen before. These paved the way for social science by making manifest the enormous diversity of human cultural forms and practices. Both missionaries and administrators - as well as eventually anthropologists - sought to understand kinship, family, the organization of household economies, hierarchies of power, specialization of religious responsibilities, and approaches to educating the young - as variables in a complex collection of social structures, and inquired into what function each might serve. Knowledge of human diversity helped to break the assumption that locally observable social organization needed no explanation.

From Classical Philosophy to Modern Social Science, from the renaissance through the 18th century, scholarship on political and social subjects remained largely commentary on ancient texts. Thomas Hobbes' Leviathan (1651) drew in important ways on classical sources, but also marked a transition to modern social science. It presented a theory of the state formulated through what Hobbes claimed were strict deductions from empirical bases. To be sure, the notion of social contract at its centre was either a thought experiment or a metaphor, not a statement of factual history. But Hobbes based his arguments about the legitimacy of government on reasoning from what he took to be facts and logical necessity, not tradition or divine inspiration. Criticism and revision could (and did) focus on both the putative facts and the reasoning without (always) going back to first principles.

John Locke made political theory depend more on an idea of society (and the benefits that language and money as well as government could bring). Among the first great works of comparative social science was Montesquieu's Spirit of Laws (1748).

Montesquieu made a more systematic effort than Locke to account for the differences in legal and governmental systems by differences in environmental context, social organization, and culture. Adam Ferguson took this further, developing the notion of "civil society" as a counterpart to government (and indeed to the derivation of social laws from theology). In 1767, Ferguson presented the history of civil society in a series of stages, prefiguring 19th century evolutionary thought. Much less empirical, Jean-Jacques Rousseau nonetheless contributed to social science a theory of learning from experience, a strong idea of the social whole, the idea of alienation, scepticism about progress, and an alternative construction of the social contract to that of Hobbes and Locke.

The Physiocrats in 18th century France introduced the powerful notion of system, suggesting hat the accumulation of wealth was based on circulation in society, not the action of the state. At least in its economic aspect, society could therefore be largely self-regulating. This paved the way for Adam Smith's (1776) suggestion that a market ordered "as though by an invisible hand" could be a model for social self-organization with minimal government interference. But Smith importantly rejected the physiocrats' notion that all wealth derived from nature, especially agriculture, insisted that human labour was itself productive, and that the social organization of production, as through the division of labour, could make it more so. Smith used the notions of division of labour and market to theorize the ways in which interactions among individuals could produce a self-regulating system. The behaviour of each conditioned that of all. The capacity of the market to turn private greed into motivation for publicly useful work was testimony to the extent to which civil society could organize itself outside the control of the state -in non-economic as well economic dimensions. Thomas Malthus (1798) gave the idea of system a different twist, arguing that the growth of population followed "natural" laws that would periodically result in social catastrophe. Like the proto-evolutionary analyses of Ferguson and other Scots, that of Malthus influenced Charles Darwin and the formulation of a theory of biological evolution. Indeed, in the 19th century, biological and sociological thought were not altogether distinct -as the career of Herbert Spencer reveals. Spencer contributed the phrase "survival of the fittest" to an evolutionary theory he thought equally applicable to biological and social life.

In every European country, the state had become a dramatically larger set of institutions, and had begun to penetrate much more basically into the daily lives of its citizens. This was among the pivotal occasions for the development of social science, not only because the state demanded knowledge to guide

its actions but because the state's very efficacy suggested the potential for remaking society. At the same time, thinking in terms of the state, rather than simply the ruler (as in Machiavelli's Prince), not only stressed the extent to which the government was a complex social organization, but also stressed the extent to which a public, political order defined a whole country. This is the sense in which the state, not the government of that state, enjoyed sovereignty. The American and French revolutions symbolized this. Alexis de Tocqueville combined the influences of the two, studying the interrelationship of an individualistic culture and an egalitarian political economy in Democracy in America, and of the nature of politics itself in The Old Regime and the French Revolution. His work formed part of the lineage of both liberal political theory and of sociology. John Stuart Mill integrated Tocqueville's insights with Jeremy Bentham's more systematic approach in his account of utilitarianism—the identification of the social good as the greatest good of the greatest number, and of social science as inquiry into how best to achieve that. But while Mill and Bentham approached society as aggregation of individuals, other social scientists stressed the importance of social structure.

The idea of structure received a dramatic articulation in Karl Marx's theory of capitalism. This cantered both on a view of history as class struggle and on an analysis of modern society as grounded in an economic structure that transformed the production and accumulation of value, more or less independent of the intentions of individuals. Marx's theory was not only more social, in the sense that it emphasized the analysis of an emergent whole not the individuals who made it up, it was more systemic, in seeking to grasp how the complex patterns of that whole could be traced to certain fundamental causal influences and their interrelationships. Marx relied on a labour theory of value and agreed with Smith about the importance of social organization to making labour more productive, but following Rousseau he was much less willing to accept private property as a given. He insisted both that capitalism (like all earlier economic formations) was a system of constraints, not the achievement of freedom that Smith had extolled, and also that it was unstable, prone to crises but also transcendable. The socialism that would follow could rely much more directly on social science (than on inefficient and unfair markets) to guide production and distribution. Building on the classical political economists, thus, Marx insisted that capitalism was a historical stage, not simply an expression of timeless natural laws.

In this, Marx shared much with other evolutionary thinkers of the 19th century (and indeed, he praised Darwin warmly). In Herbert Spencer's evolutionary theory, individualism itself would be seen as an outcome of evolutionary change in social structure. Some social scientists, like the American William Graham Sumner would however develop Spencer's thought into "social Darwinism," a rationalization for unfettered competition in capitalism, since only the fittest would survive (and in un Darwinian fashion, they equated fitness with virtue).

1.4 Social Sciences - intellectual disciplines

The differentiation of social science into a set of distinct intellectual disciplines dates only from the 19th century. This involved first a growing distinction of the social sciences from the natural and physical sciences on the one hand and what came to be called the humanities on the other. Secondly, the social sciences separated from each other.

Increasingly economics was defined by the study of market (and related) phenomena that could in principle operate independently of direct government intervention (whether or not a specific intervention might be beneficial as economists sometimes argued). Economics included studies of the relative merits

of organization through markets and through hierarchical power, of nonmarket allocation of goods, of regulative action, and of macroeconomic factors shaped by governments on nonmarket bases. It also addressed questions of the nature of rationality and purposive action and of the relations among different factors of production. Nonetheless, market exchange was definitive; nonmarket phenomena were on the frontier between economics and other fields. Economics is distinctive for much greater use of mathematics than the other social sciences, a development made possible by the development of a concept of utility (pioneered by Jeremy Bentham and made more operational by W.S. Jevons). This allowed for the modelling of otherwise disparate market phenomena, involving seemingly incommensurable goods, in terms of units of "good" as such. The very success of mathematical modelling, however, has been the occasion for recurrent debates over the trade-offs between theoretical elegance and empirical veracity.

Political science correspondingly grew up as a field studying, and sometimes advising, modern states. It developed out of a much older tradition of advice to rulers and philosophical consideration of themes like justice. Machiavelli is often taken to mark a turning point, as the first modern political theorist, though his modernity (like that of the Italian city states in which he lived) is ambiguous. Hobbes was more decisively modern.

At least as important, however, was the late 19th and early 20th century redefinition of the field in terms of the empirical study of states (and sometimes power more generally). A further "behavioral revolution" (mainly after the Second World War) refined this idea of an empirical science of politics, and the distinction from older fields like normative political theory and diplomatic history, but the discipline retains a hybrid character.

Unlike economics or sociology, it is united more by a concern for certain "dependent variables"-political outcomes- than by a focus on the effects of certain sorts of independent variables (such as supply and demand, population structure, or group dynamics). The term sociology was coined by August Comte in the 1840s, though like other social sciences sociology can claim an older ancestry. It was shaped by the growth of industrial organization, studies of population change, patterns of immigration, changes in family structure, and concern for the "social question" of how the poor would fare economically in modern societies and inequality would shape modern political systems.

Never altogether distinct from economics and political science, thus, sociology nonetheless developed a focus on the dimensions of social life that were organized at least largely on bases other than market relations and governmental dictates. Significantly engaged with empirical data collection from its origins -for example in the studies of working class families by Frederic Le Play- sociology increasingly developed a distinctive body of theory addressing questions of social structure, orientations to social action, and processes of social change. If economics, politics, and sociology constituted the core of the social sciences, this is not because they were larger or more important intellectual fields, but because they fit more squarely and completely into the social sciences, overlapping less with the natural sciences and the humanities. Most of the time, they were also more closely related to each other than to the rest of the social sciences. Anthropology, psychology, geography, history, and statistics have also been central to the growth and improvement of the social sciences. But, cultural anthropology and archaeology have always been closely linked to the humanities while physical anthropology (and certain versions of archaeology) has been more centrally involved in natural science. Social psychology has always been among the social

sciences, but the extent of emphasis on the social dimensions of human mental and emotional life in the rest of psychology has varied. In the decades after World War II, an effort to integrate psychology and social sciences in the paradigm of "behavioural sciences" flourished. More recently, much of psychology has tended away from social science and towards the natural sciences and cybernetics.

Similarly, geography has always been divided between an emphasis on social and cultural dimensions and on physical dimensions (with the two partially joined in work made possible by technologies like satellite imaging). History is part of the older intellectual tradition out of which social science emerged, and most often understood as part of the humanities; at the same time, several branches of historical research have been transformed by social science and historical research remains central to social science.

Not least, though last for this list, statistics grew up in significant part as a social science -for example in the pioneering social statistics of the Belgian, Adolphe Quetelet, who invented the notion of the "average man"--and the work of many social science disciplines is organized largely in terms of statistics. Statistics remains, however, a partially autonomous discipline and heavily influenced by biological and medical statistics, and models from the physical sciences.

Each of the social sciences was shaped importantly by the ways in which it was institutionalized during this period, and the contrasts used to distinguish it from others. Crossing all the social sciences was a struggle over methods, or methodenstreit, which pitted more objectivist, universalizing sciences against more subjectivist, particularizing humanities, dividing the social sciences between the two. Economics and psychology have been the most universalizing, while politics, sociology, and anthropology have been internally divided.

National contexts also mattered. In France, for example, Emile Durkheim fought to distinguish sociology from psychology, following Comte in claiming that each science needed its own distinct subject matter and arguing that sociology should study "social facts" that were irreducible to more individual level phenomena. In the United States, economics was the more influential counterpart discipline (and remains so to this day).

Sociology was initially organized as an interest area within the American Economic Association, and then split off to form the American Sociological Association in 1905. In Germany, sociology was commonly taught in faculties of law and the distinction between normative and empirical theory especially salient. But the most important founder of German sociology, Max Weber, was keenly interested in maintaining the relationship between economics and sociology, and approached both in with a comparative-historical historical method that remained a hallmark of German sociology. Great Britain was strong in economics but particularly weak in sociology.

Other social sciences were similarly shaped by the contexts in which they matured. Economics was stronger in relation to neighbouring disciplines, but heavily influenced by expectations that it would deliver immediately useful advice to governments and businessmen. Though the world depicted by neoclassical economic theory is generally devoid of historical specificity and political action, from the late 19^{th} century introduction of marginal utility theory, through the work of John Maynard Keynes and other efforts to address the great Depression, and on into the development of monetarism, economics has been closely linked to policy-making.

Political science had if anything a greater difficulty emancipating itself from political commentary and history, and in some national settings from law. Indeed, the very name of the discipline was contested - with some calling it politics, some government, and others political science. Though its origins are old, the last of these grew in popularity in the mid-20th century as many leaders in the field sought to stress the objective, especially quantitative study of political "behaviour" rather than more interpretative or normative relations to or preparations for politics as such.

Quantification and Comparison Across the social sciences the second half of the 20th century was an era of growing emphasis on quantification. This accompanied efforts to make the social sciences more scientific, understood largely in terms of an objectivistic orientation to knowledge and a belief in the accumulation of truth. There were innovations in techniques of empirical data collection, perhaps most prominently in sample surveys, but also in censuses, experimental research (especially in psychology), and secondary analysis of data produced as by-products of market transactions, elections, and other processes. And there were new approaches to both analytical statistics and mathematical modelling. While certain multivariate methods, like regression and path analysis, form a sort of centrepiece to this process, becoming standard in the 1960s and 70s, the overall pattern is not only advancement but proliferation of different techniques, often linked to different theoretical assumptions (but also to different empirical challenges, like handling the massive data of censuses and global surveys versus the smaller populations more common studied in psychology). Network analysis and nonlinear models have grown in importance, sometimes in competition with conventional multivariate methods.

In economics, and to a much lesser extent in other social sciences, empirical quantification was complemented (or even overridden) by theoretical mathematicization. Both trends reflected the computational power made available by the increasing improvement of electronic computers. But the competition between them represented also a return of the 18^{th} century opposition between empiricists and deductive theorizers.

The distinction was never hard and fast, but significant. Econometric statistics made major advances in the mid-20th century; theory in the form of mathematical models dominated from the 1970s to early 1990s. In economics a recent trend has been for renewed prestige to empirical inquiry, especially where this examines whether important basic assumptions are in fact valid, or perhaps operated only under restricted conditions.

Behavioral economics, produced largely by economists drawing explicitly on psychological research, but also by researchers in new fields like "decision sciences", has proved especially fertile in this regard.

Quantification and mathematics introduced a greater division among social science disciplines (and within disciplines between lines of work in which quantitative methods figured more or less). It sharply reduced the connections between anthropology and the other social sciences, for example, reinforcing closer links between cultural anthropology and the humanities.

Qualitative research methods also underwent continual improvement. Among the most significant was the development of "ethnography" within anthropology and to a lesser extent sociology. This signified efforts, usually based on long-term "participant observation" fieldwork, to document the different aspects of a way of life and how they fit together. Ethnography integrated cultural and social organizational analysis, and was typically a strongly integrative, holistic perspective. Later developments included both a growing reflexivity about the location and perspective of the ethnographer within the field site, and attention to the limits of what could be known through first-hand observation and conversation (as for example ethnography tended to explore the local thoroughly but state-level structures minimally).

The same post-war period that saw the "behavioural revolution" in political science, the rise of quantitative methods in sociology, and in economics first a growing sophistication of econometrics and then a growing interest in mathematical models (which would become dominant by the 1970s), also saw the rise of predominantly qualitative fields of international and comparative research. The most visible institutionalizations of this new emphasis on international knowledge were the "area studies" fields—African Studies, Latin American Studies, South Asian Studies, and so forth.

The area studies fields were distinctively interdisciplinary, combining not only different social science disciplines but history, literature, and other humanistic inquiries as well. Though the area studies fields flourished and produced major and influential research, they were in tension with the "core" social science disciplines. To some extent, the terms of this tension replayed the *methodenstreit*. The area studies fields were seen by many social scientists as producing particularistic knowledge while the disciplines sought universal truths. To a very large extent, mainstream economics withdrew from the area studies project. Political science and sociology were split, but by the last third of the century the non-context-specific approaches had the upper hand.

1.4.(a). Struggle and Renewal

If the growth and spread of social science looked smooth in the 1950s and early 1960s, it suffered a shock in the late 1960s and early 1970s. Each of the themes basic at the beginning came back into dispute. The idea of an objective social science was challenged by complaints about the hubris of those who believed knowledge was more perfect than the freedom of human beings or the complexity of culture and society allowed. There was accordingly a renewal of interest in interpretative approaches, in critical theories that sought to avoid generalizing from what currently existed, and in reflexivity, especially the ways in which the socio-cultural location of the researcher shaped his or her perspective.

The close relationship of many social scientists to their states was also challenged. The most visible version of this was the implication of some social scientists in American government counterinsurgency programs, but there was also a more international wave of critique of the way anthropologists had served colonial states, of how political scientists were embedded in domestic power structures, and so forth. But the debate was vigorous, for at the same time certain links of social scientists to states were challenged, there were also calls for social science to me more "relevant," less abstractly academic and more directly engaged in efforts to solve social problems. In some cases this meant that social scientists allied themselves more with social movements and less with states. In economic affairs too, on the one hand more social scientists were engaged in market research and other work done specifically on behalf of for-profit clients. On the other hand, there was a renewal of interest in Marxism and more generally in social science that challenged existing political and economic arrangements.

Not least of all, the independence struggles of former European colonies occasioned a rethinking of the relationships between power and culture. This included a critique of the ways in which Europeans had viewed the cultures of non-Western societies and constructed evolutionary schemes that implied that there was only one form of advancement—and that it called for non-Western societies to become more like the dominant countries of the West. Even the idea of modernity came under attack, partly because unilinear evolutionary ideas had been incorporated into the notion of modernization. In the 1960s and 70s, this critique came often in the framework of Marxism or other alternative modernist programs. Soon, though, postmodern thought (rooted more in the humanities) criticized the mainstream social sciences as embodiments of a modernity that was built around notions of unidirectional progress, reductions of diversity, and the imposition of power even through the forms in which knowledge was produced. False

universalisms were challenged by social scientists writing from the perspectives not of the ivory tower or what Karl Mannheim had called the "free-floating intelligentsia" but of different social locations: women, people of color, postcolonial subjects; or of engagement in one or another movement or struggle. This struggle was played out against the background of more material transformations in the social sciences. The postwar era was marked by dramatic growth in higher education generally and in the social sciences in particular. This came sooner and was more pronounced in the United States, and that accelerated a second trend which was a growing prominence of American social science on the global scene. Where the roots of social science lay mostly in Europe, at least the largest scale of new development came from the US. Without comparable resources, there was nonetheless also a growth of social science outside the Euro-American countries. India was perhaps the single most influential setting for the growth of non-Western social science, and by the late 20thcentury fields like Subaltern Studies had become influential throughout the world.

International social science associations were founded (often under the auspices of UNESCO) to complement the national societies. International social science was also significantly influenced by the major ideological and political economic struggles of the era, from the Cold War and decolonization through the non-aligned and non-proliferation movements to the intensification of capitalist globalization and opposition to it—and the American model for it--in the 1990s and early 21st century.

Even while the social sciences were most engaged in political disputes in the 1960s and 70s, the seeds were being laid for another material change that would change their engagement with practical affairs and political policy-making. This was the growth of professional schools, and with them new fields of social science organized outside the traditional disciplines, and usually with a more "applied" orientation. Business schools, for example, have departments of finance that have largely supplanted one of the core fields of economics. Sociology and psychology figure prominently in both organizational behavior and marketing programs, but each of these fields now offers its own PhD programs, making for a greater distance from the "parent" disciplines. Schools of education, public health, medicine, nursing, engineering and communications have also both employed social scientists in large numbers and in varying degree produced parallel fields of social science focused on their specific professional domains.

The social sciences have expanded enormously since their early modern origins. They have become impressively international, though in all countries they are still (with the partial exception of anthropology) disproportionately domestic in focus. They have been at once part of the spread of a dominant version of Western culture, and one of the resources for developing critical analyses of that culture-as indeed of other dominant cultural and institutional formations.

The social sciences have also become a great deal more methodologically sophisticated, and now use a variety of both quantitative and qualitative techniques to advance knowledge. Theory too has advanced from several early contending grand systems to a range of middle range theories and several theoretical frameworks with different strengths and weaknesses and potential for mutual engagement. Most importantly, substantive knowledge of different problems and empirical topics has grown exponentially. From inequality and organizational processes through market structures, voting procedures and behavior, decision-making, to kinship, family dynamics, and migration social scientists have created numerous fields of cumulative research and scholarship. Neither the creation nor the reform of modern health care or welfare systems, business organizations or trade unions, humanitarian assistance or peacekeeping operations, mass media or movements for cultural survival happens without the involvement of social scientists or intellectual tools they have created.

Social science Disciplines

1.5 Social Science Disciplines

The following are the discipline branches within the social sciences. Anthropology, Business studies, Communication studies, Criminology, Demography, Development studies, Economics, Education, Geography, History, Industrial relations, Law, Linguistics, Media studies, Methodology, Philosophy, Political science, Psychology, Public administration, Sociology, Legal Management, Paralegal studies, International studies, Library Science and Information Science.

The Social Science disciplines are branches of knowledge which are taught and researched at the college or university level. Social Science disciplines are defined and recognized by the academic journals in which research is published, and the learned Social Science societies and academic departments or faculties to which their practitioners belong. Social Science fields of study usually have several sub-disciplines or branches, and the distinguishing lines between these are often both arbitrary and ambiguous.

Anthropology is the holistic "science of man," - a science of the totality of human existence. The discipline deals with the integration of different aspects of the Social Sciences, Humanities, and Human Biology. In the twentieth century, academic disciplines have often been institutionally divided into three broad domains. The natural sciences seek to derive general laws through reproducible and verifiable experiments. The humanities generally study local traditions, through their history, literature, music, and arts, with an emphasis on understanding particular individuals, events, or eras. The social sciences have generally attempted to develop scientific methods to understand social phenomena in a generalizable way, though usually with methods distinct from those of the natural sciences.

The goal of anthropology is to provide a holistic account of humans and human nature. This means that, though anthropologists generally specialize in only one sub-field, they always keep in mind the biological, linguistic, historic and cultural aspects of any problem. Since anthropology arose as a science in Western societies that were complex and industrial, a major trend within anthropology has been a methodological drive to study peoples in societies with more simple social organization, sometimes called "primitive" in anthropological literature, but without any connotation of "inferior." Today, anthropologists use terms such as "less complex" societies or refer to specific modes of subsistence or production, such as "pastoralist" or "forager" or "horticulturalist" to refer to humans living in non-industrial, non-Western cultures, such people or folk (ethnos) remaining of great interest within anthropology.

Economics is a social science that seeks to analyze and describe the production, distribution, and consumption of wealth. The word "economics" is from the Greek οἶκος [oikos], "family, household, estate," and νόμος [nomos], "custom, law," and hence means "household management" or "management of the state." An economist is a person using economic concepts and data in the course of employment, or someone who has earned a university degree in the subject. The classic brief definition of economics, set out by Lionel Robbins in 1932, is "the science which studies human behaviour as a relation between scarce means having alternative uses." Without scarcity and alternative uses, there is no economic problem. Briefer yet is "the study of how people seek to satisfy needs and wants" and "the study of the financial aspects of human behaviour."

Economics has two broad branches: microeconomics, where the unit of analysis is the individual agent, such as a household or firm, and macroeconomics, where the unit of analysis is an economy as a whole. Another division of the subject distinguishes positive economics, which seeks to predict and explain economic phenomena, from normative economics, which orders choices and actions by some criterion; such orderings necessarily involve subjective value judgments. Since the early part of the 20th century, economics has focused largely on measurable quantities, employing both theoretical models and empirical analysis. Quantitative models, however, can be traced as far back as the physiocratic school. Economic reasoning has been increasingly applied in recent decades to other social situations such as politics, law, psychology, history, religion, marriage and family life, and other social interactions. This paradigm crucially assumes (1) that resources are scarce because they are not sufficient to satisfy all wants, and (2) that "economic value" is willingness to pay as revealed for instance by market (arms' length) transactions. Rival heterodox schools of thought, such as institutional economics, green economics, Marxist economics, and economic sociology, make other grounding assumptions. For example, Marxist economics assumes that economics primarily deals with the exchange of value, and that labour (human effort) is the source of all value.

Sociology evolved differently depending on where it was taught and who was teaching it. There are two major types of sociology that emerged, quantitative and qualitative sociology. Most of the universities in the world use both qualitative and quantitative methods for information as not one method is necessarily better than the other. Qualitative sociology is concerned with trying to obtain an accurate picture of a group and how it operates in the world. Small and his group from the University of Chicago tied to study how immigration was affecting the city and its residents. A large number of people were migrating to the United States from a variety of countries. Sociologists were fascinated by the social changes they saw taking place and began conducting qualitative studies which involved personal interviews and observations of ethnic rituals and ceremonies.

Quantitative sociology relied on statistical analysis to analyze experience and trend that people took. Some researchers talked to the people and observed them but many preferred to remain within the confines and quantify their data suitable for statistical manipulation. Apart from the two types of sociology, there are other types of sociology, they include; anthropology, political science, psychology and economics

Political Science this concerns government of various societies and it considers the kind of government a society is having. It also considers how the leader of this society attains positions of power within a particular government. It also considers the relations of people in a society to whatever form of government they have.

Psychology is the study of human brain and how it functions, considering issues such as memory, dreams, learning and perception. Psychology always takes an individual out of his/her social circumstances and examines her mental processes taking place.

Social science involves scholarly or scientific discipline that deals with such study, generally regarded as including sociology, psychology, anthropology, economics, political science, and history. The main social science disciplines include anthropology, communication, criminology, cultural studies, economics, human geography, linguistics, law, political science, psychology, sociology, development studies and women's studies. Each of these social science subjects uses a range of approaches to study society. Like all sciences, social sciences evolve through the interplay of the ideas and theories of academics and the evidence that supports or refutes them.

1.6 The Importance of Social Science

Social scientists influence our lives usually without us being aware they are doing so. For example:

- the role of governments in an increasingly market-based society has been determined by famous thinkers such as John Maynard Keynes and Karl Popper
- it was an economist who came up with the idea of the National Health Service
- the payment of billions of pounds of state benefits for the needy has been influenced by the work of social scientists.

Social science research findings continue to provide invaluable information whether you are a parent, a local councillor, a police officer, or a business executive.

Most of these fields touch everyone's lives in one way or another. The vast majority of the fields do no better than make broad generalizations to try to explain the observed facts. There are not many that can with confidence claim any absolutes. It has been said, "the unexamined life is hardly worth living!" Many of these disciplines work to examine human's collective lives and to learn lessons from them. Only as mankind learns from the many mistakes they have made can they hope to survive and have a peaceful, fulfilled society. Therefore, scientists and society as a whole must take the time to analyze the past and learn to do things better this time around. The social sciences often draw in the concepts of the natural sciences as well in efforts to achieve understanding and improve the world.

1.7 Max Weber's View of Objectivity in Social Science

1.8(a). Facts and Values

Max Weber thought that "statements of fact are one thing, statements of value another, and any confusing of the two is impermissible," Ralf Dahrendorf writes in his essay "Max Weber and Modern Social Science," acknowledging that Weber clarified the difference between pronouncements of fact and of value. Although Dahrendorf goes on to note the ambiguities in Weber's writings between factual analysis and value-influenced pronouncements, he stops short of offering an explanation for them other than to say that Weber, being human, could not always live with his own demands for objectivity. Indeed, Dahrendorf leaves unclear exactly what Weber's view of objectivity was. More specifically, Dahrendorf does not venture to lay out a detailed explanation of whether Weber believed that the social scientist could eliminate the influence of values from the analysis of facts.

Did Weber believe that, even though facts are one thing and values another, social and economic facts could be evaluated without the analysis being influenced by values? And what is the relation of objectivity to values? Could objectivity, for instance, be used to show that one value is superior to another? Or does objectivity apply only to the analysis of facts? Do one's values or perspective stem from human nature, metaphysical views, personal identity, or is it just as likely that they are a mere construct of culture?

These questions, and others like them, underlie much that has been considered ambiguous in Max Weber's writings: His methodology. Since his death, sociologists and political scientists have been disputing where Weber stood with regard to questions concerning the relationship of objectivity to facts and values. "Most of Weber's commentators," Edward Bryan Portis writes, "have assumed his advocacy of the fact-value dichotomy, despite his explicit and implicit assertions to the contrary, because of his numerous statements denying the ability of science to refute any normative position or to help one choose among contending normative orientations." Indeed, hardly a scholarly piece is written on Weber, it seems, without the preamble that Weber's views on this subject have been widely misunderstood, with the implication that the scholar at hand intends to finally set the record straight.

This essay has more humble ambitions. Although it takes issue in the final section with part of the exhaustive view laid out by Portis, this essay does not purport to set forth yet another definitive interpretation of Weber's views on objectivity. Rather it seeks to shed light on Weber's view of the applicability of objectivity by attempting to answer the overarching question that sits at the foundation of those posed above: Was Weber an advocate of value-free social science? The answer, as will be shown, is both yes and no -- because, this essay will argue, Weber maintained a two-tiered approach to value-free social science. On the one hand, he believed that ultimate values could not be justified "scientifically," that is, through value-free analysis. Thus, in comparing different religious, political or social systems, one system could not be chosen over another without taking a value or end into consideration; the choice would necessarily be dictated by the analyst's values. On the other hand, Weber believed that once a value, end, purpose, or perspective had been established, then a social scientist could conduct a value-free investigation into the most effective means within a system of bringing about the established end. Similarly, Weber believed that objective comparisons among systems could also be made once a particular end had been established, acknowledged, and agreed upon, a position that allowed Weber to make what he considered objective comparisons among such economic systems as capitalism and socialism. Thus, even though Weber maintained that ultimate values could not be evaluated objectively, this belief did not keep him from believing that social problems could be scientifically resolved -- once a particular end or value had been established.

1.8(b). Stating the Standpoints

But first, just what is Weber's own standpoint, as determined by his ultimate values? It is, no doubt, influenced by one of his key concerns: "the quality of human being in any given economic and social order." Sometimes, however, his standpoint is nationalistic. And in yet other essays, it champions individual liberty. Indeed, Weber's perspective changes, and it is likely to be driven not by one value but by levels of them, ranging from humanism to a concrete objective. But the fact that Weber had a perspective lends little support to the two-tiered interpretation, other than to show that he believed it was permissible for a social scientist to possess a value-determined standpoint. His treatment of perspective is another matter, however.

One hint that begins to shed light on Weber's view on the fact-value question is a characteristic that recurs in several of Weber's essays and speeches: Weber announces, often at the beginning of a speech or essay, the standpoint from which he plans to evaluate a given situation or set of facts. Likewise, if he changes his focus during a presentation, he often declares the new standpoint. In his opening remarks of "The Nation State and Economic Policy," one of Weber's early speeches, he sets a precedent for this pattern while unveiling a justification for his perspective. The "inaugural lecture is an opportunity," Weber says, "to present and justify openly the personal and, in this sense, `subjective' standpoint from which one judges economic phenomena," revealing that he maintained that even the examination of such seemingly hard data as economic facts were subject to the influence of a perspective determined by values. When Weber shifts course later in the speech to prescribing what should be done to deal with the problems on Germany's eastern frontier, he discloses his new perspective: "the standpoint of the German people." The solution would obviously be quite different if it were made, say, from the standpoint of the Polish workers. Similarly, in one of his later lectures, "The Profession and Vocation of Politics," Weber tells his audience near the beginning of his remarks that he will expose "the political deficiency of this system ... from the standpoint of success."

Although Weber often announces the value from which he intends to analyze a particular policy, he also acknowledges that the value may be merely a construct of one's culture or society. An example of the influence of culture upon perspective lies in Weber's comments about political economy. As soon as the method of analysis known as political economy makes value judgments, Weber says, "it is tied to the particular strain of humankind (Menschentum) we find within our own nature. ... The economic policy of a German state, and, equally, the criterion of value used by a German economic theorist, can therefore only be a German policy or criterion." Yet the perspective still must be acknowledged.

Regardless of whether a social scientist's value-orientation stems from cultural norms, nationality, or a worldview, what remains certain for Weber is that the value is neither intrinsic to the subject matter nor specific to its context -- a view that categorically separates value from facts. Weber takes care to refute such views in his discussion of the methodology of political economy in "The Nation State and Economic Policy." First, Weber assails those economists who maintain that political economy can derive its own ideals from the subject matter. The notion that there are independent or socio-political ideals shows itself to be a delusion as soon as one delves into the literature in an attempt to identify the basis for its evaluation, Weber says. "The truth is that the ideals we introduce into the subject matter of our science are not peculiar to it, nor are they produced by this science itself." Rather, the values stand above the subject matter; they are of a higher order. For Weber, it is less important what another analyst's core values are than whether he clarifies them for the benefit of both himself and his audience.

Weber also criticizes those scientists who often "unconsciously allow the starting point for our analyses and explanations of economic events to determine our judgements of these events," demonstrating that he separates the subjectivity of value-orientation from the objective evaluation that is carried out after the value orientation has been established. In other words, Weber is chastising those scientists who allow the subjectivity of their perspective to determine their analysis of the facts. As examples of the economic scientists who have made this mistake, Weber points to the historical apologists and to the Marxists.

1.8(c). Adhering to Values

What matters even more to Weber is whether one adheres unflinchingly to his values. In "The Profession and Vocation of Politics," Weber explicitly articulate show one must look at life from a chosen value: "What matters is not age but the trained ability to look at realities of life with an unsparing gaze, to bear these realities and be a match for them inwardly." The comment exposes the inherent relationship, for Weber, between value-free analysis and value-driven moral action, a dichotomy that resurfaces in Weber's discussion of an ethics of commitment and an ethics of responsibility. To be "a match for them inwardly" is to cling to one's values even in the face of the inevitable "polar night of icy darkness." "For truly, although politics is something done with the head, it is certainly not done with the head alone." Values are linked to the heart -- to subjectivity -- as much as they are linked to the head.

Weber himself seems to adhere to his own values -- or at least he argues repeatedly for the veracity of one `cause' over another. Perhaps this is among the trends that have led many Weber scholars astray, especially since "Weber feels that no cause can be `proved', simply by intellectual means, to be superior to any other." Despite his own attachment to, for example, the values of individual liberty, his "philosophical stance did not provide a mechanism for validating democratic values in and of themselves."

How can Weber's arguments for his ultimate values be reconciled with the view that value-free analysis can be conducted only after a value or purpose has been established? Lassman and Speirs, in their introduction to Weber: Political Writings, provide the answer. "Although Weber believed that values could not be given any form of `ultimate' foundation, it was possible and indeed necessary" to argue for them because "the tensions between competing values are essential in order to prevent cultural stagnation." Even though Lassman and Speirs do not explain precisely how it is possible to put forth objective arguments supporting subjective values while maintaining a commitment to truth, they do allude to one solution: Weber's "scholarly investigations and political essays have the purpose of making clear, in an objective manner as possible, the realities and possibilities given in any particular situation."

1.8(d). A Two-Tiered Approach to Value-Free Social Science

Having examined Weber's views of the role of perspective and values in social scientific analysis, the evidence, both from Weber's writings and from commentaries on them, must now be considered in support of the interpretation that Weber took a two-tiered approach to value-free social science.

First, it must be shown that held Weber believed ultimate values could not be proved scientifically, a position alluded to in several preceding remarks. Lassman and Speirs, writing in their introduction to Weber: Political Writings, address the matter directly. Weber held the belief, they say, that "there is no longer any possibility of an objective ranking of ultimate values or moral principles."

Weber's own writings support Lassman and Speirs' conclusion that Weber considered ultimate values and their subsequent political values to be subjectively determined. For instance, in "Between Two Laws" Weber writes that certain communities are able to provide the conditions for not only such "bourgeois" values as citizenship and true democracy, "but also much more intimate and yet eternal values, including artistic ones." The language that Weber uses to characterize these two types of values leads to the interpretation that he held them to be a subjective matter. Regarding the first set of values, labelling them "bourgeois" brings to light their contingent nature: They are the product of a class, a strata. Regarding the second set, the labels "intimate" and "eternal" clearly set them apart from any objective foundation. An "intimate" value is by definition personal, an opinion. Further: It carries the connotation of emotion, of mystification. Likewise with "eternal."

This element of mystification, of faith in what is ultimately unknown and unknowable, materializes in other pieces of evidence that help substantiate Weber's view that ultimate values cannot be objectively established. "The nature of the cause the politician seeks to serve by striving for and using power is a question of faith." Yet here Weber refers to the politician, not the social scientist. But could the same theorem not be applied to the social scientist? Could "social scientist" not be substituted for "politician" and, say, "facts" for "power"? And then could the social scientist not be asked to use those facts objectively while maintaining a commitment to his values? Answering these questions in the affirmative, which can be done only through an argument by extension, a frail but not hopeless step, leads to interpreting "The Profession and Vocation of Politics" as a metaphor for the actions of the social scientist, showing that the values he seeks to serve are also a question of faith.

The argument by extension notwithstanding, there is other evidence that Weber held the social scientist's values to be a subjective matter. Portis, for instance, says Weber "believed it impossible to justify ultimate values scientifically. He presumed they were derived from the metaphysical commitments that define one's general outlook." Rogers Brubaker, in The Limits of Rationality, also acknowledges that Weber's discussion of value orientations amplifies those of a long line of ethical relativists. Weber

believed that "value orientations are essentially subjective, and that conflict among them cannot be rationally resolved." Furthermore, Weber believed that value orientations could not be eliminated from social scientific work. They necessarily determine the analyst's perspective.

1.9. Social Research Theories & Methodology

Theory and methodology provide ways to study social relationships and interactions. During the 19th century, philosophers and others advocated the scientific study of human society. It was during that time that many thinkers developed theories about society, followed later by methodologies for testing theories and developing new ones. Theory and methodology go hand in hand when studying patterns of life in human society.

History: The 19th century French philosopher Auguste Comte was an important early figure in the development of social science theories. He believed society could be studied scientifically and objectively at a time when most societal changes were explained in religious terms.

Function: Theory helps social scientists make sense of patterns observed in everyday society. It also helps keep researchers from being taken in by patterns that could just be flukes. Theory also helps shape social research and gives it direction. In this way, theory acts as a guide, pointing researchers to the most interesting issues of society, including its politics, economics and other interactions. Finally, theory helps researchers understand social phenomena in such a way that can suggest actions. For example, a theory that explains why teenagers drop out of high school can provide a basis for programs and interventions aimed at reducing dropout rates.

Types: Two types of approaches to the relationship between theory and research include the deductive and inductive methods. The deductive method argues from the general to the specific. Under a deductive methodology, a researcher begins with a hypothesis, then makes observations or collects data to test that hypothesis. Based on empirical evidence from the study, the researcher then decides whether to accept or reject the hypothesis. The deductive methodology, in short, tests theories and hypotheses. The inductive method, in contrast, goes from the specific to the general. Under this methodology, social scientists observe social phenomena, identify patterns and then analyze them to reach broad conclusions and develop new theories, based on research findings.

Features: To study social, political and economic phenomena, social scientists use a variety of research designs and methodologies. Research designs include single-group case studies and quasi-experimental designs comparing two groups of subjects. Measurement methods include observations, surveys, interviews, documents, and data from other sources, such as government agencies. To compensate for the inability to randomly assign subjects to treatment and control groups, many social researchers must conduct more in-depth statistical analysis to control for differences between groups of subjects. Research methodologies include qualitative and quantitative methods. Many quantitative studies follow a deductive approach, while many qualitative studies are more inductive.

Considerations: Because social science research occurs within society itself, rather than in the controlled laboratory settings used by the natural sciences, social research methodologies have some limitations that the natural sciences do not. The main limitation is that social researchers usually cannot perform the controlled experiments with random assignment to treatment and control groups. This is usually for ethical and other reasons. A group of social scientists could not, for example, randomly assign academically low-performing children to an experiment in which one group would receive a new instructional intervention and one group would not.

1.10 Relevance of studying social sciences

Studying the social sciences is vital for many reasons. It enables the learner:

• to understand the society in which they live to learn how society is structured, managed, and governed, and also about the forces seeking to transform and redirect society in various ways.

• to appreciate the values enshrined in the Indian Constitution such as justice, liberty, equality and fraternity and the unity and integrity of the nation and the building of a socialist, secular and democratic society.

• to grow up as active, responsible, and reflective members of society.

• to learn to respect differences of opinion, lifestyle, and cultural practices.

• to question and examine received ideas, institutions, and practices.

1.11 The Relevance of Social Science Research

The field of social science encompasses the many aspects of individual and social behavior. Research in this field addresses how social patterns of behaviour affect a population's overall health status--both physical and mental--and predicts the likelihood of diseases or disorders within a population. Through understanding the principles underlying behavioural and social processes, practical applications can be used to improve health outcomes within a population.

Identification : Social science is a field that studies the many aspects of human society. Anthropology, psychology, linguistics, economics and political science are some of the areas that fall within social science studies, according to the Office of Behavioural and Social Sciences Research. Research in this field uses an assortment of tools and techniques to gather data and formulate theories on individual and group patterns and behaviours. Its relevance lies in a clearer understanding of the principles underlying human behaviour and social function and the ability to predict certain outcomes.

Function : A society's overall health and well-being can be influenced by the behavioural and social patterns that prevail, according to the Office of Behavioural and Social Sciences Research. Social science research works to formulate theoretical frameworks based on accumulated data. Research approaches use basic fundamental procedures---like surveys, questionnaires and direct observation--along with practical applications that assess health outcomes and risk factors. Practical applications may be carried out within hospital, clinical or laboratory settings with a focus on how behavioural and social influences affect existing risk factors. Risk factors are any influences that cause damage to health. Also of interest are protective factors, which are influences that work to protect health.

Social Processes: Social processes within small groups, organizations, communities and populations are areas of study within social science research, according to the Office of Behavioural and Social Sciences Research. Individual patterns of behaviour in terms of how a person perceives situations and how she interacts with her environment is also an area of interest. Interactions between different types of groups and individuals help in determining the socio-cultural factors and psychological processes that create a society's way of life. Also at issue are how environmental factors affect behavioural patterns and how a population interacts with them.

Bio-psycho-social Processes: Social science research techniques are designed to identify the behavioural and social risks associated with health conditions and the onset and course of illness, according to the Office of Behavioural and Social Sciences Research. Ultimately, researchers look to identify relationships between biological factors and behavioural-social conditions. These associations make up the bio-psychosocial processes affecting a society's or population's health status. Information gathered from treatment outcomes research can also provide data on how bio-psychosocial processes affect the types of illness prevalent within a population.

Potential: Social science research findings are used to improve upon existing methods and techniques within the field, as well as within the biomedical and behavioral fields, according to the Office of Behavioral and Social Sciences Research. Research findings incorporate a "big picture" perspective in terms of how the whole affects its parts, and vice versa. Through ongoing data collection and analysis, researchers attempt to develop frameworks for predicting health outcomes for a population. These frameworks can then be used to develop practical applications that influence existing risk and protective factors.

1.12: How do Social Sciences Differ from Natural Sciences?

The term “social sciences” encompasses a wide range of fields. Despite the diversity, it is easy to assume that any one social researcher can cover the entire range of social sciences. However, just as entomologists would not be expected to research fisheries issues, sociologists would not be expected to investigate political science questions. Social scientists recognize that the social world consists of individually as well as socially constructed knowledge, which leads to a diversity of “world views.” This diversity is best investigated and understood by means of a diverse set of research methods. Each management question and decision may benefit from the application of one or more of the social sciences. And this diversity also may explain why different fields of social science occasionally appear to be in significant disagreement with each other.

Social science research can include either quantitative or qualitative methods, and both methods are frequently used in a single study. Qualitative data collection methods provide in-depth understanding of a single topic and are unique to the social sciences. Qualitative data-collection methods include individual and group interviews, surveys, content analysis (such as document review), observation, and case studies. Quantitative methods typically include surveys and document reviews, as well as the modelling of statistical trends and behaviour.

Social science research is complicated by ethical concerns because it inevitably involves human subjects. Participation in studies needs to be voluntary and fully consensual.

What Does Social Science Research Do?

While the field is too large to present all possible applications of social science research in this paper, here we highlight 10 important contributions of social sciences in informing decision making in resource management.

1. Social Science research generates documents / data
2. Social Sciences examine relationships between factors
3. Social Sciences analyze values, attitudes, and perceptions
4. Social Sciences design, describe, and evaluate decision processes
5. Social Sciences investigate power relationships and act as agents of change

6. Social Sciences contribute to understanding indigenous knowledge systems and human culture
7. Social Sciences analyze political and institutional structures
8. Social Sciences assess social and economic impacts
9. Social Sciences helps to estimate economic valuation which is concerned with identifying and estimating the economic value associated with any particular resource, including market values and non-market values.
10. Social Sciences evaluate projects and provide decision support tools. The social sciences comprise a large array of approaches that can assist resource managers in formally structuring and making decisions. In many cases these tools are not isolated methods, but instead they rely on elements of the methods mentioned above.

1.13. Types and Methods of Research: Every Research needs lots of dedication from the researcher's part-the amount of dedication mainly depends on the subject matter of the research. Before undertaking any research in any subject areas one must be sure about the intended purpose of the research-this purpose determines what type of research one is going to undertake. Any scientific research may fall into the following three broadly categories:

Exploratory research-This type of research may generate any novel idea in the domain of knowledge. It is primarily done for the purpose of finding anything new in any subject arena and always tries to shed some light in the unknown domain of knowledge. This kind of research also
Help us to generate new discipline in sciences and help us to identify problems of those particular research areas.

Constructive research- This is mainly done by many technological corporates in order to find new/alternative solutions to any particular crisis or problems. For example-renewable energy research or development of the capacity of optical fiber may fall into this category of research.

Empirical research-This is very impressive observational type of research, where one observes or test on real-life data or analysis the pattern of some specific events in order to identify the nature or the class of trend that specific phenomenon maintains. Based on the test result, researchers try to draw lines in order to predict the result of that type of incidents with certain level of confidence.

1.14. Economy and Society: Some Theoretical Concepts

Karl Marx:

Karl Marx has distinguished between different types of societies on basis of economic system. These are primitive communism, ancient slave production, feudalism and capitalism, socialism and communism. A man is both the producer and product of society. Marx's analysis of history is based on his distinction between the means/forces of production literally those things, such as land, natural resources, and technology, that are necessary for the production of material goods, and the relations of production in other words, the social and technical relationships people enter into as they acquire and use the means of production. Together these comprise the mode of production. Marx observed that within any given society the mode of production changes, and that European societies had progressed from a feudal mode of production to a capitalist mode of production. Marx did not understand classes as purely subjective. He sought to define classes in terms of objective criteria, such as their access to resources. For Marx, different classes have divergent interests, which is another source of social disruption and conflict. Marx was especially concerned with how people relate to that most fundamental resource of all, their own labour power. Marx wrote extensively about this in terms of the problem of alienation. For Marx, the possibility that one may give up ownership of one's own labour - one's capacity to transform the world - is tantamount to being alienated from one's own nature; it is a spiritual loss. Marx considered the capitalist class to be the most revolutionary in history, because it constantly revolutionized the means of production.

Max Weber

Max Weber formulated a three component theory of social stratification with social class, status class and party class (or political class) as conceptually distinct elements. Social class is based on economically determined relationship to the market (owner, employee etc.). Status class is based on non-economical qualities like honour, prestige and religion. Party class refers to affiliations in the political domain. All three dimensions have consequences for what Weber called "life chances". According to Weber there are two sources of power. One is derived from constellation of interests that develop in a free market and the other is from an established system of authority that allocates the right to command and the duty to obey.

Emile Durkheim

Emile Durkheim sees division of labour in terms of social process. He has tried to determine the social consequences of the division of labour in the modern societies. He has made a fundamental difference between pre-industrial and industrial societies and also made difference between two types of solidarity- mechanical solidarity and organic solidarity. Mechanical solidarity prevails in simple folk societies where division of labour is restricted to family, village or small region. Here individuals do not differ much from one other and follow the same set of norms, beliefs etc. Organic solidarity holds the modern societies together with a bond. Here societies are large and people are engaged in variety of economic activities. They hold different values and socialize their children in varying patterns. The conditions of the modern society compel division of labour to reach the extreme level. This extreme form of division of labour leads to feeling of individualism or anomie. Anomie according to Durkheim refers to a state of normlessness in both the society and the individual. It is a social condition characterised by the breakdown of norms governing social interaction. People feel detached from their fellows having little commitment to shared norms people lack social guidelines for personal conduct. They are inclined to pursue their private interests without regard for the interests of society as a whole.

Karl Polyani:

According to economist Karl Polanyi, the three principles of exchange are market principle, redistribution, and reciprocity. The market principle describes the buying and selling of goods and services based on the laws of supply and demand (things cost more the scarcer they are and the more people want them), and often involves bargaining. It is associated with industrial societies and involves a complex division of labour and central government. In redistribution, products move from the local level to a hierarchical centre, are reorganized, and sent back down to the local level. Redistribution is the main form of exchange in chiefdoms and some industrial states, and works with the market system. Polyani identifies reciprocity of three kinds: generalized, balanced, or negative. Generalized reciprocity involves an exchange between closely related people in which the giver expects nothing concrete or immediate in return. Another form of reciprocity is balanced reciprocity, in which the social distance between giver and recipient increases relative to generalized reciprocity. It involves an exchange outside the immediate family, and the giver expects something in return in the future, but not immediately. If there is no reciprocation, the relationship between the two parties will be strained. The third kind of reciprocity is negative reciprocity, which is an exchange relationship in which parties do not trust each other and are strangers. The giving must be reciprocated immediately and there is very little communication, if any, between groups. Each group is trying to maximize its economic benefit, but eventually friendly relationships between the groups may develop.

1.15. Economic system of simple societies

Herbert Spencer has defined simple society as one which forms a simple working whole unsubjected to any other and of which the parts cooperate for certain public ends. Simple societies have low division of labour. The occupational differentiation being limited primarily to birth, sex and age. These societies have no specialized economic organization. The productive skills are simple and productivity is low therefore these societies cannot sustain large population size. Most of the adult members are engaged in food gathering activities. There is little or no surplus so the social inequalities are not significant and economic interaction takes place within egalitarian frame-work. The production system is simple but exchange of goods and services assume a complex form. The forms of exchange are reciprocal and redistributive type. Some of the simple societies inhabiting regions having abundant food and other resources indulge in conspicuous consumption. The members lack high degree of achievement motivation as there is neither any intense preoccupation on generation and accumulation of economic surplus. In fact most economic activities emphasize on giving rather than storing or accumulation. Private ownership of means of production is non-existent. There is no clear separation between domestic economy and community economy as they overlap to varying degrees. The economic system is dominated by sacred consisting of magic-religious ideas. The innovation is rare and change is slow. The customary practices and norms regulate production and exchange of goods and services.

1.16. Social determinants of Economic Development

Economic development implies two things: Economic growth which leads to increase in production and generation of income and equitable distribution of this income among the population to improve the quality of life. Although economic development does not necessarily imply industrialization there is no historical precedent for substantial increase in per capita income without diversion of both capital and labour from agriculture. Economic development is synonymous with industrialization. Economic development is very much influenced by various social factors. Nation states are created with common language and culture. Economic development of any country hinges on the efficient employment of factors of production such as labour, land, capital and organization. There is commercialization of production with monetization of economy. The employment of factors of production is conditioned by cultural and social factors. The people must have the required ability, experience and knowledge to make the best use of the facilities that are made available. There is decline of the proportion of the working population engaged in agriculture. The technology plays very important role when appropriate social conditions are present.

There is trend towards urbanization of society with growth of scientific knowledge. A new value system emerges which emphasis individual initiative and responsibility and enables the individual to function without any control. The exclusiveness of clan, kin or caste breaks down and provides norms of behaviour suited to the secondary group type of relationship characteristic of an industrial society. There is widespread spread of education. The social stratification emerges based on achievement criteria and permitting occupational mobility.

1.17. Social Control

Social control has been defined by Maclver as the way in which entire social order coheres and maintains itself –how it operates itself as a whole as a changing equilibrium. Mannheim defines social control as the sum of those methods by which a society tries to influence human behavior to maintain a given order. To Ogbuand an established rule is social control. Ogburn and Nimkoff the patterns of pressure which a society exerts to maintain order E Ross defines it the system of devices whereby society brings its members into conformity with the accepted standards of behavior. According to Lapiere social

control is a corrective for inadequate socialization. G.A Lundberg has defined social control as designating those social behaviors which influence individuals or groups towards conformity to established or desired norms. Kimball Young defines social control as the use of coercion, force, restraint, suggestion or persuasion of one group over another or of a group over its members or of persons over others to enforce the prescribed rules of the game. These rules may be set down by the members themselves as in a professional code of ethics or they may be those laid down by a larger, more inclusive group for the regulation of other smaller groups.

Social control may thus be defined as: Social control is an influence where influence may be exerted through public opinion, coercion, social suggestion, religion or any other method. The influence is exerted by the society which means that group is better able to exercise influence over the individual than a single individual. This group may be the family, the church, the state or the trade union etc. The effectiveness of the influence however depends on variable factors. Sometimes family may exercise more effective influence than the state or vice-versa.

The influence is exercised for promoting the welfare of the group as a whole. The person is influenced to act in the interest of others rather than in accordance with his own individual interests. Social control is exercised for some specific end in view. The aim is always the welfare of the whole. The individual is made conscious of other's existence; their interests. He is made to adhere to the appropriate social ways.

Social solidarity is essential for the existence of society. No two persons is alike in their nature, ideas, attitudes and interests. Every individual is a separate personality. There are cultural differences among the individuals. As a matter of fact society is a heterogeneous organization. If every individual is allowed unrestricted freedom to act and behave, it may create social disorder. For an orderly social life social control is necessary. The aims of social control are to bring out conformity, solidarity and continuity of a particular group or society.

Social control is necessary for maintaining order in the society. It is necessary for every society or group to maintain its social order and this is possible only when its members behave in accordance with that social order. An important objective of social control is to maintain the old order. Although enforcement of the old order in a changing society may hinder social progress, yet it is necessary to maintain continuity and uniformity in society. Without social control social unity would be a mere dream. Social control regulates behavior in accordance with established norms which brings uniformity of behavior and leads to unity among the individuals. The family maintains its unity because its members behave in a similar manner in accordance with family norms.

No two men are alike in their attitudes, ideas, interests and habits. Even the children of same parents do not have the same attitudes, habits or interests. Men believe in different religions, dress, eat differently and have different ideologies. There are so many differences in the ways of living of the people that at every moment there is the possibility of clash between them. In modern times this possibility has all the more increased because man has become too self-centered. Social control is necessary to protect social interests and satisfy common needs. If social control is removed and every individual is left to behave freely society would be reduced to a state of lawlessness.

Further reading

1. Solving the Social Sciences' Hard Problems

http://harvardmagazine.com/2010/04/social-sciences-hard-problems

2. Social Science and social problems

http://understandingsociety.blogspot.com/2008/02/social-science-and-social-problems.html

3. Abhijit Kundu. , Methodology and Perspectives of Social Science, New Delhi, Pearson Education

MODULE II
INTRODUCTION TO MICRO ECONOMIC THEORY

The purpose of Theory

A Theory is a hypothesis that has been tested and proved. A hypothesis is an intelligent guess (a supposition or an assumption) the validity of which is to be tested. The purpose of theory is to predict and explain. A hypothesis is tested by its ability to explain and predict accurately and not by the realism of its assumptions.

The function of Microeconomic Theory

Microeconomic theory (price theory) studies the behavior of individual decision making units, e.g., consumers, resource owners and business firms in a free enterprise economy.

Market

A market is the place in which buyers and sellers buy and sell goods, services and resources. In a wider sense, a market may signify any area in which buyers and sellers are in contact with one another and this area may in fact comprise the whole world. Prices in a free market are determined by the combined actions of buyers and sellers. The function of a market is to enable an exchange of goods or services to take place, a means by which buyers and sellers are brought into contact with one another.

A Function

A function shows the relationship between two or more variables. It indicates how the value of one variable (dependent variable) depends on the value of other (independent) variables.

Slope

In mathematics, the slope or gradient of a line describes its steepness, incline, or grade. A higher slope value indicates a steeper incline. The slope is defined as the ratio of the "rise" divided by the "run" between two points on a line. In other words, the ratio of the altitude changes to the horizontal distance between any two points on the line. Given two points (x_1, y_1) and (x_2, y_2) on a line, the slope m of the line is:

$$m = Y_2 - Y_1 / X_2 - X_1$$

In other words, slope is the ratio of the vertical side of the triangle to the length of the horizontal side. Thus, Slope = Vertical Side / Horizontal side equals the rise/the run. Through differential calculus, one can calculate the slope of the tangent line to a curve at a point.

Equilibrium

Equilibrium is a situation in which economic forces have no tendency to change. Thus, equilibrium is the market conditions where there is no tendency to change.

Comparative Statics

Comparative Statics studies and compares two or more equilibrium positions, without regard to what happened in the transitional period.

Dynamics

Dynamics deals with the time path and the process of adjustment itself.

Partial Equilibrium

Partial Equilibrium analysis is the study of the behavior of individual markets viewed in isolation.

General Equilibrium

General Equilibrium analysis studies the behavior of all individual markets simultaneously by considering that all markets are interrelated.

The word economics comes from the ancient Greek word *oikonomia* which means management of a household. Previously Economics was known as Political Economy. At present Economics is a social science that deals with human wants and their satisfaction. It is mainly concerned with the allocation of resources for the maximization of welfare of the people. In other words, Economics is a social science that studies about production, distribution and consumption of goods and services. The existence of human wants is the starting point of all economic activity. We cannot get all we want by sitting idle. We have to work hard for getting goods and services to satisfy our wants. Hence, due to the existence of human wants, people engage in economic activities or they make efforts to satisfy their wants. Hence, wants, efforts, satisfaction constitute the circle of economics.

Definition of Economics

Defining economics precisely is a difficult task. Different writers defined economics differently and hence we do not have a generally accepted definition of economics. However, generally definitions of economics are grouped in to four and they are (i) Wealth Definition (ii) Welfare Definition (iii) Scarcity Definition and (iv) Growth Definition. We may explain these definitions in a little detail.

Wealth Definition (Adam Smith's Definition)

The classical economist Adam Smith (1723-90) has defined economics in his well-known book "An Enquiry into the Nature and Causes of Wealth of Nations (Wealth of Nations, 1776) as the *Science of Wealth*. He is regarded as the Father of Political Economy. According to Smith, Economics was regarded as the Science which studies about production and consumption of wealth. Wealth in Economics generally means those things that are necessary to satisfy human wants. However, only those goods that are relatively scarce and have money value are generally considered wealth. In other words, according to Smith, economics is concerned with the problems arising from wealth-getting and wealth using activities of men.

Criticism of Wealth Definition

The Wealth Definition of Adam Smith has been criticized by writers in Social Sciences. Social Scientists like Carlyle and Ruskin have called economics as "dismal science", "dark science", because of its emphasis on wealth. The term 'wealth' was mistaken to be a substitute for 'money' hence; economics was criticized as the 'bread and butter' science. Economics was interpreted as a science that taught selfishness and love of money.

1. This definition has given emphasis to the material wealth. The classical economists had taken a narrow view of wealth confining to earning and spending of money. The economic activity of a man need not be selfish.
2. The wealth definition is unscientific and incomplete. The definition is incomplete in the sense that it gives an impression that human wants are satisfied only through earning and spending money. A man gets satisfaction not only from earning and spending of wealth but also through other activities like singing, dancing, painting etc. These activities have economic aspects involving choice and satisfaction of human wants which are outside economics according to Classical economics..
3. The definition gives an impression that wealth is the end of all human activities. Wealth is only a means to an end and not an end itself.

On account of the above criticisms, the wealth definition was discarded at the close of the 19th century. While evaluating the wealth definition of Smith, We should remember that Adam Smith was writing his book at a time when wealth was produced on a large scale in England. This was the reason behind Smith's emphasis on wealth in his definition of economics. In fact he was interested in increasing the total volume of production in the economy and even today it is an important aspect of the economic policies all over the world. Hence, there is an element of truth in Smith's definition.

Welfare Definition (Alfred Marshall)

Alfred Marshall (1842-1924) was the fist economist who saved economics from the vehement criticisms from social scientists during 19th century owing to the classical definition of economics. He shifted the emphasis from 'wealth' to 'welfare'. Wealth, according to him, was not the end but only a means to an end, the end being human welfare. He defined economics in accordance with his ideas of human welfare. He defined economics in his book, *Principles of Economics* (1890), thus: "*Political Economy or Economics is the study of mankind in the ordinary business of life: it examines that part of individual and social action which is most closely connected with the attainment and with the use of the material requisites of well-being".* Hence, Marshall emphasized that economics is a social science that studies social problems. Marshall accepts the view that economics studies about wealth. However, according to him the more important aspect of economics is that it studies about man. Marshall thus makes man the centre of his study. He considers the study of man as more important than the study of wealth. According to him, economics studies about mankind in the ordinary business of life meaning that earning activities. Marshall's definition is not concerned with social actions which do not bring material welfare. Besides Marhsall, Pigou, and Cannan defined economics in 'welfare' terms.

There are certain implications of Marshall's definition. Firstly, it is a study of human beings. Secondly, it studies the economic aspects of the life of human beings. An individual has several aspects of his life- social, religious, political and economic. Economics, it is evident has no concern with the social, religious, and political aspects of human life. Economics is concerned purely with the economic aspect of human life. In other words, it studies about material welfare – how one earns income and how one spends income. Thirdly, economics studies about human welfare - economic or material welfare.

While classical writers emphasized only wealth, Marshall has given importance to man and wealth. It is thus clear that Marshall laid emphasis on material welfare as the primary concern of economics.

Criticisms of Marshall's Definition

It should be noted that the welfare definition of economics has been criticized on several grounds.

1. Welfare Definition excludes non-material things from the scope of economics. Welfare definition includes only materials things as economic activity and hence it is unsatisfactory and unscientific according to Lionel Robbins.
2. According to Lionel Robbins, the inclusion of the concept of welfare makes economics normative. According to Robbins, economics has noting to do with welfare. Welfare is subjective and is an immeasurable. In our practical life, we undertake many activities which do not promote welfare. For example liquor and cigarettes are not conducive for human welfare and yet, in economics, we are concerned with the production and consumption of these items.
3. Marshall's definition is contradictory. He regards the services of singers and dancers as productive so long as they are demanded by the people. But since they are non-material, they do not promote human welfare and hence their services are not subject matter of economics.
4. Welfare definition is classificatory and not analytical.

In spite of the criticisms mentioned above, it should be remembered that the definition of Marshall has widened the scope of economics by taking into account wealth as a part of the study of economics in relation with the welfare of mankind.

Scarcity Definition (Lionel Robbins)

Lionel Robbins (London School of Economics) defined economics in his book, *An Essay on the Nature and Significance of Economic Science* (1932). as "the science which studies human behavior as a relationship between ends and scarce means which have alternative uses." The three fundamental ideas included in this definition are *ends, scarce means* and their *alternative uses*. Ends here means human wants. Scarcity is the fundamental economic problem as humans wants are unlimited whereas resources by which we can satisfy our wants are limited. It states that society has insufficient productive resources to fulfill all our wants. Goods (and services) that are scarce are called economic goods. According to Robbins economics is entirely neutral between ends as he considers economics as a pure science free from value judgments. Robbins' definition has certain implications and they are:

1. Human wants (ends) are unlimited. If a particular want is satisfied, some other wants will crop up in its place and there would be no end to our wants. Multiplicity of wants makes it very important for people to work continuously for its satisfaction.
2. The means to satisfy our wants are scarce. For example, time is limited, money is limited, and resources are limited. In short, means are scarce (limited) in relation to our wants.
3. The scarce means can be used for alternative uses which increases the scarcity. For example, a piece of land can be used for growing rice or wheat. However, use of resources for one use prevents its use for any other use.
4. Since wants are unlimited and we do not have resources to satisfy all our wants, we have to make a choice between more urgent and less urgent wants. In other words, man has to make a choice about the wants to be satisfied and the way the resources are to be utilized in the process of satisfying human wants.

Superiority of Robbins' Definition

Robbins' definition is superior to the earlier definitions.

1. Universal application: Robbins' definition is has universal applicability because scarcity is felt by everybody at every time and place. Hence, it is applicable to the economies of all countries whether capitalistic, socialistic or a mixed economy.
2. Wider Scope: Robbins' definition has widened the scope of economics. According to this definition, economics studies about all activities of man whether they are concerned with material or non-material welfare.
3. Analytical and Scientific: The definition of Robbins is considered to be more scientific and more analytical as this definition examines the economic aspect of all activities. This definition is neutral between ends. It does not take into account the ethical aspects of economic problems. According to him, an economist has no business to judge whether a particular economic action is morally right or not.

4. The scarcity definition has given rise to the concept of 'scale of preferences' which has great importance in economic analysis.

Criticisms of Scarcity Definition

Although Robbins' definition is superior to earlier definitions of economics, this definition is also criticised on the following grounds:

1. No Human Touch: Robbins' definition reduces economics to the theory of value without any human tough. He considered economics merely a science of the pricing process.
2. Does not cover Economic Growth: In recent times, economic growth (development)has become an important branch of economics. Robbins' definition does not pay attention to factors which increase national income and productive capacity of the economy.

3. Economic Problems may arise due to plenty: It is possible that economic problems arise not only due to scarcity but also due to plenty. According to Robbins', economic problems arise only due to scarcity and neglected the chances of economic problems due to plenty. During the period of the Great Depression of 1930's, the problem was not scarcity but abundance of goods.
4. The definition is static, narrow as the macro economic aspect was ignored from his definition. Keynesian economics, which studies how the levels of national income and employment are determined, has now become an integral part of economics.

Growth Definition (Samuelson)

Paul A.Samuelson has defined economics and his definition is known as growth definition of economics. According to him, "Economics is the study of how men and society end up choosing, with or without the use of money, to employ scarce productive resources that could have alternative uses to produce various commodities, over time, and distribute them for consumption, now or in the future, among various people or groups in society. It analyses the costs and the benefits of improving patterns of resource allocation".

Features of Samuelson's Definition

In Samuelson's definition, we find most of the points found in Robbins' definition. The important characteristics of growth definition are as follows:

1. Samuelson has emphasized the problem of scarcity of resources in relation to unlimited wants. He has also accepted the alternative uses of resources.
2. This definition has incorporated the time element which makes the scope of economics dynamic. This is an improvement of Samuelson over Robbins' definition.
3. His definition has stresses the importance of the problem of distribution and consumption along with that of production. The definition also emphasizes the consumption of various commodities produced overtime and on their distribution and for future economic growth.

In short, Samuelson's definition is growth oriented with universal appeal with dynamism. Hence, Samuelson's definition may be considered as a modern and general definition of economics. We have examined the popular definitions of economics. The classical economists emphasised wealth; the neo-classicals welfare, while Robbins' scarcity of resources. It is really difficult to delimit boundaries of economics by means of a definition. Generally economics is a social science that studies about human wants and their satisfactions with an aim to increase the production of goods and services over time.

Subject matter or Main sub-divisions of Economics

There are four main divisions of economics and they are consumption, production, exchange, distribution However, in modern times we add one more aspect and that is public finance.

Consumption: Consumption is the branch of economics that deals with the satisfaction of human wants. Consumption is using up of goods and services for satisfaction of wants. Consumption studies about the nature of wants, the classification of wants, laws relating to consumption like law of diminishing utility and the law of demand.

Production: Production means creation of utility or wealth. Utility refers to the capacity of a good to satisfy a want. Under production, we study about the laws which govern the factors of production

Exchange: Exchange deals with the giving sand taking of one thing for another. Goods may be exchanged for goods or for money. Under exchange we study about the functions and characteristics of money, the role of banks, price determination etc.

Distribution: Distribution studies about how the wealth produced by the four factors of production (land, labour, capital and organization) are distributed among these factors as rent, wages, interest and profit.

Public Finance: Public Finance deals with the income and expenditure of public authorities like central, state or local governments.

Micro Economics

Economic theory is broadly divided into micro economics and macro economics. These terms were first introduced by Ragnar Frisch. The term micro economics is derived from the Greek word micros which means small. The term macro economics is derived from the Greek word macros which means large. Thus, micro economics deals with a small part of the economy of a country. In other words, microeconomic theory (Price Theory) studies the economic behavior of individual decision making units (eg., consumers, resource owners, and business firms) in an economy.

Definition of Micro economics

Micro means a millionth part. Micro economics is the study of the economic actions of individuals and small group of individuals. Thus micro economics may be defined as that branch of economic analysis which studies the economic behavior of the individual unit – a person, a household or a firm. In micro economics, we study the various units of the economy and how they function and how they reach their equilibrium. In other words in micro economics, we analyse only a tiny part of the economy at a time. In microeconomics, we study the various units of the economy; how they function and how they reach their equilibrium. In other words, in micro economics, we attempt only a microscopic study of the national economy. In micro economics, we enquire about how a particular person maximizes his satisfaction or how a particular firm maximises it's profits. An important tool used in micro economics is the marginal analysis. For example, the law of diminishing utility, the law or equi-marginal utility are based on marginal analysis.

Scope of Micro economics

Micro economics studies about price and value theory, the theory of the household, the firm and the industry. It also studies about production and welfare theory. In short micro economics studies about:

a. Theory of Product Pricing with its two constituents, namely the theory of consumer's behavior and the theory of production and costs.
b. Theory of factor pricing or the theory of distribution. This aspect studies about the theories of wages, rent interest and profits.
c. The theory of economic welfare. It is sometimes referred to as Price theory. Thus the prices are the core of microeconomics.

Importance of Microeconomics

Microeconomics is an important method of economic analysis and it occupies an important place in the study of economic theory. Microeconomics tells us how a free market economy with its millions of consumers and producers works to decide about the allocations of productive resources among the thousands of goods and services. It tells us how the goods and services produced are distributed among the various people for consumption through price (Market) mechanism. Microeconomic theory explains the conditions of efficiency in consumption and production and highlights the factors which are responsible for the departure from the efficiency (optimum). Thus, microeconomics has theoretical and practical importance which can be summarized as follows:

a. Helps to Use Resources Efficiently**:** Microeconomics is helpful in the efficient employment of the limited, scarce resources of a country. The principal problem faced by the modern governments is the allocation of its scarce resources among competing uses. Microeconomics theory explains the conditions of efficiency in consumption and production, which are the vital parts of economics. .Microeconomic theory suggests suitable policies which should be adopted by modern governments to promote economic efficiency for achieving the all-round development and stability of the economy of the country.
b. Helps to Understand a Free enterprise Economy. Microeconomics is of great importance in understanding the working of free enterprise economy without any central planning and control.
c. Basis for welfare economics. The greatest advantage of micro economics is that it provides the basils for welfare economics. The entire structure of welfare economics has been built on price theory which is an ingredient of micro economics.

Limitations of Microeconomics

It should be noted that there are certain limitations to micro economics which are briefly explained below:

1. Fallacy of Composition: What is true of an individual may not be true in the case of aggregates. For example, thrift may be good for an individual but bad or harmful to the society as a whole. It the society starts saving more, effective demand will come down ad the employment will decline and the economy as a whole would suffer on account of thrift..
2. Macroeconomic analysis is based on unrealistic assumptions like full employment which is really unrealistic.
3. There are certain problems which cannot be analysed with the aid of micro economics. For example, important problems relating to public finance, monetary policy and fiscal policy etc. are beyond the purview of microeconomics.

Macroeconomics: The e word ‘macro’ is derived from the Greek word ‘makros’ meaning large and therefore, macroeconomics is concerned with the economic activity in the large. Macroeconomics may be defined as that branch of economic analysis which studies the behavior of not one particular unit, but of all the units taken together. In other words, macroeconomics may be defined as that branch of economic analysis which studies the behavior of not one particular unit, but of all the units combined together. Macroeconomics study is a study of aggregates. Hence, it is often called Aggregative Economics. It is the study of the economic system as a whole. It is the study of the overall conditions of the economy like total production, total consumption, total savings and total investment. According to K.E.Boulding, “Macroeconomics deals not with individual quantities as such but with aggregates of these quantities; not with individual incomes; but with the national incomes; not with individual price but with the price level, not with individual outputs but with the national output. As such macroeconomics deals with great averages and aggregates of the system.

Differences between Micro and Macro Economics

1. In microeconomics, we study the working of individual markets where as in macroeconomics; we study about aggregates or totals such as total output, total employment and total income.
2. In micro economics, the unit of study is a part where as in macro economics we study about the economy as a whole. For example, micro economics takes the general price level as given whereas in macro economics we take it as a variable.
3. In micro economics we are concerned with partial equilibrium where as in macro economics we are concerned with general equilibrium.
4. In micro economics we study about relative prices assuming that the general price level is given whereas in macro economics we study about changes in general price level like inflation and deflation.

Wants

Human Wants

Man is a bundle of desires. Human needs are material items people need for survival, such as food, clothing and housing. In other words, human needs are generally understood as things necessary to enable a human to continue to live. It is generally understood that humans need food, clothing, and shelter without which existence of man's life is not possible. Hence human needs are limited. On the other hand, human wants are infinite in variety and number. Wants also vary from individual to individual and they multiply with the stages of development of a country. Consumption studies about the satisfaction of human wants. All men have wants. Existence of human wants is the starting point of all economic activity. Under consumption, we study about the nature of wants, the classification of wants and the laws relating to consumption.

Characteristics of Human Wants:

1. **Human wants are unlimited:** Man is never completely satisfied and hence there is no end to human wants. When one want is satisfied another want will crop up to take its place and thus there is a never ending cycle of wants. Hence, no man is completely satisfied forever.
2. **A Particular want is Satiable:** Although we cannot satisfy all our wants, we can satisfy a particular want if there are resources..
3. **Wants are Complementary:** Sometimes for satisfying a particular want we need several things together. For example, to write one should have pen and ink. Hence, the relationship between pen and ink is complementary.
4. **Wants are Competitive:** Wants compete with each other for our limited resources (Money). For example with Rs. 25, one can have meals or can see a Cinema or buy a note book. So one has to make a choice.
5. **Wants are alternative:** There are several ways of satisfying a particular want. For example, .If a man is hungry, his hunger may be satisfied by bread, rice or fruits. The final choice depends upon availability of money and the relative prices.
6. **Wants vary with time place and person:** Wants are not always the same. It varies with individual to individual. People want different things at different times and in different places.
7. **Wants multiply with Development:** With the level of economic development wants multiply. With the development of a country, the demand for radio, T.V, motor-car etc, increases.
8. **Wants are Recurring:** Some wants are recurring in nature, e.g. food we require food again and again.
9. **Wants are influenced by income, salesmanship and advertisement:** It income is higher more wants can be satisfied. Many things we buy of particular brands due to salesmanship or advertisement.
10. **Wants are the result of custom or convention:** As a part of custom and convention we buy many thins. Really they are not required but unlikely we have to purchase it e.g. expenses on social ceremonies.

Classification of Wants

Wants may be classified into necessaries, comforts and luxuries.

A. Necessaries: Necessaries are those things which are most essential and without which we cannot live. These can be sub divided as:

- Necessaries of existence: The things without which we can not exist e.g. water, food, clothing, shelter.
- Conventional necessaries: The things which we are forced to use by social custom.

B. Comforts: After satisfying our necessaries we desire to have some comforts. Comforts give pleasure and add to the efficiency of the person. For example table and chair for a student help to increase the efficiency. But cushioned costly chair is not a comfort.

C. Luxuries: Luxury means superfluous consumption. After getting comforts, man desire luxury. Luxuries increases pleasure of a person but do not increase one's efficiency and hence some luxuries are waste. The luxury articles need not required e.g. gold and silver, costly furniture, etc.

Scarcity:

The word scarce means limited or insufficient. Scarcity of resources is the fundamental problem of every society. As the resources of every society are limited (scarce), the ability of the society to produce goods and services are also limited. In other words, human wants are unlimited and the means to satisfy them are limited, every society is faced with the twin problems of scarcity and choice. Hence, all societies face some basic (fundamental) problems - production, technique of production (combination of factors) and distribution.

The Problem of Scarcity

We live in a world where everything is scarce. People want a variety of goods and services. This implies that human wants are unlimited and the means to fulfill them are limited. At a particular time, the economy can produce only a limited amount of goods and services. This is because of the scarcity of resources like land, labour, capital and organization. These factors of production (inputs) are used to produce goods and services (Economic goods). These factors explain scarcity is the basic problem of every society. Thus the law of scarcity states that human wants are unlimited and the resources available to satisfy theses wants are limited.

The Problem of Choice

As the resources to produce goods and services are limited, a society can produce only a small portion of goods and services it want. Therefore, scarcity of resources results in the fundamental (basic) economic problem of choice. As a society cannot produce all the goods and services to satisfy all the wants of the people, it has to make a choice regarding the goods and services to be produced at present. The economic problem fundamentally revolves around the idea of choice.

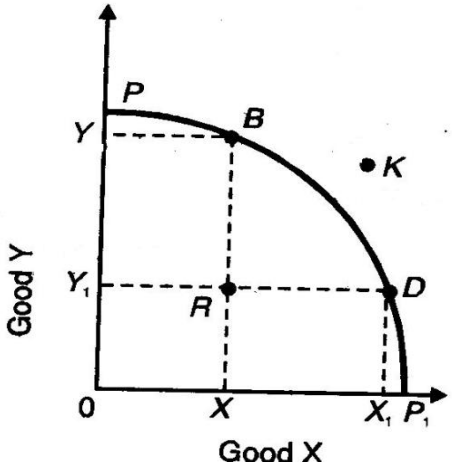
A decision to produce one good may result in a decision not to produce another good. So choice involves sacrifice. Thus every society is faced with the basic problem of deciding what it is willing to sacrifice to produce the goods it wants the most. For example, if the country decides to have more hospitals, it may have to reduce the resources available for the construction of schools. The sacrifice of the alternative (Schools) in the production of a good (Hospitals) is called the opportunity cost. Similarly there arise the problem of choice in selecting techniques of production (combination of factors) when there are alternative ways of producing goods.

The Production Possibility Curve (PPC)

Choices (alternatives) can be explained with the help of a production possibility curve. A production possibility curve shows all the possible combinations of two goods that a society can produce within a specified time period when the resources are fully (efficiently) utilised. In other words, production possibility curve is the locus of all combinations of two goods that can be produced with the given resources. PPC is also known as Production Possibility Frontier (PPF) or Product Transformation Curve.

In the figure below, PP_1 is the production possibility curve (fig.2.1) which shows the problem of choice between two goods X and Y in a country. Good X is measured on the horizontal axis and Good Y on the vertical axis. PP_1 curve shows all the combinations of X and Y that can be produced with the available resources. If the entire resources are used in the production of Y, the economy can produce OP quantity of Y and if the entire resources are used in the production of X, the society can produce O P_1 quantity of X. Point B represents OY quantity of Y and OX quantity of X. If the economy chooses to produce more of X, it would have to sacrifice the production of some quantity of Y. The sacrifice of some quantity of Y is the opportunity cost of producing some more quantity of X. The PP_1 curve is downward slopping because production of more X, involves production of less Y.

Figure 2.1



Basic Problems of an Economy:

An economy is an aggregation of institutions and institutions include religious, political, social and economic. The sum total of institutions is called an economy. In other words, economy means an economic system. So, economy or economic system refers to the organization of the economic activities of the people for getting a living. It thus involves production of goods and services and distribution of them among the people. Although there are different types of economic systems the major ones are Capitalism, Socialism and Mixed economy. Some countries adopted Capitalism or capitalist economy for solving their problems where the market solves the basic problems. Some other countries adopted Socialism for solving their basic problems where state decides and solves the major problems of the society still other countries adopted a mixed economy where state and market together solves the major problems of the economy. We know that wants are unlimited and the means (resources) to produce goods and services are limited. Hence, a society has to decide which wants are to be satisfied first.

Capitalist Economy

Capitalism is an economic system under which all the means of production – farms, factories etc. are privately owned and production is for private profit. Private individuals are free to use these resources for making profit. The instruments of production are owned and managed by their owners exclusively for their own benefit. Although state make certain restrictions on the economic freedom of citizens in the interest of general welfare, the owners of the means of production can use their resources in any manner they like to earn profit. Individuals and institutions are free to produce any thing and to enter into contracts with others to make profit. The important features of capitalism are private property and the system of inheritance, economic freedom, price mechanism and consumer's sovereignty, perfect competition, inequalities in the distribution of income, lack of central planning etc. The U.S.A and Britain are the excellent examples of capitalist countries.

Socialistic Economy

It is quite difficult to define socialism precisely. It may mean different things to different people at different times. There are different types of socialism and hence socialism is compared to hat that lost its shape because everybody wears it. However, generally socialism is defined as a social and economic system in which the major means of production are owned and operated by the society or the government. Hence social ownership of the means of production is the most important feature of socialism. All the major factories, farms and businesses are conducted by the state and profits earned go to the state. As such under socialism production is not for profit and production is for satisfying the wants of the society. In other words, the important characteristics of socialism are social ownership of the means of production, central planning, equitable distribution of income and wealth. Socialism was the economic system that prevailed in Russia, China, Holand, Hungary and China although there were certain changes in these countries recently.

Mixed Economy

Mixed economy is an economic system in which the means of production are partly owned by the private and partly owned by the state. It is neither wholly capitalistic nor wholly socialistic. Both the features of capitalism and socialism are present under mixed economy. In other words, under mixed economy there will be both public sector and private sector. Today almost all economies are mixed in reality. Here, the private enterprises are not allowed to function freely through price mechanism and to maximize profit neglecting the social interests. Excellent examples of mixed economy are India, Sweden etc. In India, we have both public sector and private sector and hence mixed economy represents a mixture of capitalism and socialism. Recently almost all the countries of the world are considered as mixed economies as we cannot see a pure capitalistic and pure socialistic economy at present.

Communism or Communistic Economy

Communism or communistic economy represents the society in which all the means of production are owned by the community. Marx has elaborated the concept of communism in which the society is classless, casteless, and there is no difference between man and man on the basis of wealth and represents an egalitarian society. The communist principle of distribution is from each according to his ability to each according to his need. According to communism, people should be self less and should work for the community. Such a society never existed before although China and Russia were termed as communist countries.

Economic System

Functions of an Economy

There are some basic (central) problems that an economy has to solve. These problems or functions arise due to the multiplicity of wants and scarcity of resources. If resources were unlimited there would not have been any economic problem in the economy. In reality resources are not only limited but these resources have alternative uses. The basic functions that an economy or economic system has to perform are What to produce, How to produce, For whom to produce and how to choose between present and future. These functions are to be performed by all economic systems whether Capitalistic, Socialistic or Mixed economy .How ever, economic systems differ in their method to solve these problems. The Capitalistic economies solve these problems using the market or market mechanism (price mechanism) whereas Socialism solves these problems using the state mechanism by planning whereas mixed economy uses market as well as state to solve these problems. The economic problem is most simply explained by

the question how do we satisfy unlimited wants with limited resources?. In short, the economic problem is the choice one must make, arising out of limited means and unlimited wants. We will discuss these basic problems in some detail.

What to Produce?

We have seen that every economy has to solve the problem of what to produce. What to produce means what goods and services are to be produced and in what quantities these goods are to be produced. As resources are limited, no economy can produce as much of every commodity or service as desired by all members of the society. As such production of more of one good means less of other goods. Therefore, every society must decide the goods and services the economy has to produce and how much of these goods are to be produced.

How to Produce?

Once the economy has decided what goods and services are to be produced and in what quantities, it must decide how the chosen goods shall be produced. Thus, how to produce refers to the choice of the combination of factors and the technique of production to produce goods and services. We know there are alternative (different) techniques of producing a commodity. Clearly, this is a problem of the choice of production techniques. Different techniques (methods) of production would use different quantities of various factors and the society has to make a choice between these techniques. Production of goods can be by using labour intensive (technique which uses more labour and less capital) or capital intensive (technique that uses less labour and more capital) techniques. For example Cloth can be produced by Power looms in mills or it can be produced by handlooms in cottages. Thus, a good can be produced by using different factor combinations and techniques. The scarcity of resources demands that goods should be produced with the most efficient method. Thus, the society would choose that technique of production which minimizes the cost of production. Clearly, the choice between different methods would depend on the factor supply and the prices of the factors of production. Thus, it needs the knowledge of relative prices of factors of production (resources) to produce with minimum cost.

For whom to Produce?

Once, the problems of what and how to produce are solved, then arises the problem for whom to produce. For whom to produce refers to how the total output is to be divided among different consumers. Thus, for whom to produce means how produced goods are to be distributed among the people. In other words, for whom to produce means how the national products to be distributed among the members of the society. Thus, it is the problem of sharing the national product. Distribution of national product depends on the distribution of national income. People who have large incomes will get a larger share from the national product. Since goods and services are scarce no society can satisfy all the wants of all the people. The distributive principle differ from economy to economy and the socialistic principle is from each according to his ability to each according to his needs where as the capitalistic principle is from each according to his resources (money).

The Choice between the Present and Future (How to achieve Economic Growth)

To a certain extent this problem is included in the problem of what to produce. The society has to decide how much goods to be produced for the present generation and how much for the generation to come (future generation). In other words, this implies how much consumer goods are to be produced and how much producer's goods are to be produced. This is a choice between the present and future. As economic resources are scarce, the economy has to decide how much resources are to be used for the present generation and how much resource for the future generation. Use of more resources at present means less of it will be available future and vice versa. An economy has to maintain a balance between present and future with regard to resource use.

Circular Flow of Economic Activity

Economists have developed a model of how an economy works and this is known as the Circular Flow Model. Circular model explains the interaction between households and firms. The flow of economic activity represents employment, production; consumption capital formation etc.. In other words, through economic activity, business firms and households are linked with each other. Thus Circular flow means movement of money and goods in the economy. It attempts to illustrate the flow of money and goods in the economy from households and business enterprises and back to households. The economic activities and money have a circular flow.

Households and business firms interact in product markets and factor markets. Households sell factors or production to business firms in the factor market and purchases goods and services from business firms from the product markets. Hence, households are sellers in the factor markets and buyers in the product markets. The consumption expenditure of households represents the income (money receipts) of business firms. The business firms purchase the services (factors of production) from households. Thus the production cost of business firms represents the money income of households.

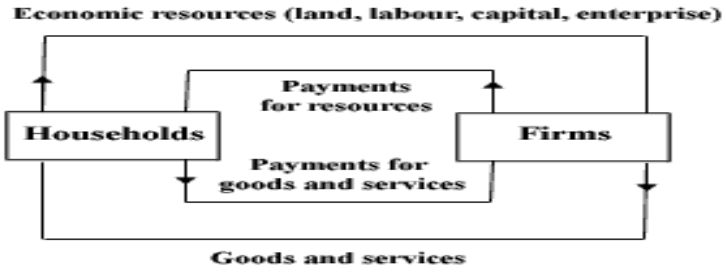
Two Sector Model

The simplest form of the model is called the two sector circular flow model. In this model, we assume that there are only two sectors - the household sector and the business sector.

Households own all economic resources: these are, as you know, land, labour, capital and enterprise. A two sector model consists of households and business firms. Households own all factors of production or resources. These resources are either labour force (human resources) or capital stock (non-human resources) or both. Households are fundamentally consumer units and their ultimate aim is satisfaction of their wants. On the other hand, business firms are production units. They employ the factors of production (resources or inputs) and produces goods (output) for sale. Business firms pay money for the purchase of factor services (scarce resources) from the markets and they receive money from households in return for the sale of goods and services. Thus, flow of goods and services in one direction are always matched by the flow of money in the opposite direction. This is explained in the below diagram.

Figure 2.2

Circular Flow of Economic Activity: Two Sector Model



Households sell their resources to firms and they use these factors of production to produce goods and services. Households are paid for their resources (see the flow "payments for resources"). It is assumed that there are no savings and investments in this model. The household sector constitutes consumers while business sector constitute producers.

In this model, there are two clear flows. The first is the outer flow; economic resources are provided to firms, who use them to produce goods and services. This flow is called the real flow of resources and production. The second, inner flow is the money flow. Firms pay households for their resources, and in turn households use this income to buy goods and services from firms. Households spend their income on goods produced by firms. Firms spend their money on production, buying resources from households. The model indicates that total demand will always equal total supply. That is, total spending equals total production. The economy is always in equilibrium and there is no tendency for a change.

Price determination and functions of prices – concept of margin, Economic models, Methodology, Value judgment, Positive and Normative analysis.

Price Determination and Functions of Prices

Prices

All are interested in prices whether they are consumers or producers. A consumer is interested to know whether the prices of the products he wanted to buy has gone up or down. Similarly a producer is also interested to know whether the prices of his products have gone up or down. Hence, economists are interested in explaining how prices are determined and the reasons behind their variation. The price theory operates in the price mechanism (market mechanism). The price system is a system of economic organization in which economic agents are engaged in economic activity with freedom. In other words, price system is a system of mutual exchanges and coordination of economic activity efficiently.

The price mechanism works in a free market (capitalist) economy through supply and demand of goods in competitive markets. The demand and supply are also determined by prices. Prices determine the production of innumerable goods and services. They organize production and helps in distribution of goods and services. They also ration the scarce goods and services.The role of prices are very important in a free enterprise economy as price or price mechanism decides what goods are to be produced in what quantities they are to be produced, how the decided goods are to be produced and how to distribute the produced goods among the contributors of factors of production (people) and how to utilize the resources fully and how how to achieve rapid economic growth in a country.

Limitations of Price Mechanism

There are certain limitations to the operations of the price mechanism. They are:

1. The government regulates the prices of many commodities to protect the social interest.
2. The price mechanism functions under the assumptions of perfect competition and perfect completion do not exist any where in the world.
3. Price mechanism has resulted in unequal distribution of income and wealth in a country.

Value and Price

Economists some times make a difference between Value and Price. Value of a commodity is the purchasing power of one commodity. Value is a real quantity. In other words, value of a commodity is the

quantities of commodities that are obtained in exchange for a commodity. Value expressed in money is called price. Households (Consumers) are interested in prices. They wanted to know whether the prices of the products they wanted to buy has gone up or down. Similarly, the producers are interested in whether the prices of the products or inputs he uses have gone up or down. In Economics, Economists are interested and tried to explain how the prices are determined and how and when they are high or low. Micro economics itself is known as Price Theory and the vital part of micro economics is explanation of the behavior of individual units and how the prices of goods and services are determined. Fluctuations in commodity prices are an important factor affecting the standard of living of the people.

Functions of Price

The function and role of prices are different in different economic systems like Capitalism and Socialism. The role of prices is very important in a Capitalist economy compared to a Socialist economy. Hence, prices are very vital in a capitalist economy because it solves the basic problems of what how and for whom to produce in a private enterprise economy. However, its role is not so critical in a socialist economy because the resource allocation in a socialist economy is primarily carried out by the central planning board. Price acts as a signal to producers in a capitalist economy. For example, higher the prices of a commodity, higher will be the demand and the profitability is also higher. Hence high price is a green signal to produce the commodity. On the other hand, lower the prices lower the demand and lower the chances of profitability. Hence, a low commodity price is a red signal to producers not to produce or to reduce the production of the commodity. Hence producers in a capitalist economy always look into prices to take production decisions.

1. Allocation of Resources: In a capitalist economy the problem of what to produce is solved (decided) by the prices or price mechanism.

2. Distribution of Resources: Factor prices allocates scarce resources (factors of production) among the production of various goods and services and also helps to produce goods and services according to the preferences of the consumers.

3. It indicates the demand in a capitalist economy. Price acts as a signal to producers to produce the commodities demanded by the Consumers.

Price Determination

One of the important aspects of microeconomic theory is the determination of market prices. We know that the price determination differ from market structure to market structure. In the market when prices increases, buyers will reduce the quantity demanded while the sellers will be ready to increase their sales. On the other hand, when the prices come down, the buyers are ready to buy more whereas the sellers will reduce their supply of goods. Changes in prices would change demand and supply. Thus price changes will change demand and supply. Thus, it is true that the demand and supply of a commodity are both affected by the price of a commodity. On the other hand, both the demand and supply of a commodity influence the price of a commodity. Hence, we can say that forces of demand and supply determine the price of a commodity. In other words, in a perfectly competitive market, the prices are determined by the intersection of forces of demand and forces of supply in the market.

Equilibrium Price

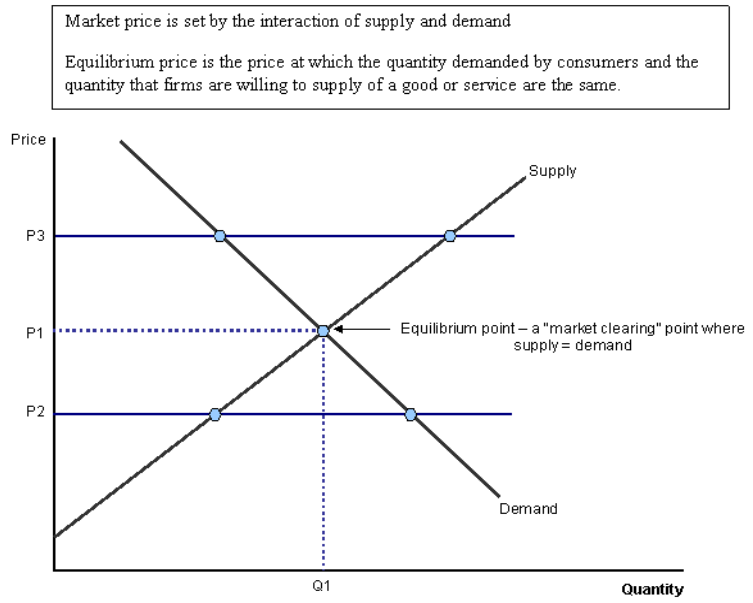
We know that buyers of a commodity demand more of it at a lower price and less of it a higher price. In other words there is an inverse relationship between price and quantity demanded which is known as the law o f demand. Hence the demand curve slopes down ward to the right showing a negative slope. On the other hand sellers of a commodity supply more of it at a higher price and less of it at a lower price. Thus, there is a direct relationship between quantity supplied and price of a commodity which is known as the law of supply. The supply curve rises upward to the right showing a positive slope. There can be a price of the commodity at which quantity demanded and quantity supplied is equal. This price of is called the equilibrium price. Thus equilibrium price of a commodity is the price at which quantity demanded and quantity supplied is equal. It should be noted that at the equilibrium price, there would be no further changes in the demand and supply. Thus equilibrium, price of a commodity is the price at which quantity demanded and quantity supplied is equal.

PRICE DETERMINATION

The market price is determined by demand and supply under perfectly competitive market situation. We know that a demand curve normally slopes downward. In other words, with the fall in price quantity demanded rises and vice versa. On the other hand the supply curve of a commodity usually slopes upward. In other words, an industry will offer to sell more quantity of a good at a higher price than at a lower price. The level of price at which demand and supply curves intersect each other will finally come to stay in the market. In other words, the price which will come to prevail in the market is one at which quantity demanded is equal to quantity supplied. The price at which quantity demanded equals quantity supplied is called equilibrium price because at this price the two forces of demand and supply exactly balance each other. The quantity of the good which is purchased and sold at this equilibrium price is called equilibrium amount. Thus the intersection of demand and supply curves determines price quantity equilibrium and hence the equilibrium price.

There will be discrepancy between quantity demand and quantity supplied if the price is not at equilibrium. In the following figure, at the price $P_{3,}$ the quantity supplied exceeds the quantity demanded. The difference between the quantity supplied and the quantity demanded is called the excess supply. This means that the firms are not able to sell all of their output and hence they will be forced to reduce the price so as to increase their sales. Reduction in price increases the quantity demanded and will reduce the quantity supplied. On the other hand, if the price is $P_{2,}$ the quantity demanded exceeds the quantity supplied. The difference between the quantity demand and the quantity supplied is called the excess demand. When the demand exceeds the supply, consumers cannot buy the quantity they would like to purchase and these results in a tendency to offer a higher price. As the price increases, the quantity demanded decreases and the quantity supplied increases. When the price is P_1, the quantity demanded equals the quantity supplied and there is no tendency to to change the price either by buyers or sellers. Thus, P_1 is the equilibrium price which prevails in the market.

Figure 2.3
Price Determination



Concept of margin

Marginal means addition to total. The marginal unit of anything is the last to be added. The marginal is the smallest increase in ones stock of a commodity. Hence marginal unit of something is the smallest additional amount the consumer considers to be worth buying. The marginal concept can be applied to satisfaction (utility) or psychological influences such as the propensity to consume.

There are different terms associated with marginal and they are marginal cost, marginal product, marginal revenue, marginal rate of substitution, marginal utility, marginal propensity to consume, marginal propensity to save, marginal efficiency of capital, marginal firm, marginal dose, marginal benefit, etc.

Marginal Cost

Marginal cost is the extra (additional) cost of producing one more unit of a product. In other words, marginal cost is the addition to the total cost of producing ***n*** units instead of ***n-1*** units where **n** is any given number.

$$MC_n = TC_n - TC_{n-1}$$

Where MC is the Marginal Cost, TC is the Total Cost. Thus, if it costs Rs.110 to produce 50 units of a commodity and 115 to produce 51 units, marginal cost is Rs.5.

Marginal Product

Marginal product is the extra output resulting from the employment of one more unit of land, labour or capital. In other words, marginal product of a factor is the addition to the total production by the employment of an extra unit of a factor. Suppose when 2 workers are employed to produce rice in an agricultural farm and they produce 100 quintals of rice per year. Now if 3 workers are employed and as a result production of rice increases to 140 quintals, then the third worker has added 40 quintals of rice to total production. Thus, 40 quintals is the marginal product.

Marginal Revenue

Marginal revenue is the extra revenue (income) obtained from the sale of one more unit of output. It is the increase in total revenue by selling one more unit of the commodity. In other words, marginal revenue is the addition made to the total revenue by selling ***n*** units of a product instead of ***n-1*** where ***n*** is any given number.

We may write marginal revenue as: $\mathbf{MR}_{n} = \mathbf{TR}_{n} - \mathbf{TR}_{n-1}$ where MR is Marginal Revenue, TR is Total Revenue.

Marginal rate of substitution

Marginal rate of substitution is the rate at which an individual will exchange successive units of one commodity for another.

Marginal utility

Marginal utility is the additional amount of satisfaction (utility) to be obtained from the consumption of one more unit of a commodity.

Marginal Dose

The terms 'dose' was used by James Mill and Alfred Marshall for amounts of labour and capital applied to land. Thus, the marginal dose would be the smallest possible amount of labour and capital that would be added to land.

Marginal Approach

The standard theoretical model used in economics is called a marginal method or approach. This is because all optimizing decisions are taken 'at the margin' under this method. Margin or marginal change means small changes in an economic entity under consideration, such as utility, cost, factor services, wage rate, quantity demanded or supplied, etc. Such a small or marginal change is in fact a mathematical tool used in calculus. In mathematics, the first derivative of any algebraic function is known as 'the rate of change.' In economics, marginal value or quality serves exactly the same purpose

Economic models

A model is a simplification of reality. It is a device used to represent a real situation. A model may be defined as a mathematical statement of economic theory. It specifies the interrelationships of the parts of a system in verbal or in mathematical terms. In other words, a model is theoretical environment assumed by an economist to work out the implications of a theory in precisely known conditions.

An economic model is a statement of relationships among economic variables. Its purpose is to illustrate causal relations among crucial variables in the real world ignoring irrelevant or minor variables. The purpose of a model is to have a clearer understanding of the relationship among variables. A model may be stated in prose or in mathematics. The variables in a model must be defined and separated in to dependent and independent variables. Moreover, the variables must be statistically measurable for empirical testing. Thus, an economic model is a method of analysis which presents an over- simplified picture of the real world. It is always clear and exact.

An economic situation in a real world may be complex and complicated consisting of a bundle of facts and figures. Hundreds of economic forces are at work in the economy and it becomes a problem for the economist to isolate the important economic variables from unimportant variables. Some variables may be major and some may be minor. The complexity of the situation may be made simple by proper sorting out of the variables on the basis of their importance. As such a simplified version of the situation may be arrived at. This may require some assumptions and on the basis of assumptions, minor variables may be eliminated and the model is prepared by studying the relationship among major variables. For example a model Aeroplane does not exhibit all the features of a real Aerorplane. However, the model Aeroplane exhibits the essentials of what an Aereoplane is.

Let us explain the meaning of the term model with an example. Suppose we have to study the pricing of coconuts. The pricing of coconuts are influenced by so many factors. If we were to take into account all the factors affecting the pricing of coconuts, our analysis would become complex. To simplify our analysis, we will have to build a model of coconut pricing. This model will include only the most essential factors influencing the pricing of coconuts either from demand side or the supply side. We may, for example, exclude weather from our model, though bad weather does affect the supply of coconuts. The point is that only the most essential and relevant relationships are included in the model.

Types of Models

Model may be classified as Aggregate Models (Macro Economic Model), Sectoral Model (Micro Economic Model) and Intersectoral Model. Aggregate models deals with the function of the entire economy such as national income, level of prices, total investment etc. Sectoral models deals with individual sectors. For example, agricultural production. Inter sectoral model is based on relationship between two sectors like agricultural production and industrial production. On the basis of time models are classified as Static model and Dynamic model. The static model explains the relationships among variables at a particular point of time. On the other hand, dynamic model considers relationship among variables over a period of time. A dynamic model may be a closed model or an open model.

Uses of Models

A model is highly useful to an economist.

1. It enables the economist to concentrate on those important factors on which he is interested.
2. It brings clarity of expression and consistency of thought.
3. It presents a concise overall picture of a complex economic activity.
4. It helps to understand the relationship among variables.
5. It helps in policy making

On the other hand, economic models have their limitations also. Since an economic model is based on some assumptions and these assumptions may not be true in real life. Moreover, a model excludes so many factors; the conclusions reached on the basis of models must be accepted with caution. It may happen that some of the omitted variables may be very important and hence the conclusions may be misleading. The real world is much more complicated and a mathematical model cannot incorporate all the details.

Value judgment

A value judgment is an opinion as to what course of action should be taken. Hence, all statements (judgments) which are recommendatory are value judgments. It is concerned about the notions of the people about what is good and what is bad. These ideas about values of the people are based on ethical, political philosophical and religious beliefs of the people and are not based on any scientific logic or law. A value judgment is a judgment of the rightness or wrongness of something based on a personal view. As a generalization, a value judgment can refer to a judgment based upon a particular set of values or on a particular value system. A value judgment also can refer to a tentative judgment based on the information at hand. Most commonly the term value judgment refers to an individual's opinion. Value judgments judge things to be good or bad in some respect. Moral or ethical values are only one type of value, and moral evaluation is only one type of value judgment. In other words, value judgments are assessments that reveal more about the values of the person making the assessment than about the reality of what is assessed.

Value judgments describe facts in an emotional way and tend to influence people by altering their beliefs and attitudes. Statements like ‘inequalities in incomes need to be reduced’ are value judgments. According to the values of the rich the existing unequal distribution of income and wealth are just whereas according to the values of the poor unequal distribution of income and wealth are very bad and unethical. Hence, ethical judgments or statements which perform influential and persuasive functions are value judgments.

Positive and Normative Analysis

Whether economics is a positive or a prescriptive (normative) science is a frequently debated issue. Economists make a distinction between positive and normative analysis. There is a controversy regarding the nature of economics. Economists disagree with regard to whether economics is a positive science or a normative science. According to some economists, economics is a positive science whereas according to some others, economics is a normative science. We may define what a positive science is and what a normative science is before going to explain positive economics and normative economics.

Positive Science

Positive means 'value free'. Positive statements are often referred as descriptive statements. The most common usage refers to analysis or theories which only attempt to describe how things 'are'. A positive statement is a statement about what is and that contains no indication of approval or disapproval. Hence, a positive science explains and describes the existing relationships or explains what is. As such a positive science may be defined as a body of systematised knowledge concerning what it is and with the actual. The objective of a positive science is the establishment of uniformities. Thus, in positive economics, we derive propositions, theories and laws following certain rules of logic. These theories, laws and propositions explain the cause and effect relationship. Hence, positive science deals with things as they are, and it simply explains causes and effects without passing any moral judgment on the desirability of having certain ends. When a subject confines itself to statements about causes and their effects and to statement of functional relation, the subject is said to be positive. According to Lional Robbins and N. Senior economics is a positive science. According to Senior, an economist is not authorized to add even a word of advice.

Normative Science

A normative science is based on norms. Hence, a normative statement expresses a judgment about whether a situation is desirable or undesirable. Therefore, a normative science makes prescriptions and recommendations about what ought to be. In other words, a normative (regulative) science is a body of systematized knowledge relating to criteria of what ought to be and concerned with the ideal. When a subject embraces norms and standards, mixing them with cause effect analysis, the subject is said to be a normative. Normative statements mainly depend on value judgments and consequently there is a lot of scope for disagreement as the ideas of good and bad become subjective. According to A.C.Pigou, Paul Streeton and Alfred Marshall economics is also concerned about normative statements and value judgments. According to Streeton, Economics cannot and should not refrain from making value judgments if their studies are to be more than a purely formal technique of reasoning, an algebra of choice.

Positive Economics

It is quite useful to know the difference between positive economics and normative economics. Positive economics is devoid of any ethical position or value judgments. It is primarily empirical or statistical in nature and is independent of normative economics. Hence, positive economics is the study of what is, or how the economic problems facing a society are actually solved. In other words, positive economics is concerned with explaining what it is. Hence, it describes theories and laws to explain observed economic phenomena. Positive economics states that monopolist will fix a price which will equate marginal cost with marginal revenue. Positive economics explains how national income is distributed among different individuals. The study of the actual effect of minimum wage regulations on the economy is a study in positive economics. It involves the examination of which occupations will be affected by the regulations, the extent of substitution of capital for labour and what happens to the displaced workers.

Normative Economics

Normative economics is based on positive economics and the value judgments of the society. It provides guidelines for policies to increase and possibly maximize the social welfare.. Thus, normative economics is the study of what ought to be. It studies how the economic problems facing the society should be solved. Normative economics is concerned with what should be or what ought to be the things. It is also called prescriptive economics. What price for a product should be fixed, what wage rate should be paid, how income should be distributed etc. fall within the purview of normative economics. Normative economics involve value judgments. Normative economics is concerned about welfare propositions.

Methodology

Economic analysis attempts to find relationship among variables and reach conclusions on the basis logical reasoning from the data collected. Hence, economic analysis helps us to understand the operations of an economy. Every science adopts certain methods for analysis and Economics being a science also adopts scientific methods for analysis. Economics adopts two important methods for investigations and formulations of laws and principles. These two methods are ‘deductive’ and

'inductive' methods. There was a controversy among economists regarding the method to be used for economic analysis as some advocated deductive method while others suggested inductive method. The controversy was ultimately settled by Marshall who stated that both induction and deduction are needed for scientific thought as right foot and left foot both are needed for walking. We may briefly explain these methods ion a little detail.

Deductive Method

Deductive method is also called analytical, abstract of a priori method. Deductive method was made use of by classical economists. Deduction descends from the general to the particular or we reach inferences from universal to the individual. Here we start from a few indisputable facts about human nature and draw inferences from them about individual cases. For instance, it is assumed that businessmen aim at maximum profit. It follows from this that businessmen buy the materials in the cheapest market and sell it in the dearest market. In deduction, we start from a few indisputable facts about human nature and draw inferences from them about individual cases. For example, the law that the utility derived by an individual from a commodity goes on diminishing with every successive increase is a self evident truth from which we may reach many conclusions. From this, we conclude that larger the stock of money , the lower shall be the utility of money utility or rich persons have lesser marginal utility of money than the poor people etc.

Steps in Deduction

The major stages involved in the process of deduction are (i) Perception of the problem (ii), Making Assumptions, (iii) Formulation of Hypothesis and ()iv) Testing of Hypothesis.

Merits of Deductive Method

The deductive method has the following advantages

1. Deductive method is very simple. From a few basic facts of human nature, we can derive a large number of inferences in a short time.
2. This method avoids the necessity of experimentation. The scope of experimentation is very much limited in Economics and hence the next best alternative is the method of deductive reasoning.
3. This method is very accurate and exact. As this method uses logic and mathematics, the chances of mistakes or inaccuracies are very much limited.

Demerits of Deductive Method

There are certain dangers associated with the deductive method and they are:

1. Deductive method is based on assumptions and hence the generalization reached is true only if the assumptions up on which they are based are correct.
2. There are too much of abstraction in deduction and this may result in the creation of theories lacking reality.
3. This method is dangerous when economic policies are arrived on the basis of generalizations based on incorrect assumptions.

Inductive Method

The inductive method works from the particular to the general and is based on facts. The inductive method was adopted by Historical school. The inductive method is also known as historical or empirical method. Induction is the process reasoning from a part to the whole or from individual to the universal. It is an ascending process in which facts are collected; arranged and then general conclusions are drawn. There are two forms of induction and they are experimentation and Statistical approach. The experimentation is concerned with testing the validity of laws (generalisations) arrived at as a result of deductive reasoning by resorting to the study of actual facts. The statistical method is concerned with the framing of laws (generalizations) on the basis of a large number of facts collected from the various sectors of the economy. However, the inductive method is generally associated with the statistical form of induction.

Merits of Inductive Method

The inductive method has the following advantages:

1. This method is highly practical and realistic. The generalizations (laws) reached on the basis of carefully collected data leads to precise conclusions.
2. This method makes use of statistical method which is most suited in analyzing economic problems.
3. The inductive method is dynamic.

Demerits of Inductive Method

The important limitations of inductive method are the following"

1. There is the possibility of hurried conclusions. More over, as this method makes use of statistical method for analysis, there is the possibility of misuse or misinterpretation of this method.
2. The inductive method is time consuming and costly.
3. Collection of facts and data are very difficult.

Suggested readings:

1. Dominick Salvatore, Micro Economics-Theory and Applications, New Delhi, OUP.

2. Koutsoyannis, A, Modern Micro Economics, Macmillan

3. Abhijit Kundu. , Methodology and Perspectives of Social Science, New Delhi, Pearson Education.

Module III

Basic Demand and Supply Analysis

Market Analysis

A market is a network of communications between individuals and firms for the purpose of buying and selling of goods and services. A market can, but need not, be a specific place or location where buyers and sellers actually come face to face for the purpose of transacting their business. That is, the idea of a particular locality or geographical place is not necessary to the concept of market. What is required for the market to exist is the contact between sellers and buyers so that transaction at an agreed price can take place between them. There is a market for each good or services or resource bought or sold in the economy. Some of these markets are local; some are regional while others are national or international in character.

Thus, market is a collection of buyers and sellers that determine the prices that are established and quantities that are transacted. Markets provide the framework for the analysis of the forces of *demand and supply* that, together, determines prices of goods and services. **The Nature of Demand**

In economics, demand refers to the various quantities of a good or service that people will be and able to purchase at various prices during a period of time. It is important to note that a mere desire for a good or service does not constitute demand. Demand implies both the desire to purchase and ability to pay for the good. Unless demand is backed by purchasing power, it does not constitute demand. Further, demand does not refer to the specific quantity that will be purchased at some particular price, but refer to a series of quantities and their associated prices.

Demand Function

Demand for a commodity is determined by several factors. An individual's demand for a commodity depends on the own price of the commodity, his income, prices of related commodities, his tastes and preferences, advertisement expenditure made by the produces of the commodity, expectations etc. Thus, individual's demand for a commodity can be expressed in the following general functional form,

$$Q_x^d = f(P_x, I, P_r, T, A, E) \text{ where,}$$

Q_x^d = Quantity demanded of commodity "x"; P_x = Price of commodity x; I = Income of the individual consumer; P_r = Price of related commodities; T = Tastes and preferences of individual consumer; A = Advertisement expenditure;E = Expectations

The demand function is just a short hand way of saying that quantity demanded , which is recorded in the left hand side depends on the variables that are recorded on the right hand side. For many purposes in economics, it is useful to focus on the relationship between quantity demanded of a good and its own price, while keeping other determining factors constant. Thus, we can write the demand function as:

$$Q_x^d = f(P_x)$$

This implies that the quantity demanded of the commodity x is a function of its own price, other determinants remaining constant.

Law of Demand

Law of demand expresses the functional relationship between price and quantity demanded. According to the law of demand, other things being equal, if the price of the commodity falls the quantity demanded of it will rise and if the price of the commodity rises, its quantity demanded will decline. Thus, according to law of demand, there is an inverse relationship between price and quantity demanded, other things remaining the same. The other things which are assumed to be constant are tastes and preferences of the consumer, the income of the consumer, prices of related commodities etc. Thus, the law of demand assumes that all things other than price remain constant.

The law of demand can be illustrated through a demand schedule and through demand curve. Demand schedule shows various quantities of good or service that people will buy at various possible prices during some specified period, while holding constant all other relevant economic variables on which demand depends. A demand schedule is presented below.

Price	Quantity Demanded
10	20
8	40
6	60
4	80
2	100

We can convert the demand schedule into demand curve by graphically plotting the various price-quantity combinations, as shown below.

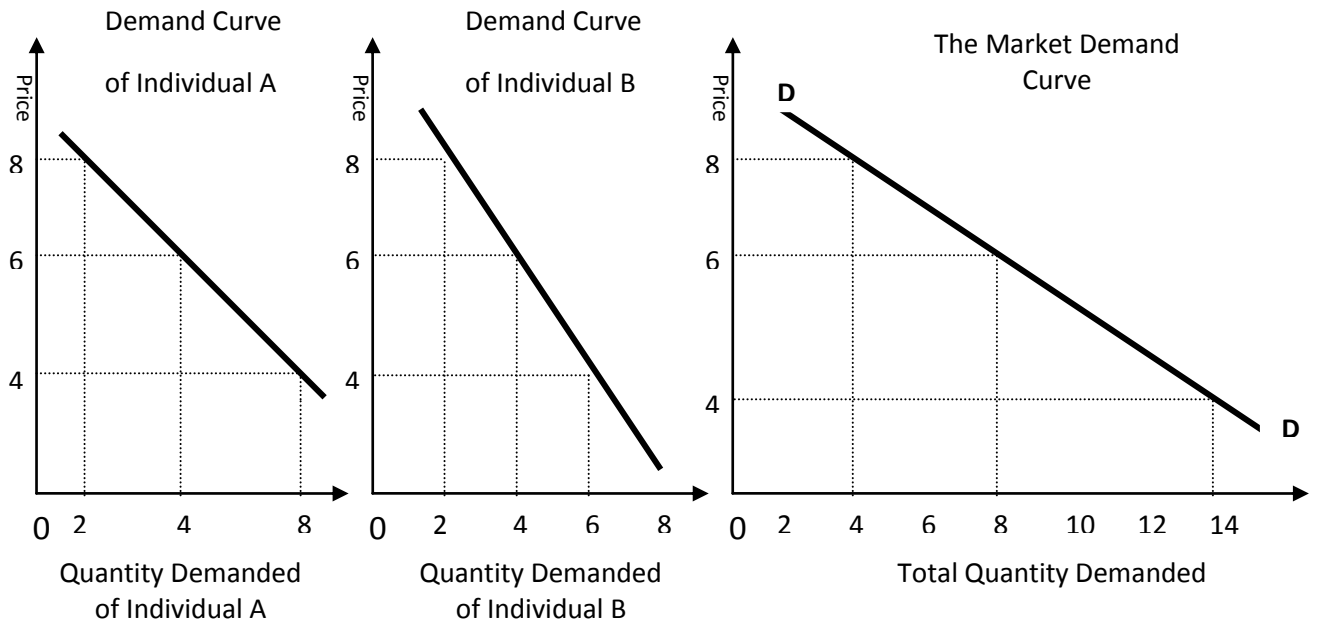
It is common practice in economics to measure price on vertical axis and quantity demanded per unit of time on the horizontal axis. Thus, the demand curve is a graph showing the various quantities of a good or service that the people will be willing and able to buy at various possible prices. Demand curve slopes downwards from left to the right. The downward sloping demand curve is in accordance with the law of demand, which describes inverse price-quantity demanded relationship. The various points on the demand curve represents alternative price -quantity combinations.

The Market Demand

As mentioned above, the quantity of a product demanded by one individual depends on the product's price, other things being equal. To explain the market behaviour, we need to know the total demand of all individuals. The market demand for a commodity gives the alternative amounts of the commodity demanded at various prices by all individuals in the market during a period of time. To obtain the market demand, we sum the quantities demanded by each individual at a particular price to obtain the total quantity demanded at that price. We repeat the process for each price to obtain market demand schedule at all possible prices. The market demand for a commodity depends on the all the factors that determine the individual's demand. In addition, it also depends on the number of buyers of the commodity in the market. Geometrically, the market demand curve for a commodity is obtained by the horizontal summation of the entire individual's demand curve for the commodity. For sake of simplicity, let us assume that there are only two individual consumers in the market, individual A and individual B. The individual demand schedules for these two consumers along with the market demand schedule is given below.

Individual A		Individual B		Market Demand	
Price	Quantity Demanded	Price	Quantity Demanded	Price	Quantity Demanded
8	2	8	2	8	4
6	4	6	4	6	8
4	8	4	6	4	14

At price Rs 8, individual A will buy 2 units and individual B will also buy 2 units. The total quantity demanded at Rs 8 is therefore 4 units. This is shown in the market demand schedule. Similarly, the total quantity demanded in the market at Rs 6 is 8 units and 14 units are demanded at Rs 4 in the market. It can be seen that the market demand schedule is the sum of the demands of the individual consumers in the market. A graph of this market demand schedule is called the market demand curve. The market demand curve is shown below



The above figure illustrates the proposition that the market demand curve is the horizontal sum of the demand curves of all the individuals who buy in the market. The market demand curve will also slope downwards from left to the right because the individual demand curves whose lateral summation gives the market demand curve normally slope downward from left to the right.

Reasons for law of Demand

Let us analyse the reasons for the inverse relationship between price and quantity demanded. This is due to both "income effect" and "substitution effect".

When the price of the commodity falls, the consumer can buy more quantity of the commodity with his given income. If he chooses to buy the same amount of the commodity as before, some money will be left with him. That is, consumer's real income or purchasing power increases. This increase in real income induces the consumer to buy more of the commodity. This is called the income effect of the change in price of the commodity. This is the reason why a consumer buys more of a commodity whose price falls. Similarly, an increase in the price of the commodity results in the reduction of real income of the consumer. Hence, the consumer buys less of a commodity whose price rises.

Again, when price of the commodity falls, it becomes relatively cheaper than other commodities. This induces the consumer to substitute the commodity whose price has fallen for other commodities which have now become relatively dearer. This change in quantity demanded resulting from substituting one commodity for another is referred to as substitution effect of the price change. As a result of this substitution effect, the quantity demanded of the commodity whose price has fallen rises. For normal commodities, the income and substitution effect of a price decline are positive and reinforce each other leading to a greater quantity demanded of the commodity. Apart from the income effect and substitution effect, there is an additional reason why the market demand curve for a commodity slopes downwards.

When the price of the commodity is relatively high, only few consumers can afford to buy it. When the price of the commodity falls, a greater number of consumers will be able to afford to buy it. In other words, the size of the market expands. Thus, the quantity demanded increases. This is called the "market size effect".

Exceptions to the Law of Demand

Law of demand is generally believed to be valid in most situations. However, some exceptions have been pointed out. According to Thorestein Veblen, some consumers measure the utility of a commodity entirely by its price. That is, for them, the greater the price of the commodity, the greater it's utility. These consumers demand more of such commodities the more expensive these commodities are in order to impress people. E.g. Diamonds. This form of conspicuous consumption is called "Veblen effect". When the price of such commodities goes up, their prestige value also goes up. Consequently, quantity demanded also will rise and law of demand breaks down.

Another exception to the law of demand is the case of some inferior commodities and was pointed out by 19th century English economist Sir Robert Giffen. He introduced the case of some inferior goods in which there is a direct price-quantity demanded relationship. If the price of an inferior good falls, consumer's real income increases. So, instead of buying more inferior goods, consumers substitute other superior goods. In such case, quantity demanded of inferior goods falls as price falls. After the name of Robert Giffen, such goods are called "Giffen Goods". In the case of Giffen goods, positive substitution effect is smaller than negative income effect when the price of such goods falls. With the rise in the price of such goods, its quantity demanded increases and with the fall in the price, its quantity demanded decreases. Thus, the demand curve will slope upwards to the right and not downward in the case of Giffen goods. It should be noted that Giffen good is an inferior good but all inferior goods are not Giffen goods. Though occurs rarely in the real world, Giffen goods represent an exception to the law of demand.

Determinants of Demand

Income, prices of related goods, taste and preferences of the consumer, expectations, number of buyers in the market, distribution of income etc are likely to affect the demand for the product. These factors are called "non-price determinants" and are assumed to be constant while deriving the demand schedule and demand curve. But any change in these non-price determinants will change the demand schedule and demand curve. Let us analyse how these factors can affect the demand for the product.

(1) Income: The demand depends up on income of the people. The greater the income of the people, the greater will be their demand for goods and services. If their income increases, people will tend to buy more goods and services than they did before the increase in income. This is the case of most goods and services. Hence economists refer to goods whose demand varies directly with income as "normal goods". Although most commodities are normal goods, there are cases when consumers may not buy some goods more as their income increases. Instead they buy less. Such goods are called "inferior goods" because as people's income increases they actually reduce the purchase of such goods.

(2) Prices of related goods:Goods and services may be related to each other in two ways; they may be substitutes or they may be complements. One good is said to be substitute for a second good if it can be used in the place of second good. Example: tea and coffee, beef and chicken. Two goods are said to be complementary if they are used together. Complementary goods are demanded jointly. Example: scooter and petrol, computer and computer software. In general, if the price of a substitute commodity increases, consumers tend to increase their purchases of the substitute in question. Goods are substitutes when an increase in the price of one leads to an increase in the quantity demanded of the other. For instance, if the price of coffee increases, people will substitute tea for coffee and as a result demand for tea increases. On the other hand, if the price of complement falls, people will tend to increase their purchases of the commodity in question. Two goods are complements if a fall in the price of one leads to increase in the quantity demanded of the other. For instance, if the price of scooter falls, the demand for them will increase which in turn will increase the demand for petrol.

(3) Taste and Preferences:The quantity of a commodity that people will buy will be affected by the taste and preferences. Companies spend millions of Rupees in advertisement in an attempt to influence consumer's tastes in favour of their products. Consumer's taste and preferences often change and as a result, there is a change in the demand for products. A good for which consumer's tastes are greater, its demand would be larger. On the contrary, any good goes out of fashion or people's taste and preferences no longer remain favourable to them, the demand for them decreases.

(4) Expectations: The expectations of the consumers regarding the price in the future will affect present purchases of goods and services. If consumers expect the price of the product to increase in the future, they are likely to increase their present purchases to stock up on the good and thus postpone paying the ensuing higher price for as long as possible. Conversely, if the price is expected to fall in future, consumers will attempt to delay their present purchases in order to take advantage of the lower future prices. The expectations of the consumer about the future change in income will also affect the purchases of goods and services. If people expect substantial increase in their income sometime in the near future, they are likely to buy more goods and services even before the increase in income materialises. If the people expect decrease in their income, they are likely to buy fewer goods and services.

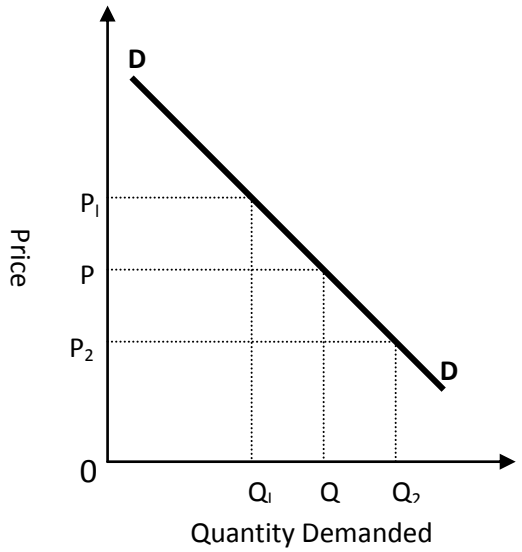
(5) Number of buyers in the market: The quantity of the commodity that people will buy depends on the number buyers in the market for that particular commodity. The greater the number of buyers of a good, the greater the market demand for it. If population increases we can expect the demand for most goods and services to increase as a consequence.

(6) Distribution of income:Distribution of income in the society also affects demand for goods. If the distribution of income is more equal, then the propensity to consume of the society as a whole will be higher which results in greater demand for goods. On the other hand, if the distribution of income is more unequal, then the propensity to consume of the society will be relatively less because propensity to consume of rich people is less than that of poor people.

Extension and Contraction in Demand

When as a result of change in price, the quantity demanded rises or falls, extension and/or contraction in demand is said to have taken place (change in quantity demanded). When the quantity demanded of a good rises due to a fall in price, it is called extension of demand. When the quantity demanded falls due to rise in price, it is called contraction in demand.

It should be remembered that extension and contraction in the demand takes place as a result of changes in the price alone when other non-price determinants of demand such as income, prices of related goods etc remain constant. The extension and contraction in demand is shown below.



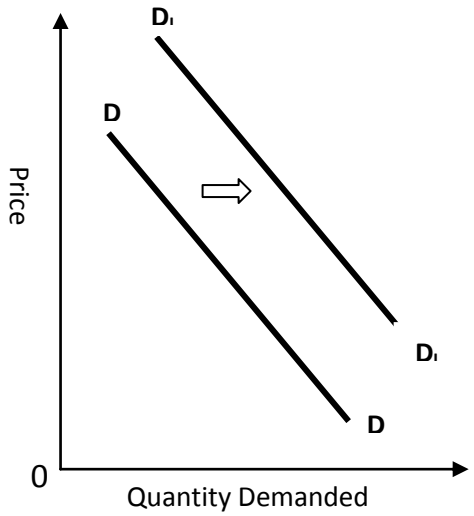
It can be seen that when the price is OP, the quantity demanded is OQ. If the price falls to OP_2, the quantity demanded rises to OQ_2. Thus, there is extension in demand by the amount QQ_2. On the other hand, if the price of the commodity rises from OP to OP_1, there is a contraction in demand equal to the amount QQ_1. Thus, as result of changes in the price of the good, the consumers move along the same demand curve. Thus, the movement along the given demand curve is referred to as a change in quantity demanded (extension or contraction in demand). A movement down the demand curve is an increase in quantity demanded (extension in demand). A movement up the demand curve is called a decrease in quantity demanded (contraction in demand).

Increase and Decrease in Demand (Shifts in Demand)

If the non-price determinants of demand such as income of the consumer, prices of related commodities etc change, the whole demand curve will change. The demand curve will shift to a new position in response to changes in any of the factors or variables that were held constant when original demand curve was drawn. When as a result of changes in these factors, the demand curve shifts upwards to the right, an increase in demand is said to have occurred. Increase in demand means the consumer buys more of the goods at various prices than before. An increase in demand occurs due to the following reasons:

(a) A rise in consumer's income; (b) A rise in price of the product's substitutes; (c) A fall in the price of complementary goods; (d) Change in the taste in favour of the product; (e) Increase in the number of buyers in the market; (f) Changes in distribution of income favouring those having high propensity to consume.

An increase in demand is shown below:

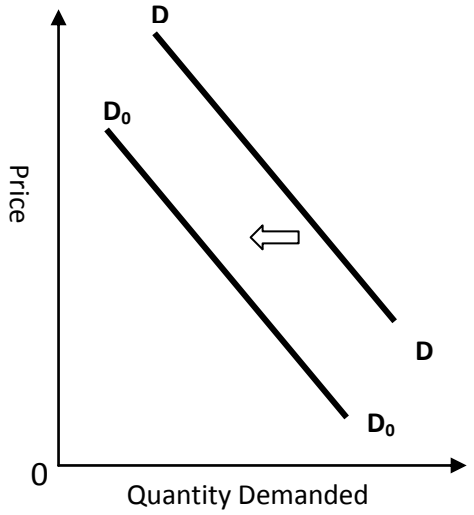


In the figure, DD is the original demand curve and D_1D_1 is the new demand curve. An increase in demand is shown by a shift in the demand curve to the right. The location of the demand curve has now changed. Now at any given price greater quantity is purchased.

On the other hand, a decrease in demand means entire demand curve shifts to a lower position to the left. Decrease in demand does not occur due to the rise in price but due to changes in other determinants of demand. A decrease in demand occurs due to the following reasons:

(a) A fall in consumer's income; (b) A fall in price of the product's substitutes; (c) An increase in the price of complementary goods; (d) Change in the taste away from the product; (e) number of buyers in the market declines; (f) a redistribution of income favouring those having high propensity to save and away from those who favours the commodity

A decrease in demand is shown below



In the figure, DD is the original demand curve and D_oD_o is the new demand curve. A decrease in demand is shown by the leftward shift in the demand curve. A decrease in demand would mean that at any given price smaller quantity would be purchased. Thus, a change in the price of the commodity will not cause change in the demand, it will cause a change in the quantity demanded. Only a change in the non-price determinants can cause a change in demand, that is, cause the entire demand curve to shift. These non-price determinants are often referred to as "demand shifters" or "shift factors".

Nature of Supply

Supply refers to the various quantities of a good or service that sellers will be able to offer for sale at various prices during a period of time. It shows how price of a good or service is related to the quantity which the sellers are willing and able to make available in the market. As in the case of demand, supply refers not to a specific quantity that will be sold at some particular price, but to a series of quantities and a range of associated prices. Supply is a desired flow. That is, it shows how much firms are willing to sell per period of time, not how much they actually sell.

Supply Function

Like demand, supply also depends on many things. In general, quantity supplied of a product is expected to depend on own price, prices of related products, prices of inputs, state of technology, expectations, number of producers (sellers) in the market etc. This list can be summarised in a supply function

$$Q_X^S = f(P_x, P_r, P_i, T, E, N)$$

Where.,

Q_X^S = Quantity supplied of commodity x

P_x = Price of the commodity x

P_r = Prices of related products

P_i = Prices of inputs

T = State of technology

E = Expectations

N = Number of producers in the market

For a simple theory of price, we need to know how quantity supplied varies with the product's own price, all other things being held constant. Thus we can write the supply function as

$$Q_X^S = f(P_x)$$

That is, quantity supplied of commodity x is a function of its own price, other determinants are assumed to remain constant.

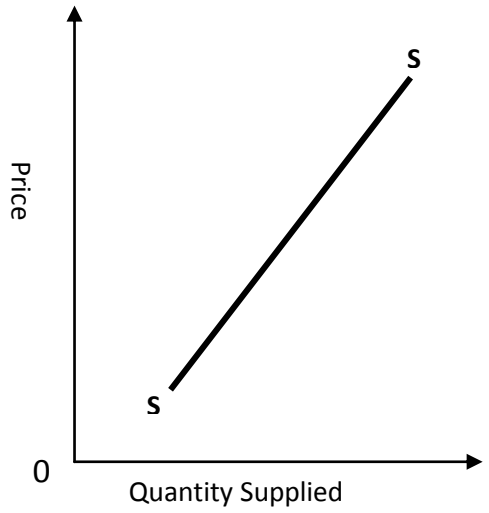
Law of Supply

The functional relationship between price and quantity supplied is called the law of supply. According to the law of supply, as the price of the commodity falls, the quantity supplied decreases or alternatively, as the price of the commodity rises the quantity supplied increases, other things being equal. Therefore, there is a direct relationship between of the commodity and quantity supplied.

The law of supply can be illustrated through a supply schedule and supply curve. Supply schedule is a table that shows various quantities of a good or service that sellers are willing and able to offer for sale at various possible prices during some specified period. A supply schedule is presented below

Price	Quantity Supplied
5	40
10	60
15	80
20	100
25	120

Supply schedule shows that as price rises, a greater quantity is offered for sale. By plotting the information contained in the supply schedule on a graph we can derive the supply curve as shown below.



The supply curve is a graph showing various quantities of a good or service that sellers are willing and able to offer for sale at various possible prices. The supply curve slopes upwards because of the direct relationship between price and quantity supplied. Note that the entire supply curve represents supply while a point on the supply curve represents quantity supplied at some specific price.

Why there is a direct relationship between price and quantity supplied? The main reason is that higher prices serve as an incentive for sellers to offer greater quantity for sale. The sellers or producers can be induced to produce and offer a greater quantity for sale by higher prices. It is assumed that sellers or producers aim to maximise profit from the production and sale of the commodity. The higher the prices of the commodity, other things being equal, the greater the potential gain producers can expect from producing and supplying it in the market. Moreover, increases in price may invite new suppliers in the market.

Determinants of Supply

The quantity of a good or service that sellers are willing and able to offer for sale depends on the price of good or service. The non-price factors such as prices of related products, input prices, technology, expectations and number of producers in the market are likely to affect supply. Let us analyse each of them.

(1) Prices of related products

Goods can be substitutes or complements in production. Goods are substitutes in production if they are produced as alternatives to each other. Example: rice and vegetables (as farmer can produce one or the other on the same piece of land). Goods are complements in production if they are produced together; the production of one good implies the production of the other. Complements in production are also called 'joint products'. Example sugar and molasses, beef and hides. In general, if the price of a substitutable product increases, sellers will tend to reduce the supply of the substitute in question. At the same time, if the price of a complement in production falls, the supply of the good in question will also falls.

(2) Prices of inputs

An increase in the production cost will results in a reduction in the supply of the product. Payments for factor inputs represent a significant part of production cost. The higher the prices of these inputs, the greater the cost of production will be and the supply will be less. On the other hand, a reduction in input prices will cause an increase in supply.

(3) Technology

Overtime, knowledge and production technologies change and it will affect the supply of the product. A technological change that decreases cost will increase profits earned at any given price. Since increased profitability leads to increased production, it will cause an increase in supply.

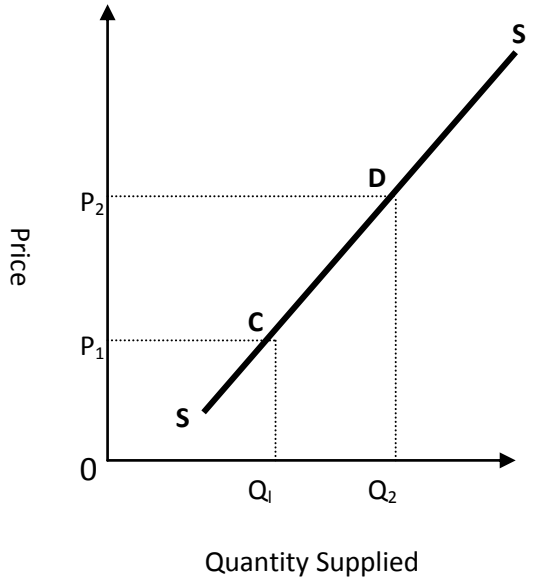
(4) Expectations

If producers expect prices to rise in the future, now they might begin to expand their productive capacity and thus increase their present output levels. However, it is also possible that expectations of higher future prices may lead producers into building up stocks now so that they will have larger quantity to sell at future higher prices. Such action will reduce current supply. Therefore, generalisation should not be made about the effects of expected price changes on supply.

(5) Number of Producers : Obviously, the number of sellers in the market will have some effect on the total market supply. This is so because the market supply of a good or service is the sum of the quantities offered for sale by all individual sellers in the market. We can expect market supply to increase as number of sellers' increases and to decrease as number of sellers' decreases.

Changes in quantity Supplied

A change in quantity supplied refers to the change in the quantity that would be offered for sale as a result of a change in price, other factors being held constant. That is, change in the price of the commodity will not cause a change in supply, it will cause a change in quantity supplied. Since there is a direct relationship between price and quantity supplied; at a higher price more will be supplied and vice versa. Change in quantity supplied involves a movement from one point on the supply curve to another point on the same curve, as shown below.



At price P_1, Q_1 quantity supplied. If price rises to P_2, quantity supplied rises Q_2. This change in quantity supplied is represented by a movement along the same supply curve from point C to point D in the figure.

Changes in Supply

A change in supply refers to the change in the supply curve due to changes in factors other than product's own price like prices of related goods, input prices, technology, expectations, number of producers etc. That is, only a change in non-price determinant can cause a change in the supply of the commodity. Changes in these factors cause the entire supply curve to shift to a new position. Therefore, these non-price determinants are called "supply shifters".

An increase in supply means an increase in quantity supplied at each price of the commodity. Increase in supply causes a rightward shift in the supply curve. That is, producers supply more of the commodity at each price. Major factors that causes increase in supply is listed below.

(a) Decrease in the price of production substitutes

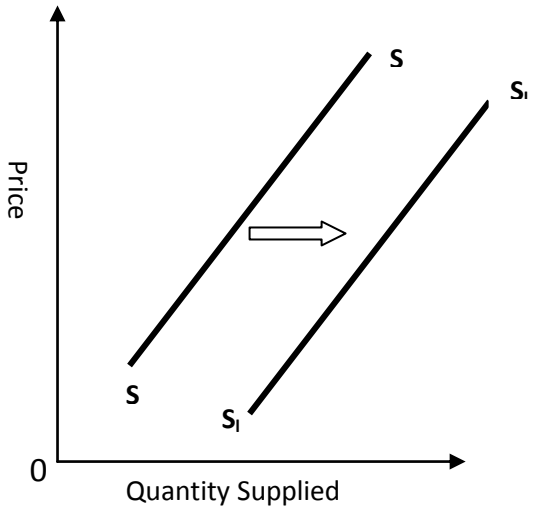
(b) Increase in the price of production complements

(c) Fall in the price of inputs

(d) Technological change that decreases cost

(e) Increase in the number of producers

An increase in supply is shown below



SS is the original supply curve and S_1S_1 is the new supply curve. The increase in supply is shown by shifting the entire supply curve to the right. The location of the curve has now changed. At any given price, greater quantity is supplied.

Decrease in supply means a reduction in quantity supplied at each price of the commodity. Decrease in supply causes a leftward shift in supply curve. That is producer's supply less of the commodity at each price. Major factors that causes decrease in supply is listed below.

(a) Increase in the price of production substitutes

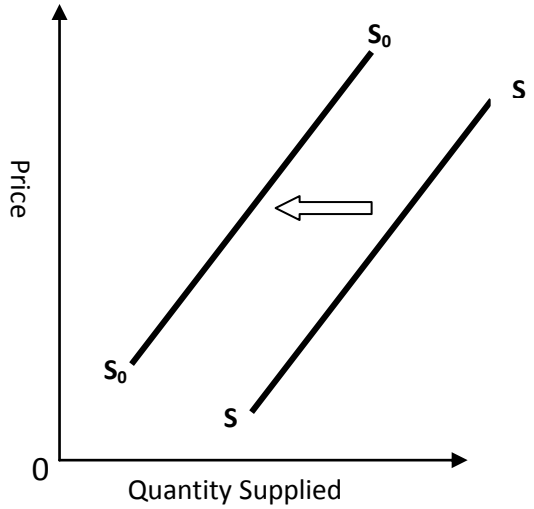
(b) decrease in the price of production complements

(c) rise in the price of inputs

(d) decrease in the number of producers

(e) Imposition of tax on the sales and/or production of commodity by the government.

A decrease in the supply is shown below



SS is the original supply curve and S_oS_o is the new supply curve. Decrease in supply is represented by a leftward shift in the supply curve. A decrease in supply curve would mean that, at any given price, a smaller quantity is supplied.

Market Equilibrium

The market equilibrium occurs when the prevailing price equates quantity demanded to quantity supplied. It refers to the price-quantity pair at which this takes place. Consumers bring demand to the market for buying goods to satisfy their wants. Producers or sellers bring supply of their goods to the market to sell them and earn profit. The market demand and supply determine prices of goods and services exchanged between buyers and sellers. Thus, market equilibrium is reached when market demand for and market supply of a good are equal and as a result, equilibrium prices and equilibrium quantities are determined. At such equilibrium, buyers find that they are able to buy exactly the same amount that they are demanding at the prevailing price and sellers are able to sell exactly the amount they are willing to supply at the prevailing price. In other words, there is no incentive for anyone in the market to change their behaviour. Thus equilibrium is the condition, which once achieved tends persist in time.

By bringing together the market demand and supply schedules we can see how market forces determine equilibrium price and quantity of the good. The following table presents a hypothetical demand and supply schedules of commodity X.

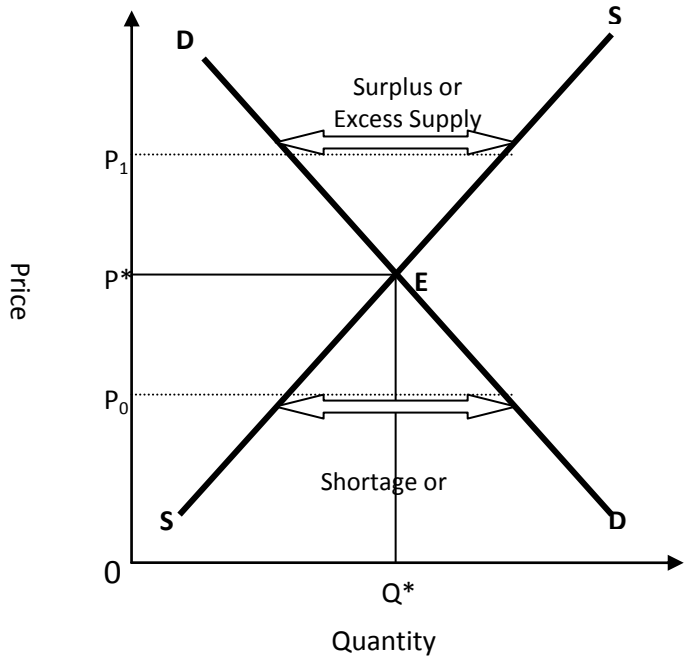
Price of commodity X (P_X in Rupees)	Quantity Supplied (Q_X^S)	Quantity Demanded (Q_X^D)	Surplus (+) Shortage(-)	Pressure on Price
5	140	20	120	Downward
4	100	40	60	Downward
3	60	60	0	Equilibrium
2	40	80	-40	Upward
1	20	100	-80	Upward

When the price of commodity X is Rs 1, buyers are willing and able to purchase 100 units but sellers are willing and able to offer only 20 units for sale. Therefore, there is a shortage of 80 units. At price of Rs 5, buyers are willing and able to purchase only 20 units while sellers are willing to offer 140 units. Therefore, there will be a surplus of 120 units in the market. Let us now consider a price of Rs 3. At this price, buyers are willing to purchase 60 units and sellers are willing to offer 60 units for sale. That is, at this price, there is neither a surplus nor a shortage. Quantity supplied of commodity is equal to the quantity supplied. Thus P_X = Rs 3 is the equilibrium price and $Q_X^S = Q_X^D$ =6o is the equilibrium quantity.

At any other price other than the equilibrium price of Rs = 3, market forces are set in motion to raise or lower the price. At the prices above the equilibrium price, the quantity supplied exceeds the quantity demanded. For example, at P_X = Rs 4, sellers are willing to put 100 units of commodity X on the market but buyers are willing to take only 40 units. There will be surplus or excess quantity supplied of the commodity. Then the sellers will attempt to dispose this surplus by lowering the price. As price falls, a greater quantity will be demanded. At lower prices sellers supply smaller quantities and buyers demand larger quantities until the equilibrium price of Rs 3 is reached, at which the quantity supplied of 60 units of commodity X equals the quantity demanded and market clears.

On the other hand, at prices below the equilibrium price, the quantity supplied fall short of quantity demanded. For example, at P_X = Rs 2, buyers are willing to purchase 80 units but sellers will be able to offer only 40 units. There is a shortage or excess quantity demanded. Unhappy with the shortage, and wanting more commodity X, buyers will bid up the price to induce sellers to supply them the desired amount. Then the sellers offer a greater quantity at higher prices. The price will again settle at P_X = Rs 3, because at this price, the quantity demanded equals quantity supplied. Note that, price of Rs 3 is the only price that will prevail in the market. There will be no tendency of this price to change. Such a price is referred to as equilibrium price and quantity traded or exchanged at this price is called equilibrium quantity. The market for the product is said to be in equilibrium when the quantity demanded equals the quantity supplied at a specific price.

The determination of equilibrium price and quantity can also be shown graphically by bringing together the market demand and market supply curve on the same graph, as shown below.



The intersection of market demand curve DD and market supply curve SS at point E defines the equilibrium price P* and the equilibrium quantity Q*. At the equilibrium price, quantity demanded is equal to the quantity supplied. Because there is no excess demand or excess supply there is no pressure for the price to change further.

As said above, the equilibrium between demand and supply is not reached at once. There is the process of changes and adjustments which ultimately results in equilibrium price and quantity. Suppose that price is above the equilibrium level, say at P_1. At such higher price, there is excess supply or surplus of the commodity. Then the sellers would begin to lower prices in order to sell their excess suppliers. This surplus is eliminated as prices fall, quantity demanded increases and quantity supplied would decrease until the equilibrium price P* is reached, at which quantity demanded = quantity supplied. The opposite will happen if the price is below the equilibrium price, say at P_O. There will be excess demand or shortage. Consumers are unable to purchase the entire commodity they want at below-equilibrium prices and they bid up the price. This would put upward pressure on price and quantity supplied increases and until price eventually reach the equilibrium price P*, and the market clears.

Thus, through the process of adjustment in price and quantity, eventually equilibrium price and quantity are determined at which quantity demanded and supplied are equal. As long as demand and supply do not change, the equilibrium point remains the same. But in should be noted that, at a particular point in time, the observed market price may or may not be the equilibrium price. All we know is that market forces always push the market price towards the equilibrium price when they are not equal. We can also assume that, in the absence of price controls, the market price is the equilibrium price.

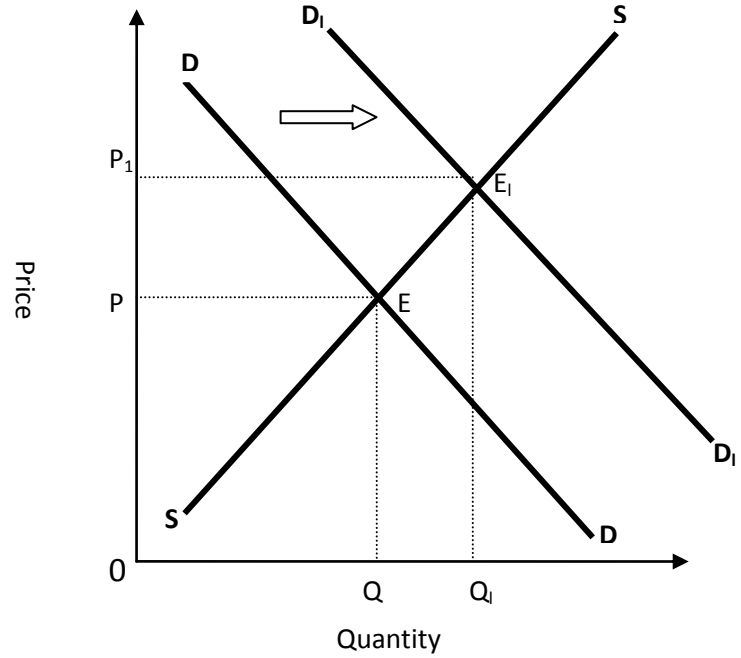
Changes in Market Equilibrium: Comparative Static Analysis:

Changes in either demand or supply cause changes in market equilibrium. Several forces bringing about changes in demand and supply are constantly working which will affect the equilibrium price and quantity of the commodity. Increase or decrease in demand, supply being the unchanged will change the market equilibrium. Similarly, increase or decrease in supply, the demand being unchanged would also affect the equilibrium price and quantity. Both demand and supply of goods may change simultaneously causing a change in market equilibrium. In order to study the shifts in demand and supply curve affect the equilibrium, we use the method known as "comparative statics". That is, we start from a position of equilibrium and then introduce the changes to be studied. The new equilibrium position is determined and compared with the original equilibrium position. The differences between the two equilibriums must result from the change that was to be introduced.

Adjustments to changes in demand

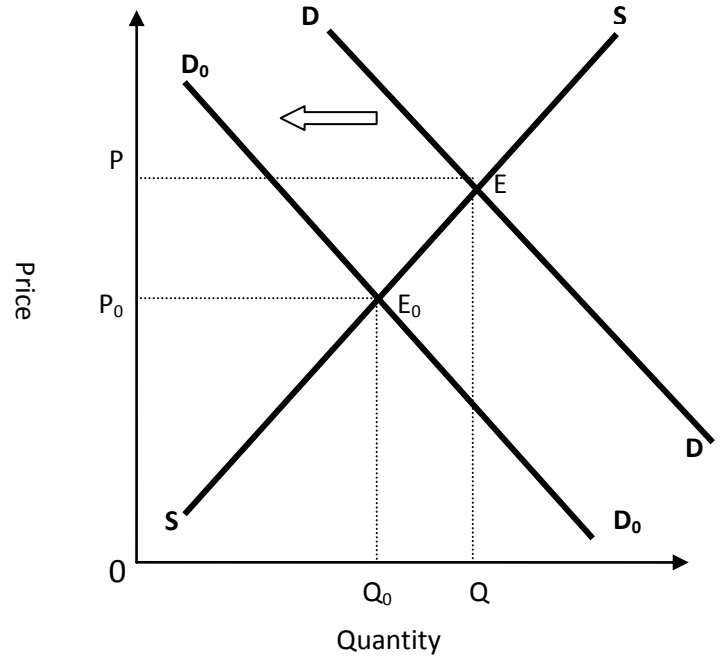
We have seen that the market demand curve of a commodity shifts as a result of a change in consumer's income, their tastes and preferences, prices of related goods (that is, substitutes and complementary goods), the number of buyers in the market and as a result of change in the distribution of income. Given the market supply curve of a commodity, an increase or decrease in demand will cause change in market equilibrium.

Let us begin with the increase in demand. As a result of change in the non-price determinants of demand, demand curve shifts upwards to the right, an increase in demand is said to have occurred. An increase in demand, resulting from for example, an increase in income of the consumer, will cause an increase in both the equilibrium price and quantity bought and sold, as shown below.



DD and SS is the original market demand and supply curve. Both curves intersect at point E and equilibrium price is OP and equilibrium quantity of OQ. As a result of an increase in demand, the entire demand curve shift to the right to the new position D_1D_1, while supply curve remaining the same. It can be seen that, at old price OP, there is shortage or excess demand. This will put upward pressure on the price and it rises to OP_1, where quantity demanded equals quantity supplied and new market equilibrium is attained at E^1. thus as result of an increase in demand we could expect that the buyers pay higher price (OP_1) and sellers produce a greater quantity (OQ_1) than before the change.

Now let us examine the case with the decrease in demand. As said above, a decrease in demand means entire demand curve shifts to a lower position to the left due to factors such as fall in income of the consumer, fall in the price of substitutes etc. A decrease in demand for the product causes a decrease in both equilibrium price and equilibrium quantity bought and sold as shown below.

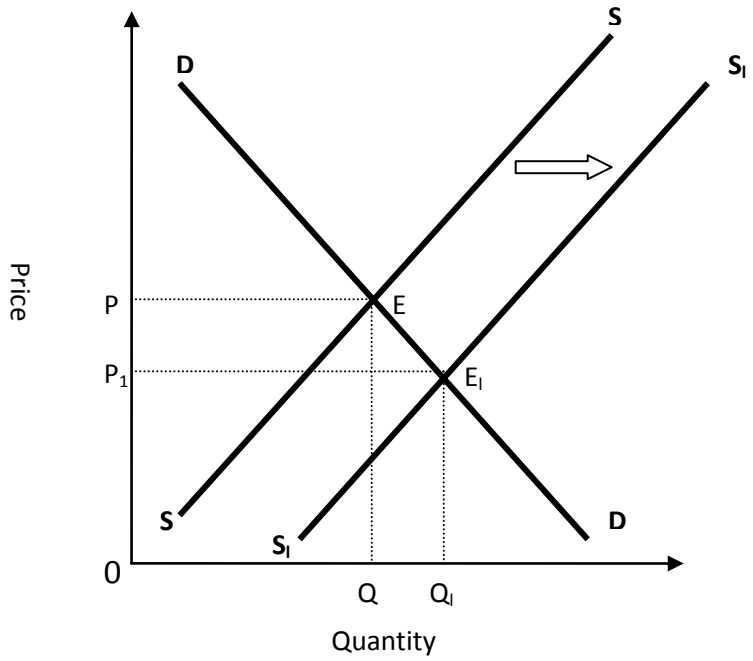


The intersection of original market demand curve DD and supply curve SS defines the equilibrium price OP and equilibrium quantity OQ. As a result of decrease in demand, demand curve shifts to the left to take the new position D_OD_O while the supply curve remaining the same. Both curves intersect at point E_O and equilibrium price is OP_O and equilibrium quantity is OQ_O. Both equilibrium price and quantity are lower than the initial equilibrium position. This is due to the fact that, when there is a decrease in demand and supply remaining the same, there will be a surplus or excess supply. This would put a downward pressure on the price until the surplus is eliminated from the market.

Adjustments to changes in supply

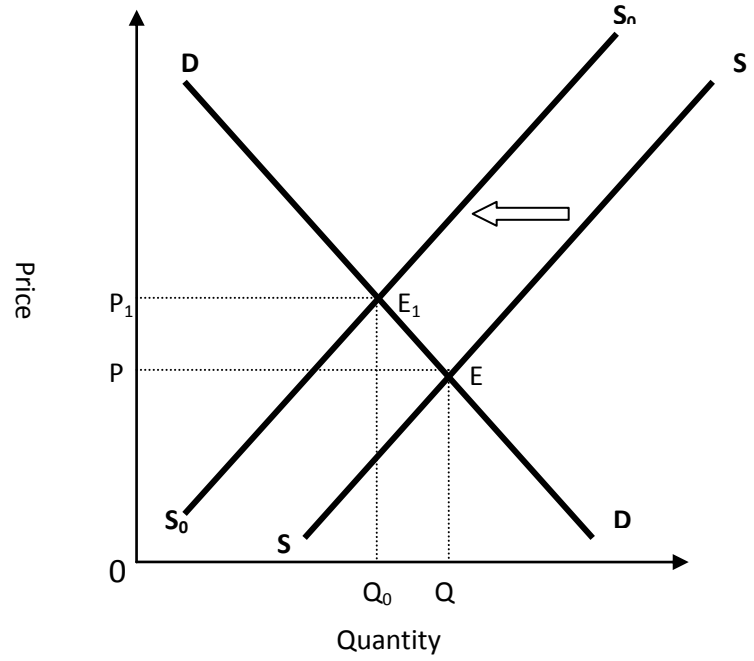
Now let us analyse the impact of changes in supply on the equilibrium price and quantity of the commodity, the demand for the commodity remaining the same. We have seen that the market supply curve for a commodity shifts as a result of changes in the price of related commodities, input prices, technology, number of producers in the market, expectations etc. An increase in supply causes a rightward shift of the supply curve while decrease in supply causes a leftward shift of supply curve.

Let us first examine the case of increase in supply. An increase in supply implies producer's supply more of the commodity as a result of factors like decrease in the price of inputs, improvement in technology etc. An increase in the supply of the commodity, demand being the same, causes a decrease in the equilibrium price and an increase in equilibrium quantity bought and sold, as shown below.



DD and SS is the original market demand and supply curve. Both intersects at point E and determine equilibrium price at OP and equilibrium quantity at OQ. Now as a result of decrease in price of inputs, for example, supply increases causing a rightward shift in supply curve to S_1S_1. When supply increases, demand being the same, there will be a temporary surplus or excess supply in the market. In order to get rid of their surplus, sellers reduce the prices and prices drops to OP_1, which is new equilibrium price defined by the intersection of new supply curve S_1S_1 and demand curve DD. As price falls quantity demanded increases and new equilibrium quantity is OQ_1. Thus, an increase in supply results in lower equilibrium price but larger equilibrium quantity.

Let us now analyse the effect of decrease in supply on the equilibrium price and quantity. A decrease in supply can occur due to factors like increase in the price of inputs used in the production, rise in the price of production substitutes, imposition of tax on the sales/production by the government etc. A decrease in the supply for the commodity, demand being the same causes an increase in the equilibrium price and decrease in the equilibrium quantity bought and sold. This is shown below.



DD and SS is the original demand and supply curve and OP is the equilibrium price and OQ is the equilibrium quantity. A decrease in supply, as a result of increase in price of inputs for example, is shown by the leftward shift in the supply curve. The new supply curve SoSo intersects the given demand curve DD at E_l. New equilibrium price is OP_l, which is higher than old price OP and new equilibrium quantity is OQ_l, lower than old quantity OQ. Thus, decrease in supply leads to the rise in price and fall in equilibrium quantity.

Algebraic Explanation to Market Equilibrium

Let us analyse demand, supply and market equilibrium by using elementary algebra. We have seen that demand function can be expressed as

$$Q_d = f (P, I, P_r, T, A, E) \text{ where,}$$

Q_d = quantity demanded

P = price of commodity

I = income of the individual consumer

P_r = price of related commodities

T = tastes and preferences of individual consumer

A = advertisement expenditure

E = expectations

Let us assume that income, prices of related commodities, tastes, advertisement and expectation are constant. The demand function now becomes a relation between price and quantity demanded and may be expressed as

$Q_d = f(P)$

If we assume that a linear relationship between price and quantity demanded, we can express the demand function as

$Q_d = a - bP$ …….. $(a>0, b>0)$

Where 'a' represents constant intercept term and value of 'a' shows the quantity demanded when price is zero. '–b' represents the slope of demand function, which is approximately negative since the demand curve is downward sloping (so that when price rises quantity demanded falls). For example, a demand function could take the form of an equation such as :

$Q_d = 100 - 2P$

Similarly supply function can be expressed as

$Q_S = f(P, P_r, P_i, T, E, N)$

Where ;

Q_S= quantity supplied of commodity

P = price of the commodity

P_r = prices of related products

P_i = prices of inputs

T = state of technology

E = expectations

N = number of producers in the market

If we assume that prices of related products, prices of inputs, technology, expectations and number of producers are constant, then the supply function can be expressed as

$Q_S = f(P)$

As in the case of demand function, let us assume a linear relationship between price and quantity supplied, we can express the supply function as

$Q_S = -c + dP$ ……$(c>0, d>0)$

Where –c is the constant intercept term so that supply curve crosses the price axis at a positive price. The term'd' is the slope of the supply function. The positive slope means that the supply curve is upward sloping so that when price rises quantity supplied also rises. For example, a supply function could be of the form

$Q_S = -60 + 5P$

To determine equilibrium price and quantity, we must bring the demand and supply function together. The demand and supply function gives us two equations and three unknowns (namely Q_d Q_S and P). We know that, at equilibrium, quantity demanded of a commodity (Q_d) is equal to the quantity supplied (Q_S) so that there is neither a shortage nor a surplus in the market.

Hence,

$Q_d = Q_S$

This equation completes the model and allows us to obtain a unique solution. The complete model is:

$Q_d = a - bP$

$Q_S = -c + dP$

$Q_d = Q_S$

By solving the system of equations for P and Q we will obtain the market equilibrium price and quantity

Example 1:

Consider the following demand and supply equations

$Q_d = 50 - 4P$

$Q_S = -10 + 8P$

Since in equilibrium $Q_d = Q_S$,

We have

$50 - 4P = -10 + 8P$

$60 = 12P$

$P = 60/12 = 5$

By substituting P=5 in either the demand or supply equation we will get Q=30. Hence the equilibrium price is 5 and equilibrium quantity is 30.

Example 2:

Consider the following demand and supply equations

$Q_d = 80-10P$

$Q_S = -40 + 20P$

At equilibrium $Q_d = Q_S$,

So we have

$80-10P = -40+ 20P$

$120 = 30P$

$P = 120/30 = 4$

By substituting P=4 in either the demand or supply equation we will get Q=40. Hence the equilibrium price is 4 and equilibrium quantity is 40.

Example 3:

Consider the following demand and supply equations

$Q_d = 32-3P$

$Q_S = -12 + 8P$

Since in equilibrium $Q_d = Q_S$,

We have

$32-3P = -12+ 8P$

$44 = 11P$

$P = 44/11 = 4$

By substituting P=4 in either the demand or supply equation we will get Q=20. Hence the equilibrium price is 4 and equilibrium quantity is 20.

Elasticity of Demand

We have seen that the demand for a commodity is determined by its own price, income of the consumer, prices of related goods etc. Quantity demanded of a good will change as a result of a change in the size of any of these determinants of demand.

Elasticity measures the sensitivity of one variable to another. Specifically, it is a number that tells us the percentage change that will occur in the variable in response to one percent increase in another variable. Therefore, elasticity of demand refers to the sensitiveness or responsiveness of quantity demanded of a good to a change in its own price, income and prices of related goods. Accordingly, there are three kinds of elasticity of demand .They are

1. Price elasticity of demand
2. Income elasticity of demand
3. Cross elasticity of demand

Price elasticity of demand measures the sensitivity of quantity demanded to change in own price of g good. Income elasticity of demand measures the sensitivity of quantity demanded to change in income of the consumer. While cross elasticity of demand analyses the responsiveness of quantity demanded of one good to changes in the price of another good.

Price elasticity of demand

Price elasticity of demand refers to the responsiveness or sensitiveness of quantity demanded of a good to changes in its own price. In order to have a measure of the responsiveness of quantity demanded of a good to change in its price that is independent of units of measurement, Alfred Marshall, defined in terms of percentage or relative change in quantity demanded to price. As such, price elasticity of demand is given by the percentage change quantity demanded of a good divided by the percentage change in its price. The elasticity is usually symbolised by Greek letter eta (η). Thus, we have

$$\eta = \frac{\text{Percentage change in quantity demanded}}{\text{Percentage change in price}}$$

Now denoting ΔQ for change in quantity demanded and ΔP for the change in price (the symbol Δ is Greek letter delta; it means "the change in") we have the formula for the price elasticity of demand as

$$\eta = \frac{\Delta Q/Q}{\Delta P/P}$$

That is, $\eta = \frac{\Delta Q}{Q} \cdot \frac{P}{\Delta P}$

Or

$$\eta = \frac{\Delta Q}{\Delta P} \cdot \frac{P}{Q}$$

Since, price and quantity demanded are inversely related the coefficient of price elasticity of demand (η) is a negative number. In order to avoid dealing with negative values, a minus sign is often introduced into the formula of price elasticity of demand. That is

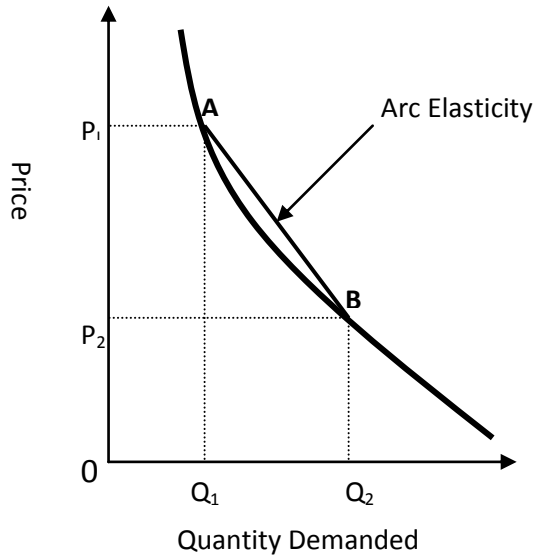
$$\eta = -\frac{\Delta Q}{\Delta P} \cdot \frac{P}{Q}$$

Thus price elasticity of demand is measured by a ratio; the percentage change in quantity demanded divided by the percentage change in the price that brought it about. For normal negatively slopped demand curves, price elasticity will be negative, but two Elasticities are compared by comparing their absolute values. As such, price elasticity of demand is a pure number that is it has no units of measurement attached to it. This allows meaningful comparison between the price elasticity of demand of different commodities.

The above formula is called point elasticity formula of demand because it measures elasticity at a point on the demand curve. The value obtained for η is just a number like 2 or 5 or ½ and is referred to as the coefficient of elasticity. Since price elasticity is being measured at a point on the market demand curve we are assuming that all other factors that affect market demand remain fixed. The demands for some goods are more responsive to changes in price than those of others. That is, demands for some goods are more 'elastic' than those for others or the price elasticity of demand of some goods is greater than those of others. It should be noted that the terms elastic and inelastic are used in relative sense. In other words, elasticity is a matter of degree only.

The Arc Elasticity Formula

Formula of point elasticity of demand measures the elasticity at particular point on the demand curve. It can be conveniently used when the changes in the price and resultant quantity demanded are infinitesimally smaller. However, when the price change is large, we have to measure elasticity over an arc of the demand curve rather than at a specific point on it. The arc elasticity measures elasticity of demand between two points on the demand curve. That is, arc elasticity is a measure of average elasticity. Consider the following figure.



The initial price is P_1 and corresponding quantity is Q_1. When price falls to P_2, quantity demanded increases to Q_2. The arc elasticity measures elasticity at the point of the cord that connects the two points A and B on the demand curve defined by the initial and new price level. By taking the average of the two prices and average of two quantities, we can obtain the following formula for the price elasticity of demand

$$\eta = - \frac{\Delta Q}{\Delta P} \cdot \frac{(P_1 + P_2)/2}{(Q_1+Q_2)/2}$$

Or

$$\eta = - \frac{\Delta Q}{\Delta P} \cdot \frac{(P_1 + P_2)}{(Q_1+Q_2)}$$

The new formula is called the arc elasticity of demand formula or average elasticity of demand formula because it measures η between two points on the demand curve. Arc elasticity of demand treats the price and quantity as if they were midway between the initial and new prices and quantities and then uses the point elasticity at this midpoint

Arc elasticity formula should be used when the change in price is somewhat large, but not very large. On the other hand, when the two points on the demand curve are very close together, arc becomes almost identical with the true demand curve and the arc elasticity measurement becomes almost identical with the point elasticity measurement on the demand curve.

Total Outlay Method

Another method to measure price elasticity of demand is known as total outlay or expenditure method. In this method, changes in the total expenditure made on the good as a result of change in its price is analysed to measure price elasticity of demand. But with the total outlay method, we can know only whether price elasticity is equal to one, greater than one or less than one. With this method, we cannot find out the exact coefficient of price elasticity of demand.

If as a result of the change in price of the commodity total expenditure remains the same, then elasticity of demand for the commodity will be equal to unity. This is so because total expenditure made on the commodity can remain the same only if the proportional change in the quantity demanded is equal to proportional change in price. On the other hand, due to fall in price of the commodity, quantity demanded rises and, as result, total expenditure made on the commodity increases, then price elasticity of demand is said to be greater than unity. This is so because with the fall in price of the commodity, total expenditure can increase only if the proportional change in quantity demanded is greater than the proportional change in the price.

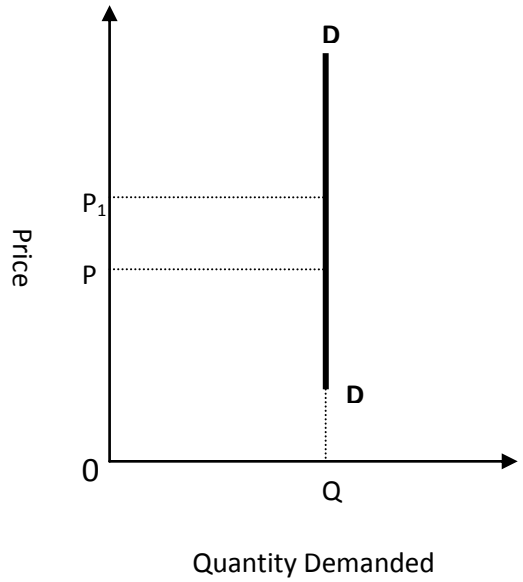
If as a result of fall in the price of the commodity total expenditure decreases, then price elasticity of demand will be less than unity. This is for the reason that with the fall in price, total expenditure can decrease only if proportional increase in quantity demanded is less than proportional change in price. Thus, through the total outlay method, we can find out whether prices price elasticity is equal to unity or greater than unity or less than unity. Note that with this method, we cannot know the precise value of the price elasticity.

Degrees of Elasticity of Demand

The value of price elasticity of demand ranges from zero to infinity. That is, $0 \leq \eta \leq \infty$. Based on the value of elasticity or degree of responsiveness of quantity demanded, price elasticity of demand is classified into five categories. They are: (i) Perfectly inelastic demand, (ii) Inelastic demand, (iii) Unitary elastic demand, (iv) Elastic demand, (v) Perfectly elastic demand. Now let us analyse each of them in detail:

(1) Perfectly inelastic demand

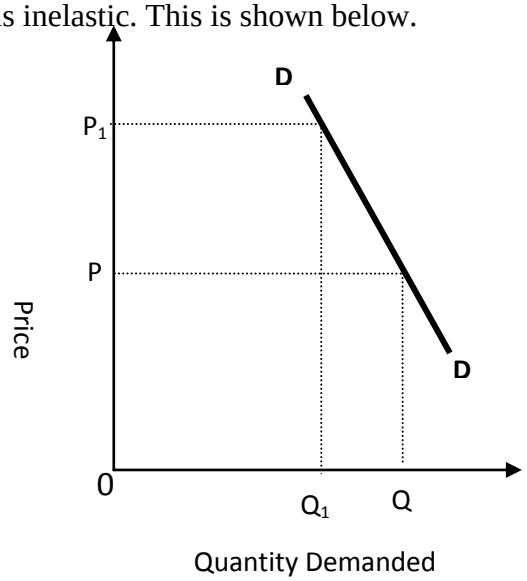
When quantity demanded does not change as a result of change in price, demand is said to be perfectly inelastic. Quantity demanded is unchanged when price changes or demand shows no response to change in price. In other words, same quantity will be bought whatever the price may be. Numerical value of elasticity will be zero ($\eta = 0$) when there is perfectly or completely inelastic demand. The following figure illustrates the case of perfectly inelastic demand.



A change in price from P to Pl leaves quantity demanded unchanged at Q units. That is, quantity demanded does not change at all when price changes.

(2) Inelastic Demand

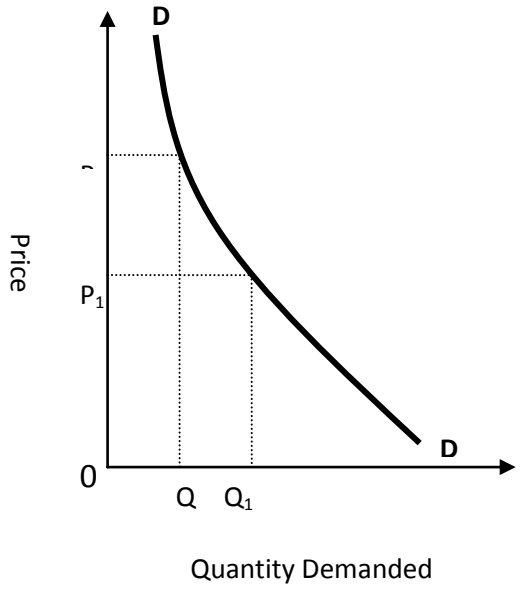
As long as there is some positive response of quantity demanded to change in price, the absolute value of elasticity will exceed zero. The greater the response, the larger the elasticity. However, when percentage change in quantity demanded is less than percentage change in price, demand is said to be inelastic. That is, a certain percentage change in price leads to a smaller percentage in quantity demanded. The coefficient of elasticity will be less than one but greater than zero ($0< \eta <1$) when demand is inelastic. This is shown below.



When change in price from OP to OPl causes a less than proportionate change in quantity demanded. That is, quantity demanded changes by a smaller percentage than the change in price.

(3) Unitary Elastic Demand

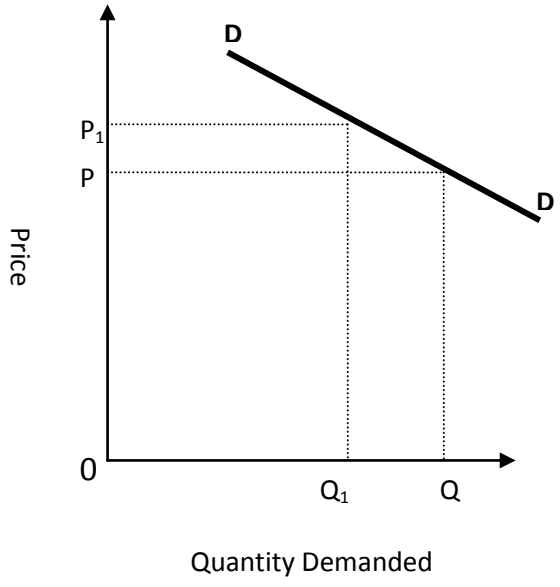
If a certain percentage change in price leads to an equal percentage change in quantity demanded, then demand said to have unitary elasticity. Unitary elasticity is the boundary between elastic and inelastic demand. The coefficient of elasticity will be equal to one when demand is unitary elastic ($\eta=1$). The demand curve having unitary elasticity over its whole range is shown below



OP and OQ are the initial price and quantity. A fall in price from OP to OP_1 causes an equal proportional change in quantity demanded from OQ to OQ_1.

(4) Elastic Demand

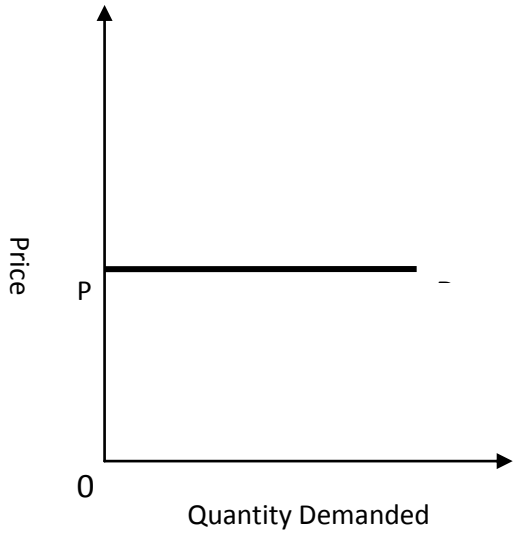
When the percentage change in quantity demanded exceeds the percentage change in price, the demand is said to be elastic. That is, a certain percentage change in price leads to a greater percentage change in quantity demanded. The value of coefficient of elasticity will be greater than one but less than infinity when demand is elastic ($1<\eta<\infty$). This is shown below.



An increase in price from OP to OPl causes a more than proportionate increase in quantity demanded as shown by the change in quantity demanded from OQ to OQl. Thus, a small rise in price brings in more than proportionate fall in quantity demanded.

(5) Perfectly Elastic demand

If a small change in price leads to an infinitely large change in quantity demanded, we can say that demand is perfectly elastic. When demand is perfectly elastic, small price reduction will raise demand to infinity. At the same time, a slightest rise in price causes demand to fall to zero. At the going price, consumers will buy an infinite amount (if available).above this price, they will buy nothing. The coefficient of elasticity will be infinity when demand will be infinite when demand is perfectly elastic ($\eta = \infty$). The graph for perfectly elastic demand is shown below.



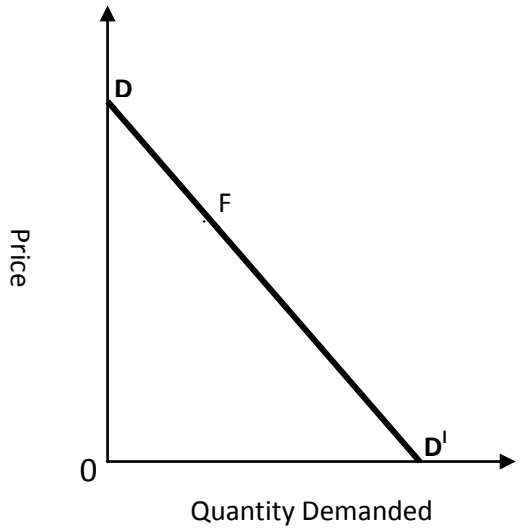
When it is perfectly elastic, demand curve is a horizontal straight line. In his case an infinitely large amount can be sold at the going price OP. A small price increase from OP decreases quantity demanded from an infinitely large amount to zero (hyper sensitive demand).

The following table summarises the terminology of price elasticity of demand

Term	Numerical Measure of elasticity	Shape of the demand curve	Verbal description
Perfectly inelastic	Zero	Vertical (parallel to Y-axis that measures price)	Quantity demanded does not change an price changes
Inelastic	Greater than zero but less than one	Steeper	Quantity demanded changes by a smaller percentage than does price
Unitary elastic	One	Rectangular hyperbola	Quantity demanded changes exactly the same percentage as does price
Elastic	Greater than one but less than infinity	Flatter	Quantity demanded changes by a larger percentage than does price
Perfectly elastic	Infinity	Horizontal (parallel to X-axis that measures quantity)	Buyers are prepared to buy all they can at some price and none at all at higher prices.

Price elasticity on a linear Demand Curve

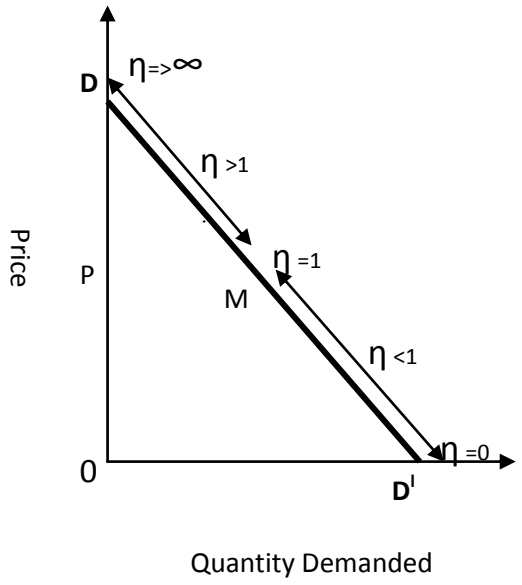
When demand curve is linear, the price elasticity can be computed graphically. Consider the following linear demand curve.



Graphically, the point elasticity of a linear demand curve is the ratio of the segment of the line to the right and to the left of the particular point. That is,

$\eta = \frac{\text{Lower segment}}{\text{Upper segment}}$

In the above demand curve, the elasticity at point F is the ratio FD^1/FD. Given this graphical measurement of point elasticity, it is obvious that at mid-point of a linear demand curve, price elasticity will be equal to unity ($\eta = 1$), as shown below.



In the figure, at the mid-point of the curve M, the coefficient of elasticity $\eta = 1$. At points to the right of M, elasticity is less than unity ($\eta<1$). At any points to the right of M, elasticity will be greater than unity ($\eta>1$). At point D the elasticity approaches infinity ($\eta=>\infty$). At point D^1, the coefficient of elasticity is zero ($\eta=0$).

Thus, elasticity of a negatively slopped demand curve varies between infinity at the price axis to zero at the quantity axis. At the price axis, since Q=0, elasticity is infinity. As we move down the demand curve price falls and quantity demanded rises steadily, elasticity falls. At the quantity axis, where P=0, elasticity is also zero.

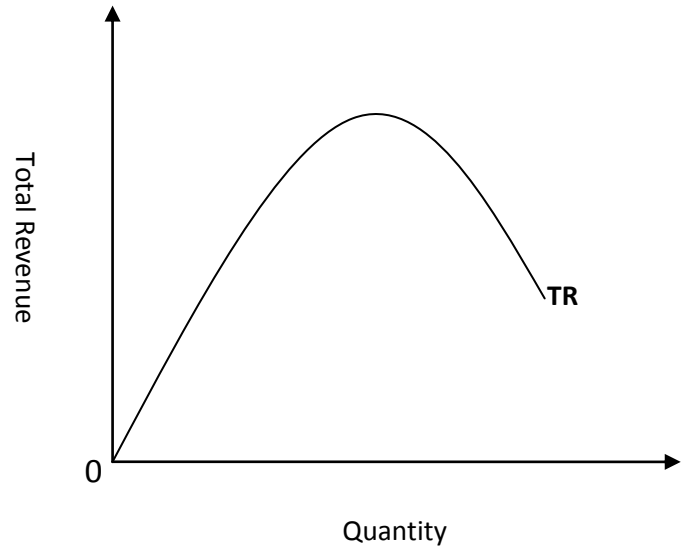
Marginal Revenue, Price and Elasticity

So far we analyzed price elasticity of demand from the consumer's side only. However, consumer's expenditures on a commodity are the receipts or the total revenues of the sellers of the sellers of the commodity. Thus, another way of looking at the price elasticity of demand for a good or service is to see what happens to the total revenue as the price of the good or service changes.

The total amount of money earned by sellers of a commodity is called total revenue (TR). Total revenue is derived from the sale of any commodity is the price of the product (P) multiplied by the quantity sold (Q). That is,

$$TR = P.Q$$

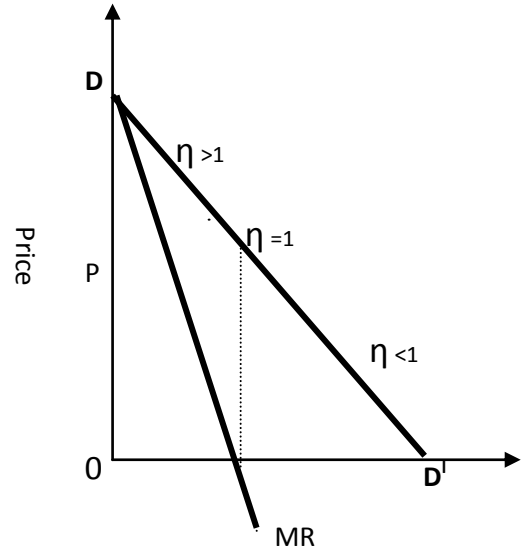
If the market demand is linear, the total revenue curve will initially slopes upwards, reaches a maximum point and then start declining, as shown below.



The marginal revenue (MR) is the changes in total revenue resulting from selling an additional unit of a commodity. That is, marginal revenue is the addition made to the total revenue by selling an additional unit of the commodity. Marginal revenue is calculated by dividing the changes in total revenue (ΔTR) by the change in quantity sold (ΔQ)

That is, $MR = \Delta TR / \Delta Q$

Graphically, the MR curve is the slope of TR curve at any one level of output. If the demand curve is linear, it is obvious that in order to sell additional units of the commodity price must fall. Thus MR curve will lie below the demand curve. It starts at the same point on the vertical axis as demand curve and is everywhere else below the demand curve. This is illustrated in the following figure.



The important relationship among marginal revenue, price and price elasticity of demand is given by:

$MR = P (1-1/\eta)$

This is a crucial relationship for the theory o pricing. If $\eta = 1$, then MR is equal to zero and TR reaches its maximum point. That is, if

$$\eta = 1, \text{ hence } MR = P (1-1/\eta)=0.$$

If $\eta > 1$, MR will be positive and TR curve has a positive slope and hence has not reached its maximum point. That is,

$$\eta > 1, \text{ then } MR = P (1-1/\eta)>0 \text{ since } P>0$$

If $\eta < 1$, MR will be negative and TR curve has a negative slope. That is, TR curve is falling. If

$$\eta < 1, \text{ then } MR = P (1-1/\eta) <0 \text{ since } P>0$$

We can summarise these results as follows.

(1) If the price elasticity of demand has a unitary elasticity ($\eta = 1$) total revenue is not affected by changes in the price. This is because if $\eta = 1$, then MR=0.
(2) If demand is elastic ($\eta > 1$) an increase in price will results in decrease in the total revenue while a decrease in price will results in an increase in total revenue (MR>o).
(3) If demand is inelastic ($\eta < 1$) an increase in price lead to increase in total revenue and a decrease in price leads to fall in total revenue (MR >0).

Determinants of price Elasticity of Demand

Important determinants of elasticity of demand of a commodity with respect to its own price are explained below.

(1) Availability of Substitutes

One of the most important factors likely to influence the price elasticity of demand for a commodity or service is whether or not substitutes are available. If a commodity has many close substitutes, its demand is likely to be elastic. This is so because if the price of that commodity rises buyers will switch to some of many close substitutes available. Hence quantity demanded of that commodity will tend to fall significantly. The greater the possibility of substitution, the greater the price elasticity of demand for it. On the other hand, if there are not many substitutes quantity demanded will tend to fall as a result of the higher price, but not by much. That is, if there are few or no close substitutes, demand tend to be inelastic.

(2) Nature of the commodity

Whether the commodity is a luxury or a necessity has some effect on its price elasticity of demand. In general, necessities are price inelastic. If the price of a basic necessity increases, say by 10%, quantity demanded will not probably fall by that proportion. Consumers tend to sacrifice some other commodities rather than a substantial reduction in the quantity of necessities. On the other hand, luxury goods are price elastic. An increase in the price of luxury good is likely to cause a more than proportionate decrease in the quantity bought, other things being equal.

(3) Time Period

The time period being considered will also have some effect on the elasticity of demand for the product. In general, the longer the time period being considered, the more elastic the demand is likely to be. This is largely due to the fact that it takes time for people to substitute one commodity for another. At the same time, in the short run, substitution of one commodity by another is not so easy. Hence demand tends to be relatively inelastic.

(4) Number of Uses

In general, the greater the number of uses of a commodity has, the more price elastic the demand for that commodity is likely to be. A decrease in the price of a commodity that has large number of uses (milk, for example) more of it will be bought to allocate to different uses. On the other hand, if the commodity has only one or two uses, it is unlikely that a fall in its price will cause a significant increase in quantity demanded.

(5) Proportion of income spent on the commodity

Another factor that is likely to affect price elasticity of demand is the proportion of income spent on the commodity. If only a negligible percentage of consumer's income is spend on the commodity, the demand for that commodity is likely to be inelastic. An increase in the price of such commodity has no appreciable effect on the consumer's budget. Example, matches, soap. The greater the proportion of income spent on the commodity the greater will be its price elasticity of demand.

Income Elasticity of Demand

The responsiveness or sensitiveness of quantity demanded of a commodity to changes in income of the consumer is termed as income elasticity of demand. It is the proportionate or percentage change in quantity demanded resulting from proportionate change in income. Thus we have

η_y = Percentage change in quantity demanded
Percentage change in income

Now denoting ΔQ for small change in quantity demanded and ΔY for the small change in income we may symbolically write the formula for the income elasticity of demand as

$\eta_y = \Delta Q/Q \ / \Delta Y/Y$ That is, $\eta_y = \Delta Q/ Q \ . Y/ \Delta Y$ Or $\eta_y = \Delta Q/ \Delta Y \ . Y/Q$

For the most commodities, increase in income leads to increase in quantity demanded. Therefore, income elasticity is positive. If the resulting percentage change in quantity demanded is larger than the percentage change in income, income elasticity will exceed unity ($\eta_y > 1$). Then the commodity's demand is said to be income elastic. If the percentage change in quantity demanded is smaller than the percentage change in income, income elasticity will be less than unity ($\eta_y < 1$). Then the commodity's demand is said to be income inelastic. If the percentage changes in income and quantity demanded are equal, income elasticity will be unity ($\eta_y = 1$). The commodity's demand is said to have unitary income elasticity of demand. Unitary income elasticity represents a useful dividing line.

There is also a relationship between income elasticity for a commodity and proportion of income spent on it. If the proportion of income spend on the commodity increases as income increases, then the income elasticity of demand for the commodity is greater than unity ($\eta_y > 1$). If the proportion of income spend on the commodity decreases as income rises, then the income elasticity of demand for the commodity is less than unity ($\eta_y < 1$). At the same time, if the proportion of income spend on the commodity remains the same as income rises, then the income elasticity of demand for the commodity is equal to unity ($\eta_y = 1$).

If the commodity is normal, a rise in income causes more of it to be consumed. Other things being equal, this means a rightward shift in the commodity's demand curve. Thus, income elasticity will be positive for normal commodities. In the case of such commodities, an increase in income leads to an increase in quantity demanded. On the other hand, if the commodity is inferior, a rise in income causes less of it to be demanded. This implies a leftward shift in the commodity's demand curve. Thus income elasticity for inferior commodities will be negative. In the case of inferior commodities increase in income will lead to fall in quantity demanded. The boundary case between normal and inferior commodities occurs when a rise income leaves quantity demanded unchanged so that income elasticity is zero. Zero income elasticity implies that quantity demanded of the commodity is quite unresponsive to changes in income. Zero income elasticity is significant because it represents a dividing line between positive income elasticity on one side and negative income elasticity on the other.

A normal commodity can be further classified as necessities and luxury using income elasticity. A commodity is considered as necessity if the income elasticity is less than unity. That is, in the case of necessities, the proportion of income spend on it falls as income rises. A commodity is considered to be luxury if its income elasticity is greater than unity. The proportion of consumer's income spend on luxuries rises as his income increases.

It should be said that, sometimes, the same commodity can be regarded as a luxury by some individuals or at some income levels and as a necessity or even as inferior commodity by other individuals or at other income levels. The terminology of income elasticity is summarised in the following table.

Type of Goods	Numerical Measure of Income elasticity	Verbal description
(1)Inferior Goods	Negative	Quantity demanded decreases as income increases
(2)Normal goods	Positive	Quantity demanded increases an income increases
2.1 Necessity	Less than one	Quantity demanded increases less than proportion to increase in income
2.2 Luxury	Greater than one	Quantity demanded increases more than proportion to increase in income

Cross Elasticity of Demand

The responsiveness of quantity demanded of one commodity to changes in the prices of other commodities if often of considerable interest. The responsiveness or sensitiveness of quantity demanded of one commodity to the changes in the price of another commodity is called cross elasticity of demand. Thus, cross elasticity of demand can be defined as percentage or proportionate change in quantity demanded of commodity X resulting from a proportionate change in the price of commodity Y. the cross elasticity of commodity X with respect to the price of Y (η_{XY}) can presented as

η_{XY} = Percentage change in quantity demanded of X
Percentage change in price of Y

We may symbolically write the formula for the cross elasticity of demand as:

$\eta_{XY} = \Delta Q_X/Q_X \ / \ \Delta P_Y/P_Y$; That is, $\eta_{XY} = \Delta Q_X \ / \ Q_X \ . \ P_Y \ / \ \Delta P_Y$

Or

$$\eta_{XY} = \Delta Q_X / \Delta P_Y \ . \ P_Y / Q_X$$

Where ΔQ_X is the change in quantity demanded of X, ΔP_Y is the change in price of Y, P_Y is the original price of Y and Q_X is the original quantity of X. The coefficient of cross elasticity can vary from minus infinity to plus infinity. Substitute goods have positive cross elasticity and complementary goods have negative cross elasticity.

If η_{XY} is positive, the commodities X and Y are said to be substitutes. X and Y are substitutes if more of X is purchased when price of Y goes up. That is, an increase in P_Y leads to an increase in Q_X as X is substituted for Y in consumption. For example, consumers usually purchase more coffee when price of tea rises. Thus coffee and tea are substitutes or competing goods. In response to the rise in the price of one good, the demand for the other good rises.

On the other hand, if η_{XY} is negative, X and Y are said to be complementary goods. When X and Y are complementary goods, less of X will be purchased when the price of Y goes up. That is, an increase in P_Y leads to a reduction in Q_X (and Q_Y). For example consumers usually purchase fewer scooters when the price of petrol goes up. Thus scoter and petrol are complements. Other examples of commodities that are complements are bread and butter, tea and sugar and so on. In the case of complements, a rise in the price of one good brings about a decrease in demand for the other, as they are consumed together. If η_{XY} is zero, X and Y are independent commodities. A change in price of Y has no effect on the quantity demanded of X. this may be the case with cars and pencils, telephones and chewing gum and so on.

It should be noted that the value of η_{XY} is not equal to the value of η_{YX} since the responsiveness of Q_X to the change in P_Y need not be equal to the responsiveness of Q_Y to the change in P_X. For example, a change in the price of tea is likely to have a greater effect on the quantity of sugar (a complement of tea) demanded than the other way around, since tea is more important of the two in terms of total expenditure.

The concept of cross elasticity of demand is very significant in economic theory. The classification of commodities into substitutes and complementary is in terms of cross elasticity of demand. Again, a high positive cross elasticity of demand is often used to define an industry since it indicates that various commodities are similar. Besides we can also classify different market structures on the basis of cross elasticity of demand.

Following table summarizes terminology of cross elasticity of demand

Type of goods	**Numerical measure of cross elasticity**	**Verbal description**
Substitutes	Positive	Quantity demanded of a good increases if the price of substitutes increases
Complementary	Negative	Quantity demanded of a good decreases if the price of complements increases
Independent	Zero	Quantity demanded of a good remains unchanged to change in the price of other good

Elasticity of Supply

The concept of elasticity of supply closely parallels that of elasticity of demand. Though quantity supplied is influenced by a number of factors, we will focus on the commodity's own price as a factor influencing supply. That is, we will be concerned with price elasticity of supply.

Price elasticity of supply measures the responsiveness or sensitiveness of quantity supplied of a commodity to a change in its price. It is given by the percentage change in the quantity supplied of a commodity divided by the percentage change in price. Letting ε (Greek letter epsilon) stand for the coefficient of price elasticity of supply, we have,

$$\varepsilon = \frac{\text{percentage change in quantity demanded}}{\text{percentage change in price}}$$

Being expressed in terms of relative or percentage changes, the price elasticity of supply is a pure number. That is, it has no units attached to it. The value of price elasticity of supply does not change when the units of measurement are changed. This allows meaningful comparisons in the price elasticity of supply of different commodities.

Using the point elasticity formula

$$\varepsilon = \frac{\Delta Q/Q}{\Delta P/P}$$

That is,

$$\varepsilon = \frac{\Delta Q}{\Delta P} \cdot \frac{P}{Q}$$

Where ΔQ represents change in the quantity supplied and ΔP represents change in price. Since quantity supplied and price move in the same direction, supply curves normally have positive slope. Therefore, supply elasticity is normally positive. It will be anything between zero and infinity ($0 \leq \varepsilon \leq \infty$).

Point elasticity of supply measures elasticity at a particular point on the supply curve. More frequently, we measure elasticity of supply by using arc elasticity formula between two points on the supply curve. Arc elasticity of supply measures the average of two prices and the average of two quantities. Letting P_l refer to the lower of the two prices and Q_l being quantity and P_2 to the higher of the two prices and Q_2 the corresponding quantity, we can measure arc elasticity of supply by

$$\varepsilon = \frac{\Delta Q}{\Delta P} \cdot \frac{(Pl + P2)/2}{(Q1+Q2)/2}$$

Or

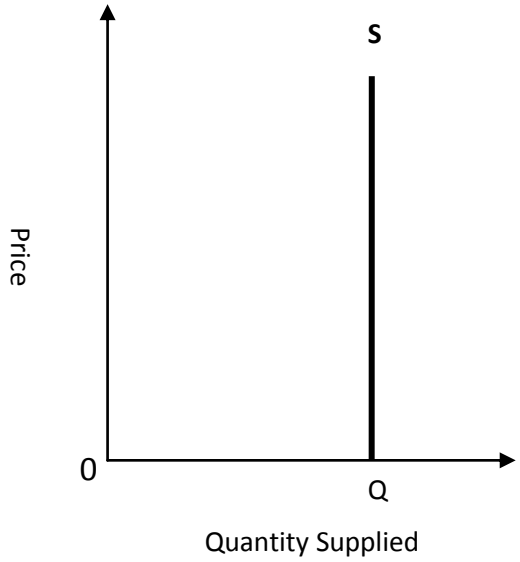
$$\varepsilon = \frac{\Delta Q}{\Delta P} \cdot \frac{(Pl + P2)}{(Q1+Q2)}$$

Degrees of Supply Elasticity

When the supply curve is upward sloping, the elasticity of supply will be anything between zero and infinity. On the basis of the value of the coefficient of elasticity of supply we can classify it into the following five categories: (i) Perfectly inelastic supply, (ii) Inelastic supply, (iii) Unitary elastic supply , (iv) Elastic supply, (v) Perfectly elastic supply. Let us each one of them in detail.

(1) **Perfectly Inelastic Supply**

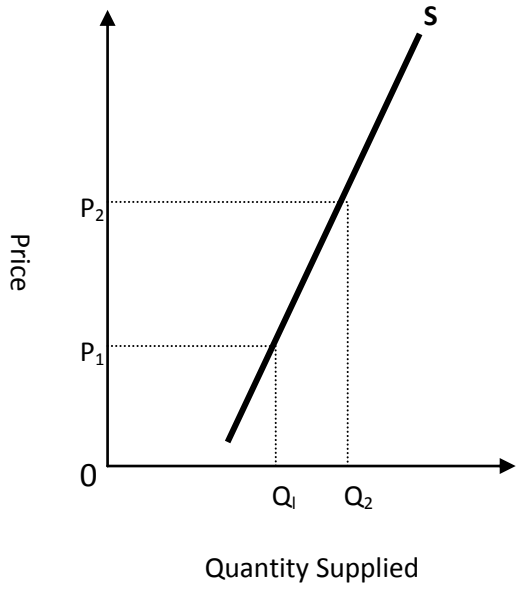
When the quantity supplied of a commodity does not change at all in response to the change in price, elasticity of supply is said to be perfectly inelastic. This is the case of zero elasticity (ε = o) and the supply curve will be vertical straight line, as shown below.



The supply curve has zero elasticity since the same quantity Q is supplied whatever the price.

(2) **Inelastic Supply**

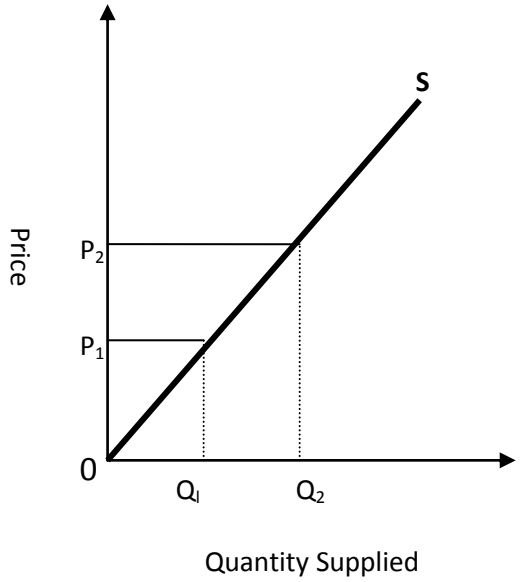
If the percentage change in quantity supplied is smaller than the percentage change in price, supply is said to be inelastic. The value of the coefficient of supply will be greater than zero but less than unity ($0<\varepsilon<1$). If a linear supply curve crosses or cuts the horizontal (quantity) axis, supply is inelastic, as shown below.



A change in price from P_1 to P_2 causes less than proportional change in quantity supplied from Q_1 to Q_2.

(3) Unitary Elastic Supply

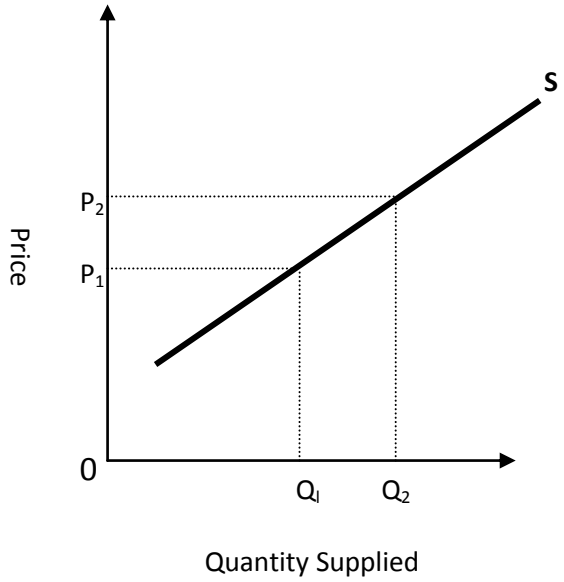
If the percentage change in quantity supplied is equal to percentage change in price, supply is said to be unitary elastic. The value of coefficient of elasticity will be equal to one ($\varepsilon = 1$) when supply is unitary elastic. If linear supply curve passes through the origin, supply is unitary elastic regardless of its scope. This is illustrated below.



The figure shows that any straight line has a unitary elasticity indicating that the percentage change in quantity equals the percentage change in price between any two points on the curve.

(4) Elastic Supply

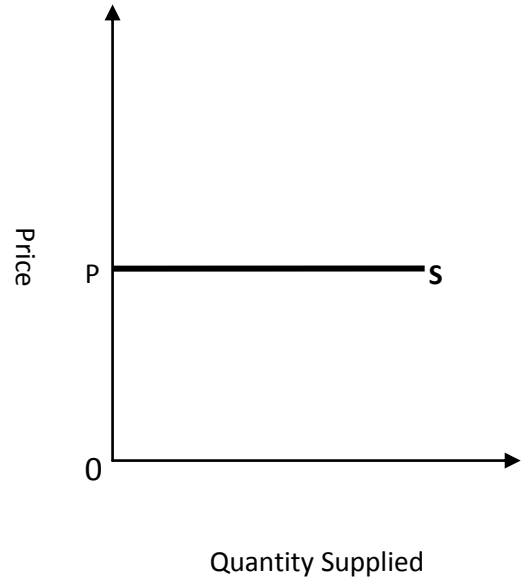
If the percentage change in quantity supplied is greater than percentage change in price, supply is said to be elastic. The value of the coefficient of elasticity will be greater than unity ($1<\varepsilon<$) when the supply is elastic. A linear supply curve indicates an elastic supply if it cuts the vertical (price) axis.



An increase in price from P_1 to P_2 causes more than proportionate increase in quantity supplied from Q_1 to Q_2.

(5) Perfectly Elastic Supply

At any given price infinite quantity is supplied, supply is said to be perfectly elastic. The coefficient of elasticity will be infinity ($\varepsilon=\infty$) when supply is perfectly elastic. Perfectly elastic supply curve is depicted by a horizontal supply curve parallel to quantity axis.



The supply curve has infinite elasticity at price OP. Nothing at all will be supplied at price below OP, while an infinitely large quantity will be supplied at price OP.

Determinants of Elasticity of Supply

The following are important factors that affect elasticity of supply

(1) Time

The time period under consideration has a significant effect on the price elasticity of supply. If the time period is very short, an increase in price does not significantly affect the quantity offered for sale. If a certain quantity of a commodity has already been produced and brought to the market, an increase in price does not cause a large quantity to be offered for sale, because the quantity is fixed. As the time period under consideration becomes longer, supply tends to become more elastic. Sellers will be able to respond more easily to changes in the prices of their products.

(2) Change in cost of Production

Elasticity of supply of a commodity depends upon the ease with which increases in output can be obtained without bringing about a rise in cost of production. It also depends on how easily producers can shift from the production of other products to the one whose price has risen. For example, if agricultural land and labor can be readily shifted from one crop to another, the supply of any one crop will be more elastic than if labor cannot be shifted.

(3) Storage Cost

The elasticity of supply for goods that are not perishable and can be stored at relatively low cost tends to be greater than that for perishables and goods with high storage cost. If price of item that can be stored cheaply falls, sellers may respond by withdrawing the items from the market and storing it. If the price of such an item rises, suppliers may be in a position to release some extra quantities from storage onto the market. These options may not be feasible in the face of high cost of storage.

(4) Responsiveness of Producers

The elasticity of supply for a product depends on the responsiveness of producers to change in price. If producers do not respond positively to the increases in prices, the quantity supplied of the product would not increase as a result of rise in its price. A producer is said to be rational if he raises the supply following the rise in price to maximize his profit. However, producers do not always exhibit profit maximizing behavoiur and as r result do not raise the supply in response to the rise in price.

(5) Production substitutes and complements

If a product has large number of substitutes in production, its supply is likely to be elastic. If the price of such a product falls, producers can shift resources into the production of any of the many substitutes. On the other hand, production complements are goods that are produced together. These are joint products. The supply of relatively minor joint product is likely to be inelastic.

Uses of Elasticity

Knowledge of elasticity of demand is very useful and often necessary in reaching correct decisions in business and government. The following are the important uses and applications of elasticity of demand.

(1) Pricing Decisions of Business Firms

The business firms must take into account the elasticity of demand when they take decisions regarding pricing of products. This is because change in price of a product will bring about change in quantity demanded depends upon the coefficient of elasticity. If the demand for the product of the firm happens to be elastic, then any attempt on the part of the firm to raise the price will bring about a fall in total revenue. Thus instead of gaining from the increase in price, it will lose substantial part of its revenue. On the other hand, if the demand for the product happens to be inelastic, then increases in price will lead to increase in total revenue. Thus, for fixing an optimum or profit maximizing price, the firm cannot ignore the elasticity of demand for the product.

(2) Government Tax Policy

The elasticity of demand is also significant in the field of tax policy by the government. Government must take into account the elasticity of demand for the product before imposing and/or tax on it. This is because it is only when tax is imposed on the commodity with price inelastic demand, which will raise a great deal of revenue of the government. At the same time, if the demand for the commodity is price elastic, a rise in price caused by tax will bring about large decline in quantity demanded and as a result government revenue will also decline

(3) Importance in international trade

The concept of elasticity of demand is also important in the field of international economics. The decisions of country's to undertake devaluation of their currencies or not to the improve balance of payment depends upon the coefficient of elasticity of exports. As a result of devaluation, rice of imports will increase and price of exports will fall. If the world demand for country's exports is inelastic, the fall in the prices of exports as a result of devaluation will lower their export earnings rather than increasing it. On the other hand, if the world demand for country's exports is elastic, then the fall in the prices of exports due to devaluation will bring about large increase in their quantity demanded which increase export earnings and will improve balance of payment position of the country.

Suggested Readings:

1. Dominick Salvatore: Micro Economics, Harper Collins Publishers, Chapters 2 and 5.
2. Dominick Salvatore: Micro Economic Theory- Third Edition, Schaum's Outline Series, Chapters 1, 2 and 3.
3. Koutsoyiannis: Modern Micro Economics – Second Edition, McMillan Press Ltd, Chapter 2.
4. Anindya Sen : Micro Economics; Theory and Applications, Oxford University Press, Chapters 2,3 and 5
5. Richard G Lipsey and Alec Chrystal : Principles of Economics – Ninth Edition, Oxford University Press, Chapters 3 and 4
6. Robert S Pindyck, Daniel L Rubinfeld and Prem L Mehta: Micro Economics – Sixth Edition, Pearson Education, Chapters 1 and 2.
7. Ahuja: Modern Micro Economics – Third Edition, S Chand and Company Ltd, Chapters 4, 7, 13 and 20.
8. Elijab M James: Economics; A Problem Solving Approach, Fourth Edition, Prentice Hall, Chapters 4 and 5.

Module IV

The Consumer Behaviour and Demand

The purpose of the theory of demand is to determine the various factors that affect demand. Demand is a multivariate relationship, that is, it is determined by many factors simultaneously. The traditional theory of demand has concentrated on four determinants, viz, Price of the commodity, prices of the related goods, income and tastes. The market demand is assumed to be the summation of the demands of individual consumers. Thus the traditional theory of demand starts with the examination of the behaviour of the consumer. The consumer is assumed to be rational. Given his income and the market prices of the various commodities, he plans spending of his income so as to attain the highest possible satisfaction or utility (This is the axiom of utility maximization). Further it is assumed that the consumer has full knowledge of all the information relevant to his decision, ie., he has complete knowledge of all the available commodities, their prices and his income.

Utility Analysis

The word utility denotes the want satisfying power of a commodity or service. An individual demands a particular commodity because of the satisfaction he could receive from consuming it. A commodity possesses utility if it satisfies an economic want. Goods may be poisonous or dangerous to one's health, but it possesses utility for those who want them. Utility is thus subjective and does not carry any ethical connotation.

Total and Marginal Utility

An individual demands a particular commodity because of the satisfaction or utility that can be received from consuming it. When the consumer buys a commodity, he receives them in units 1, 2, 3, 4, etc. when the unit increases, the utility he receives also increases. The sum total of utilities obtained by the consumer from different units of a commodity is the total utility. In the illustration that follows the total utility of 2 units is 18 (10 + 8) utils. It is 24 (10 + 8 + 6) utils from 3, 28 (10 + 8 + 6 + 4) from 4 etc.

Table 1: Total Utility and Marginal Utility

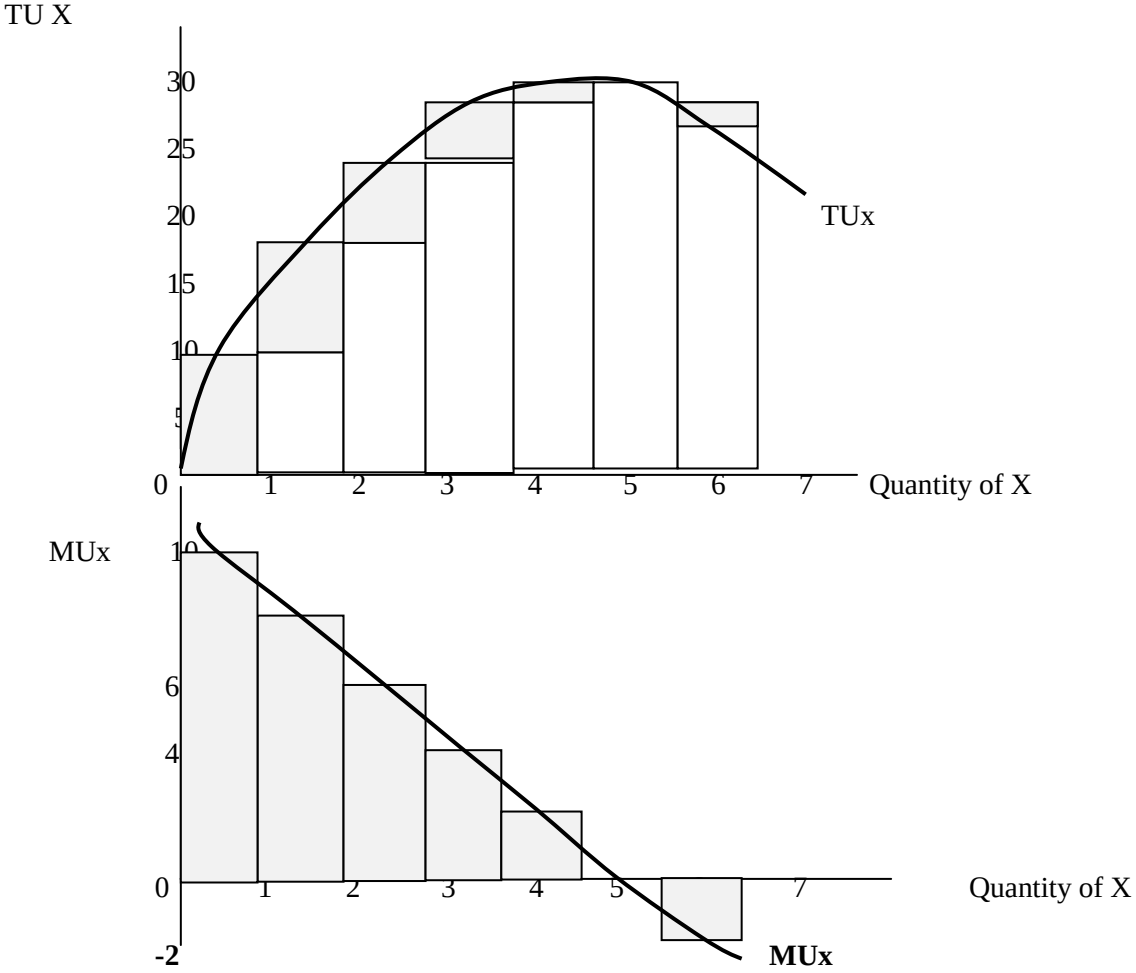
(1) Units of commodity X	(2) TU in Utils	(3) MU in Utils
0	0	--
1	10	10
2	18	8
3	24	6
4	28	4
5	30	2
6	30	0
7	28	−2

Marginal utility is the net addition to total utility resulting from one unit change in consumption or stock. It is the addition to total utility by having an additional unit of the commodity. In other words marginal utility of a commodity is the loss in utility if one unit less is consumed. When the consumer consumes the 3rd unit, the total utility becomes 24 utils. Thus the marginal utility of the 3rd unit is 6 utils (24−18). Algebraically, the marginal utility (MU) of n^{th} unit of a commodity is the total utility of n units minus the total utility of n−1units.

Thus **MUn = TUn − TU(n−1)**

The relation between total and Marginal Utility is shown in the table-1. The first two columns of the table give an individual's hypothetical total utility (TU) schedule from consuming various alternative quantities of commodity X per unit of time. As shown in the table as the individual consumes more units of commodity X per unit of time, total utility increases. Column (1) and (3) of the table show the individual's marginal utility (MU) schedule for commodity X. The value in column (3) is obtained from column (2), ie., the difference between successive values. For example, if the individual's consumption of commodity X goes from zero units to 1 unit, the TU goes from zero utils to 10 utils. Then the marginal utility is 10 utils (MUx =10 – 0). Similarly if the consumption of commodity rises from 1 unit to 2 units, total utility rises from 10 to 18, giving a marginal utility of 8 (18 – 10).

Figure-1



Up to a point, the more units of a commodity the individual consumes per unit of time, the greater be the total utility he receives. Though total utility increases, the extra or marginal utility received from consuming each additional unit of the commodity usually decreases. As shown in the table, so long as total utility is increasing, marginal utility is decreasing up to the 5th unit. When total utility is maximum at the 6^{th} unit, MU is zero. This is the saturation point. When consumption goes beyond this point, total utility decreases and marginal utility becomes negative (7th unit in the table). Such additional units give disutility or dissatisfaction to the consumer. So it is of no use having them. This relationship is shown in the figure: 1.

Cardinal and Ordinal Utility

In order to attain his objective of maximisation of utility, the consumer must be able to compare the utility of the various 'baskets of goods' which he can buy with his income. There are two basic approaches to the problem of comparison of utilities – the Cardinalist approach and the Ordinalist approach.

The cardinalist school postulated that utility can be measured. It means that an individual consumer attach specific values or number of utils from consuming each quantity of good or combination of goods. While some economist suggested the measurement of utility in monetary unit, others suggested its measurement in subjective units called utils. The ordinalist school postulated that utility is not measurable, but is an ordinal magnitude. As per this notion, a consumer need not know in specific units the utility of various commodities to make his choice. While ordinal utility only ranks various consumption bundles, cardinal utility provides an actual index or measure of satisfaction.

Cardinal Utility Theory

The Marshallian utility analysis is based on cardinalism. The cardinal utility theory is based on certain assumptions.

1. The consumer is rational. He aims at the maximisation of his utility subject to his income constraint.
2. Cardinal Utility: The utility is a cardinal concept which assumes utility is measurable and additive.
3. Utility is measurable in terms of money.
4. Constant Marginal Utility of Money: this assumption is necessary if monetary unit is used as the measure of utility.
5. Diminishing Marginal Utility: The marginal utility of a commodity diminishes as the consumer acquires larger quantities of it. (Utility gained from successive units of a commodity diminishes).
6. Consumer has full knowledge of the availability of commodities and their technical qualities.
7. Consumer possesses perfect knowledge of the choices of commodities open to him.
8. Consumer knows the exact prices of various commodities, and their utilities are not influenced by variation in their prices
9. There are no close substitutes

The Law of Diminishing Marginal Utility (LDMU)

Cardinal utility analysis explains two basic laws of consumption.

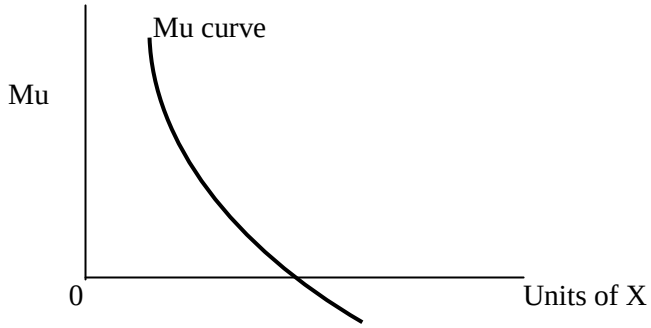
1. Law of Diminishing Marginal Utility and
2. Law of Equi-Marginal Utility

The Law of Diminishing Marginal Utility

This is one of the laws related to consumer satisfaction and refers to common experience of every consumer. This law is known as Gossen's first law (due to Javons).

Marshall states the law thus, "the additional benefit which a person derives from a given increase of his stock of a thing diminishes with every increase in stock that he already has". It means that, under certain conditions, the additional satisfaction which a person derives from successive units of a thing goes on diminishing. The additional satisfaction means the marginal utility. The basis of the law is underlined in the explanation of marginal utility concept whose diagram follows the pattern such as in figure: 2.

Figure-2



The Law of Equi-Marginal Utility

The law of equi-marginal utility (Known as Gossen's second law, law of maximum satisfaction, law of substitution etc) is another important law of consumption. Marshall defined it as: "If a person has a thing which he can put to several uses, he will distribute it among these uses in such a way that it has the same marginal utility in all" This means that, in the process of utility maximization, consumer substitutes goods with higher utility for those with lesser utility. In doing this, he relates the price of various goods with their utility and equates the marginal utility –price ratio of various goods (Thus known as proportionality rule). In the case of two commodities, the proportionality rule sets the conditions of consumer's equilibrium as $\frac{MU_x}{P_x} = \frac{MU_y}{P_y}$

In the actual market money is taken as a composite commodity. So the action of the consumer is to balance the marginal utility of a particular commodity with that of money. Thus the above condition will become $\frac{MU_x}{P_x} = \frac{MU_y}{P_y} = MU_{Money}$.

Consumer's Equilibrium

Consumer will be in equilibrium when he attains a position of maximum satisfaction and would have no further incentive to make any change in his purchase.

Equilibrium with one commodity

The Law of Diminishing Marginal Utility tells us the position of a consumer's equilibrium in the case of a one commodity model. In such cases, consumer will go on buying successive units till the marginal utility of the commodity becomes equal to price. If price falls, he will buy more and MU will come down to the level of price. On the other hand, if price rises, the consumer will buy less and MU goes up to the level of price. Thus equilibrium is defined as the equality of price and marginal utility. Ie., MUx = Px (It is to be noted that in the case of a free good, the consumer will be in equilibrium when MU is zero. (Since Px = 0, equilibrium is at the point where MUx = 0).

Two Commodity Model

In the case of a two commodity model, the consumer will be in equilibrium when he equalises the marginal utility-price ratio of two goods. Ie., $\frac{MU_x}{P_x} = \frac{MU_y}{P_y}$. In general, the equilibrium condition for 'n' commodity model, is $\frac{MU_x}{P_x} = \frac{MU_y}{P_y} = \cdots = \frac{MU_n}{P_n}$ or ., $\frac{MU_x}{P_x} = \frac{MU_y}{P_y} = \cdots = \frac{MU_n}{P_n} = MU_{Money}$

Demand

Meaning of Demand: Demand is the desire supported by purchasing power and willingness to pay. The term demand for a commodity (ie., quantity demanded) has always a reference to 'a price', 'a period of time' and 'a place'. Thus, demand means the various quantities of a given commodity or service which consumers would buy in one market in a given period of time at various prices.

Demand for a commodity ($Q^{\uparrow}d$) depends on many factors like price of the commodity (P_x) income of the consumer (Y), prices of related goods (P_r), taste and preference (t), etc. In functional notation, it is expressed as: $Q \square_x^d$ **= f (P_x, Y, p_r, t)**

Types of Demand: Three kinds of demand may be distinguished

(a) **Price demand**: Price demand shows the relationship between price and quantity demanded of a commodity. It refers to various quantities of a commodity demanded at its various prices. (ie., relationship between Q and $P_{\downarrow}x$)

(b) **Income demand**: Income demand refers to the quantities of a commodity demanded at various levels of income of the consumer. (ie., relationship between Q and Y)

(c) **Cross Demand**: Cross demand refers to the change in quantity demanded of a commodity owing to a change in the price of related goods - Complements or substitutes. (Ie., relationship between Q and p_r.)

Demand Schedule and Demand Curve

A demand schedule is a list of prices and quantities. The market demand schedule shows the amounts of the commodity per period that buyers are prepared to buy at different prices.

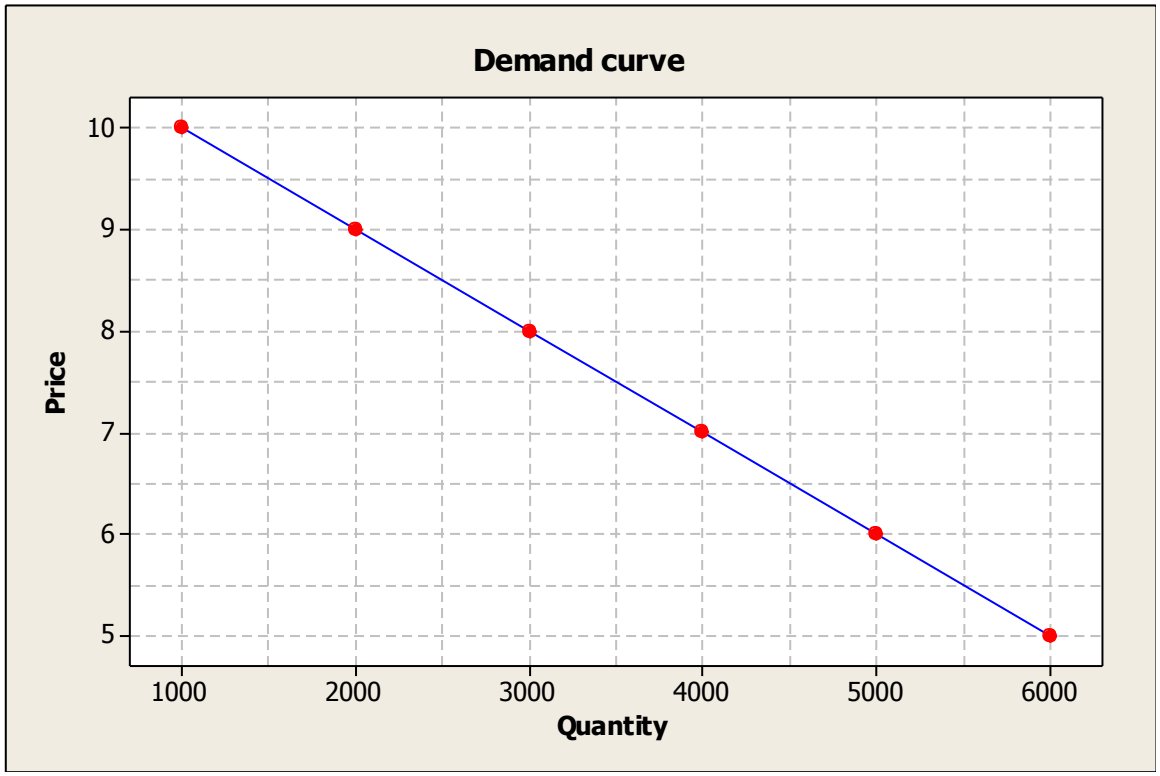
Table.2

Demand Schedule

Price (Rs)	Quantity (Units)
10	1000
9	2000
8	3000
7	4000
6	5000
5	6000

When the relationship between price and quantity demanded is plotted as a graph, we get a demand curve. In other words, demand curve is the diagrammatic representation of a demand schedule. Demand schedule and curve for a hypothetical data are given in the table-2, and figure-3.

Fig.3



Law of Demand

Other things remaining the same, the quantity demanded of a commodity varies inversely with its price. Ie., a greater quantity of the commodity is demanded at a lower price and a smaller quantity at a higher price. This inverse price-quantity relationship is called the law of demand.

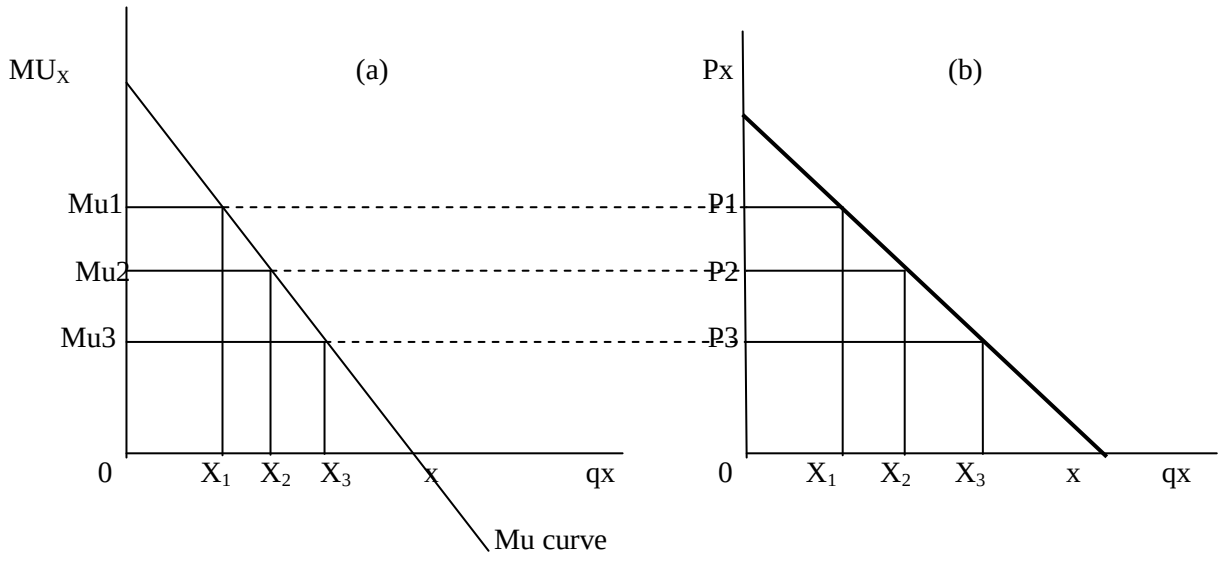
Derivation of the Demand Curve

Within the frame work of cardinal utility approach, demand curve (and the law of demand) may be derived by using either the law of diminishing marginal utility or the law of equi-marginal utility.

1. Derivation of Demand Curve from the Law of Diminishing Marginal Utility

Due to the operation of the law of diminishing marginal utility, the marginal utility curve of a good is a downward slopping curve as shown in the figure 4(a). Here the marginal utility of X declines continuously, and becomes negative beyond quantity X. Marginal utility is MU_1 for X_1 unit. When unit increases to X_2 and X_3, Marginal Utility falls to MU_2 and MU_3. Marginal Utility is zero at X^{th} unit, and negative then. The consumer attains equilibrium when he equates marginal utility with price. Thus, by definition, MU_1 is equal to P1, MU_2 is equal to P2 and MU_3 is equal to P3. It follows that, consumer purchase X1 unit at P1 (MU_1), X2 unit at P2 (MU_2) and X3 units at P3 (MU_3). Figure 4(b) represents this relationship where price is measured on the vertical axis. Thus, the positive segment of the marginal utility curve forms the demand curve (negative portion does not form the part of demand curve, since the negative quantities do not make sense in economics).

Fig. 4



Derivation of demand curve from the Law of equi-marginal Utility

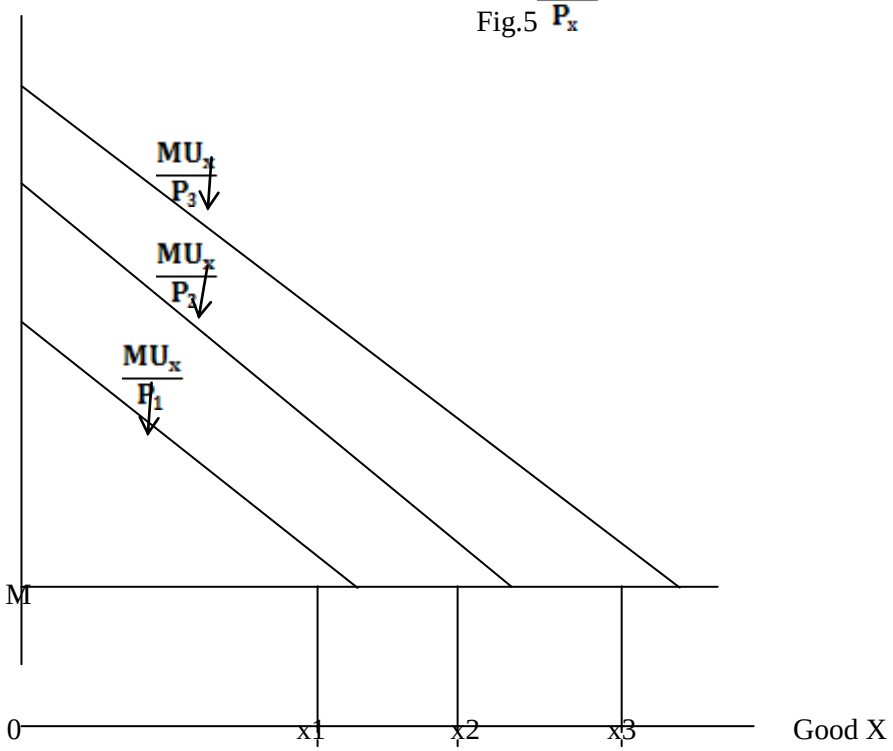
In terms of law of equi-marginal utility, the consumer's equilibrium is defined by the proportionality rule: $\frac{MU_x}{P_x} = \frac{MU_y}{P_y} = \cdots = \frac{MU_n}{P_n} = MU_{Money}$. This shows that the consumer equates the marginal utility of money (expenditure) with the ratio of marginal utility and price of each commodity he purchases. Now, suppose that the price of one good, say X, falls keeping all other determinants ($P_{r,Y}$ and t) the same. This will distort the equilibrium condition. Now, $\frac{MU_x}{P_x}$ is greater than $\frac{MU_y}{P_y}$ or MU_{Money}.

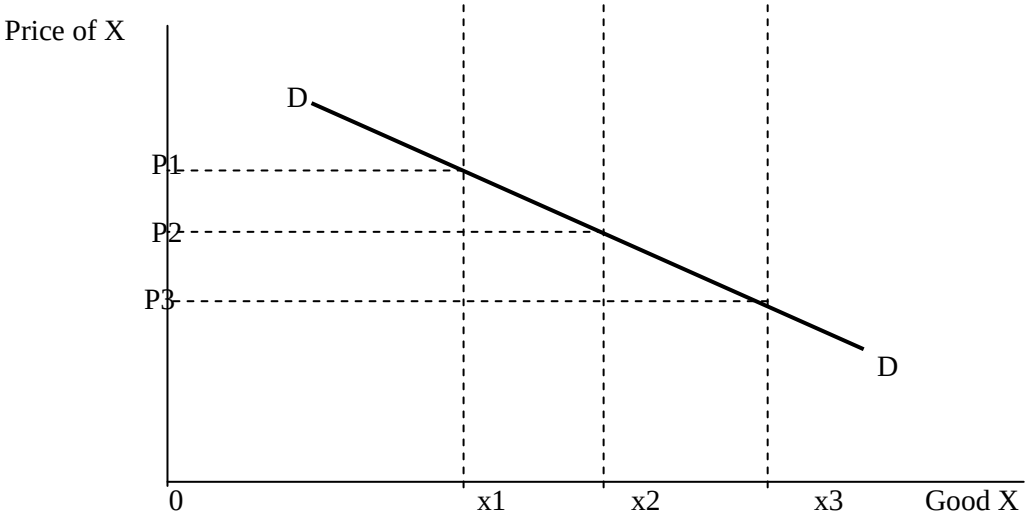
In order to restore the equality, the value of the ratio $\frac{MU_x}{P_x}$ is to be reduced to the level of MU_{Money}. The nominator value MU_x will be reduced if the consumer purchases more units of X. It follows that

when price of X falls, the equality can be retained only by purchasing more units of the commodity. It shows the inverse relation between price and quantity demanded of a commodity. The process of derivation is shown in the figure: 5.

In the figure.5, X axis measures quantities of good x demanded. The vertical axis of the upper portion shows the MU – price ratio (Ie., $\frac{MU_x}{P_x}$). The marginal utility of money is OM. When the price of good X is p1 the consumer buys ox1, where marginal utility of money (OM) is equal to Marginal Utility -price ratio $\frac{MU_x}{P_1}$. If the price of X falls from P1 to P2, the demand will increase. Ie., the demand curve is shifted upwards. The quantity demanded of X must increase to OX2 so as to make $\frac{MU_x}{P_2}$ equal to the MU_{Money} (OM). Thus the upper portion of the figure shows that when price of X falls, the demand curve will shift upwards showing an increase in the quantity demanded.

Fig.5 $\frac{MU_x}{P_x}$





In the lower portion of the figure Y axis measure various prices corresponding to various marginal utility – price ratios: When P1 is the price ($\frac{MU_x}{P_1}$), OX_1 is the quantity demanded. When price falls to P2 (from $\frac{MU_x}{P_2}$), OX_2 is the quantity and, when P3 is the price (from $\frac{MU_x}{P_3}$), OX_3 is the quantity of X demanded. The curve DD represents the demand curve indicating the inverse relation between price and quantity demanded.

Indifference Curve Analysis

The indifference curve analysis is one of the ordinalist approaches to demand analysis. It explains consumer behaviour in terms of his preferences or ranking for different combinations of two goods, say X and Y. An indifference curve is a graphic device for showing consumer's preferences. It is the locus of points of particular combinations of goods which yield the same utility to the consumer so that he is indifferent as to the particular combination he consumes. Thus, an indifference Curve represents various combinations of commodity X and Y that provide same level of satisfaction to the consumer.

The basic assumptions of the theory

1. Rationality: The consumer is assumed to be rational. He aims at the maximisation of his utility, given his income and market prices
2. Utility is ordinal: The consumer can rank his preferences according to the satisfaction of each basket. (only ordinal measurement is required for comparison purposes)
3. Diminishing Marginal rate of Substitution(DMRS): The IC is convex to the origin and shows the diminishing rate of MRS.
4. Consistency and Transitivity of Choice: It is assumed that the consumer is consistent in his choice. If in one period, the consumer chooses bundle A over B, he will not choose B over A in another period where A and B are present. This consistency assumption may be symbolically written as If A > B, then B is not > A .

   The transitivity character suggests that if bundle A is preferred to B, and B is preferred to C, then bundle A is preferred to C. Symbolically, If A > B and B > C, then A > C.
5. Completeness: The consumer's scale of preference is complete so that he can state whether he prefers one combination to the other, or is indifferent between them.
6. Non satiation : A consumer prefers more to less
7. Continuity or substitutability: An Indifference Curve is smooth and continuous which means that the two goods are highly divisible and that leads to satisfaction also change in a continuous manner.
8. There are two goods X and Y
9. The consumer possesses complete information about the prices of the goods in the market
10. The prices of the two goods are given
11. The consumer's tastes, habits and income remain the same throughout the analysis

Indifference Curve

An Indifference Curve is the locus of points of particular combinations of goods, which yield the same utility or level of satisfaction to the consumer. An indifference curve is drawn from the indifference schedule of the consumer. The latter shows the various combinations of the two commodities such that the consumer is indifferent to those combinations. An imaginary indifference schedule representing various combinations of goods X and Y is given in table:3.

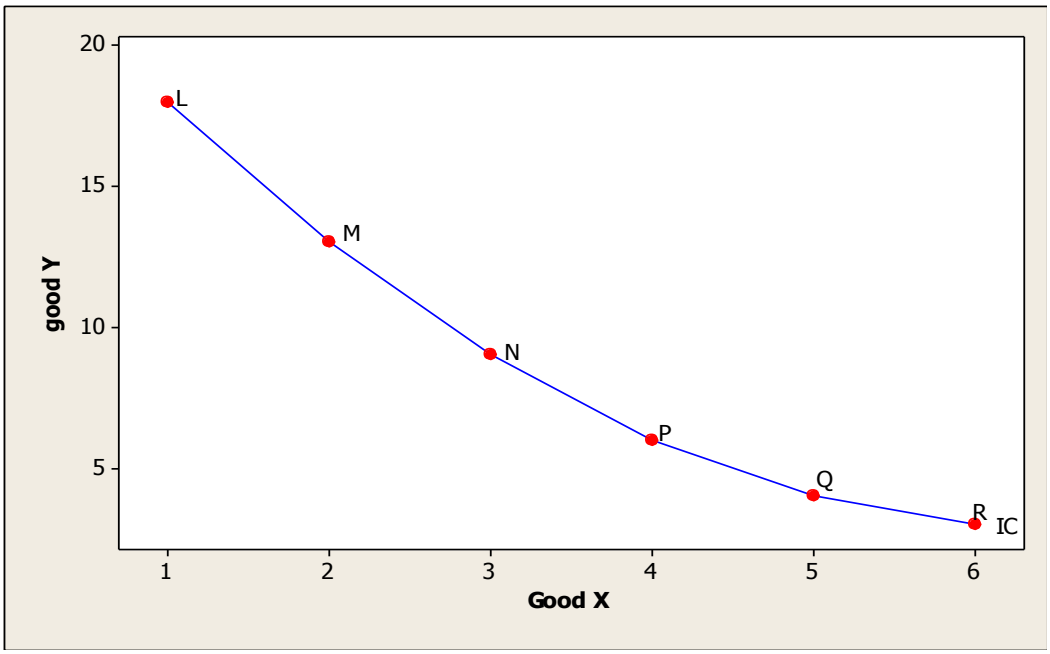
Table.3

Indifference Schedule

Combination		Good X	Good Y
1	L	1	18
2	M	2	13
3	N	3	9
4	P	4	6
5	Q	5	4
6	R	6	3

In the above schedule, the consumer is indifferent whether he buys the first combination (18 units of Y and 1 unit of X) or second combination (13 units of Y and 2 units of X) or any other combination. All combinations give him equal satisfaction. Now we shall represent this schedule by a diagram which give an indifference curve figure:6).

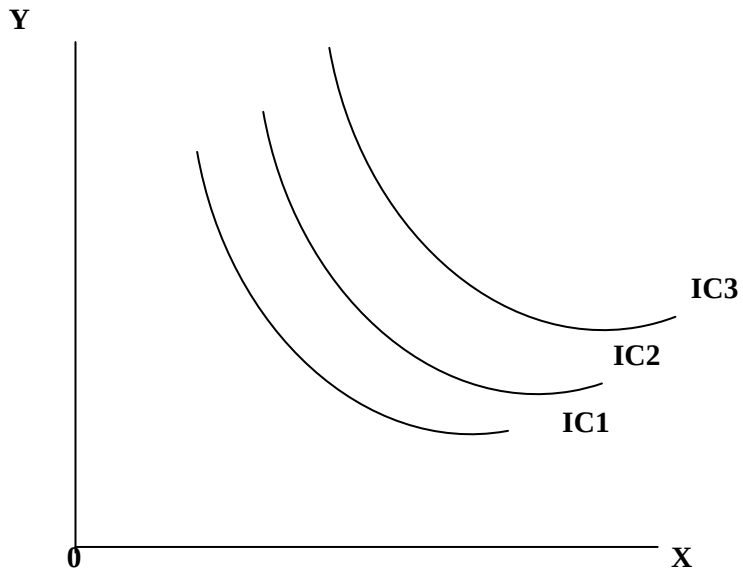
Fig.6



The Indifference Curve IC is the locus of the points L,M,N,P,Q and R showing the combinations of goods X and Y between which the consumer is indifferent. This is an Iso-utility curve showing equal satisfaction at all its points. A single indifference curve concerns only one level of satisfaction. But there are a number of ICs representing different levels of satisfaction. A diagram comprising a set of

indifference curves form an indifference map. An indifference map shows all the indifference curves which rank the preferences of the consumer. An indifference map is depicted in figure7. The curves that are farther away from the origin represent higher levels of satisfaction as they have larger quantities of X and Y. In the figure all points on IC_3 are preferred to all the points on IC_2 or IC_1. Thus IC_3 is preferred to IC_2. It indicates higher level of satisfaction than IC_2 and so on.

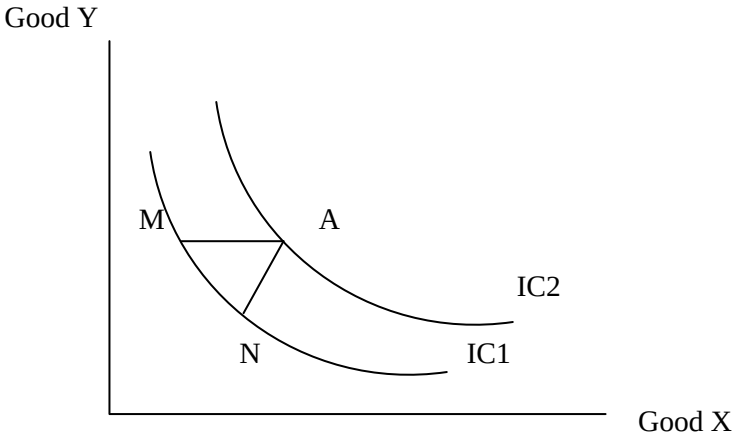
Fig.7



Properties of Indifference Curves

1. A higher IC to the right of another represents a higher level of satisfaction and preferable combinations of goods.

Fig.8

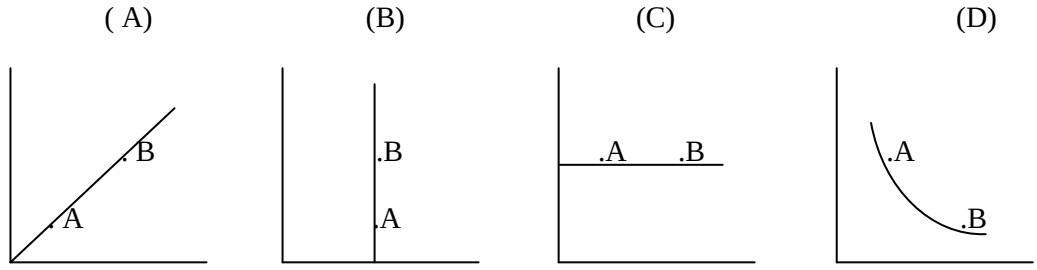


In the figure:8, IC_2 represents higher levels of satisfaction. For example, combination A on IC_2 is preferable to any combinations like M or N on IC_1.

2. Indifference Curves are negatively slopped: The indifference curves are slopping downward from left to right. It denotes that if the consumer is to stay on the same level of satisfaction, decrease in the quantity of one commodity Y must be followed by an increase in the quantity of other commodity X in the combination. In other words, to make the consumer indifferent, a decline in the utility due to a decrease in the quantity of one commodity in the combination must be compensated by an increase in the utility of the other commodity.

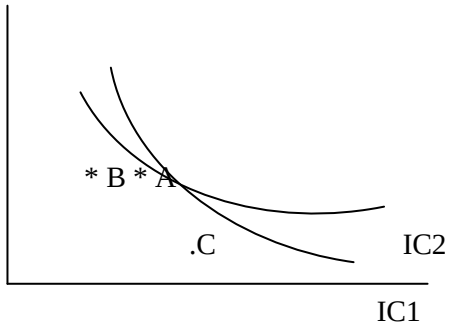
In the figures 9(a), 9(b) and 9(c), the combination B is preferred to A as it represents more quantities of at least one commodity. In the set, figure: 9(d) only represent an indifference curve since A and B can be indifferent.

Fig.9



3. Indifference curve can neither touch nor intersect each other. If they did so, the point of intersection would imply two different levels of satisfaction, which is impossible. The absurdity of the intersection can be presented in the figure.

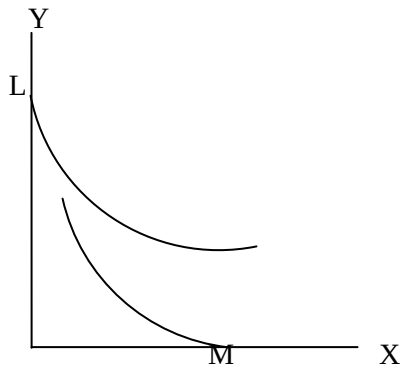
Fig.10



Point A on IC_1 indicates a higher level of satisfaction than B on IC_2 since it lies farther away from the origin. But point C which lies on both curves yield same level of satisfaction as point A and B. This is absurd since A is preferred to B. The same absurdity can be proved if two ICs are touched each other. Ie, A is preferred to B, but B = C and A = C.

4. An IC cannot touch either axis. By definition IC represents locus of points representing combinations of two goods. Touching of an IC with any axis means, it has no units of the commodity representing that axis.

Fig. 11

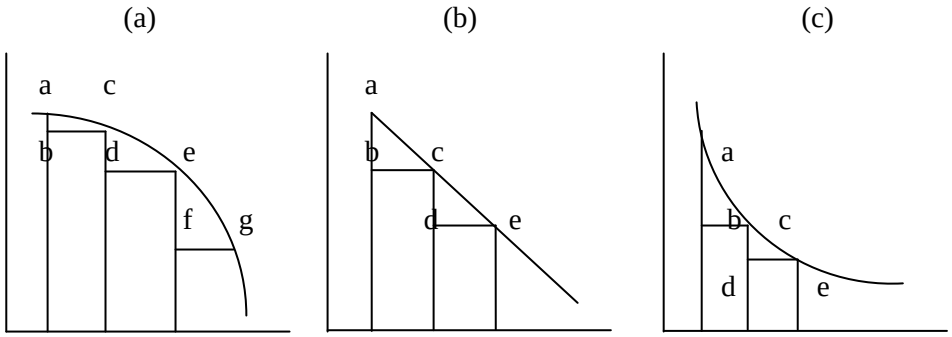


For example, at L there is no unit of X in the combination. Similarly, at M there is no unit of Y in the combination. Such curves are In contradiction to the assumption that the consumer buys two goods in combination.

5.Indifference curves are convex to the origin: The convexity rule implies that as the consumer substitutes X for Y, the marginal rate of substitution diminishes. It means that as the amount of X increases by equal amounts that of Y diminishes by smaller amounts. This implies that the slope of the curve becomes smaller as we move to the right

In a concave curve such as in the figure: 12a, MRS (Marginal Rate of Substitution) of X for Y increases. Ie., more of Y is giving up to have additional unit of X: ab < cd < ef unit of Y for bc = de = fg.

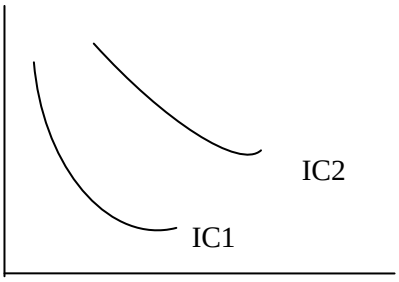
Fig.12



In the case of a straight line (Fig 12b), MRS between the two goods will be constant. Ie., ab of Y = bc of X and cd of Y = de of X. In a convex curve (Fig 12c), the consumer is giving up lesser and lesser amounts of Y in order to have equal additional units of X. Ie., ab > cd of Y for bc = de of X. Therefore an IC is always convex to the origin because the marginal rate of substitution between two goods decline.

6.Indifference curves are not necessarily parallel to each other: Though falling negatively to the right, the rate of fall will not be the same for all ICs. The slope may vary according to the difference in Marginal Rate of Substitution between the two goods.

Fig. 13



Marginal Rate of Substitution (MRS)

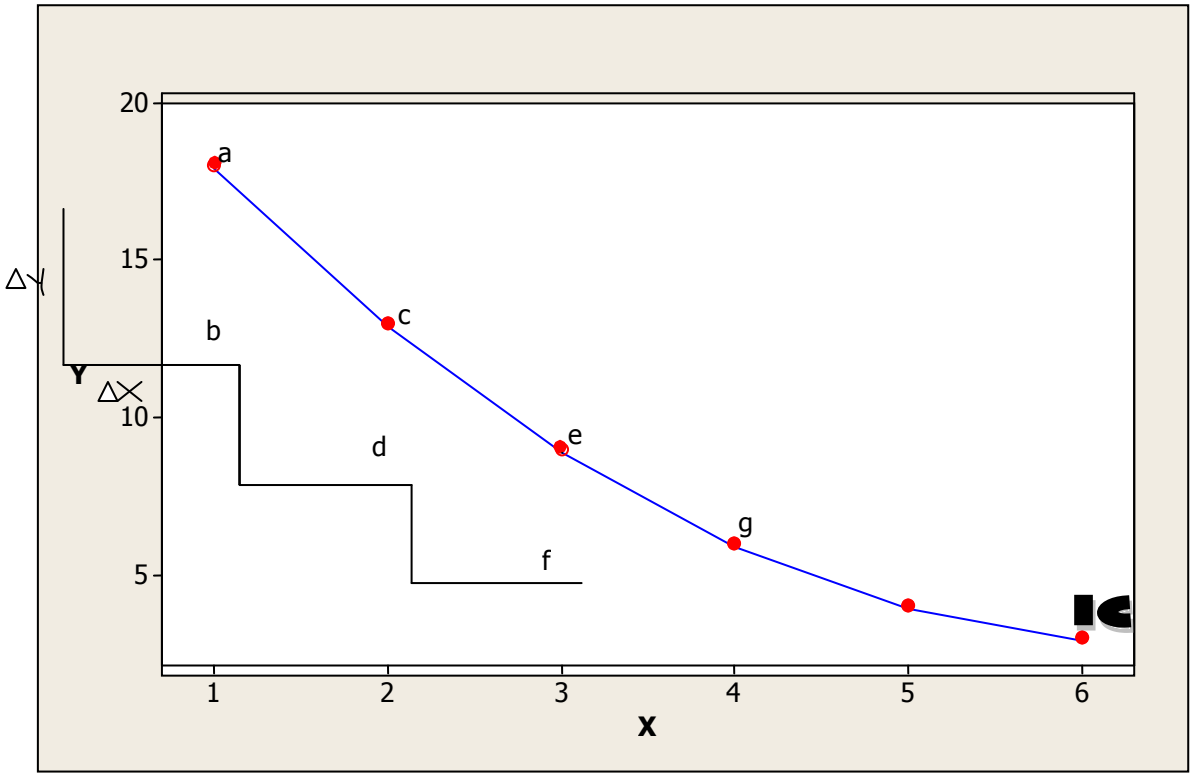
The marginal rate of substitution of X for Y is defined as the number of units of commodity Y that must be given up in exchange for an extra unit of commodity X so that the consumer maintains the same level of satisfaction. MRS is the rate of exchange between some units of goods x and Y which are equally preferred. It shows the rate at which the consumer is willing to substitute one commodity for another so as to remain on the same IC. This rate is explained in table: 4

Table.4: Marginal Rate of Substitution between Commodity X and Y

Combination	X	Y	MRSxy
1	1	18	_
2	2	13	5 : 1
3	3	9	4 : 1
4	4	6	3 : 1
5	5	4	2 : 1
6	6	3	1 : 1

In the table to have the second combination and to remain at the same level of satisfaction, the consumer is prepared to forgo 5 units of Y for obtaining an extra unit of X. At this point, the MRS of X for Y is 5: 1. But, to have the 3rd combination, the consumer is prepared to give up only 4 units of Y for an extra unit of X. Here the MRS of X for Y is 4: 1. As moving from left to right, the consumer's stock of X increases while that of Y decreases. Accordingly, the consumer's readiness to forgo units of Y becomes lesser and lesser while moving from left to right. This behaviour of willingness to forgo lesser and lesser units of Y for having additional units of X is known as the principle of Diminishing Marginal Rate of Substitution (DMRS). The principle could be cleared by the figure14.

Fig.14



The MRS of X for Y is in fact the slope of the IC. Ie., MRSxy = $^{\Delta Y}/_{\Delta X}$. It means that MRSxy is the ratio of change in good Y to a given change in good X. In the figure, the vertical sides of the triangle represent ΔY and the horizontal sides represent ΔX . Accordingly at point c MRSxy = ab/bc, at point e, MRSxy= cd/de and at point g MRSxy = ef/fg. More over it could be seen that ab > cd > ef. It follow that, as the consumer moves from left to right, he possesses additional units of X and then give up lesser and lesser units of Y for extra units of X. ie., MRSxy diminishes.

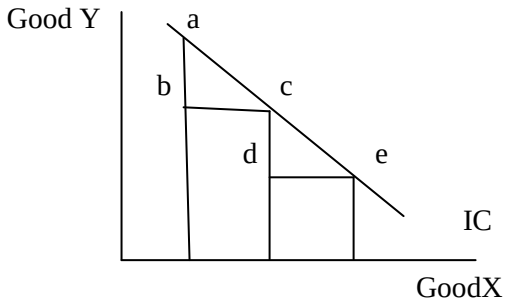
The principle of diminishing MRS is superior to the law of DMU. Prof. Hicks regards the replacement of the principle of diminishing marginal utility by the principle of diminishing marginal rate of substitution as a positive change and not a mere translation in the theory of consumer demand. The principle of diminishing MRS is based on the ordinal measurement of utility and considers commodities in combination. Thus it is scientific and realistic compared to the Marshallian utility analysis which measures utility cardinally in a single commodity analysis.

Some Special Types of IC

Indifference curves are usually negatively slopped and convex to the origin. But there may be exceptions to the general principle of DMRS. This may lead to the special types of ICs having straight, concave or L shaped (right angled)

1. **Straight line IC :** If MRS between two commodities is constant, the indifference curve will be a straight line slopping
   downwards to the right at 45^0 angle to either axis. In figure: 15,
   ΔY is the same throughout (ab = cd)corresponding to same ΔX, ie., bc = de)

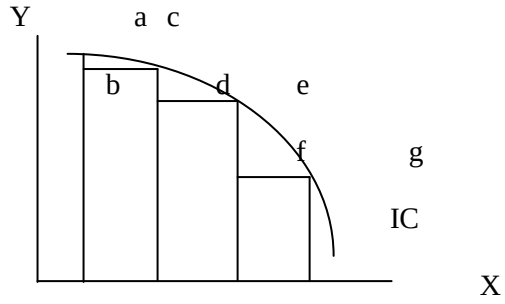
Fig.15



2. Concave IC

If MRS between the commodities is increasing, the indifference curve will be concave to the origin. In the figure, Δ Y : ab < cd < ef and ΔX : bc = de = fg

Fig.16



3. L Shaped IC

In the case of perfect complementary, the MRS between two goods is zero. The indifference curves for such goods are L shaped (Figure: 17A). In the case of ordinary complementary, the rate of substitution is low on or near the curvature of the curve as shown in figure: 17B.

Fig.17 (A)

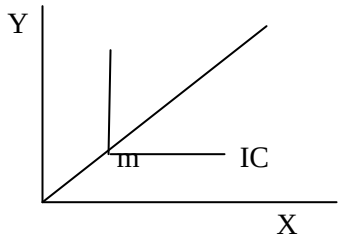
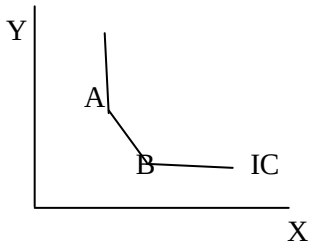


Fig.17 (B)

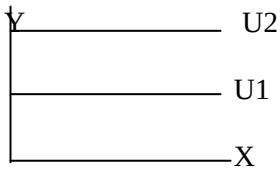


4. Horizontal or Vertical IC

If one of the goods in combination is a neuter, ie., consumer is indifferent between having more or less of the commodity, the indifference curve may be either horizontal or vertical.

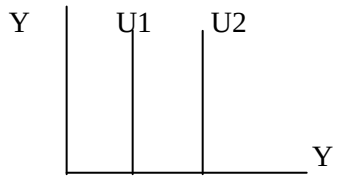
Horizontal indifference curve would indicate that good X is a neuter (Figure: 18A), Whereas Vertical IC would indicate that commodity Y is a neuter (Figure: 18B)

Fig.18A



Good X is neuter

Fig.18B



Good Y is neuter

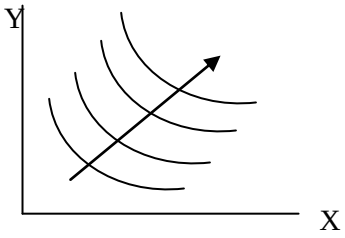
The Consumer's Income and Price Constraints

A rational consumer always tries to maximize his satisfaction. Ie., a consumer always tries to reach the highest IC (as denoted by the arrow in the fig.19). But there are constraints or limitations faced by a consumer in satisfying his or her wants. Two constraints are important in this regard.

1. The consumer has to pay price for the commodity and
2. The consumer has a given income to spend on goods.

Thus, the amount of goods that a consumer can purchase over a given period of time is limited by the consumer's income and the price of the goods. It implies that the consumer faces a budget constraint due to his or her limited income and the given prices of goods. This constraint is represented by the tool of Budget Line.

Fig.19

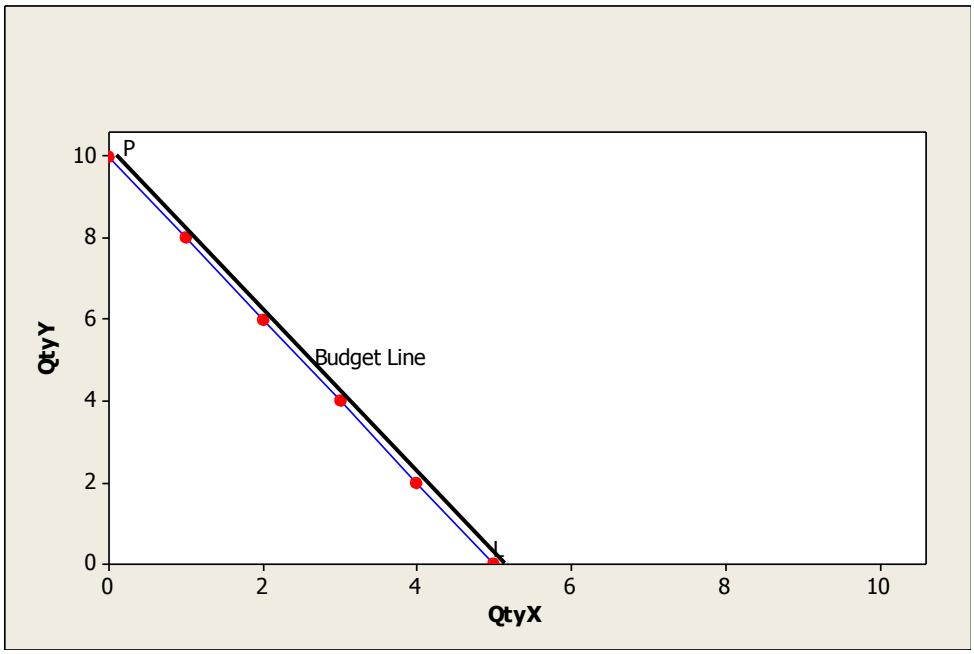


Budget Line

Consider two commodities X and Y. Let the price of X be Rs.2 per unit and the price of Y Rs.1. Also let the consumer's budget be Rs.10 at a time period. Then 5 unit of X could be bought by the consumer if he spent the whole budget on X, or 10 units of Y could be bought if the consumer spent the Rs.10 for Y. Now a straight line between these two points 5X and 10Y shows every possibility of spending the budget on the two commodities at their given prices.

This straight line between 5X and 10Y is the budget line (also known by various names: budget restraint, consumption possibility line, expenditure line, price line, price opportunity line and price income line).

Fig.20



The consumer can buy 5X and 0Y, 4X and 2Y or 3X and 4Y and so on. (An illustration of the some possible combinations on which Rs. 10 can be allocated is shown in the table 5.

Table.5: Feasible combinations of Goods X and Y with a budget of Rs.10.

Combination	Good X (Units)	Good Y (Units)
A	5	0
B	4	2
C	3	4
D	2	6
E	1	8
F	0	10

The price line PL shows the combinations of goods X and Y, given their prices, when he spends his income totally on them. The budget equilibrium of the line is represented algebraically as $I = P_x X + P_y Y$ .Where I represent income, P_x is the priceof goods X and P_y is the price of good Y.

The budget equation is the equation of the line connecting the points P and L where $P = \frac{I}{P_y} and\ L = \frac{I}{P_x}$

The consumer can buy any quantity inside the triangle. But the purchase of any quantity within the budget line means the consumer is not spending all of the income Rs.10. (In the diagram prices and the budget are represented indirectly by the physical quantities). The slope of the budget line is the ratio of the prices of the two commodities. The ratio is Px/Py, the price of X divided by the price of Y. Thus for any budget line, Slope = $\frac{QtyY}{QtyX} = \frac{\frac{I}{P_y}}{\frac{I}{P_x}} = \frac{Px}{Py}$. In our example, the slope = 2/1 = 2.

If both prices were equal, the slope of the budget line would be unity. If the slope is less than unity (Px/Py < 1), x has the lower price. If the slope is higher than unity (Px/Py >1), Y has the lower price.

For many purpose it is convenient to let the horizontal axis represent amounts of a commodity and the vertical axis an amount of money income per time period. Then X is a commodity and Y is money income. In this case Y stands for other commodities (represented by money income).

By assuming that a consumer spends all of his incomes on goods X and Y, we have the budget equation $P_xX + P_yY = I$ ..... **(1)**

By rearranging, $P_yY = I - P_xX$ .... **(2)**

Dividing both sides of equation (2) by P_y, $Y = \frac{I}{P_y} - \frac{Px}{Py}X$ **....(3)**

The first term on the right hand side of equation (3) is the vertical or Y-intercept of the budget line and $-\frac{Px}{Py}$ is the slope of the budget line. In our example, the slope of the budget line refers to the rate at which the two goods can be exchanged for one another in the market. Ie., 2Y for 1X.

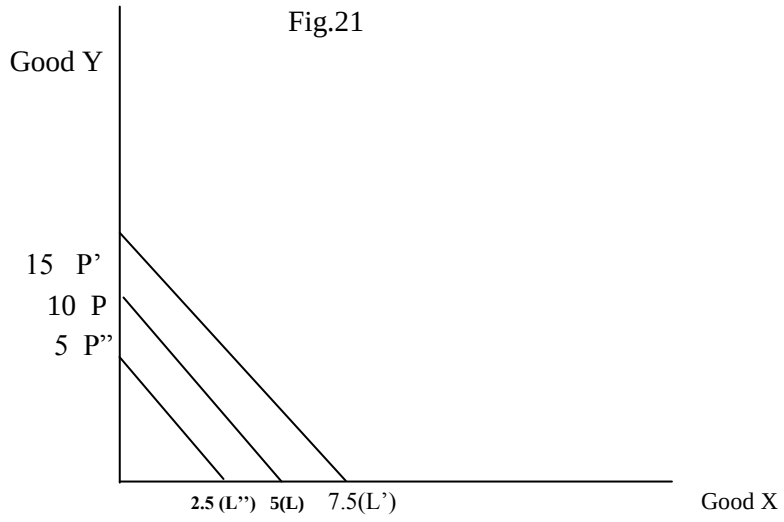
Changes in Income, prices and Budget Line

From the foregone discussion, it is clear that the position of the budget line depends on the size of the budget as well as the relative prices. A particular budget line refers to a specific level of the consumer's income and specific prices of the two goods. Thus, any change in income or price of either goods cause a change in the budget line. The consequence of such changes may be analysed as under.

Change in Income

When only the consumer's income changes, the effect will be a parallel shift in the budget line. Here, the slope of the budget line remains unchanged. When income increases, budget line shift upwards parallel to the right. Converse to this, budget line shifts downwards when income decreases.

Fig.21



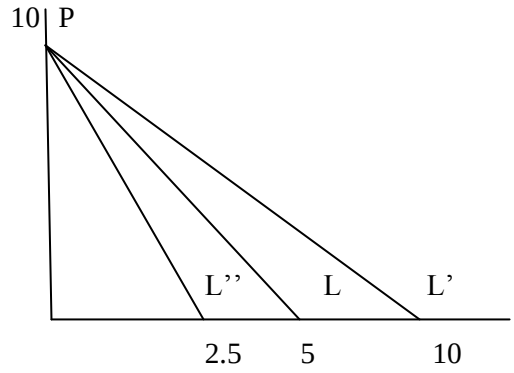
In the figure, Budget line PL denotes 10Yor 5X when income is Rs.10. When income increase to Rs.15 the consumer can purchase 15Y or 7.5X. Following a decrease in income, the consumer can purchase either 5Y or 2.5X at an income of Rs.5. Thus the three budget lines corresponding to three levels of income are parallel and their slopes are equal.

Changes in prices of good X

If only the price of good X changes, the vertical or Y intercept remains unchanged. If price of X falls, the budget line rotates upward or counter clockwise. Converse to this, the budget line rotates downwards

or clockwise if price of X rises. In the figure: 22, PL is the budget line where, Px is Rs.2 and income is Rs.10. When price of X decrease to Rs.1, the budget line rotates upwards as PL'. But the budget line will rotates downward to PL'' when the price of X increases. Here, though the vertical intercept remains the same, the horizontal intercept and the slope of the budget line changes.

Fig.22

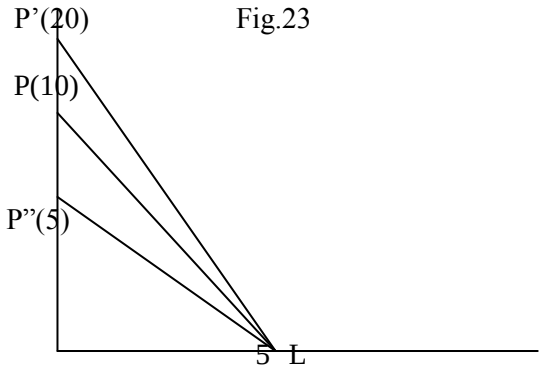


For the original budget line (PL), the slope is $-\frac{Px}{Py} = -\frac{2}{1}$. When Price of X falls to Rs.1 slope is -1/1 and when Price of X increases to Rs.4, slope is -4/1.

Changes in Price of good Y

In the case of a change in the price of Y alone, the horizontal or X intercept will be the same. At the same time budget line will rotate upwards if Price of y falls and downwards if Price of y rises.

Fig.23



In figure: 23, PL is the original budget line where Price of y is Rs.1 per unit. When price of Y decreases to Rs. 0.5, the new vertical intercept is QY=20 (P') and the slope of budget line is $-\frac{Px}{Py} = -\frac{2}{.5} = -4.$ **When** Price of y = 2, the new Y intercept is QY=5 (P") and the slope is -2/2= -1 .

Consumers Choice

A rational consumer tries to maximise utility or satisfaction. Thus the consumer's problem is to choose the most preferred affordable bundle ('best' bundle) among the alternatives open to him.

In this context, our attempt is to examine how the consumer determines which goods to purchase and in what quantities to achieve this end. Now we have all the tools required to analyse the consumer's choice problem. The budget line shows the income and price constraints faced by the consumer. The indifference curve represents the tastes and preference of the consumer (preference ordering).

Utility Maximization

A consumer is in equilibrium when he maximises his satisfaction from given income and given prices of goods. A rational consumer maximises utility by trying to attain the highest indifference curve possible, given his income and prices of goods (budget line). This occurs where an indifference curve is tangent to the budget line. At the point of tangency, the slope of indifference curve (the MRSxy) and the slope of the budget line (Px/Py) are equal. The point of tangency satisfies three conditions for utility maximisation (the conditions for the constrained utility maximisation, consumer optimisation or consumer equilibrium).

1. The consumer spends all his income. Ie., $I = P_x X + P_y Y$. This means that the consumer is on the budget line, (2) The slope of indifference curve and the slope of budget line should be the same. Ie., MRSxy = Px/Py and (3) Indifference curve should be convex to the origin.

The utility maximisation is illustrated in the figure: 24, where consumer indifference curves and budget line are brought together.

Fig.24

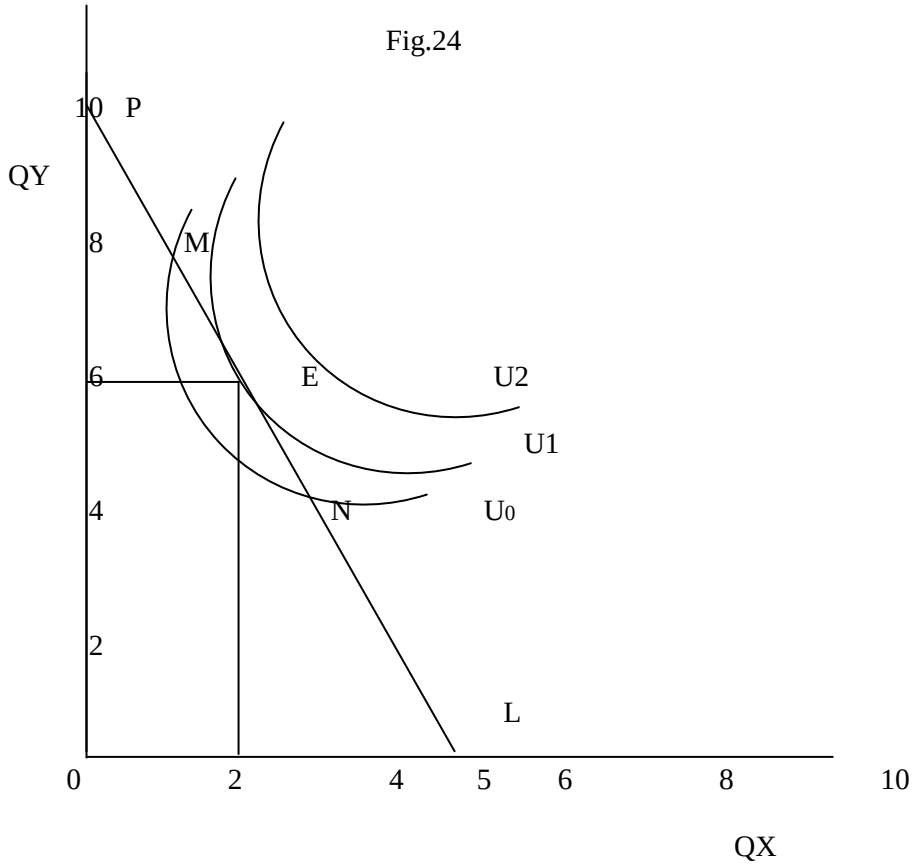


Figure shows that the consumer maximises utility at point E where indifference curveU1 is tangent to budget line PL. At point E, consumer is on the budget line (thus satisfies first condition of given income) and slopes of the budget line and indifference cure are equal, MRSxy = Px/Py, (satisfied second condition).

Indifference curve U1 is the highest that the consumer can reach with his budget line. Thus to maximise utility, the consumer should spend Rs.4 to purchase 2 units of X and the remaining Rs.6 to purchase 6 units of Y. Any other combinations of good X and Y that the consumer could purchase (whether it is on the budget line or below the budget line) provides lesser utility than the combination E.

Consider the combination M and N. They are on the budget line PL and so satisfy the budget equation. But, M and N lies on a lower indifference curve U0 and less preferred than any combinations on U_1. At point M, MRSxy (slope of indifference curve U_0) is higher than the price ratios, Px/Py, (slopes of budget line).

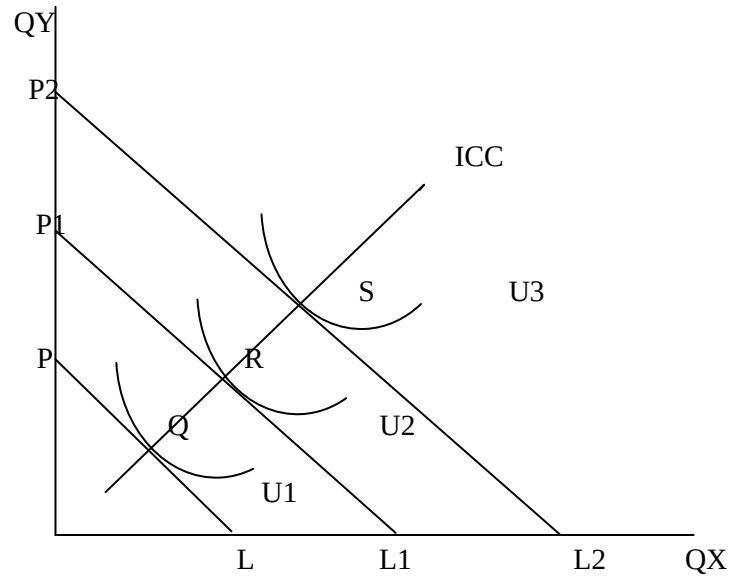
Thus the consumer can increase his satisfaction by purchasing less of Y and more of X until he reaches at point E. On the other hand, at point N, MRSxy is < Px/Py. So the consumer can increase his satisfaction by purchasing less of X and more of Y until he reaches point E. The consumer cannot reach indifference cure U_2 with the given income and prices of two goods X and Y. Thus E is the only combination that corresponds to utility maximisation or consumer equilibrium.

Changes in Income and Engels's Curve

So far we have discussed that utility maximisation point or equilibrium point is characterised by tangency of the budget line with an indifference curve. The position of this point depends on (a) the budget line and (b) preferences that determine the shape and location of indifference curves. The budget line in turn depends on prices and income of the consumer. A change in consumer's income shifts his budget line. This shift affects consumer purchases and leads to a new optimum (equilibrium) position.

If the income of the consumer alone changes, while prices and consumer's tastes remain constant, the effect it will have on his purchases is known as the income effect. If the income of the consumer increases, his budget line will shift upward to the right, parallel to the original budget line. On the contrary, a fall in the income will shift the budget line inward to the left. These lines are parallel to each other because relative prices remain unchanged.

Fig.25

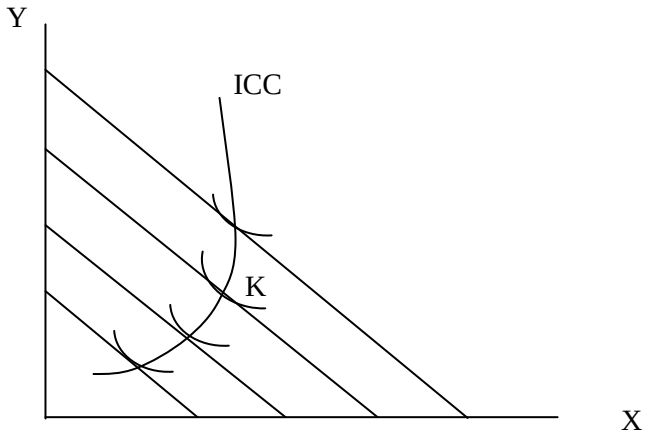


In the figure: 25, when the budget line is PL, the utility maximisation point is Q where budget line is tangent with indifference curve U1. If income of the consumer alone increases, price line will move to the right as budget line $P_1L_1$and the new utility maximisation point is R where the budget line P_1L_1 touch the indifference curve U2. As income increases further, P_2L_2 becomes the budget line and S is the utility maximisation point. By joining the optimum points Q, R and S we get the income consumption curve

for the consumer. Thus the income consumption curve is the locus of consumer optimum points resulting when only the consumer's income varies. Thus ICC shows the income effect, ie., the effect of changes in consumer's income on the purchases of the two goods. In order to give more attention on the consumer's purchase of just one good, say commodity X we may take X on as commodity X and Y as the money income (measured on the vertical axis). Accordingly the indifference curves will show the trade off between various quantities of good X and various amounts of money.

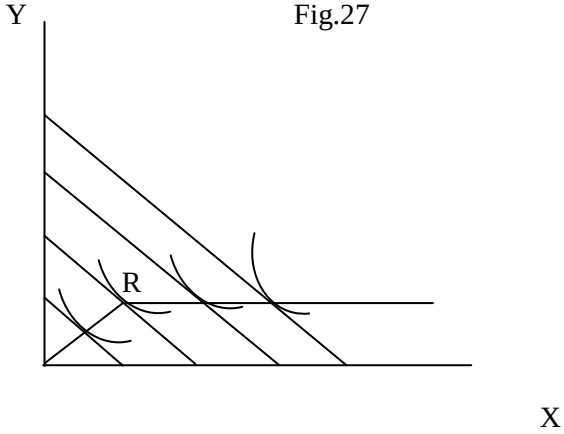
In general, the higher a consumer's income, the more will be the quantity of a commodity the person will buy. Commodities bought in larger quantities when income rises are called normal goods. Thus, in the case of normal goods, the income consumption curve slopes upward to the right (as shown in the figure: 25). The direction of ICC depends on the nature of goods (normal, inferior or giffen).

Fig.26



In figure: 26, the income consumption curve has a positive slope up to point K and curve is negatively increased beyond the point. This indicates that consumption of X increases along with increasing in income up to a certain level, but less of X is consumed after this level. This type of goods is inferior goods. In the figure: 27, ICC slopes upward up to point R and then becomes horizontal to the X axis. This signifies that R is the saturation point with respect to commodity X. This case represents a necessity.

Y Fig.27



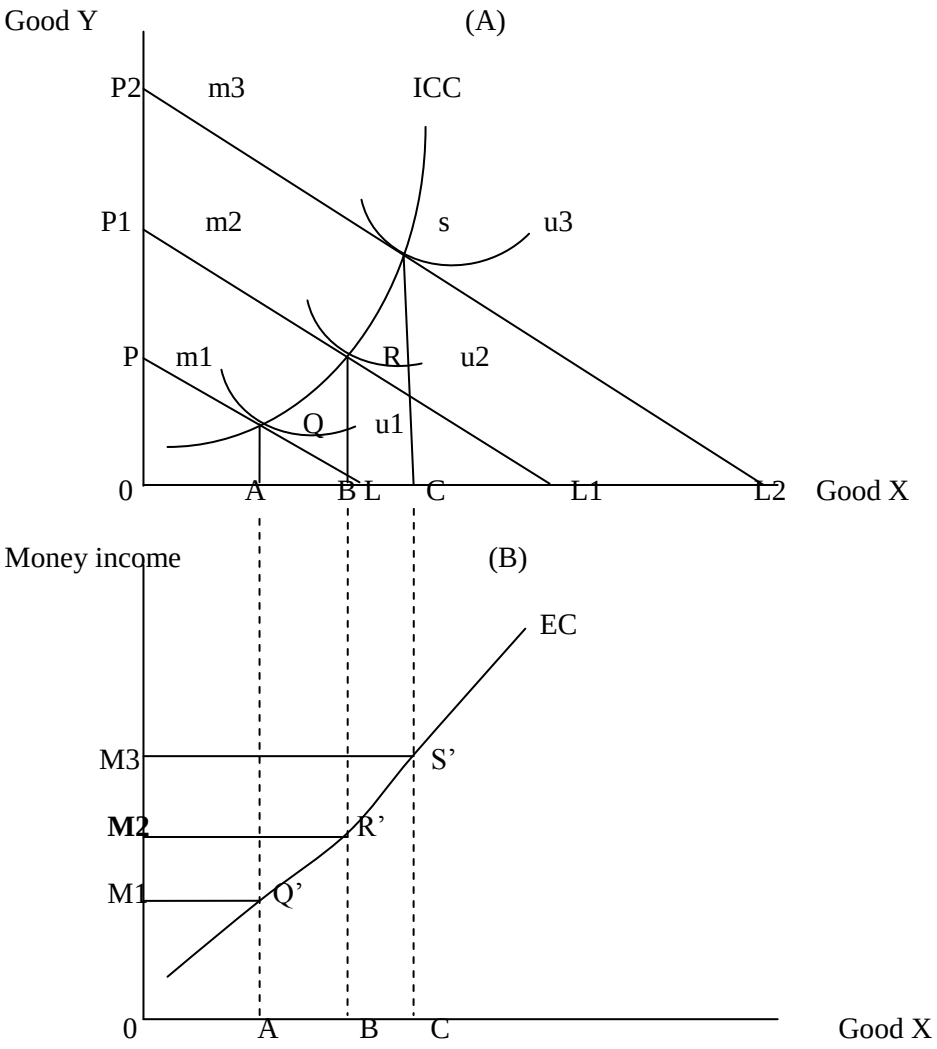
Income Consumption Curve and Engel Curve

The Engel curve shows the amount of a good that the consumer would purchase per unit of time at various consumption levels. (Curves are named after Ernest Engel, the German statistician of the nineteenth century who pioneered studies of family budgets and expenditure patterns). Sometimes, Engel

curves show the relationship between income and expenditures on various goods rather than the quantity purchased of various goods. We can derive Engel curve from income consumption curves. Now, let us derive an Engel curve for good X from the ICC of figure 28:A..

With income M1 (represented by budget line PL) consumer maximises utility at point Q in the upper panel of figure. Accordingly, he purchases OA unit of commodity X. As income increases from M1 to M2 and M3 the budget line shifts upward to P_1L_1 and P_2L_2 respectively. Following this shifts, consumer equilibrium points should shifts from Q to R and to S. The amount of commodity X that the consumer would purchase also move from points A to B and to C.

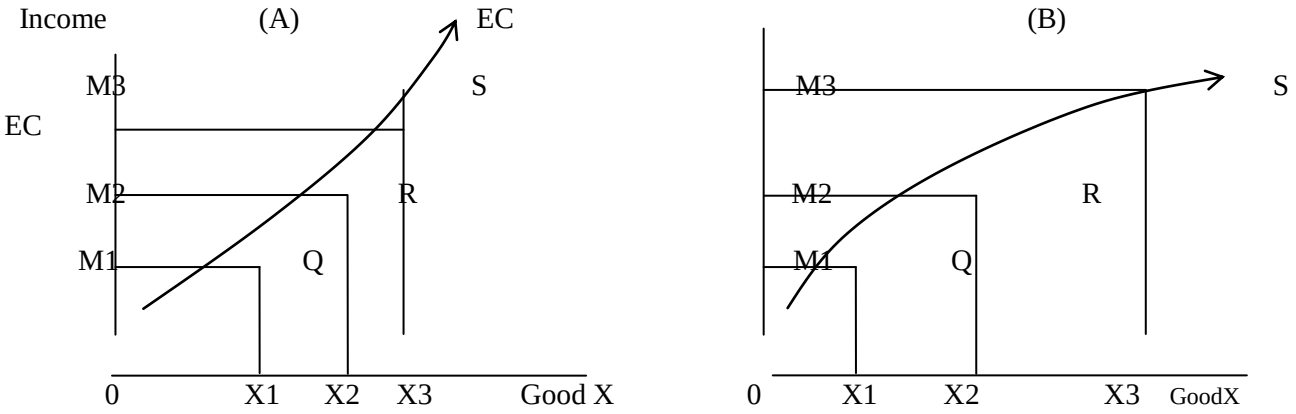
Fig.28



In panel 28B, vertical axis measures money income of the consumer. Now we transfer various combinations of income and quantities of X purchased in the lower diagram. For this, we trace utility maximisation point Q in the bottom panel as Q' representing m1 level of income and OA quantity of X. Similarly R and S are brought directly in the bottom panel as R' and S' representing increase in quantity levels of X (m2 and OB & m3 and OC respectively). By joining points Q', R' and S', we get the Engel curve in the bottom panel. Thus, the Engel curve is derived from the income consumption curve and shows the quantity of X that the consumer would purchase at various income levels. The slopes of Engel curve depends on the nature of goods. Engel curve rises rapidly in the case of necessities. The case

is shown in figure: 29A, where a given increase in income (OM1 = M1M2 = M2M3) leads to a proportionately smaller increase in the quantity purchased of X (OX1 > X1X2 > X2X3).

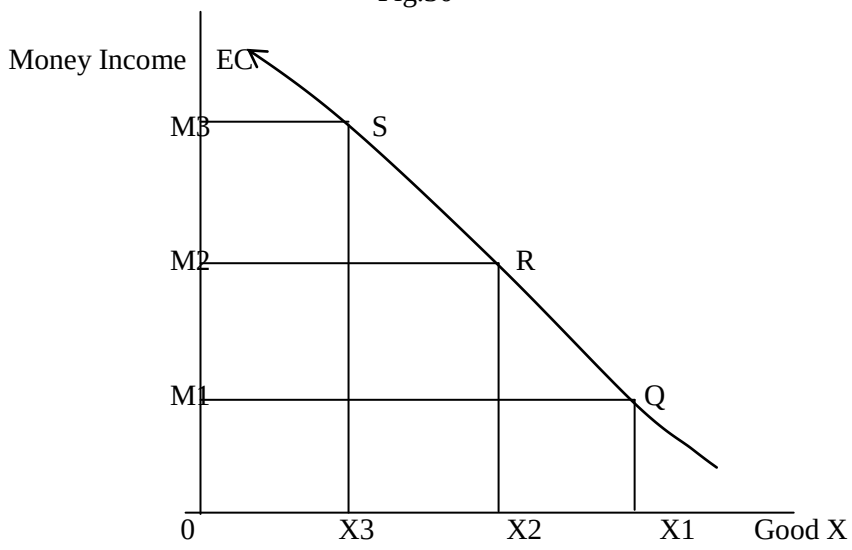
Fig.29



Engel curve rise only gently in the case of luxuries. Figure 29: B corresponds to such goods where a given increase in income ($OM_1 = M_1M_2 = M_2M_3$) leads to a proportionately larger increase in the quantity purchased of the good ($OX_1 < X_1X_2 < X_2X_3$).

Necessities and luxuries together refer to normal goods. In the case of normal goods, Engel curve is upward slopping from left to right, showing the positive relationship between income and purchase of the commodity. In the case of an inferior good, the consumer purchases less quantity of the commodity as income increases. The Engel curve for an inferior good is slopping backward from right to left. In the figure: 30, as income increases from M1 to M2 and to M3 the quantity purchased of X falls from X_1 to X_2 and to X_3. If the good is a neutral one (such as salt), the Engel Curve is a vertical line.

Fig.30



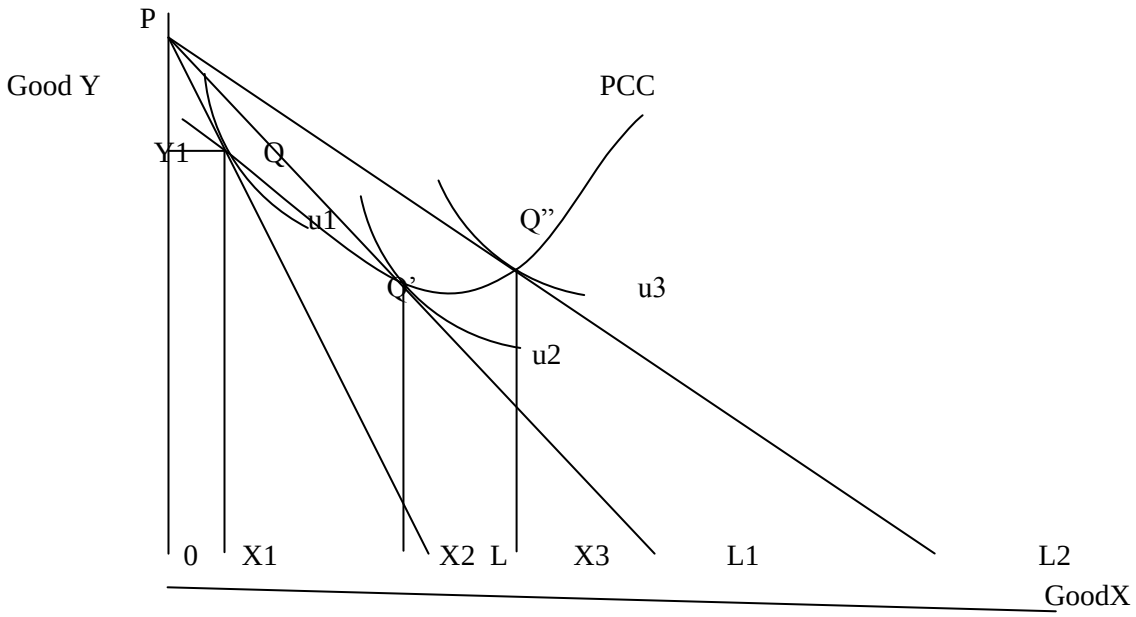
Changes in Price

A change in commodity prices changes the consumer's budget line, (as the price ratio alone changes) and this affect consumer's optimum point and the purchases.

Now, we examine how the consumer reaches a new optimum point when the price of a good changes, but the price of other good, income of the consumer and tastes remain unchanged. The effect on the purchase due to a change in the price of a commodity is the price effect and is represented by the price consumption curve.

We can derive the consumer's price consumption curve for good X by changing the price of good X while holding the price of good Y, income and tastes constant. The price-consumption curve for good X is the locus of consumer optimum points resulting when the price of good X only varies. This is shown in the figure: 31.

Fig.31



In the figure, with budget line PL consumer is in optimum at point Q, where indifference curve u1 is tangent to the budget line PL and the consumer purchase X_1 units of X and Y_1 units of Y. Now, suppose the price of good X falls from the initial level. This reduction would cause the consumer's budget line to become flatter or to extend further out to the right as PL1 showing that the consumer would purchase more units of X than before as X has become cheaper. Now Q' will be the equilibrium. The budget line PL2 shows a further fall in the price of X. The new optimum point would be the tangency point Q". By joining three optimum points Q, Q', and Q", we get the price consumption curve for this consumer.

In our illustration, at income Rs.10 and price of Y Rs.1 variation in price of X leads to variation in the quantity. For eg. If price of X = Rs.2, optimum is 2X + 6Y (2 **X** 2 + 6 **X** 1 = 10). Again, if price of X falls to Rs.1 optimum is 6X + 4Y (6 **X** 1 + 4 **X** 1 = 10). If price of X falls to Rs.0.5 optimum is 10X + 5Y (10 **X** 0.5 + 5 **X** 1 = 10)

Substitution Effect and Income Effect

As we have seen, the fall in the price of good X alone (given the price of Y) increases its demand. This is the price effect which is the result of two separate forces at work called the substitution effect and income effect

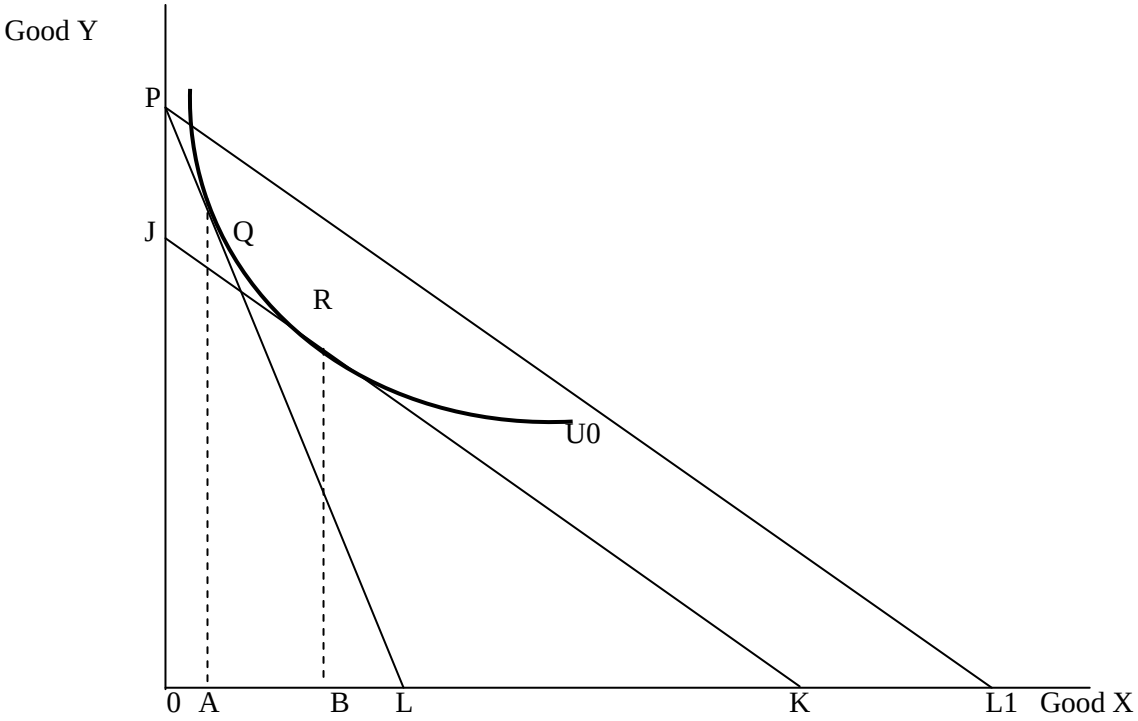
Substitution effect relates to the change in the quantity demand of a good when its price changes resulting only from the relative price change and independent of the change in income. When the price of good X falls, X becomes cheaper relative to other commodities and consumers tend to substitute X for other commodities. This always tends to increase in the quantity demanded of good X.

The income effect relates to the change in the quantity purchased of good resulting only from the change in real income that accompanies a price change. There are two approaches to the measurement of the substitution effect, one by Hicks and the other by Slutsky.

Hick's Substitution Effect

Prof. Hicks explanation of substitution effect is in terms of compensating variation method. In Hicksian version, "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". Thus, the increase in the real income, following a reduction in the price is withdrawn so as to leave the consumer neither better off nor worse off than before. The substitution effect is explained in the figure32. PL is the original budget line which is tangent to the indifference curve U0. Q is the equilibrium point where the consumer purchases OA unit of X and AQ unit of Y. Suppose the price of X falls so that the new budget line is PL1. With the fall in the price of X the real income of the consumer increases.

Fig.32



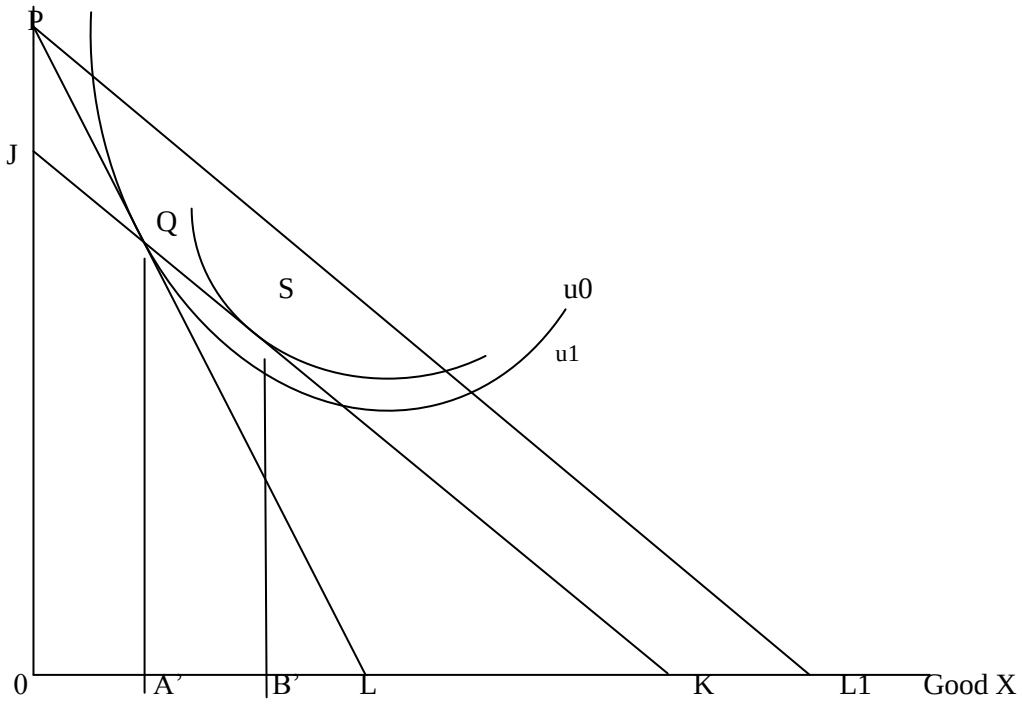
To make the compensating variation in income or to keep the consumer's real income constant, take away the increase in his income so as to bring him back to the original level of satisfaction (indifference curve U0). In terms of the figure, this is equal to PJ or KL1 of income. Then the new budget line after adjusting the income is JK which is tangent with the original indifference curve at R (not at Q). Now the

consumer buys OB unit of X and BR unit of Y. Now, as a result of fall in the price of X, consumer substitute X for Y and moves from point Q to R (ie, A to B). This movement is called the substitution effect. The substitution effect is always negative and the price quantity relation is inverse.

Slutsky's Substitution Effect

Slutsky explained the substitution effect by taking the apparent real income of the consumer as constant. In this case the adjustment in real income is made so as to leave the consumer on the same bundle of the two goods as before the price change. The consumer's movement to a higher indifference curve after this adjustment represent a substitution effect. Hicks calls this method as the cost difference method. The method is shown in the figure33.

Good Y Fig.33



PL is the original budget line and Q is the optimum point where OA' of X is purchased. Now PL1 is the new budget line following a price reduction. The increase in real income is taken away by adjustment to leave him on the same bundle Q. Thus the budget line PL1 is shifted parallely to JK which passes through point Q. Now the consumer would not be in equilibrium at point Q since the indifference curve is not tangent with the budget line JK. He will move to point S on a higher indifference curve u1, where JK is tangent to it. Since both line PL and JK have the same purchasing power, the difference between their equilibrium position Q and S is due to the substitution effect. Thus an increase in the quantity of X represented by a movement from A' to B' is the Slutsky's Substitution Effect.

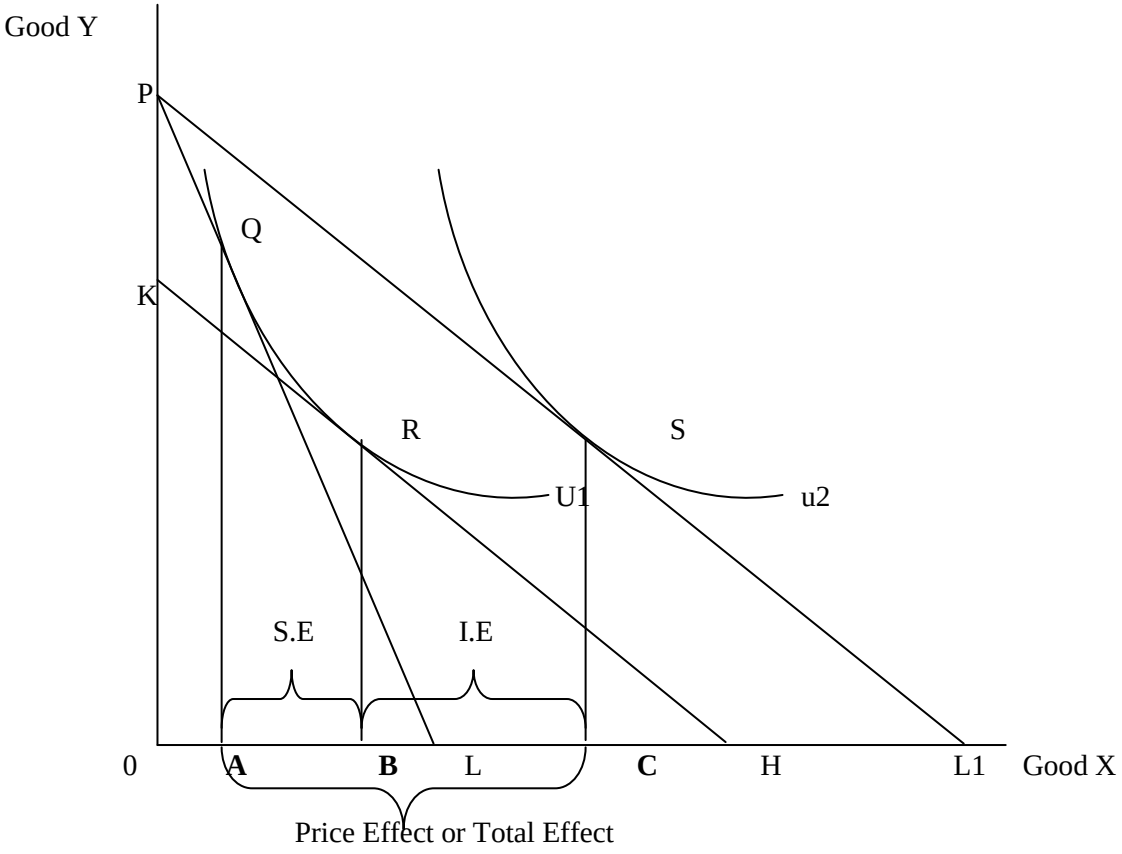
The Income Effect

The income effect relates to the change in the quantity purchased of a good resulting only from the change in income of the consumer. A fall in the price of good X represents an increase in the income of the consumer, because the consumer can now reach the same indifference curve with a smaller income. The extra income thus earns can be allocated between this good and other commodities.

Separation of Price Effect in to Substitution and Income Effects

The total effect of a price change (Price effect) can be separated in to two components as substitution effect and income effect. The process of decomposition is explained in the figure34.

Fig.34



PL is the original budget line where the consumer maximises utility at point Q. Now, if the price of X falls, the budget line takes new position as PL_1. This decrease in the price of X causes an increase in the real income of the consumer. Now, to bring the consumer back to the original level of satisfaction (indifference curve U_1), a compensating variation in income is to be made. This is done by taking a portion of money income equivalent to PK in terms of Y or L_1H in terms of X. KL is the new budget line after this compensating variation. Now the consumer maximises utility at the combination represented by point R which is tangent with the budget line U_1. Now the shift from point Q to R on the indifference curve U_1 (movement from A to B on the horizontal axis) is due to the substitution effect.

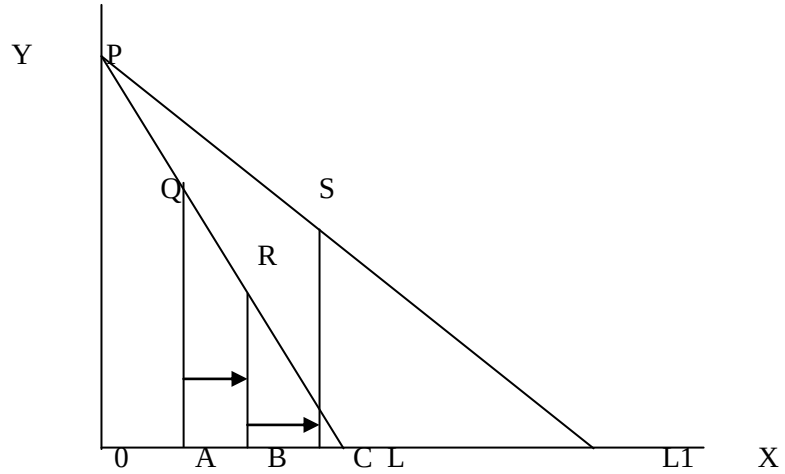
Now, if the income, taken as compensating variation is given back to the consumer, he could purchase more of X and Y and could attain a higher level of satisfaction. So an increase in purchase following this increase in (real) income is the income effect. In the figure, movement from R to S (on a higher indifference curve U_2) is due to the income effect (movement from B to C on the horizontal axis). Thus the total effect of a price change (price effect) is the sum of two effects (Substitution Effect and Income Effect). In the figure, AC = AB + BC. Or movement from Q to S = Q to R + R to S. Ie., Price Effect = Substitution Effect + Income Effect.

Income and Substitution Effect for different types of goods

1. Normal Goods

Normal goods are those goods whose purchases increase with increase in income. In the case of such goods, income effect works in the same direction to that of substitution effect. Thus income effect strengthens substitution effect and the total effect leads to an increase in quantity following a fall in price. The case is shown in the figure35. Movement from Q to R is due to substitution effect and the movement from R to S is due to the income effect. Thus Price Effect is AC that is equal to AB (Substitution Effect) + BC (Income Effect).

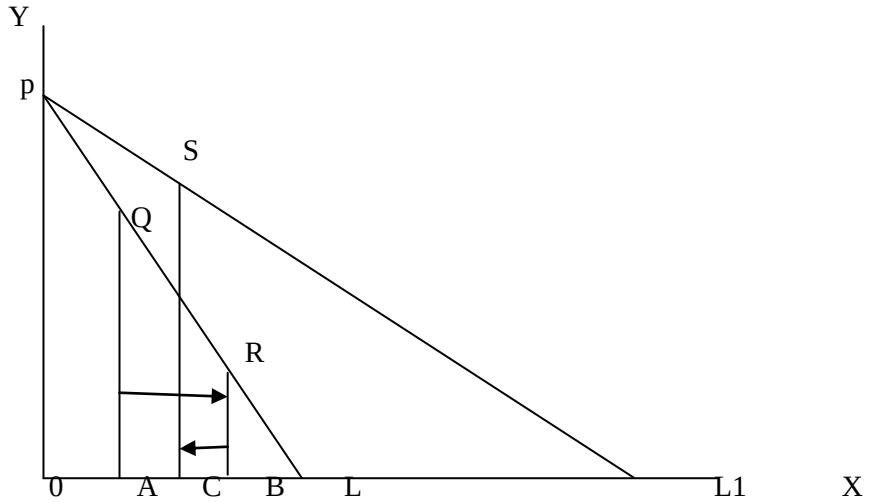
Fig.35



2. Inferior Goods

In the case of inferior goods, income effect works in the direction opposite to that of substitution effect. But the strength of income effect is not enough to cancel out substitution effect totally. As a result, when price of X falls, more units of X will be bought. In figure36, movement from A to B is substitution effect and B to C is income effect. Here, AB > BC. The demand law (inverse relationship between quantity demanded and price) operates in such goods.

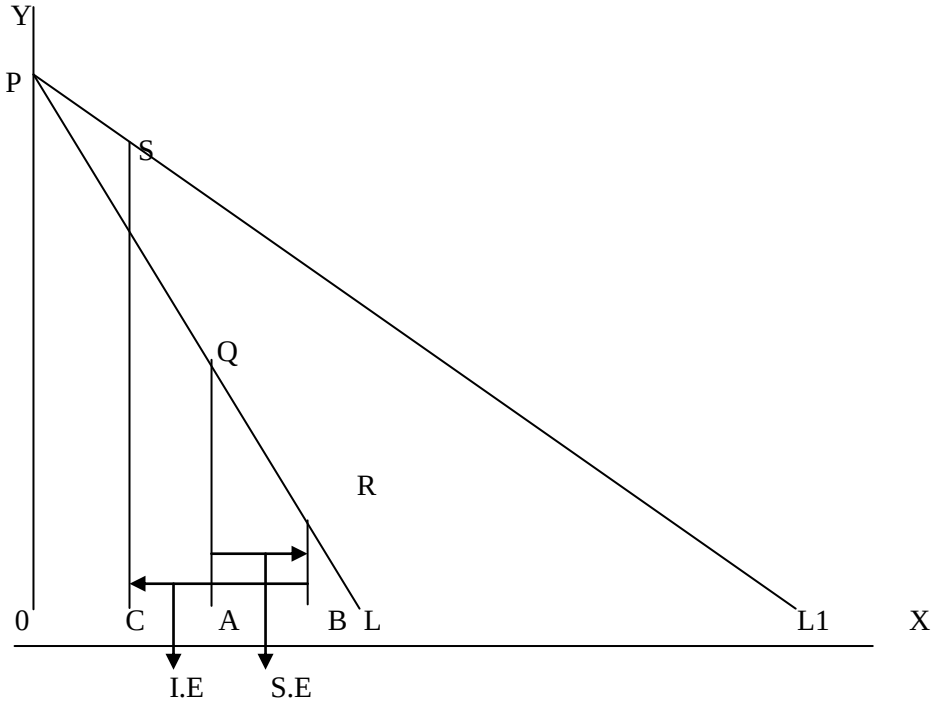
Fig. 36



3. Giffen Goods

Giffen goods are special types of inferior goods. In such goods income effect works opposite to substitution effect and is higher than it.

Fig.37



As shown in the the figure: 37, when price falls substitution effect causes an increase in the quantity purchased (movement from A to B). But the income effect causes a reverse movement and the quantity that would purchase decline with increase in real income (movement from B to C). Thus, the net effect to price change is a decrease in quantity demanded of X following a fall in the price of it. This case depicts the violation of demand law.

Derivation of Demand Curve Using Indifference Curve approach

In the case of indifference curve analysis (ordinal utility approach), the demand curve is derived from the price consumption curve. A price consumption curve represents the points of tangency of successive budget lines and higher indifference curves. The process of derivation of demand curve is illustrated with an example.

Suppose a consumer has a given income of Rs.10 (OP in the figure: 38) and the price of commodity X is falling from 5 to 2 and to 1. In the upper portion of the figure, money income is taken on the vertical axis and good x on the horizontal axis. PL, PL_1 and PL_2 are the budget lines on which R,S and T are the equilibrium position forming the Price consumption Curve, PCC. The consumer buys OA,OB and OC units of X respectively at these points on the PCC. If the total money income of the consumer is divided by the number of goods to be bought with it, we get per unit price of the good.

For OA units of X, he pays OP/OL price, for OB units OP/OL_1: and for OC units OP/OL_2 . This is the consumer's price demand schedule for good X which is shown in the table: 06.

Table 6

Price Demand Schedule for good X

Budget Line	**Price of X (Total Money income/ No of units of X**	**Quantity of X Demanded**	**Combination**
PL	OP/OL = (10 /5 = Rs.5)	OA = 1 unit	R
PL_1	OP/OL_1 (10/5 = Rs.2)	OB = 4 units	S
PL_2	OP/OL_2 (10/10 = Rs. 1)	OC = 7 units	T

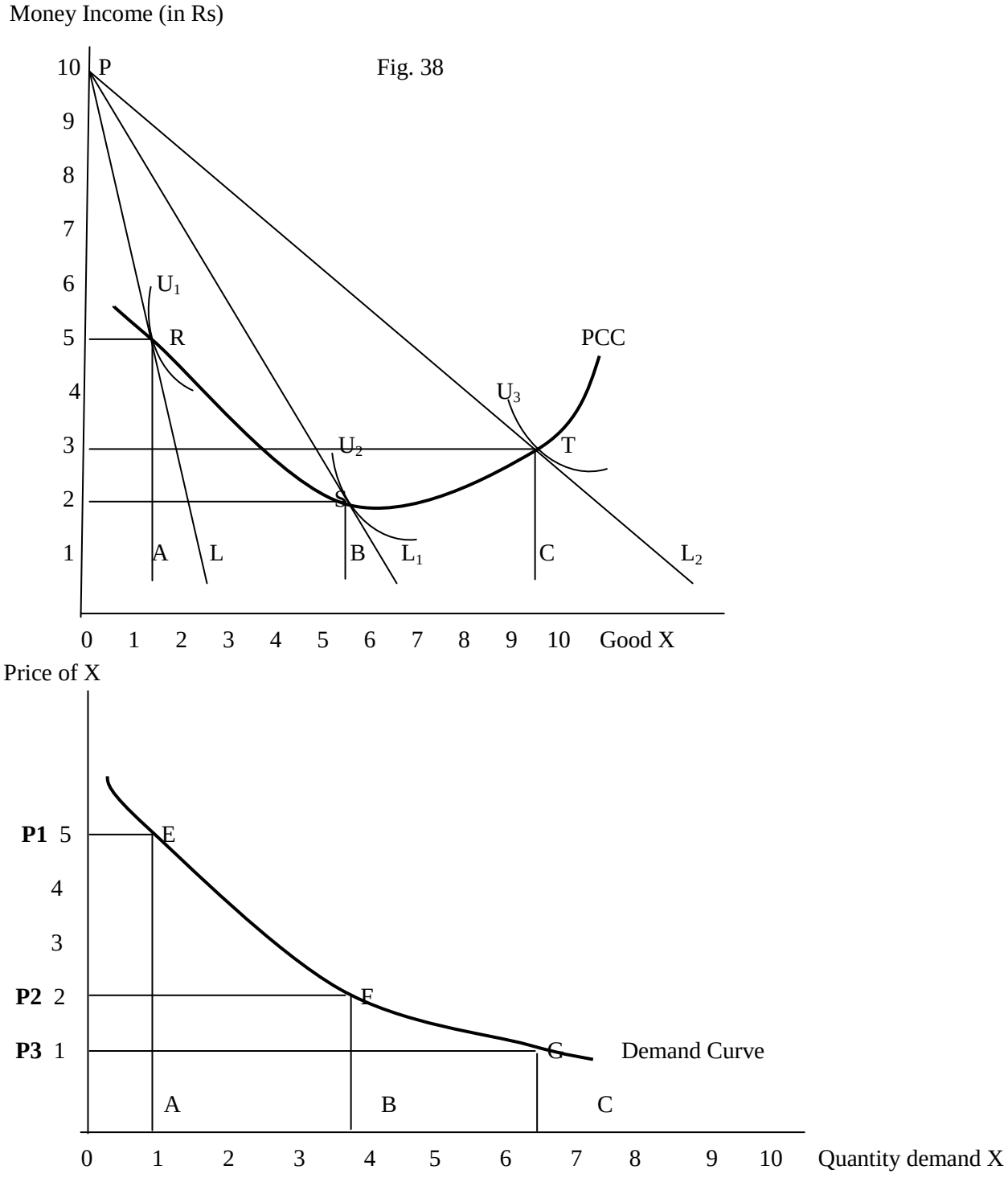


Fig. 38

The combination of price and units of commodity given in the table are represented by points R, S and T in the upper part of the figure 38. These points in the PCC are plotted on the lower portion of the diagram where prices of X are taken on the vertical axis and units of X on the horizontal axis. Point E is got by drawing perpendicular from point R through point A. This represent OA unit (1 unit) at Rs.5.(OP_1). Similarly, draw perpendiculars from S and T through B and C for having F and G on the lower figure. These points are joined by a curve which forms the demand curve. This curve (demand curve) shows the inverse relationship between price and quantity demanded.

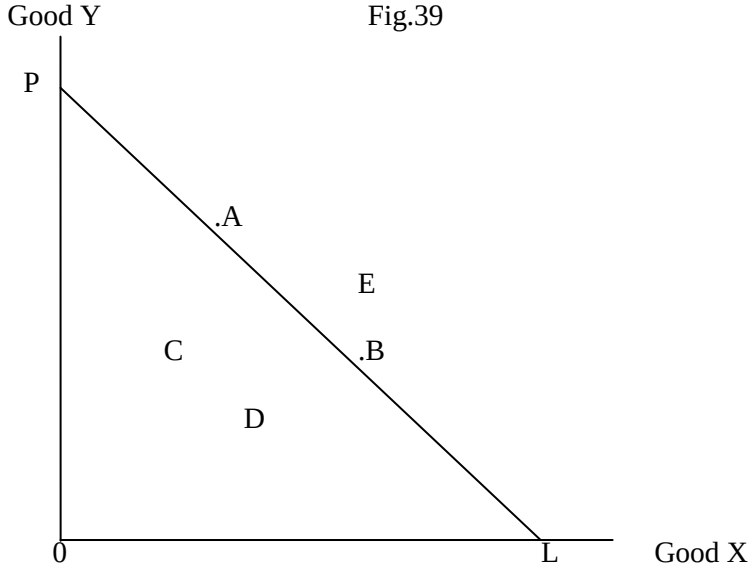
Revealed Preference Theory

Indifference Curve analysis shifted the approach from cardinalism to ordinalism. At the same time, it retained the assumption of introspection. Prof. Samuelson attempted to shift the basis from introspection to observation. Thus revealed preference theory is a behaviourist ordinal utility analysis. The revealed preference hypothesis has made possible the establishment of the 'law of demand' directly on the basis of the revealed preference axiom. Thus this hypothesis (R.P hypothesis) is considered as a major breakthrough in the theory of demand.

Revealed Preference Axiom

Prof. Samuelson's theory of demand is based on revealed preference axiom which states that choice reveals preference. According to this hypothesis, the consumer is supposed to reveal the nature of his preferences. He shows the good he would prefer to purchase in a given situation even though he may not be able to show his scale of preference on an indifference map.

Fig.39

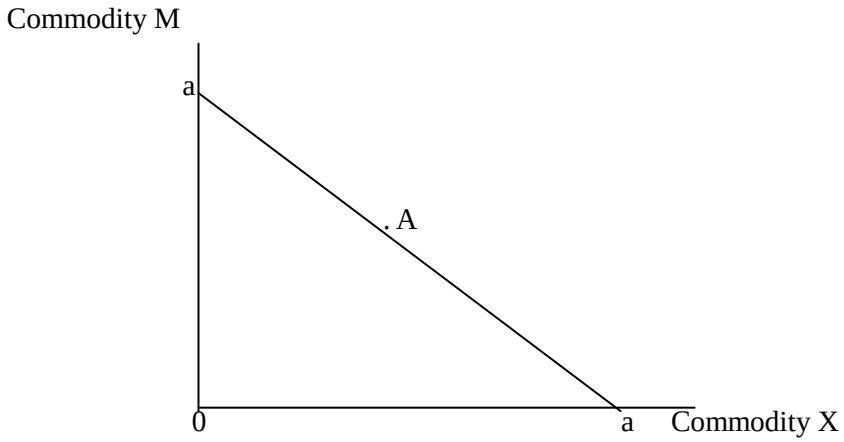


In the figure 39, with the price line PL the consumer has many alternative bundles defined by the area of triangle POL. If the consumer chooses point A, this reveals consumer's definite preference for this combination over all other combinations available (on the budget line as well as within the budget line)

Strong and Weak Ordering

Strong ordering is a distinguishing feature of Samuelson's theory. In a strong ordering, each item in a consumer's scheme of purchase is assigned a definite place or number and at each number there is only one item so that the consumer definitely reveals his preferences. In the revealed preference case, while a consumer is choosing a point among alternatives he shows his definite preference over all other points. This is a case of strong ordering. Hicks explains the strong ordering diagrammatically. He assumed 2 commodities X and M, where X is an individual good and M is a composite good representing all goods and services other than good X.

Fig.40



Given the income of the consumer and the prices of goods, the price income situation is represented by budget line aa and the choices open to the consumer are shown by points in or on the triangle a0a. The point A on the line aa represents the actual choice of the consumer. This choice shows that the consumer has a definite preference for A over all other points in or on the triangle a0a (letters of the alphabet are strongly ordered).

Weak Ordering

In weak ordering, there may be some items which cannot be arranged in order or preference. In such cases, some items may be incapable of being arranged in front of one another. A weak ordering consists of division in to groups where ordering is not possible within the group. But sequence of groups may be strongly ordered. For eg., ordering of people on the basis of birth days without regard to years.

In the indifference curve analysis, consumer preference is based on weak ordering. The different combinations on the same indifference curve are equally desirable.

Fundamental theorem of Consumption theory

Samuelson has tried to demonstrate the inverse relationship between price and the quantity demanded of a commodity using Revealed Preference hypothesis. He states the demand theorem under the title "Fundamental Theorem of Consumption Theory". For the explanation of the theory, he made certain assumptions.

1. Rationality: The consumer is assumed to behave rationally. He prefers the bundle of goods having more quantities of the commodities. The consumer seeks to maximise his satisfaction from the resources he has.

2. Consistency: The consumer behaves consistently. If he chooses combination A in a situation in which combination B was also available to him, he will not choose B in any other situation in which A is also available.

   Symbolically, If **A > B**, then **B is not greater than A**. This is the two term consistency.

3. Transitivity: If, in any particular situation, A is preferred to B and B is preferred to C, then the consumer must prefer A to C. Symbolically, **if A > B and B > C, then A > C.** This is the three term consistency.

4. The income elasticity of demand is positive: The consumer demands more commodities when income increases and less when income decreases.

5. Consumer's Choice reveals his preference for the chosen combination among the alternative combinations.

6. Strong Ordering: Strong ordering is a distinguishing feature of Samuelson's theory.

The Theory

Based on the assumptions, Samuelson stated his fundamental theorem of consumption theory (also known as demand theorem), thus: "any good (simple of composite) that is known always to increase in demand when income alone rises must definitely shrink in demand when its price alone rises". It means that when income elasticity of demand is positive, price elasticity of demand is negative. This can be shown for rise as well as for a fall in the price of a good.

Rise in Price

Consider a consumer who spends his entire income on 2 goods X and Y. Now, given the income and prices of goods, consumer's budget line is PL. Suppose that the consumer chooses point Q where he he spends all income on X and Y.

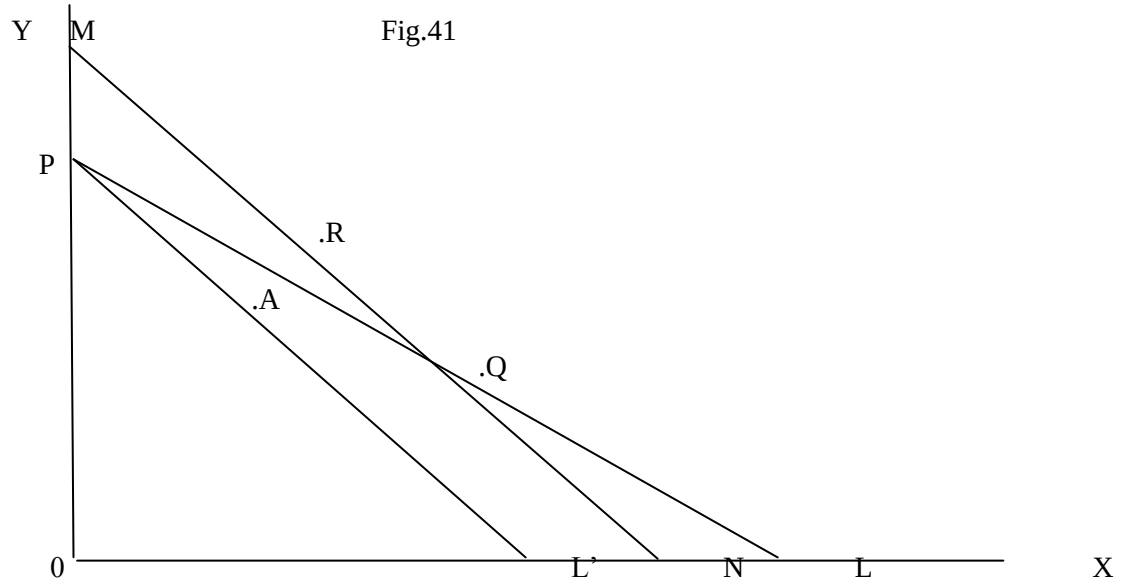


Fig.41

If the price of X rises, the consumer could purchase less units of X at given prices. With rise in price, real income decreases and the budget line will change as PL'. Now, given the triangle of choice OPL', the equilibrium Q is beyond the budget line and the consumer will choose a combination like A which contain less units of X (since price of X has risen).

In order to compensate the loss in real income as a result of rise in the price of X let us give L'N amount of money to enable him to buy original combination Q. Then MN becomes new budget line which is parallel to PL' and passing through point Q. Prof. Samuelson calls this extra money 'over compensation'.

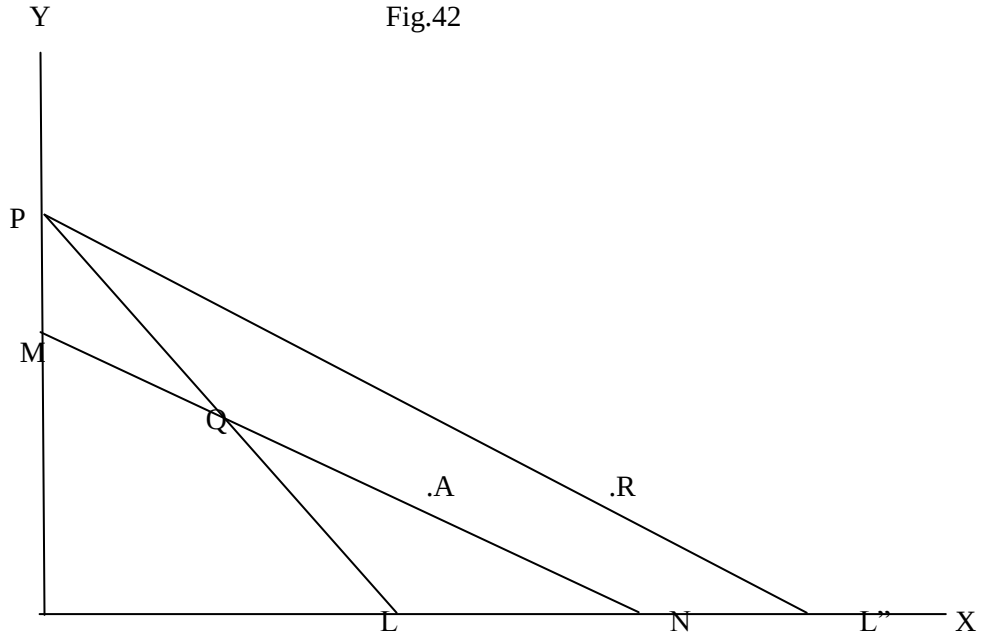
The new area of choice is OMN. Since point Q was preferred in the original budget line PL, the consumer will not choose any points lying below Q on the QN segment of budget line MN (represent lesser satisfaction and its choice will prove inconsistent behaviour. He cannot have more of X when its price has risen). The consumer will, therefore, reject all combinations below Q and choose either Q or any higher combination (such as R) lying on MQ part of the price line MN. If he selects Q, it will mean that he is buying the same amount of goods X and Y as before price rise. On the other hand, if he chooses any combination above Q on the MQ portion of MN (such as R), it will mean that he is buying less of X and more of Y. This represents a substitution effect of a price rise, (since some units of Y have been substituted for some units of X which has become dearer).

If the extra money once given in the form of compensatory variation is taken back, the consumer's choice combination will be to the left of Q (say point A) on the budget line PL', showing a reduction in the purchase of X when there is a reduction in income of the consumer (since income elasticity of demand X is positive). Thus, the theory establishes the inverse relationship between price and the quantity demanded when price of good X has risen (when income elasticity is positive, price elasticity is negative).

Fall in price

The demand theorem can also be proved in the case of a fall in price. It can be defined thus: "Any good (simple of composite) that is known always to decrease in demand when money income alone falls must definitely expand in demand when its price alone falls". The case is explained in figure 42:

Fig.42



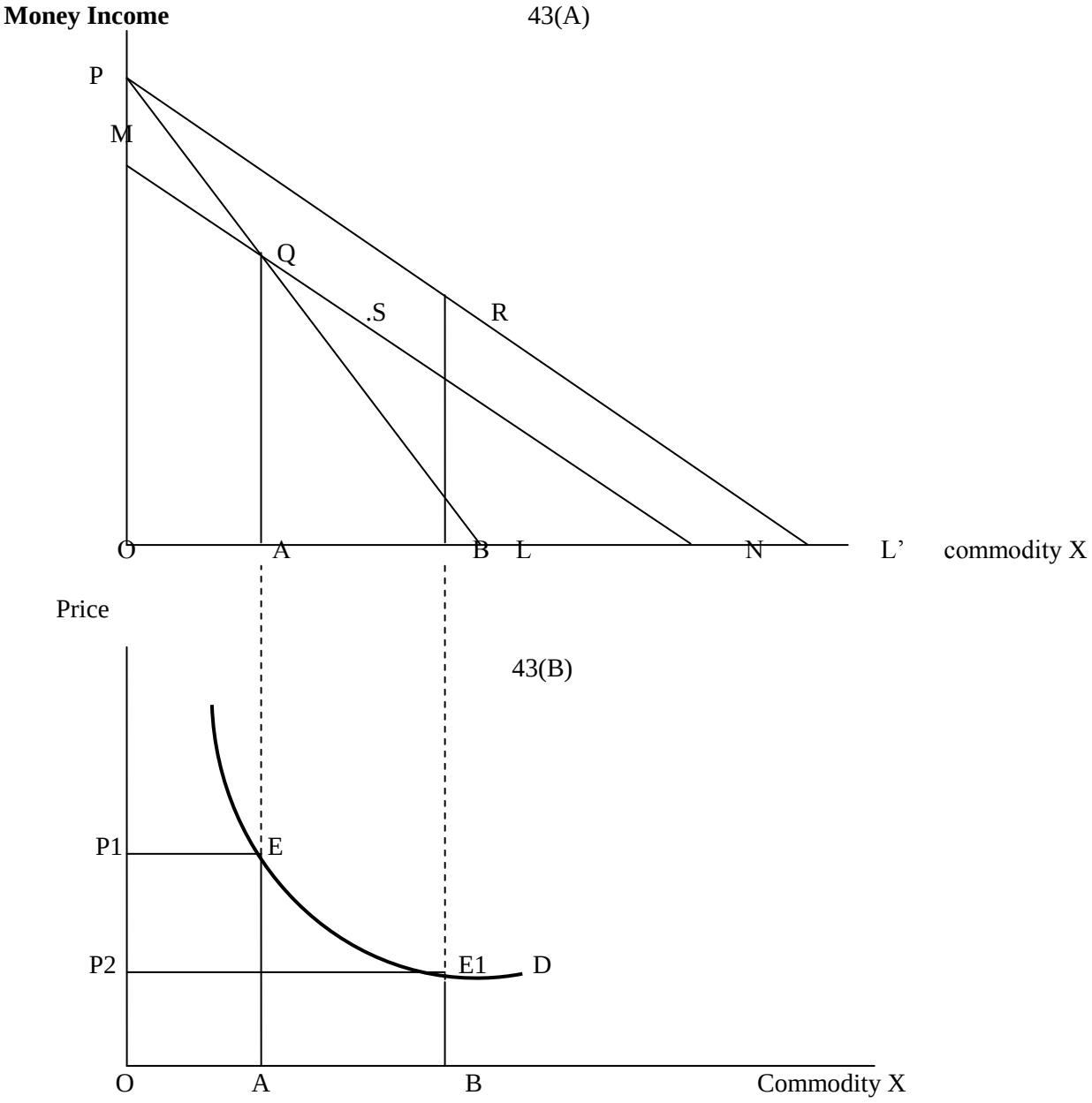
In the original price- income line PL, consumer chooses a combination Q. When price of X falls, the real income increases so that the consumer could purchase more units of X and Y. This is shown by the budget line PL". The consumer now feels better off than before. If he is to purchase the original combination Q, we have to take away from him some money to leave him neither better off nor worse off than before. This is shown by the new budget line drawn in the figure which passes through point Q (a

reduction of money income equivalent to NL"). Now original combination (before fall in price) Q is available. The MQ portion represents the combinations which are rejected for choosing Q. Thus in the new situation; the consumer would choose either Q or any point to the right of Q representing more units of X. Now, if the money withdrawn from the consumer is given back, the consumer will choose combination representing more of X because income elasticity is positive. Thus the movement will be to a point like R where the units of X increase with a reduction in the price of X. The inverse relation between price and quantity establishes again.

Derivation of Demand Curve from Revealed Preference Theory

We can derive the demand curve of an individual from the Revealed Preference Hypothesis. The process is shown in the figure 43. Assume that the consumer has the budget line PL and choose the combination of goods denoted by Q. Suppose the price of X falls so that the new budget line facing the consumer is PL'. This represents an increase in real income.

In panel (A) of figure money is measured on the vertical axis and good X on the horizontal axis. In panel (B) vertical axis measure price of the commodity. We can show that the consumer's new combination will include a larger quantity of X.



Firstly, we make a 'compensating variation' of the income. Ie., reduction of income so as to reduce the income of consumer just enough to purchase combination Q if he wishes to buy it. Following the compensating variation, the consumer's income is just enough to purchase the original combination if he wishes so. The new budget line MN passes through point Q and is parallel to PL'. The combination Q, which was selected at the original situation, is available at the present situation. Thus the consistent behaviour requires the selection of either Q or any point to the right of Q in the new budget line MN (say point S). If the money taken from the consumer is returned to him, the consumer will again be at a point such as R to the right (sine income elasticity is positive).

In figure: 43A, money is measured on the vertical axis. The price of good X can be calculated by dividing total money income by the number of units of X that can be bought. When the price of X is OP/OL=(OP_1), the quantity demanded is OA. When the price of X falls to OP/OL (=OP_2), the quantity demanded increases to OB. This relation is brought in the lower portion of the figure. Point E represents combination OA units at P1 price and E_1 represent OB units at P_2 price. Ie., when price decrease from P1 to P_2, quantity of X that would buy increases from OA to OB

Suggested Readings

1. A Koutsoyiannis: Modern Microeconomics
2. Watson and Getz : Price theory and its uses
3. Dominick Salvatore: Microeconomics
4. Anindya Sen: Microeconomics, Theory and applications
5. Dominick Salvatore: Microeconomic Theory

MODULE V

Production, Cost and Revenue

Production Theory and Production Function

1. Production

Firms are economic organizations that purchase inputs and sell outputs. Production refers to the transformation of inputs to outputs, i.e., raw materials to goods. Here production refers not to just the production of goods but the production of services as well. Production creates the supply that allows our needs and wants to be satisfied. Production is one of the main focuses in economics. Production theories have existed long before Adam Smith, but were only refined during the late 19th century. The theory of production is an effort to explain the principles by which a business firm decides how much of each commodity that it sells (its outputs or products) it will produce, and how much of each kind of labour, raw material, fixed capital good, etc., that it employs (its inputs or factors of production) it will use. The theory involves some of the most fundamental principles of economics. These include the relationship between the prices of commodities and the prices (or wages or rents) of the productive factors used to produce them and also the relationships between the prices of commodities and productive factors, on the one hand, and the quantities of these commodities and productive factors that are produced or used, on the other. Production is also an activity that creates or adds to current or future utility. Production creates three types of utility.

Form utility: This refers to production process of changing the form of the inputs, i.e. converting the raw material into items possessing utility. For eg. changing the form of clay to a pot, wood to table, iron to furniture.

Place utility: Change the place of resources from the place where they have no use or limited use to another place where they have greater use. For e.g. transporting gold ore from the mine to the factory and to markets, apples transported from Shimla (production centre) to other parts of the country. Here utility is created or enhanced by the production process of transporting one point where they have limited utility to another place where they have more utility.

Time utility: This refers to the production process of making available materials at times when they are not normally available. For example, industries involved in dried fruits make them available in a time when it is normally not available (off season).

Hence we see that the production activity that converts raw wool is converted to woollen cloth is creation of form utility, the production activity that transports cloth to markets is creation of place utility and the production activity (service of a textile shop) that make woollen clothes in winter is creation of time utility.

Thus the production process involves the use of factors of production. Of the various factors of production, land is a natural resource and labour, capital and entrepreneur are considered as human resources in the sense that they secondary in nature. Land in economics does not mean soil or earth's surface alone, but refers to all free gifts of nature which would include natural resources, fertility of soil, water, air etc. All production involves some dealing of man with nature (e.g. law of variable proportions). As John Stuart Mill put it, 'man can only move matter, not create it'. The utilisation of natural resources is indeed indispensable in production. There can be no doubt about this. When in certain theoretical

conceptualizations this fact is not visible, then this does not mean that it is not there. It only means that the authors have for simplicity set aside the problem by assuming that natural resources are available in abundance. This amounts to assuming that their services are 'free goods'. Labour in economics refers to 'mental and physical effort directed to produce goods or services'. However, note that, the work done for the sake of pleasure or affection does not represent labour in economics. For example, if a person performs a musical show for his friends, it will not be counted as labour since it is done for the sake of pleasure. Capital is that part of wealth of an individual or community which is used for further production of wealth. Note that capital is a 'stock' concept in economics and it helps to yield a periodical return called income which is a 'flow' concept. Capital is often called 'produced means of production' or 'man made instruments of production', e.g. factories, bridges etc. Capital formation is a term that refers to a sustained increase in the stock of real capital in a country. In other words, capital formation means investment creation. Entrepreneur mobilises all the factors of production, i.e., land, labour and capital, and combines them in the right proportion to initiates the process of production. In this process the entrepreneur bears the risk involved in it. An entrepreneur takes key decisions like (a) whether to produce or not (b) how much to produce (c) what input combination to use (c) what type of technology to use.

Probably the most important challenge in the production function is the combination of the factors of production in the right proportion. This requires not only entrepreneurial skills, but also knowledge about the best available technology. While entrepreneurial skill ensures economic efficiency, availability of technology ensures technical efficiency. (Details of parameters of economic efficiency and technical efficiency are discussed later in the chapter)

1(a). Practical applications of the production theory

A theory of production is the statement of technical and technological relationships between inputs and output. We know that when more and more factors of production -labour, capital, land, time, space and raw materials- are employed, total production increases. For example, when a farmer uses more and more of land, labour, agricultural equipments and machinery, irrigation and fertilizers, his total farm production increases. Production increases even if some factor (say, land) is held constant and other factors are increased, till production capacity of land is reached. These facts indicate that there is a relationship between the quantity of factors used and the quantity produced. Production theory analyses this relationship. More precisely, the theory of production explains and formalises the nature of relationships between the factors (inputs) used and output. Thus, the main function of the production theory is to analyse and make generalisations about the relationship between the inputs and the output. In simple words, production theory seeks to answer the following queries: When more and more units of a variable factor (say, labour) is used with a fixed factor (say, capital), how does the total output behave? When all the factors are increased by some proportion, does the output increase in the same proportion? In other words, if the inputs are doubled, will the output be doubled or increase at a different rate? Information on such items will be of immense practical use to producers in taking decisions.

The various decisions a business enterprise makes about its productive activities can be classified into three layers. The first layer includes decisions about methods of producing a given quantity of the output in a plant of given size and equipment. It involves the problem of what is called short-run cost minimization. The second layer, including the determination of the most profitable quantities of products to produce in any given plant, deals with what is called short run profit maximization. The third layer, concerning the determination of the most profitable size and equipment of plant, relates to what is called long-run profit maximization. All these complexities are described in the study of the production function.

2. Production Function

Before we get into the discussion on production function, let us see the brief history of production function

2.(a). History of Production Function

Various Roman and Greek authors have addressed many issues in economics included cursory attention to production and distribution. The Scholastics, including Saints Augustine and Thomas Aquinas, also devoted substantial time to economic matters including discussion and inquiries into production. Several authors associated with the Mercantilist and Physiocratic schools of thought also paid even more careful attention to matters of production in the economy. For example, Anne Robert Jacques Turgot, a member of the Physiocrats, is credited with the discovery around 1767 of the concept of diminishing returns in a one input production function. Of course Adam Smith himself devoted much time to issues concerning productivity and income distribution in his seminal 1776 book The Wealth of Nations.

The Classical economists who immediately followed Smith expanded on his work in the area of production theory. In 1815 Thomas Malthus and Sir Edward West discovered that if you were to increase labour and capital simultaneously then the agricultural production of the land would rise but by a diminishing amount. They both in effect rediscovered the concept of diminishing returns. David Ricardo later adopted this result in order to arrive with his theory of income distribution when writing his economic classic the Principles of Political Economy. The Marginalists also dabbled in the area of production. During the late 1800's W. Stanley Jevons, Carl Menger and Leon Walras all incorporated ideas of factor value into their writings. What these early post-Smith economists all had in common is that they all used production functions that were in fixed proportions. In other words the capital to labour ratios were not allowed to change as the level of output changed. Although interesting, in practice most production functions probably exhibit variable proportions.

In the 1840's J. H. von Thunen developed the first variable proportions production function. He was the first to allow the capital to labor ratio to change. Von Thunen noticed that if we were to hold one input constant and increase the other input then the level of output would rise by diminishing amounts. In other words he applied the concept of diminishing returns to a two input, variable proportions production function for the first time. An argument could definitely be made that he is the original discoverer of modern marginal productivity theory. His work never received the attention it deserved though. Instead during 1888 American economist John Bates Clark received credit for being the founder of marginal productivity theory based on his speech at the American Economic Association meetings that year.

Shortly after in 1894 Philip Wicksteed demonstrated that if production was characterized by a linearly homogeneous function (in other words one that experiences constant returns to scale) then with each input receiving its marginal product the total product would then be absorbed in factor payments without any deficit or surplus. Around the turn of the century Knut Wicksell produced a production function very similar to the famous Cobb-Douglas production function later developed by Paul Douglas and Charles W. Cobb.

In 1961, Kenneth Arrow, H.B. Chenery, B.S. Minhas and Robert Solow developed what became known as the Arrow-Chenery-Minhas-Solow or ACMS production function. Later in the literature this became known as the constant elasticity of substitution or CES production function. This function

allowed the elasticity of substitution to vary between zero and infinity. Once this value was established it would remain constant across all output and/or input levels. The Cobb-Douglas, Leontief and Linear production functions are all special cases of the CES function. In 1968 Y. Lu and L.B. Fletcher developed a generalized version of the CES production function. Their variable elasticity of substitution function allowed the elasticity to vary along different levels of output under certain circumstances.

Recently there have been many developments with flexible forms of production functions. The most popular of these would be the transcendental logarithmic production function which is commonly referred to as the translog function. The attractiveness of this type of function lies in the relatively few restrictions placed on items such as the elasticity of scale, homogeneity and elasticity of substitution. There are still problems with this type of function however. For example, the imposition of separability on the production function still involves considerable restrictions on parameters which would make the function less flexible than originally thought. The search for better, more tractable production functions continues.

In microeconomics, a production function expresses the relationship between an organization's inputs and its outputs. A production function summarizes the relationship between inputs and outputs. It indicates, in mathematical or graphical form, what outputs can be obtained from various amounts and combinations of factor inputs. In particular it shows the maximum possible amount of output that can be produced per unit of time with all combinations of factor inputs, given current factor endowments and the state of available technology.

The production function is a purely technical relation which connects the inputs and outputs. It is a functional relationship between inputs and outputs. The terms technological or functional are used to indicate the fact that the prices of the factors of production or cost of production are not included in the production function. A production function describes the laws that govern the processes of transforming inputs into outputs. Unique production functions can be constructed for every production technology.

In general, a production function is represented as

$$\mathbf{Q = \mathit{f}(x_1, x_2, x_3, ...,x_n)}$$

Where,

Q is the maximum quantity of output,

x_1, x_2, x_3, ...,x_n are the quantities of various inputs.

The mathematical form of production function contains more details than the above general form. This is because production function can provide measurement of concepts in economics like the marginal productivity of factors of production, the marginal rate of substitution, elasticity of substitution, factor intensity, the efficiency of production, technology, and the return to scale. The general mathematical form of the production function is

$$\mathbf{Q = \mathit{f}(L, K, R, S, \mathit{v}, \mathit{e})}$$

Where,

L	is labour input,
K	is capital input,
R	is raw material,
S	is land input,
v	is returns to scale and
e	is efficiency parameter.

Here the efficiency parameter *e* is included to represent the fact that two firms with same factor inputs and same returns to scale can have different output due to efficient entrepreneurship or management. Production function has been used as an important tool of economic analysis in the neoclassical tradition. It is generally believed that Philip Wicksteed (1894) was the first economist to algebraically formulate the relationship between output and inputs as $p = f(x_1, x_2, \ldots x_n)$ although there are some evidences suggesting that Johann Von Thunen first formulated it in the 1840's .

There are several ways of specifying a production function.

One is as an additive production function, where the variables are added to each other. For example, $Q= a + b X_1 + c X_2 + d X_3$, where a, b, c, and d are parameters that are determined empirically.

Another is multiplicative production function, where the variables are expressed in a multiplicative relation. For example $Q = AL^{\alpha}K^{\beta}$ (the Cobb-Douglas production function, details given later) . Other forms include the constant elasticity of substitution production function (CES) which is a generalized form of the Cobb-Douglas function, and the quadratic production function which is a specific type of additive function. The best form of the equation to use and the values of the parameters (a, b, c, and d) vary from company to company and industry to industry. In the short run production function, at least one of the Xs (inputs) is fixed. In the long run all factor inputs are variable at the discretion of management. There are two special classes of production functions that are frequently mentioned in textbooks but are seldom seen in reality.

2.1. Homogeneous production functions: The production function $Q=f(X_1,X_2)$ is said to be homogeneous of degree n, if given any positive constant k, $f(kX_1,kX_2)=knf(X_1,X_2)$.

When $n>1$, the function exhibits increasing returns, and decreasing returns when $n<1$. When it is homogeneous of degree 1, it exhibits constant returns.

2.2. Homothetic functions: are a special class of homogeneous function in which the marginal rate of technical substitution is constant along the function.

2.3. Aggregate production functions: Production functions are normally built for a firm or industry. But, in macroeconomics, production functions for whole nations are sometimes constructed. In theory they are the summation of all the production functions of individual producers, however this is an impractical way of constructing them. This is because for the economy as a whole there are many types of outputs and services. Also, there are many types of capital goods including office buildings, factory equipment, airplanes, and other durable goods. Finally, there are many types of labor, from unskilled workers to brain surgeons. There are also methodological problems associated with aggregate production functions. Economists use a process called aggregation to come up with a single measure of output that summarizes all of the different goods and services produced in the economy. The weights are closely related to the relative prices of the goods. That is, an expensive surgery will have a higher weight in the aggregation process than an inexpensive pen. (The aggregate measure of output is called real gross domestic product, or real GDP.)

2.4. Fixed Coefficients Production Function: A production function associates the maximum level of output producible with given amounts of inputs. If the inputs must be combined in fixed proportions, like the ingredients of a recipe in a cookbook, the function is a fixed coefficients production function. It is also called a Leontief function, after its inventor, the economist and Nobel Prize winner, Wassily Leontif. For example, call centers require a one-to-one proportion between workers and telecommunication equipment. The isoquants for such a production function are L-shaped (with the kink on the 45 degree line).

We will now see in detail three types of production functions.

2.5. Linear Production function

The simplest possible production function is a linear production function. A Linear Production Function is a production function that assumes a perfect linear relationship between inputs and total output.

Example: Given a Linear Production Function Q = 20 K + 40 L. If this firm employed 8 units of capital and 17 workers then how much output would they produce?

20 x 8 + 40 x 17 = 840 units of output

2.5 (a). Homogenous Production Function

A production function is said to be homogenous if it satisfies the following condition.

Consider a production function Q = f(X, Y). As you know, in the long run all the factors of production can be increased. Suppose we increase both the factors x and y by the same proportion, k. To effect this change we multiply each input factor is by a positive real constant k. The new level of output is Q* and can be represented as

$Q^* = f(kX, kY)$.

Now, if we can take k out of the brackets as a common factor (if k can be completely factored out from Q*), then the new level of output Q* can be expressed as a product of k (to any power v) and the initial level of output.

That is $Q^* = k^v f(X, Y)$ or $Q^* = k^v Q$

In such cases, where k can be completely factored out is called a homogeneous production function. The formal definition is as follows.

A homogenous production function is a function such that if each of the inputs is multiplied by a real constant k, then k can be completely factored out of the function.

If k can not be completely factored out, the production function is non homogenous.

The power v of k is called the degree of homogeneity of the function and is used to measure the return to scale of a function.

As you know returns to scale is the long run analysis of production. The Law of Returns to scale postulates that when all inputs are increased by 1 %: if output increases by 1% it is constant returns to scale; if output increases by less than 1% it is decreasing returns to scale; if output increases by more than 1% it is increasing returns to scale.

Using the power v of k, we can state:

If $v = 1$, it is constant returns to scale

If $v < 1$, it is decreasing returns to scale

If $v > 1$, it is increasing returns to scale

A production function for which $v = 1$ and so returns to scale are constant is called a linearly homogenous production function. Since this implies that when we increase all the factors of production in

equal proportions, the output is also increased by the same ratio. Hence such production functions are also called Constant Returns to scale (CRS) production functions. Returns to scale are measured mathematically by the coefficient of the production function. For example, given a production function $X = b_0L^{b1}K^{b2}$, the returns to scale are measured by the sum b1 + b2.

Examples:

Mathematically, a function z = *f* (x,y) is homogenous of degree n if for all positive values of *k*, *f* (*k*x,*k*y) = $k^n f$ (x,y).

1. Examine whether the following production function is homogenous. Also find the returns to scale z = 5x + 4y.

Multiply each input by a real constant *k* and check whether it can be completely factored out.

f (*k*x,*k*y) = 5*k*x + 4*k*y = *k*(5x + 4y)

There is no *k* inside the bracket. So the function is homogenous. The degree of homogeneity is given by the power of *k*. Here the power of *k* is 1. So the given function is homogenous of degree less than one and has decreasing returns to scale.

2. $z = \frac{3x}{y}$

$$f(kx,ky) = \frac{3kx}{ky} = \frac{k}{k}\left(\frac{2x}{y}\right) = 1\left(\frac{2x}{y}\right) \qquad \text{since } \frac{k}{k} = k^0 = 1$$

There is no *k* inside the bracket. So the function is homogenous. The power of *k* is 0. So the given function is homogenous of degree less than one and has decreasing returns to scale.

3. $Z = x^3 + 7xy + y^3$

$$f(kx,ky) = (kx)^3 + 7(kx)(ky) + (ky)^3$$

Factoring k out, $k^2(k\,x^3 + 7xy + k\,y^3)$

Here *k* can not be completely factored out. There is *k* reaming inside the bracket. So the function is not homogenous.

2.6. Cobb Douglas Production Function

Out of all the production functions used in economics, the most popular production function is the Cobb Douglas production function, also known as the C-D function. In economics, the Cobb Douglas form of production functions is widely used to represent the relationship of an output to inputs. Though similar functions were originally used by Knut Wicksell (1851–1926), the Cobb-Douglas form was developed and tested against statistical evidence by Charles Cobb and Paul Douglas.

In the 1920s the economist Paul Douglas was working on the problem of relating inputs and output at the national aggregate level. A survey by the National Bureau of Economic Research found that during the decade 1909-1918, the share of output paid to labour was fairly constant at about 74% , despite the fact the capital/labour ratio was not constant. He enquired of his friend Charles Cobb, a mathematician, if any particular production function might account for this. This gave birth to the original Cobb-Douglas production function which they propounded in their 1928 paper, 'A Theory of Production'.

The general form of a C-D function is stated as $\mathbf{Q = AL^{\alpha}\ K^{\beta}}$

where: Q = total production (the monetary value of all goods produced in a year), L = labour input, K = capital input and A = total factor productivity or technology, which is assumed to be a constant. Here α and β are the output elasticities of labour and capital, respectively. These values are constants and are determined by available technology. Output elasticity measures the responsiveness of output to a change in levels of either labour or capital used in production, ceteris paribus. For example if α = 0.15, a 1% increase in labour would lead to approximately a 0.15% increase in output.

It is generally said that a strict C-D function assumes constant returns to scale as α + β = 1.(since the returns to scale are measured mathematically by the coefficient of the production function).

Cobb and Douglas were influenced by statistical evidence that appeared to show that labour and capital shares of total output were constant over time in developed countries. However, now many economists doubt whether constancy over time exists. But at your level of understanding, we say that for a C-D function, α + β = 1

2.6(a). Properties of a Cobb-Douglass function

1). CD function is linearly homogenous of degree one. This means that when input is increased by λ, output also increases by λ.

2) Average products of capital and labour can be expressed in terms of ratios of inputs.

Average product of labour can be obtained by dividing the production by the amount of labour.

$Q = A\ L^{\alpha}\ K^{\beta}$

$$AP_L = \frac{Q}{L} = \frac{\mathbf{A\ L^{\alpha}\ K^{\beta}}}{L}$$

$$= \frac{AK^{\beta}}{L^{-\alpha}L}$$

$$= \frac{AK^{\beta}}{L^{\mathbf{1}-\alpha}}$$

$$= \frac{AK^{\beta}}{L^{\beta}} \quad \text{since } \alpha + \beta = 1, \mathbf{1} - \alpha = \beta$$

$$= A\left(\frac{K}{L}\right)^{\beta}$$

Thus we have shown that the AP_L can be expressed as the raion of the two inputs K and L

Similarly

$$AP_K = \frac{Q}{K} = \frac{\mathbf{A\ L^{\alpha}\ K^{\beta}}}{K}$$

$$= \frac{AL^{\beta\alpha}}{K^{-\beta}K}$$

$$= \frac{AL^{\beta\alpha}}{K^{\mathbf{1}-\beta}}$$

$$= \frac{AL^{\alpha}}{K^{\alpha}} \quad \text{since } \alpha + \beta = 1, \mathbf{1} - \beta = \alpha \quad = A\left(\frac{K}{L}\right)^{\alpha}$$

3). Marginal product of capital and labour can be expressed in terms of ratios of inputs.

$$MP_K = \beta\ A\ \left(\frac{L}{K}\right)^{\alpha}$$

Thus the marginal product of capital (MP_K) can be expressed in terms of ratios of inputs L and K. It is also equal to β times AP_K. That is, $MP_K = \beta\ AP_K$

Similarly the marginal product of capital (MP_L) can be expressed in terms of ratios of inputs L and K. Symbolically $MP_L = \alpha\ A\ \left(\frac{K}{L}\right)^{\beta}$ It is also equal to α times AP_L. That is, $MP_L = \alpha\ AP_L$

4. CD function satisfies Euler's theorem.

$$L\ \frac{\partial Q}{\partial L} + K\ \frac{\partial Q}{\partial K} = Q$$

5. Elasticity of substitution of a CD function is unity*.

(*The elasticity of substitution was introduced independently by John Hicks (1932) and Joan Robinson (1933) to measure the degree of substitutability between any pair of factors. Elasticity of substitution is the elasticity of the ratio of two inputs to a production function with respect to the ratio of their marginal products. It measures the curvature of an isoquant and thus, the substitutability between inputs, i.e. how easy it is to substitute one input for the other.)

6. Factor intensity: In C-D function $Q = A\ L^{\alpha}\ K^{\beta}$, the factor intensity is measured by the ratio $\frac{\alpha}{\beta}$. The higher this ratio, the more labour intensive the technique; the lower the ratio, the more capital intensive the technique.

7. A strict CD function represents constant returns to scale since $\alpha + \beta = 1$.

2.6(b). Importance of a C-D function

The C-D function is an analytical tool commonly used in economics which has the following uses.

1. C-D function can be used to determinate marginal productivity of labour and capital. Hence it can be used in the determination of wages and interest.
2. The parameters α and β of the function represents elasticity coefficient. These elasticity coefficients are helpful in inter-sectoral comparison in an economy and for the long run analysis of production i.e. returns to scale. As in the usual case of a C-D function, when $\alpha + \beta = 1$, we have constant returns to scale and the function is linear homogenous.
3. C-D function helps to compute elasticity values for inter sectoral comparisons.
4. C-D function is widely used in econometrics.
5. This production function helps us to study the different laws, of returns to scale.
6. This function is used to test laws of returns and substitutability of factors

2.6(c). Limitations of a C-D function

Cobb and Douglas were influenced by statistical evidence that appeared to show that labour and capital shares of total output were constant over time in developed countries; they explained this by statistical fitting least-squares regression of their production function. However, there is now doubt over whether constancy over time exists. Neither Cobb nor Douglas provided any theoretical reason why the coefficients α and β should be constant over time or be the same between sectors of the economy. Remember that the nature of the machinery and other capital goods (the K) differs between time-periods and according to what is being produced. So do the skills of labor (the L).

The Cobb-Douglas production function was not developed on the basis of any knowledge of engineering, technology, or management of the production process. It was instead developed because it had attractive mathematical characteristics, such as diminishing marginal returns to either factor of production. Crucially, there are no micro foundations for it. In the modern era, economists have insisted that the micro-logic of any larger-scale process should be explained. The C-D production function fails this test. It is thus a mathematical mistake to assume that just because the Cobb-Douglas function applies at the micro-level, it also applies at the macro-level. Similarly, there is no reason that a macro Cobb-Douglas applies at the disaggregated level.

1. A C-D function contains only two inputs labour and capital, but actually there may be more capital.
2. The parameter α and β can represent the labour and capital share only if there is perfect competition for labour and capital.
3. In most of the case this production function represents constant returns to scale. Other possibilities are sidetracked.
4. This function assumes that, technological conditions remain constant. But the production change due to change in technology is reality. Thus this function is based on the unrealistic assumption of stagnant technology.

2.7. Economic Efficiency and Technical Efficiency

It is relevant to note that among others there are two leading concepts of efficiency relating to a production system as propounded by Libenstein: the one often called the 'technical efficiency' and the other called the 'allocative efficiency' The formulation of production function assumes that the engineering and managerial problems of technical efficiency have already been addressed and solved, so that analysis can focus on the problems of allocative efficiency. That is why a production function is generally defined as a relationship between the maximum technically feasible output and the inputs needed to produce that output. However, in many theoretical and most empirical studies it is loosely defined as a technical relationship between output and inputs, and the assumption that such output is maximum (and inputs minimum) is often implicit.

We say that a firm is technically efficient when it obtains maximum level of output from any given combination of inputs. The production function incorporates the technically efficient method of production. A producer can not decrease one input and at the same time maintain the output at the same level without increasing one or more inputs. When economists use production functions, we assume that the maximum output is obtained from any given combination of inputs. That is, they assume that

production is technically efficient. On the other hand, we say a firm is economically efficient, when it produces a given amount of output at the lowest possible cost for a combination of inputs, provided that the prices of inputs are given. Therefore, when only input combinations are given, we deal with the problem of technical efficiency; that is, how to produce maximum output. On the other hand, when input prices are also given in addition to the combination of inputs, we deal with the problem of economic efficiency; that is, how to produce a given amount of output at thelowest possible cost.

One has to be careful while interpreting whether a production process is efficient or inefficient. Certainly a production process can be called efficient if another process produces the same level of output using one or more inputs, other things remaining constant. However, if a production process uses less of some inputs and more of others, the economically efficient method of producing a given level of output depends on the prices of inputs. Even when two production processes are technically efficient, one process may be economically efficient under one set of input prices, while the other production process may be economically efficient at other input prices.

Let us take an example to differentiate between technical efficiency and economic efficiency. A company is producing ready made garments using cotton fabric in a certain production process. It is found that 10 per cent of fabric is wasted in that process. An engineer suggested that the wastage of fabric can be eliminated by modifying the present production process. To this suggestion, an economist reacted differently saying that if the cost of wasted fabric is less than that of modifying production process then it may not be economically efficient to modify the production process. This example clearly shows that there could be distinctions between the conditions of economic efficiency and technical efficiency.

Mathematical Characteristics of production functions

In explaining some of the history regarding production functions we mentioned several characteristics that these functions possess. In this section several of the important characteristics will be explained. The first one that will be covered is the duality between the production function and the cost function. For well behaved functions we can produce a cost function from a production and vice versa. This is important due to the fact that production functions are much harder to estimate econometrically than cost functions. Cost functions depend on factor prices and output levels which are relatively easy to observe. Another key characteristic of production functions relate to homogeneity and homotheticity. All homogeneous functions are homothetic, but not all homothetic functions are homogeneous. Homogeneity can be of differing degrees. In economics we typically work with functions that are homogeneous of degree zero or one. If a production function is shown to be homogeneous of degree k then the first partials of that function would be homogeneous of degree k-1. For example, if we have a production function exhibiting linear homogeneity (degree one) then the marginal product functions would be homogeneous of degree zero meaning that they are functions of the relative amounts of inputs, but not the absolute amount of any one input used in the production process. Homogeneity also implies that the isoquant curves will be radial blowups of one another. In essence the curves will be parallel to one another, thus if a ray was constructed from the origin the slope of the isoquants along that ray would all be the same. The famous Euler's Theorem also follows from the assumption of homogeneity. The more general homotheticity has an even more important role in economics. Since all homogeneous functions are homothetic everything just stated above would hold true for homothetic functions as well. Homothetic production functions imply that the output elasticities for all inputs would be equal at any given point.

This common value can be represented by the ratio of marginal cost to average cost. Firms with increasing average cost would have output elasticity values greater than one; firms with decreasing average cost would have output elasticity less than one. Under the assumption of homotheticity all inputs would have to be normal. Separability is another key potential feature of a production function. Not all production functions can be viewed as being separable. Many production processes use many more than two inputs. This makes studying such a multi-input function rather difficult. It would be beneficial if we could break the production process down into various stages where intermediate inputs are produced and then combined with other intermediate inputs to produce the final output. If we can specify these separate production functions then the technology is assumed to be separable. This separability feature has many valuable implications for an economist including the fact that its presence greatly reduces the number of parameters to be analyzed in an applied economic analysis of cost or production functions.

2.8. Uses of a production function

A production function is helpful in the following ways.

1. When the physical quantities of inputs are specified, production function helps to estimate the level of production.

2. When the Q (quantity of output) is fixed, production function gives the different combination of inputs which yields the same level of output.

3. Production function helps to determine the technically efficient combination of inputs and also to select the least cost combination of inputs when the budget constraint is given.

4. Production function helps to estimate the degree of returns to scale prevailing in the process of production.

5. The marginal product of different factors can be obtained from production function.

Production function can be fitted to a particular firm or to a sector or industry or to an economy as a whole. Generally, for a given technology, production function remains the same. As the technology changes, production function will also change. The nature and type of production function depends upon data, time period of investigation and the type of technology employed.

Starting in the early 1950's until the late 1970's production function attracted many economists. During the said period a number of specifications or algebraic forms relating inputs to output were proposed, thoroughly analyzed and used for deriving various conclusions. Especially after the end of the 'capital controversy', search for new specification of production functions slowed down considerably. We will first concentrate on short run production function. Then we would move to long run or multi-output production function.

Short Run and Long run Analysis of Production

Having seen a production function, we now move to the analysis of production functions. The analysis of production is usually divided into two distinct categories, the short run analysis of production and the long run analysis of production. This distinction is necessary due to the fact that the relation between various inputs and outputs vary at different time intervals. For example, some factors of production like labour can be varied in very short span of time while to vary a factor of production like land requires much longer span of time. The distinction between short run and long run is made more clear in the following section.

1. Short Run and Long Run

All inputs can be divided into two categories: i) fixed inputs and ii) variable inputs. A fixed input is one whose quantity cannot be varied during the time under consideration. The time period will vary depending on the circumstances. Although any input may be varied no matter how short the time interval, the cost involved in augmenting the amount of certain inputs is enormous; so as to make quick variation impractical. Such inputs are classified as fixed and include plant and equipment of the firm.

On the other hand, a variable input is one whose amount can be changed during the relevant period. For example, in the construction business the number of workers can be increased or decreased on short notice. Many builder firms employ workers on a daily wage basis and equal change in the number of workers is made depending upon the need. The amount of milk that goes in the production of butter can be altered quickly and easily and is thus classified as a variable input in the production process.

Whether or not an input is fixed or variable depends upon the time period involved. The longer the length of the time period under consideration, the more likely that the input be variable and not fixed. Economists find it convenient to distinguish between the short run and the long run. The short run is defined to be that period of time when some of the firm's inputs are fixed. Since it is most difficult to change plant and equipment among all inputs, the short run is generally accepted as the time interval over which the firm's land and equipment remain fixed. In contrast, the long run is that period over which all the firm's inputs are variable. In other words, the firm has the flexibility to adjust or change its environment.

Production processes of firms generally permit a variation in the proportion in which inputs are used. In the long run, input proportions can be varied considerably. For example, at Tata Motors Limited, an automobile dye can be made on conventional machine tools with more labour and less expensive equipment, or it can be made on numerically controlled machine tools with less labour and more expensive equipment i.e. the amount of labour and amount of equipment used can be varied. On the other hand, there are very few production processes in which inputs have to be combined in fixed proportions. Consider, Ranbaxy or Smith-Kline-Beecham or any other pharmaceutical firm. In order to produce a drug, the firm may have to use a fixed amount of aspirin per 10 gm of the drug. Even in this case a certain (although small) amount of variation in the proportion of aspirin may be permissible. If, on the other hand, no flexibility in the ratio of inputs is possible, the technology is scribed as fixed proportion type.

To conclude we may put it like this. The short run is a time period where only the variable factors of production can be altered, e.g. labour. Long run is a time period where all the factors of production may be varied. All factors become variable here. The firm can make changes in the amount of both the factors – fixed as well as variable. Hence, supply can be adjusted here according to demand in the long run.

Basis	Short Period	Long Period
1. Implication	This is a time period which is less than the time period required to make changes in fixed factors.	This is a time period in which all factors of production can be changed.
2. Output	Output can only be increased by making changes in the quantity of variable factors of production.	Output can be increased by making changes in the quantity of both variable and fixed factors of production.
3. Nature of factors of production	Factors of production here can be grouped into two categories (i) fixed factors (ii) variable factors	In the long run, there is no distinction between fixed factors and variable factors.
4. Effect on price	Here demand plays a dominant role in the determination of price of a commodity.	In the long period, supply can be adjusted according to any change in demand. So, demand and supply play equal role in price determination.

Now let us proceed to the analysis of production. First we the short run analysis of production and then the long run analysis of production.

2. **The short run production function / Production with a single variable input / The law of variable proportions / The law of diminishing returns / Law of returns to a factor.**

The short run analysis of production assumes that labour is the only factor of production which is variable. Consider a production function in the agricultural sector where the quantity of land is fixed at one acre. We examine what happens to production as we add more and more units of labourers to this fixed piece of land. Before we move on the example let us familiarise ourselves with certain concepts. Product implies the amount of goods produced by a firm during a given period of time. Generally, three concepts are used in this context, viz, (a) total product (b) marginal product and (c) average product. Total Product (TP) or Total Physical Product (TPP) refers to the total amount of goods produced by a firm with the given inputs during a specified period of time. It is the total physical output corresponding to each set of inputs. As the quantity of an input is increased, total product increases. But the rate of increase in total product varies at different levels of factor employment. In the short run, a firm can expand its total output only by increasing the variable factors. Average Product (AP) or Average Physical Product (APP) is the per unit production of the variable factor. Thus $= \frac{TP}{Q}$. For example, suppose 20 units of a variable factor (labour) produces 100 units of output, then the Average Product of labour is $AP_L = \frac{100}{20} = 5$. Thus the average product of labour in this example is 5 units of out put. Marginal Product (MP) or Marginal Physical Product (MPP) is the addition to total product as we use one more unit of the variable factor. For example, suppose the output level is 100 kgs of wheat when we use 3 employees in an acre of land. Now assume that in the next agricultural cycle one more labourer is employed to cultivate the same one acre of land. Now the output increases to 120 kgs of wheat. So we can say that the additional contribution of the additional labourer is 120 – 100 = 20 kgs of wheat. In other word the Marginal Product of labour (MP_L) is 20 units. Thus $MP_L = \frac{\Delta TP}{\Delta L}$. Note that here we used AP_L and MP_L to indicate that it is representing the contribution of labour alone, since the other factor, land, is assumed to be constant.

Now let us see what happens to TP, AP_L and MP_L as more and more units of the variable factor (labour) is added to the fixed factor (capital).

Table 1 : The law of variable proportions

	Fixed Factor (Land)	Variable Factor (Labour)	Total Product	Average Product of Labour $AP_L = \frac{TP}{Q}$	Marginal Product of Labour $MP_L = \frac{\Delta TP}{\Delta L}$.	Stage of production
A	1	0	0	0	0	I Stage
B		1	8	8	8	
C	1	2	18	9	10	
D	1	3	30	10	12	
E	1	4	48	12	18	
F	1	5	65	13	17	
G	1	6	78	13	13	
H	1	7	84	12	6	II Stage
I	1	8	88	11	4	
J	1	9	90	10	2	
K	1	10	90	9	0	
L	1	11	88	8	−2	III satge
M	1	12	84	7	−4	

The table shows that as we add more and more units of labourers to one acre of land, the following changes take place.

1. We start with one acre of land zero labourers, as a hypothetical case. Then we add one labourer to the piece of land. Now total production is 8 units, average production is 8 units and marginal production is also 8 units. MP = 8 implies that the additional / marginal contribution of the first labourer is 8 units. When we add the second labourer to this agricultural land, the TP increases to 18. AP is 9 and MP is 10. Note that AP and MP etc are all exclusive contributions of labourer as land is fixed. As we add the 3rd labourer, TP increased to 30. Now the MP is 12. When TP increases to 48, MP is 18 and so on. And finally when TP reaches maximum (90), MP becomes 0. When TP declines (from 90 to 88 and to 84), MP becomes negative. All these show the clear relationship between TP and MP. The MP shows the rate of change of TP.

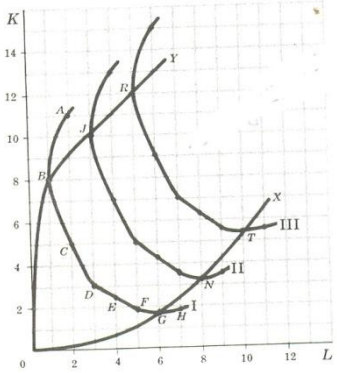
We may also see a peculiarity in the behaviour of total product, if we examine the relation between TP and MP. TP increases at an increasing rate initially (from A to G). You may note that in this range TP is increasing. MP is also increasing. But you can observe a change in the pattern of MP from H to K. In this range, TP continues to increase. But MP declines in this range. This implies that TP is increasing at a decreasing rate. Finally from L to M, TP declines. As a result the rate of change of total product, MP becomes negative.

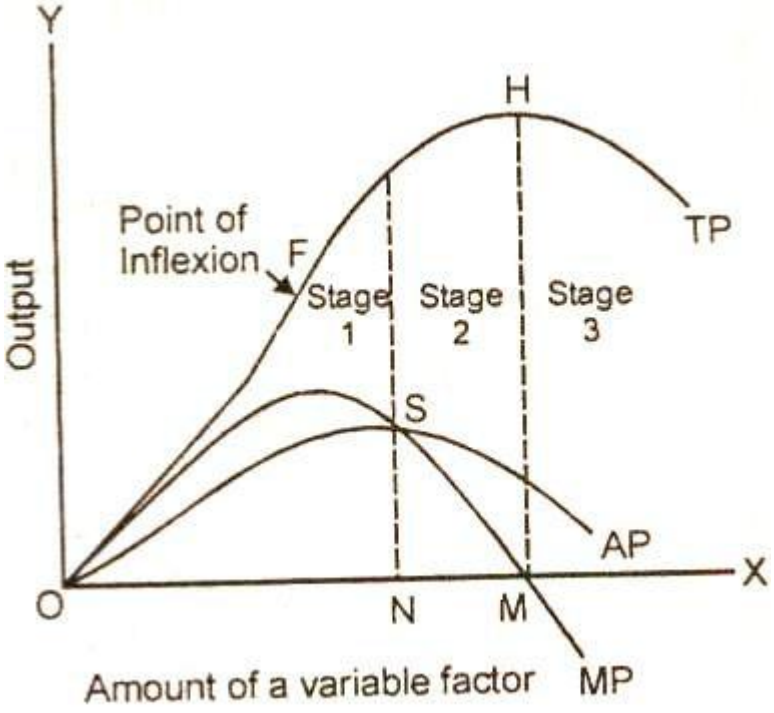
Before moving further, you may note one more thing from this table. You have seen from the headings above that the short run analysis of production is also known as the law of variable proportions. The reason can be made clear from the table. The table shows what happens to output as the proportion in which the inputs are mixed varies. In other words, it means that initially we mixed the two inputs in the ratio 1:1, then 1:2, 1:3 and so on, that is the ratio varies as we proceed forward in the production process.

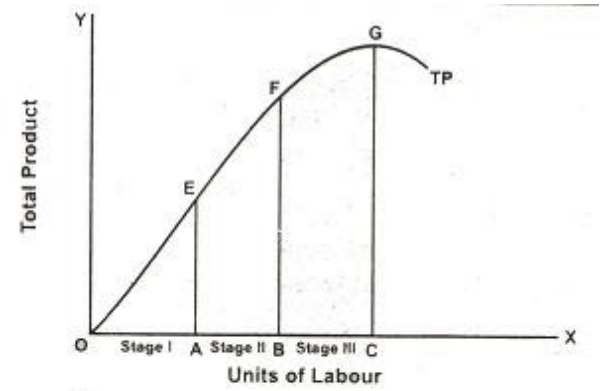
These changes can be grouped into three stages for effective analysis. These stages are shown in the following diagram, where we plot the TP, AP and MP curves.

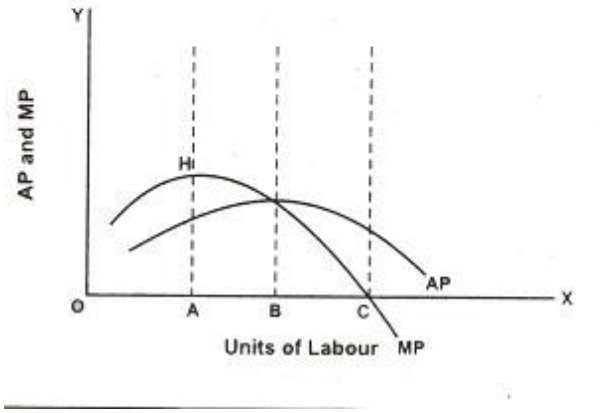
FIGURE 1. (Law of variable proportions)

This can be represented in two ways. One, with the TP, AP and MP curves in the same panel. Two, with the TP curve in the top panel and the AP and MP curves in the bottom panel. We have represented both diagrams below.Note that these diagrams are not drawn based on the data given in the table.









Note that the shape of the AP_L and MP_L curves are determined by the shape of the corresponding TP curve. The MP_L curve reaches a maximum before the AP_L curve. As long as the AP_L is rising, the MP_L is above it; when the AP_L is falling, the MP_L is below it; when AP_L is maximum, the MP_L is equal to AP_L. This behaviour is due to the following reason. For the AP_L to rise, the addition to TP (the MP_L) must be greater than the previous AP_L; for the AP_L to fall, the addition to TP (the MP_L) must be less than the previous average; for the AP_L to remain unchanged, the addition to TP (the MP_L) must be equal to the previous average.

Stages of Production

Stage I

Stage one is from the point of origin to the point where AP is maximum. This phase contains two distinct stages. The first stage is where the TP curve increases at an increasing rate and the second stage is when TP curve increases at a decreasing rate. The border between the two stages is marked by point F on the TP curve (first digram). We have already shown the reason for this behaviour in the table. Geometrically point F is called the inflection point, where the curvature of the curve changes from concave to convex or convex to concave. Stage I also witnesses the point where MP reaches its maximum. Note that at the end of stage I where AP is at its maximum MP cuts AP. This implies that at this point MP is equal to AP. You may verify all this from the Table 1. Though this diagram is not drawn based on the data from the table, all the basic relations in the table are depicted in the graph. MP reaches maximum at point E where the MP is at 18 (and AP is at 12). MP starts declining from this point and at G it becomes equal to AP at the value of 13. Note that 13 is the maximum of AP.

Stage II

Stage II goes from the point where the AP is maximum to the point where the TP is maximum or MP is zero. Here TP continues to increase at a diminishing rate and reaches a maximum. AP starts declining. MP continues to fall and reaches zero.

Stage III

Stage III covers the range where the TP is declining or MP is negative. Here TP starts to decline. AP continues to decline. But note that AP never reaches zero or becomes negative. MP continues to fall, but within the negative zone.

Now let us see why the different stages occur.

Stage I occurs because here better utilisation of the fixed factor takes place as additional units of the variable actor are employed. The short run analysis of production, as you have seen in the heading, is also known as the law of diminishing returns. This is because the diminishing returns start operating at the point where MP begins to decline. This occurs because too much labour is used to work in one acre of land.

Stage II is an ideal stage because here TP is increasing. AP and MP, though are decreasing, remains positive.

Stage III is a stage where TP starts declining and MP is negative. Let us go back to Table 1. Consider point M where the TP is 84 with 1 acre of land and 12 labourers. Suppose the farmer decides that in the next agricultural cycle, we will reduce the number of labours by 1, but all other factors (the duration of work, inputs used, implements used etc.) remain the same. We can read from the table that in the next agricultural cycle when the number of labourers was reduced by 1, the TP increased to 88. The question is what was the 12^{th} labourer doing in the field, what was his actual contribution. This is given by his MP, which we can read from the table as −4. The same trend repeats with the 11^{th} labourer. When the 10^{th} labourer is removed, the output did not increase, but remained constant at 90. It implies that the

MP of the 10^{th} labourer is zero. Such situations where the marginal contribution of a labourer is either 0 or –ve are called disguised unemployment.

Now the question is which of the three stages will be chosen by a rational producer. A profit maximising producer will not choose stage I. This is because in stage I by adding one more unit of labour, the producer can increase the average productivity of all units. Thus, it would be unwise on the part of the producer to stop production in this stage. It will also be equally irrational to produce in stage III. In this stage the producer can increase output by reducing labour input and thus reduce cost of production. We may say that stage I for labour corresponds to stage III for land (the MP of land is negative). Thus the economically meaningful stage is stage II.

3. The long run production function / Production with two variable inputs / The law of returns to scale

In the long run all the factors are variable. So we need not make an assumption that one input is fixed and the other is variable. As a result of this flexibility in the use of inputs, we can change the scale of operation of a firm; a small scale firm can become medium scale, a medium scale firm can become large scale and so on. So the long run analysis of production which shows what happens to output as the scale of operation itself is changed is called the laws of returns to scale.

Laws of Returns to scale

The laws of returns to scale may be explained like this.

Suppose all inputs are increased by 10%. There are three possible out comes:-

: if output increases by 10 %, it is constant returns to scale

: if output increases by more than 10 %, it is increasing returns to scale

: if output increases by less than 10 %, it is decreasing returns to scale

Thus,

Constant returns to scale refers to the situation where output changes by the same proportion as inputs. i.e., if all inputs are doubled, output also doubles.

Increasing returns to scale refers to the situation where output changes by a larger proportion than the inputs. i.e., if all inputs are doubled, output more than doubles.

Increasing returns to scale refers to the situation where output changes by a smaller proportion than the inputs. i.e., if all inputs are doubled, output less than doubles.

A numerical example of long run returns to scale

Units of Capital	Units of Labour	Total Output	% Change in Inputs	% Change in Output	Returns to Scale
20	150	3000			
40	300	7500	100	150	Increasing
60	450	12000	50	60	Increasing
80	600	16000	33	33	Constant
100	750	18000	25	13	Decreasing

In the example above, we increase the inputs of capital and labour by the same proportion each time. We then compare the % change in output that comes from a given % change in inputs.

In our example when we double the factor inputs from (150L + 20K) to (300L + 40K) then the percentage change in output is 150% - there are increasing returns to scale.

In contrast, when the scale of production is changed from (600L + 80K0 to (750L + 100K) then the percentage change in output (13%) is less than the change in inputs (25%) implying a situation of decreasing returns to scale.

The returns to scale can be represented in terms of diagram using isoquants. This diagram is given later in this book after we discuss the equilibrium of the producer.

4. The Equilibrium of the Producer

The equilibrium of the producer is analysed using the tools of isoquant and isocost. Let us familiarise ourselves with these tools.

Isoquant

The word 'iso' is of Greek origin and means 'equal' or 'same'. An isoquant is a curve along which quantity is the same. So the isoquant is also called equal product curve. An isoquant is equivalent o the concept of Indifference curve you studied under the analysis of consumer equilibrium. The only difference is that instead of two goods x and y used by a consumer, here we use two inputs labour and capital used by a producer. Here since we are dealing with physical quantity of production, the output can be expressed in terms of numbers. For eg. we may say the output from isoqunat I is K or 100, output from isoquant II is 2K or 200, output from isoqunat III is 3K or 300 and so on. Remember that this was not possible under indifference curve analysis.

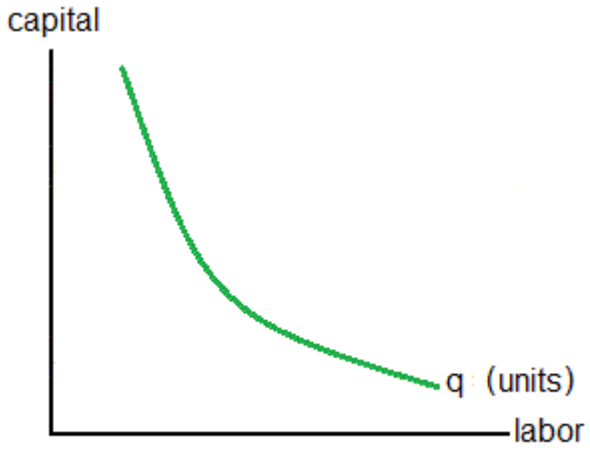
We shall get into further details. Consider a production function which uses two inputs labour and capital.

Table 2: A production function with two inputs

	Isoquant	
	Labour (L)	Capital (K)
A	2	12
B	1	10
C	2	5
D	3	3
E	4	2.3
F	5	1.7
G	6	1.2
H	7	0.8
I	8	1

Table 2 shows a production situation where both the inputs labour and capital are varied simultaneously. For instance, at situation A we use 2 L and 12 K. When we move to situation B, there is change in the quantities of both labour and capital. Now we use 1 unit of labour and 10 units of capital. In situation C, we increase quantity of labour to 2 and reduce the quantity of labour to 5 and so on. Since all factors are variable, we are dealing with the long run. Now the assumption here is that all the different combinations of labour and capital from A to I give the same level of output. If we plot the different combinations of labour and capital (A to I) that give the same quantity of output on a graph, we get an isoquant.

Thus an isoquant is the locus of different combinations of labour and capital that gives the consumer the same level of output. In other words, an isoquant shows the various combinations of two inputs that can be used to produce a specific level of output. Since output remains the same at all points on an isoquant, they are also called equal product curves.



Properties of isoquant

(1) Isoquants are negatively slopped in the relevant range [we will explain what is meant by the relevant range later in this chapter] (2) Isoquants are convex to the origin (3) higher the isoquant, higher is the level of production (4)two isoquants never intersect.

Now we explain what is meant by property number (3) and (4). Consider Table 3. Here we have added data for two more sets of isoquants. As you can see in Isoquant I, point A corresponds to 2 L and 12 K, the same point in Isoquant II corresponds to 4 L and 11 K which represents higher quantity of input and hence naturally higher level of output. Similarly point A corresponds to 6 L and 13 K in Isoquant III which represents a still higher quantity of input and hence naturally still higher level of output. Thus it is clear that higher the isoquant, higher is the level of production. Due to the same reason, two isoquants can not intersect. If they intersect, it would mean that the output is the same at that point. If we plot the information in Table 2 on a graph, we get a family of isoquants representing different levels of output which is often referred to as an isoquant map.

Table 3:

	Isoquant I		Isoquant II		Isoquant III	
	Labour (L)	Capital (K)	Labour (L)	Capital (K)	Labour (L)	Capital (K)
A	2	12	4	11	6	13
B	1	10	3	10	5	12
C	2	5	4	7	6	9
D	3	3	5	5	7	7
E	4	2.3	6	4.2	8	6.2
F	5	1.7	7	3.5	9	5.5
G	6	1.2	8	3.2	10	5.2
H	7	0.8	9	3	11	5
I	8	1	10	3.7	12	5.9

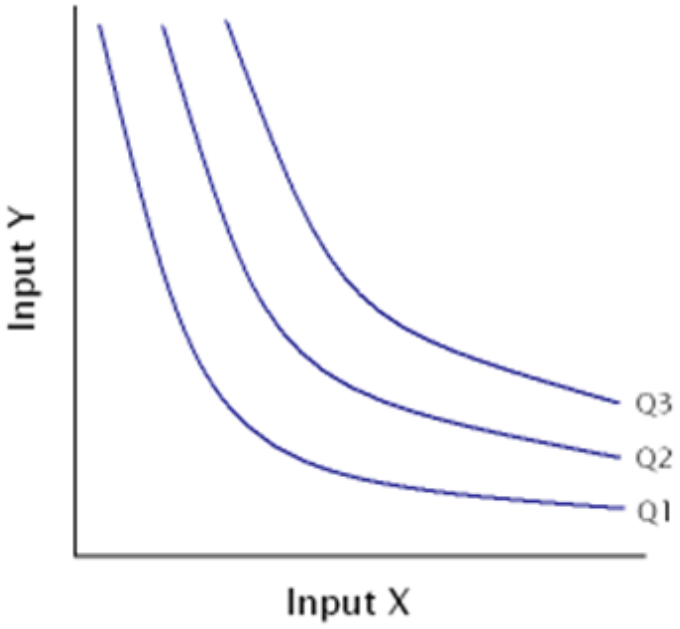


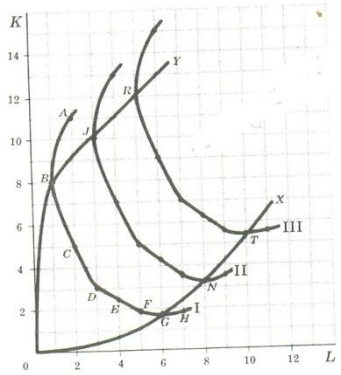
Fig. An Isoquant Map

(Note that the above figure is not drawn strictly based on the data in table 2.)

Now let us concentrate on the first property we stated in the properties of isoquants, that is, isoquants are negatively slopped in the relevant range. What is the relevant range? Go back to Table 2. Consider points A and B for Isoquant I. If move in reverse direction, you may observe that at B we are using 1 L and 10 K. As we move to A, the quantities of both the inputs increase; that is; units of labour increases to 2 and units of capital increases to 12. This is economically inefficient because efficiency means using less of one input as we use more of the other input. The same can be observed at the other end of the Isoquant

also, that is, as we move from H to I. At H, the producer uses 7 L and 0.8 K. At point I, quantities of both the inputs increase, L increases to 8 and K increases to 1. Such economically irrelevant points could be seen at both extremes of all the isoquants. If we draw a line to separate the economically relevant regions from economically irrelevant regions, such a line is called a ridge line. We can have an upper ridge line on the top and a lower ridge line at the bottom. The area between the ridge lines are economically relevant regions and the area outside the ridge lines are economically irrelevant regions.

The figure below shows the upper and lower ridge lines.



The property also says within the relevant range the isoquants are negatively sloped. This implies that if the firm wants to use more L, it must to use less K to produce the same level of output (same level of output means remaining on the same isoquant).

Another assumption is that in the relevant range the isoquants are convex to the origin. The property of convexity could be explained using the concept of MRTS.

MRTS

Marginal Rate of Technical Substitution indicates the amount of one input the producer is willing to give up to get an additional unit of another input and still produce the same level of output. MRTS is equivalent o the concept of MRS you studied under the analysis of consumer equilibrium. When defined in terms of the two inputs, labour and capital, we used in this example; MRTS can be defined like this. The Marginal Rate of Technical Substitution of L for K ($MRTS_{LK}$) refers to the amount of K that a firm can give up by increasing the amount of L used by one unit (and still produce the same level of output, i.e., remain on the same isoquant). Note that as we move down an isoquant, the MRTS diminishes. It is because of the diminishing nature of MRTS that in the relevant range the isoquants are convex to the origin.

The $MRTS_{LK}$ can be found using the concept of Marginal Physical Product or Marginal Product we have already studied. $MRTS_{LK} = MP_L / MP_K$, which is equivalent to the slope of the isoqunat.

In Table 4 you can see that in isoquant I, (ignoring the irrelevant regions), as the producer moved from B to C, to get one additional unit of labour, he gave up 5 units of capital. i.e. to increase the quantity of L from 1 to 2, he reduced the quantity of capital by 5 units from 5 to 6. Again to increase L from 2 to 3, he reduced K by 2 units from 5 to 3. This can be observed for other points also. The same is applicable for isoquant II. Thus the table clearly shows that the $MRTS_{LK}$ diminishes as we move down an isoquant.

Table 4

	Isoquant I			Isoquant II		
	Labour (L)	Capital (K)	$MRTS_{LK}$	Labour (L)	Capital (K)	$MRTS_{LK}$
A	2	12	-	4	11	-
B	1	10	-	3	10	-
C	2	5	5	4	7	3
D	3	3	2	5	5	2
E	4	2.3	0.7	6	4.2	0.8
F	5	1.7	0.6	7	3.5	0.7
G	6	1.2	0.5	8	3.2	0.3
H	7	0.8	0.4	9	3	0.2
I	8	1	-	10	3.7	-

We can represent diminishing MRTS between two factors diagrammatically using an isoquant.

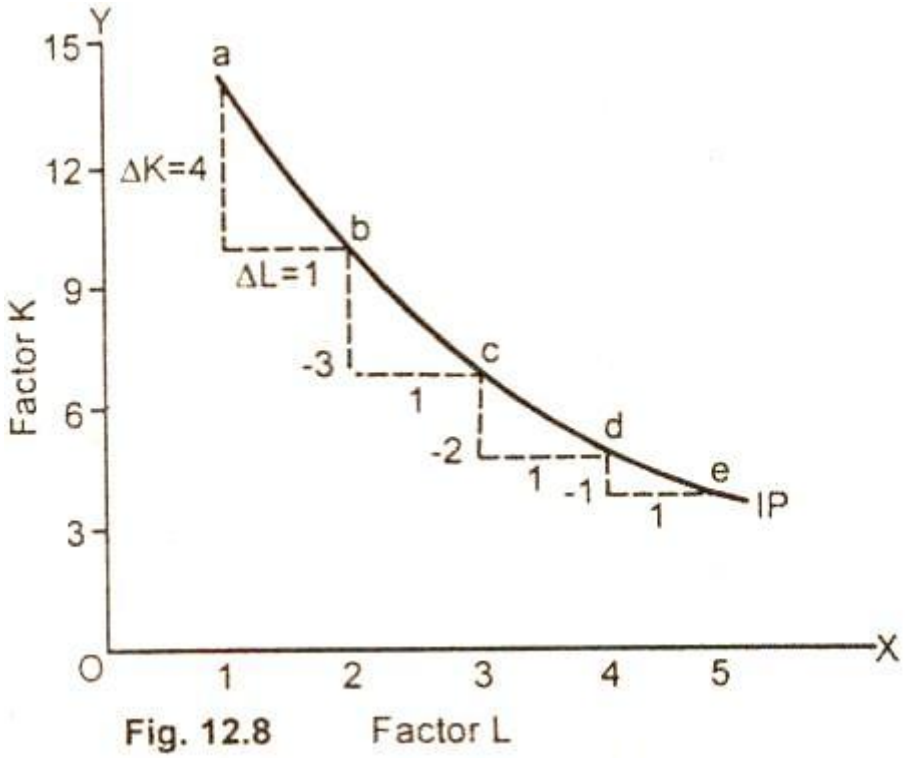


Fig. 12.8

As we can see in the figure, when the firm moves down from point (a) to point (b) and it hires one more labor, the firm gives up 4 units of capital (K) and yet remains on the same isoquant at point (b). So the MRTS is 4. If the firm hires another labor and moves from point (b) to (c), the firm can reduce its capital (K) to 3 units and yet remain on the same isoquant. So the MRTS is 3. If the firm moves from point (C) to (D), the MRTS is 2 and from point D to e, the MRTS is 1. The decline in MRTS along an isoquant as the firm increases labor for capital is called Diminishing Marginal Rate of Technical

Substitution. (Note that in drawing this figure students often make a mistake. You have to keep constant the length of the line representing ΔL. At the same time, the length of the line representing ΔK needs to be steadily reduced. This implies that as we move down an isoqunat to get an additional unit of L the amount of K that we are willing to give up reduces. Quite often I have seen students reducing the length of both lines as we move down the isoquant.)

Elasticity of factor substitution

The marginal rate of technical substitution as a measure of the degree of substitutability of factors has a serious limitation. The limitation is that it depends on the units of measurement of the two factors used. A better measure of the substitutability of the factors is provided by the concept of elasticity of substitution. The elasticity of substitution is defined as the percentage change in the capital – labour ration divided by the percentage change in the rate of technical substitution.

$$\textbf{Elasticity of factor substitution} = \frac{percentage\ change\ in\ {}^{K}/_{L}}{percentage\ change\ in\ MRS}$$

$$\textbf{Elasticity of factor substitution} = \frac{\frac{\Delta({}^{K}/_{L})}{{}^{K}/_{L}}}{\frac{\Delta\ MRS}{MRS}}$$

The elasticity of substitution is a pure number independent of the units of measurement of K and L, since both the numerator and the denominator are measured in the same units. It is this feature that makes 'elasticity of factor substitution' a much better measure of substitutability.

Isocost

Since the highest isoquant represents the highest level of output, the producers always want to reach the highest possible isoquant. But the aspirations of the producer need not match with the real world situation as in the process of maximisation of output the producer has to face certain constraints. For example, suppose our producer wants to reach figure X on isoquant III in the figure.

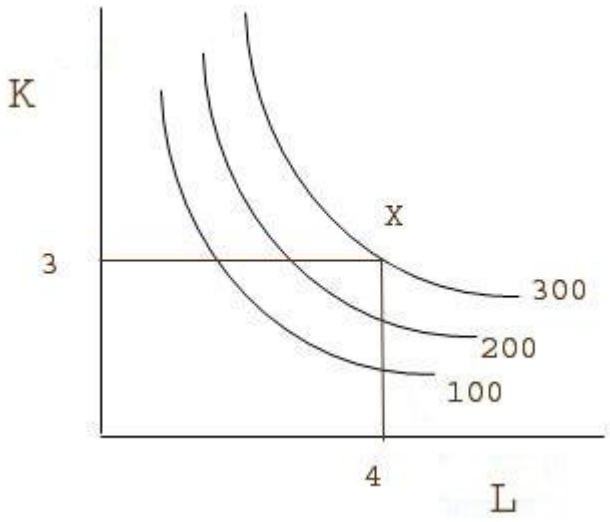
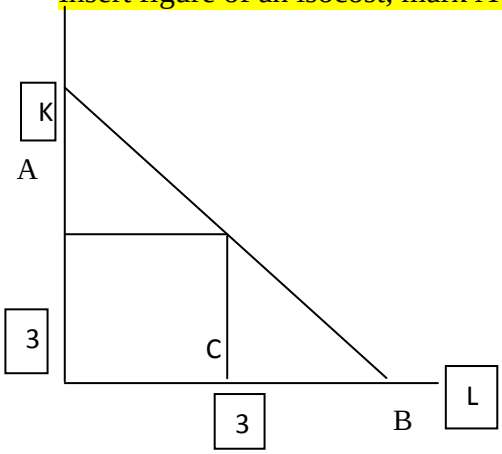


Figure : three isoqunts 100, 200, 300 (without isocost) Mark X on the 3rd isoquant. Mark K and L for this output . K is 3 and L is 4

The problem is the figure shows that to produce 300 units of output on isoquant III, the producer should use 3 units of capital and 4 units of labour. We have not examined whether the producer has enough funds to procure these inputs. For this we should have information on the per unit price of labour (wage) and capital (interest). We should also know the amount of money (total outlay) available with the producer to invest in this project.

Assume that the producer has Rs.6 to invest in the project and the price of one unit labour is Rs. 1 and also the price of one unit capital is Rs. 1. In this situation our producer can not produce on isoquant III because to produce on isoquant III he requires Rs. 7 as total outlay(Rs. 3 for capital + Rs. 4 for labour). Let us see how the situation can be tackled. We can plot a line using this information on the same graphical plane using the following procedure. Suppose the producer uses completely capital intensive technique of production and purchase only capital. In this case he can buy 6 units of capital (6/1 or TO/P_K) and 0 units of labour. (point A in figure ....). On the other hand the producer may also think of using completely labour intensive method of production and use only labour. In this case the producer will buy 6 units of labour (6/1 or TO/P_L) and 0 units of capital (point B in figure ...).

Insert figure of an isocost, mark A and B at the ends. And C in the middle (3L and 3K)

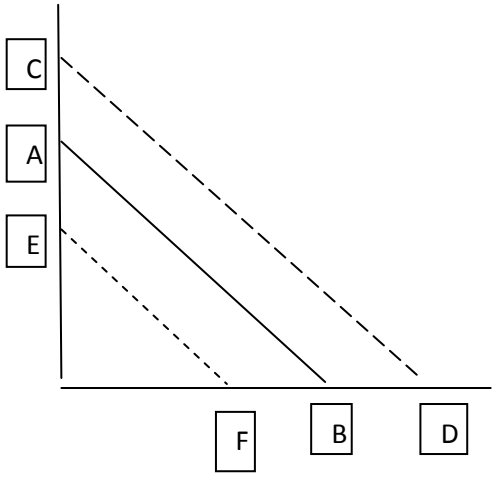


However these two are extremes and practically impossible. If we draw a line joining these two extreme possibilities, on this line will lie all other possible combinations of L and K. So in the real world situation the producer will be using a combination of these inputs, say 3 units of labour and 3 units of capital shown by point C in the figure. The line joining the two extreme possibilities of A and B is called an isocost. An isocost shows the constraints faced by the producer due to his limited total outlay (funds) and the prices of the two inputs in his attempt to reach the highest isoquant or level of production. Thus in our example an isocost may be defined like this. An isocost shows all the different combinations of labour and capital that a firm can purchase, given the total outlay (TO) of the firm and the factor prices. The slope of an isocost is given by the negative of the ratios of the prices of the factors of production, i.e., $-P_L/P_K$, where P_L refers to the price of labour and P_K refers to the price of capital.

General definition: The isocost line is the locus of all combinations of factors the firm can purchase with a given monetary cost of total outlay. $C = P_K(Q_K) + P_L(O_L)$

Note that a change in the total outlay will result in a parallel shift in the isocost. If the total outlay increases with no change in prices of the two factors of production, the isocost shifts to the right parallel to itself. For example, in the above example we assumed that the total outlay of the producer is Rs. 6.

Assume that the total outlay increases to Rs. 10. Now the new isocost will be line CD. On the other hand if the total outlay decreases to Rs. 3, the isocost shifts back parallel to itself. The new isocost will be EF.



We will discuss the impact of such shifts later in this chapter when we discuss 'expansion path'.

Least Cost Factor Combination: Producers Equilibrium or Optimal Combination of Inputs

The analysis of production function has shown that alternative combinations of factors of production, which are technically efficient, can be used to produce a given level of output. Of these, the firm will have to choose that combination of factors which will cost it the least. In this way the firm can maximise its profits. The choice of any particular method from a set of technically efficient methods is an economic one and it is based on the prices of factors of production at a particular time. The firm can maximise its profits either by maximising the level of output for a given cost or by minimising the cost of producing a given output. In either case, the factors will have to be employed in optimal combination at which the cost of production will be the minimum. There are two ways to determine the least cost combination of factors to produce agiven output. That is,

(a) Finding the total cost of factor combinations

(b) Geometrical method

(a) Finding the Total cost of Factor Combinations

Here we try to find the total cost of each factor combination and choose the one which has the least cost. The cost of each factor combination is found by multiplying the price of each factor by its quantity and then summing it for all inputs. This is illustrated in the Table.

Table: Choosing the Lowest Cost of Production Technique

Technique	Capital(units)	Labour (units)	Capital Cost Rs.	Labour Cost Rs.	Total Cost Rs.
1	2	3	4	5	6
A	6	10	500×6 = 3000	400×10 = 4000	7000
B	2	14	500×2 =1000	400×14 = 5600	6600

It is assumed that 100 units of output is produced per week and the price of capital and the wage of labour are Rs. 500 and Rs. 400 per week respectively. We assume that there are only two technically efficient methods of producing the output and they are labelled A and B. The table demonstrates that the total cost of producing 100 units of output is Rs. 7000 per week using technique A and Rs. 6600 per week using technique B. The firm will choose technique B, which is an economically efficient (or lowest cost) production technique at the factor prices assumed in the above example. If either of the factor prices alters the equilibrium proportion of the factors will also change so as to use less of those factors that display a price rise. Therefore, we will have a new optimal combination of factors. This can again be found out by calculating the cost of different factor combinations with the new factor prices and choosing the one that costs the least.

2. Geometrical method

Equilibrium of the producer

We have thus seen that in an isoquant map the producer's ambition is to reach the highest isoquant. But his aspirations are constrained by the limited income and prices of the two commodities. If we superimpose his aspirations (isoquant map) with the reality (isocost), we get the equilibrium of the producer.

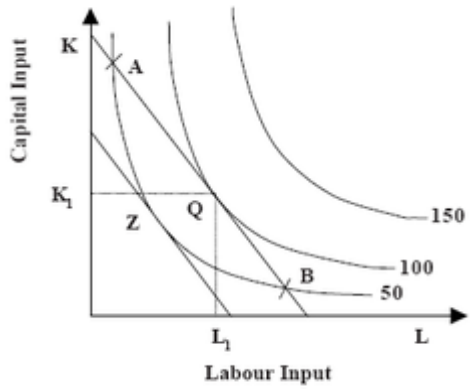
A producer is in equilibrium when he attains maximum output with given total outlay. In other words, the producer is in equilibrium when the highest possible isoquant is tangent to the isocost. At the point of tangency the absolute slope of the isoquant is equal to the absolute slope of the isocost. (Absolute slope means slope without considering –ve or +ve sign). So at equilibrium the slope of isoquant $MRTS_{LK}$ is equal to the slope of isocost P_L/P_K.

At equilibrium $MRTS_{LK} = P_L/P_K$.

Since $MRTS_{LK} = MP_L/MP_K$ we may also write

$$\frac{MP_L}{MP_K} = \frac{P_L}{P_K} \quad OR \quad \frac{MP_L}{P_L} = \frac{MP_K}{P_K}$$

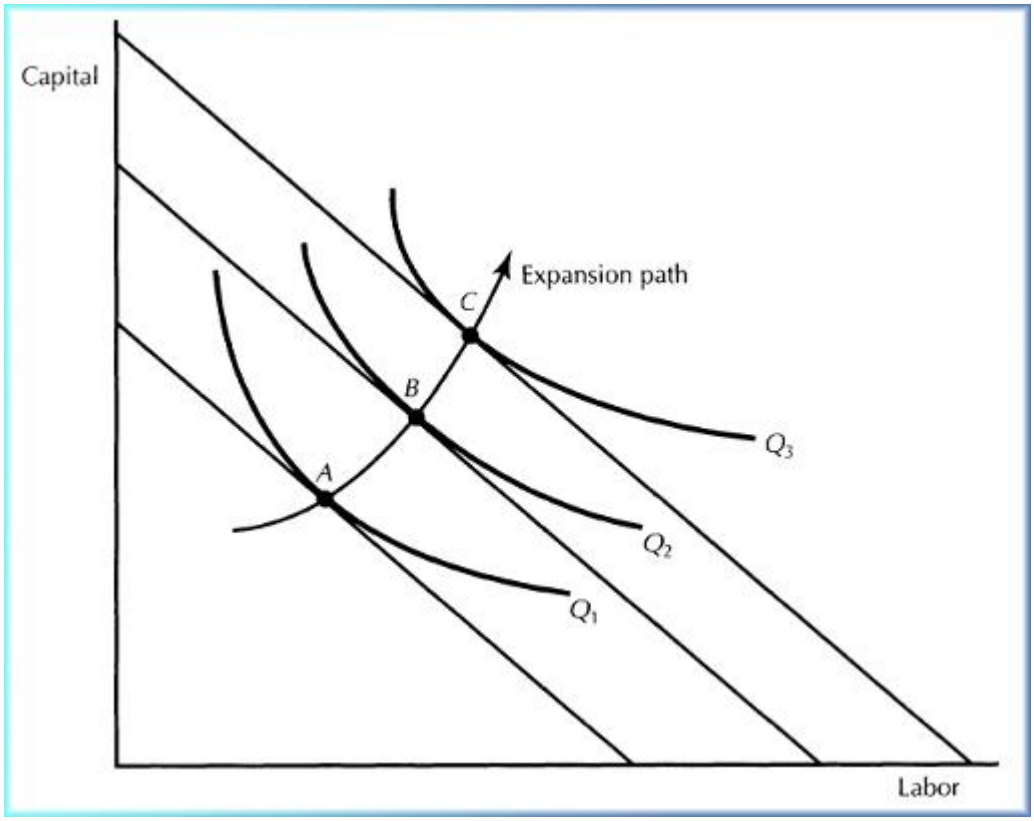
Figure of equilibrium



Suppose in the example we discussed for developing the isocost, the total outlay of the producer increases from Rs. 6 to Rs. 10. This will result in a rightward shift of the isocost to 10 L and 10 K. On the other hand if the total outlay of the producer declines to Rs. 3, the isocost will shift to the left to 3 K and 3

L. (We have already given graphical representation of such shifts in isocost lines.) These shift in the isocost lines over the isoquant map results in shift in points of equilibrium. For instance a shift in the budget line from 3 K – 3 L to 6 K – 6L shifts the equilibrium from E to E_1 in the figure ..... Similarly a shift from 6 K – 6 L to 10 K – 10 L, will result in a right ward shift of equilibrium point to E_2. If we join these different points of equilibrium by a line, we get the expansion path. Put in simple language, an expansion path shows the path through which the output will expand when the total outlay changes. The expansion path is same as the income consumption curve under the indifference curve analysis.

Figure of shift in equilibrium and expansion path

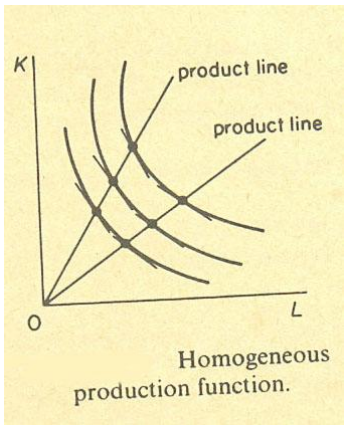


The expansion path shown here is a straight line through origin. This means that as output is expanded, the K/L ratio remain the same. The K/L ratio is the slope of the expansion path. This remains the same because the shift in the isocost lines does not imply change in input prices or the input price ratio P_L/P_K. That is why the isocost lines are shifted parallel to itself.

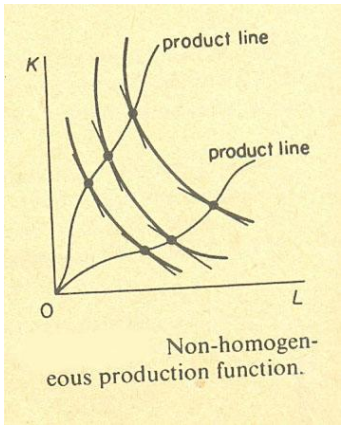
Isocline

An isocline is the locus of points of different isoquants at which the MRS of factors is constant. In other words, the line joining points on different isoquants at which the MRTS (the slope of isoquant) is constant is called an isocline. Thus, an expansion path is a particular type of isoclines along which output expands while factor prices remain constant.

If the production function is homogenous, the isoclines are straight lines through the origin. Along any one isoclines the K/L ration is constant. (fig .....). If the production function is non homogenous, the isoclines will not be straight lines, but their shape will be twiddly. The K/L ratio changes along each isoclines. (fig .....)



Homogeneous production function.



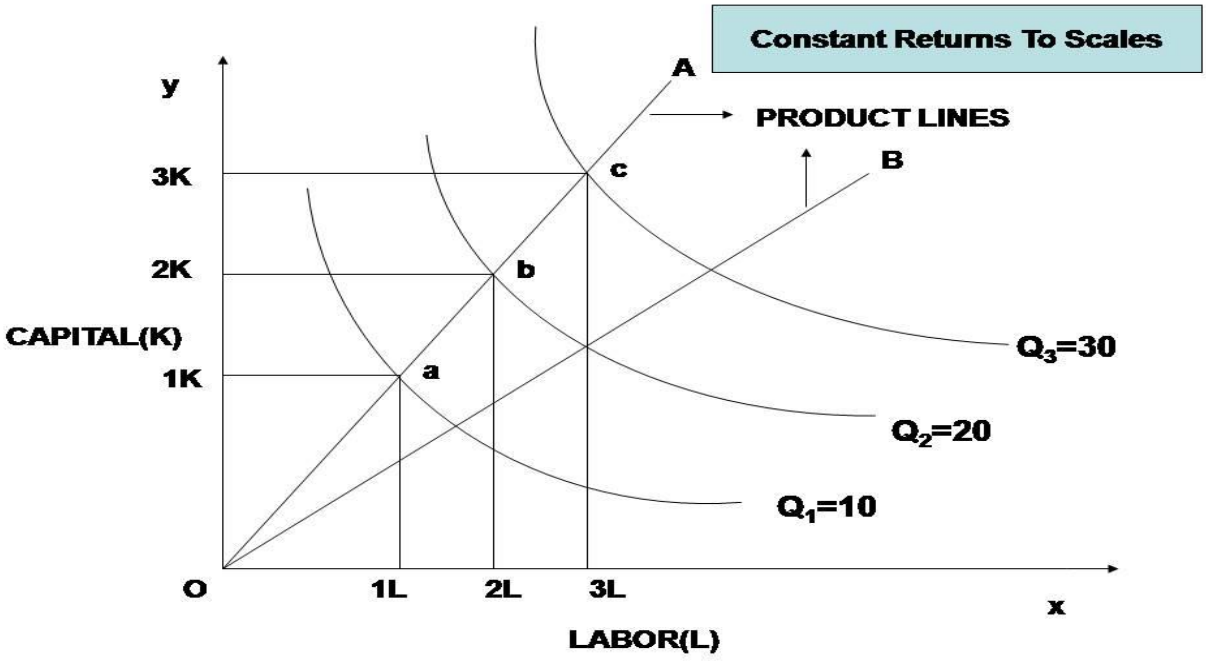
Non-homogeneous production function.

Graphical representation of Laws of returns to scale

The long run analysis of production or the 'Laws of returns to scale' we discussed earlier in this unit, can now be represented in terms of graphs using isoqunats and the expansion path. First we shall show the concept in three different diagrams and then we shall use one single diagram to represent the three concepts.

Constant Returns to scale

In this stage the scale of inputs and outputs change (increase or decrease) proportionately. This can be explained with the help of the following figure;

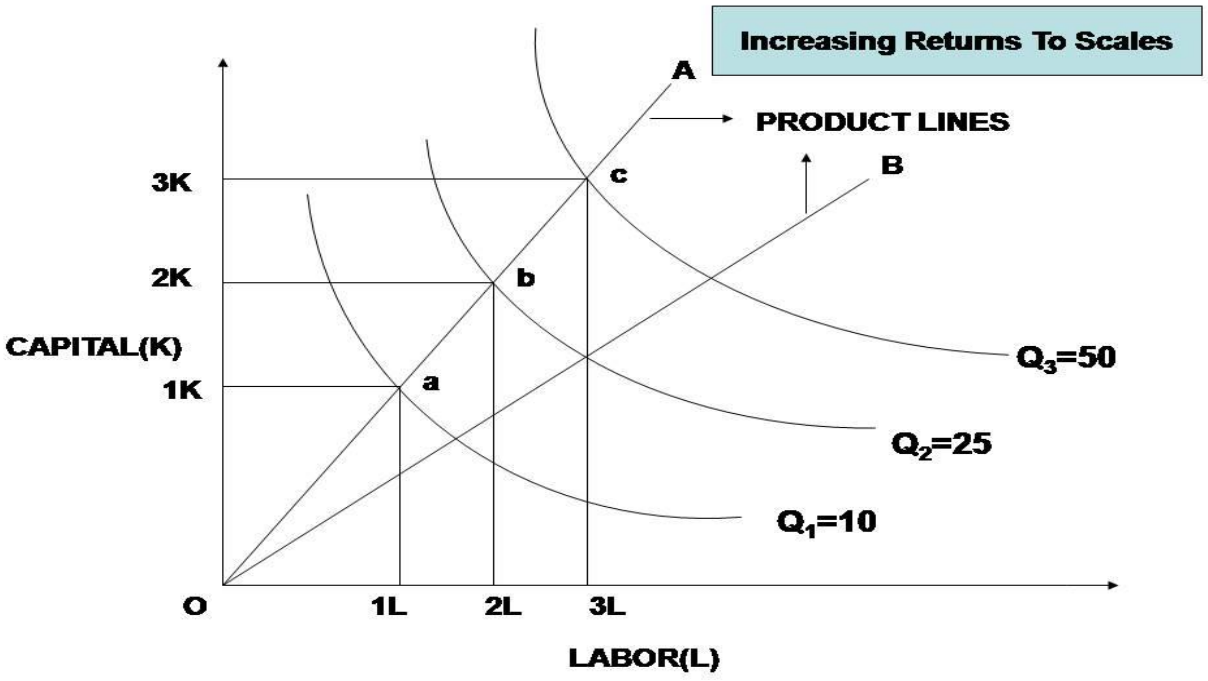


In the above figure product lines or expansion paths OA and OB indicate two hypothetical techniques of production and isoquants Q1 (10 units), Q2 (20 units) and Q3 (30 units) indicate three different levels of output. In the figure movement from point a to b indicates the doubling of both the inputs, from 1K + 1L to 2K + 2L. When inputs are doubled the outputs are also doubled, i.e. from 10 units to 20 units. Similarly, the movement from point b to c shows the increment in inputs from 2K + 2L to 3K + 3L, which is 50% increment. This 50% increment leads to the increment of output from 20 units to 30 units, which is also 50%. This kind of input output relationship exhibits the constants returns to scale.

Reasons for constant returns to scale : The constant returns to scale arise due to the limits of economies to scale (Detailed explanation of the concept of economies of scale / diseconomies of scale is given later in this chapter where we discuss cost of production. Please read that section for a proper understanding). The producers are unable to efficiently manage the inputs with gradual increase in scale. After certain time period when economies of scale end and diseconomies are yet to begin, the returns to scale appear to be constant. Various communication and coordination, management (personnel, financial, marketing) problems increase with increase in input and output, which leads to diseconomies. Constant returns to scale are transitional stage between increasing and decreasing returns to scale.

Increasing returns to scale

Increasing Returns to Scale prevails when output increases faster than inputs, i.e., percentage increase in output exceeds percentage increase in inputs. This implies that output increases more than proportionately to the increase in input and the rate of increase in output goes on increasing with each subsequent increase in input. For e.g. if all the inputs of production are increased by 100% the output increases by 150% and so on. In this kind of input-output relationship Increasing Returns to Scale exists. This can be explained with the help of the following diagram;



In the above diagram Q1, Q2 and Q3 are the isoquants showing three different levels of output – 10 units, 25 units and 50 units respectively. Product lines OA and OB show the relationship between inputs and outputs. The movement from point a to b indicates the increment of combination of inputs (labour and capital) from 1K+1L to 2K+2L. The movement also shows the increment in output from 10 units to 25 units. This shows that when inputs are increased by double the output increases by more than double, which explains the concept of increasing returns to scale. The case is same in the case of movement from point b to c as well.

Reasons for increasing return to scale: The increasing returns to scale is possible because of "economies of scale". (Detailed explanation of the concept of economies of scale / diseconomies of scale is given later in this chapter where we discuss cost of production. Please read that section for a proper understanding). The possible economies to scale are;

(a) Higher Degree of Specialisation: Due to increase in number of inputs, for e.g. labor and machines, higher degree of specialization of both labour and managerial cadre is possible. The use of specialized labour and management helps in increasing productivity per units of inputs by utilizing their cumulative efforts and thus contributes in increasing returns to scale.

(b) Technical and managerial indivisibilities: Most of the machines and equipments can be better used only in certain range of output. Such inputs, used in production process are given in a definite size and which cannot be divided into small parts to suit small scale productions. For example, half a turbine cannot be used, a part of locomotive machine cannot be used and similarly, half of a manager cannot be employed. Because of the indivisibility of these inputs, they have to be employed in a minimum quantity even if scale of production is much less than their capacity output. Therefore when scale of production is increased by increasing all inputs, the productivity of indivisible factor increases exponentially, this results in increasing returns to scale.

(c) Dimensional relations: In some cases, due to increased dimensions, output rises faster than inputs, which leads to increasing returns to scale. For instance let us consider an example of a tank manufacturer. When he uses 6 metal plates of 1 square feet each, he can produce a water tank of capacity 1 cubic feet. But when he uses 6 metal plates of 2 cubic feet each he can produce a water tank of capacity 8 cubic feet. In this example when inputs are doubled the output is 8 times, i.e. the concept of increasing returns to scale prevails in this example.

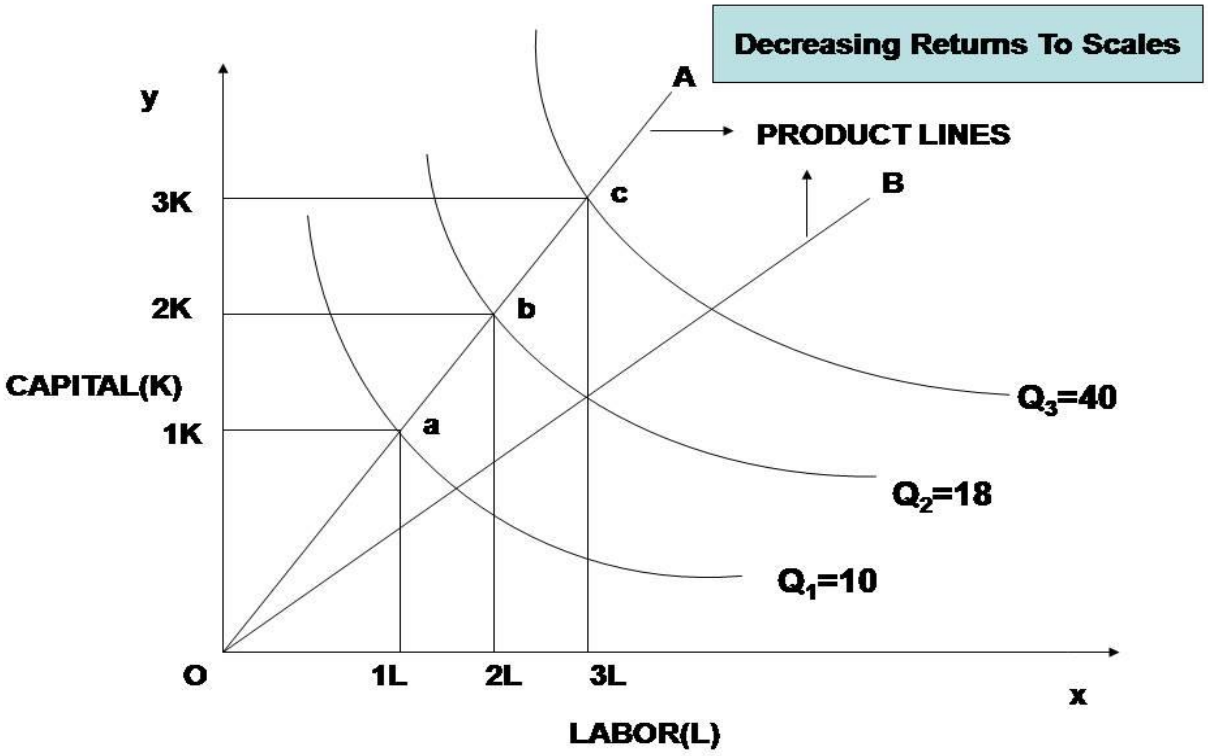
(d) Marketing economies: The greater requirements of inputs and the corresponding increase of outputs lead to various marketing economies. For e.g. when raw materials are purchased in bulk, the purchaser can purchase them at a cheaper price. Similarly suppliers also favour the bulk purchaser and the good quality raw materials are delivered timely. These factors finally help to increase output fast.

(e) Risk bearing economies: Big producers can bear more business risks than small producers. With increase in scale of inputs and outputs, risk bearing capacity also increases. Big firms can plan and diversify products and markets fast that are helpful to raise output fast.

Decreasing Returns to scale

The decreasing return to scale prevails when the output increases slower than inputs and vice-versa. Or we can say that when output increases less than proportionately to increase in inputs (capital and labour)

and the rate of rise in output goes on decreasing, it is called decreasing return to scale. This can also be explained with the help of the following figure;



In the above figure OA and OB are the product lines indicating two hypothetical techniques of production and isoquants Q1 (10 units), Q2 (18 units) and Q3 (40 units) indicate three different levels of output. When both the inputs are doubled, i.e. from 1K + 1L to 2K + 2L the output increases from 10 units to 18 units (that is 80% increase), which is less than the proportionate increase in inputs. Similarly the movement from point b to c indicates the increment in the inputs by 50%, whereas the increment in output is only 33.33%. This shows decreasing returns to scale.

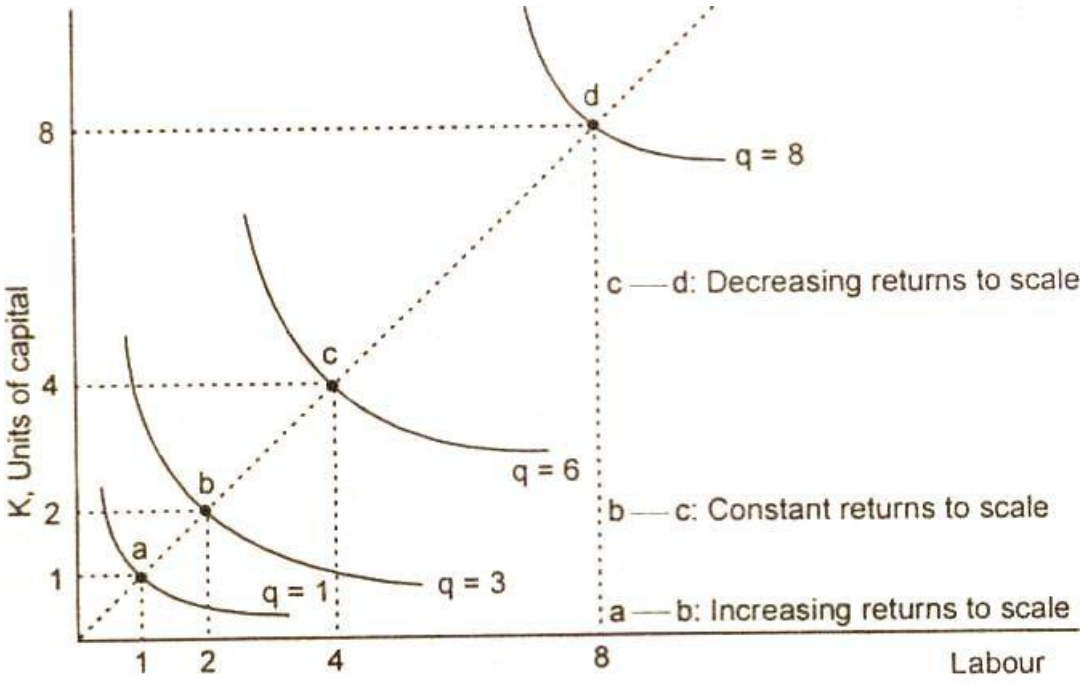
Reasons for decreasing returns to scale: Decreasing returns to scale arises mainly because of diseconomies of scale. Some of the diseconomies which cause decreasing returns to scale are;

(a) Managerial inefficiency: Diseconomies begin to start first at the management level. Managerial inefficiencies arise from expansion of scale itself, which eventually decreases the level of output.

(b) Exhaustibility of natural resources: It also leads to the decreasing returns to scale. For e.g. doubling the size of the coal mining plant does not double the coal output because of limitedness of coal deposits or difficult accessibility to coal deposits.

(c) Inefficient control: When the size of the firm is small the owner can efficiently handle and control all the departments individually. With increase in size of the firm (increase in inputs and outputs), various departments are created. Thereby controlling efficiency may decrease creating hindrances in production.

The following figure represents the three concepts of increasing, decreasing and constant returns to scale in a single diagram.



The figure shows that when a firm uses one unit of labour and one unit of capital, point a, it produces 1 unit of quantity as is shown on the q = 1 isoquant. When the firm doubles its outputs by using 2 units of labour and 2 units of capital, it produces more than double from q = 1 to q = 3.

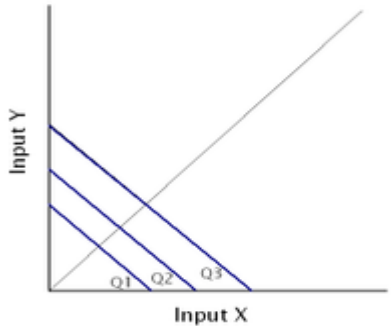
So the production function has increasing returns to scale in this range. Another output from quantity 3 to quantity 6. At the last doubling point c to point d, the production function has decreasing returns to scale. The doubling of output from 4 units of input, causes output to increase from 6 to 8 units increases of two units only.

Exceptional isoquants

The normal isoquant is convex to the origin. It is often referred to as the 'smooth convex isoquant', because it assumes continuous substitutability of the two factors K and L. However there are some exceptional isoquants. An isoquant is the locus of all the technically efficient methods for producing a given level of production. The production isoquant may assume various shapes depending on the degree of substitutability of factors.

(a) Linear Isoquant: In this case, the isoquant would be straight lines as in the figure. This type assumes perfect substitutability of factors of production. In this case, labour and capital are perfect substitutes, that is, the rate at which labour can be substituted for capital in production is constant. This implies that the MRTS between the two factors remain constant all along the isoquant, i.e, the assumption of DMRTS is not applicable here.

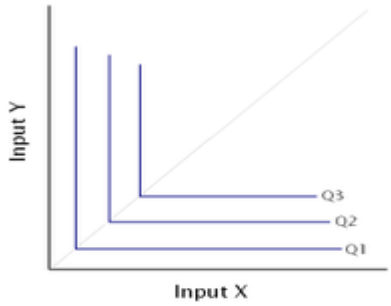
FIG



This isoquant envisages that a given commodity may be produced by using only capital or only labour or by an infinite combination of labour and capital. At point A on the isoquant the level of output can be produced with capital alone (i.e. without labour). Similarly, point B indicates that the same level of output can be produced with labour alone (i.e. without any capital). Though this may look unrealistic because capital and labour are not perfectly substitutable, there could be other situations where it is applicable. For example consider a production situation in a bakery where the oven can be heated using either LPG or electricity. Here LPG can be completely substituted by electricity.

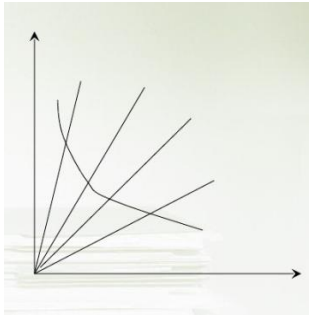
(b) Fixed Proportion Isoquant: This isoquant assumes zero substitutability of the factors of production. There is only one method of producing any one commodity. In this case, the isoquant takes the form of a right angle as in Fig.

Fig. 3.10: Right Angled Isoquant



In this case, labour and capital are perfect complements, that is, labour and capital must be used in fixed proportion shown by point C. The output can be increased only by increasing both the quantity of labour and capital in the same proportion depicted at the point C. This isoquant is called input-output isoquant or Leontief isoquant after Leontief, who invented the input-output analysis.

(c) Kinked Isoquant:



This isoquant assumes only limited substitutability of capital and labour. There are only a few processes for producing any one commodity. This is shown in Fig. where A1, A2, A3 and A4 show the production process and Q is the kinked isoquant. In this case, the substitutability of factors is possible only at the kinks. This type of isoquant is also called 'activity analysis iso quant' or 'linear programming isoquant' because it is basically used in linear programming problems. This is more realistic type of isoquant because engineers, managers and production executives consider the production process as a discrete rather than continuous process.

Now, our usual isoquant: - the Smooth Convex Isoquant: This type of isoquant assumes continuous substitutability of capital and labour over a certain range, beyond which the factors cannot substitute each other. (Recall the relevant regions and irrelevant regions we separated using ridge lines). The traditional economic theory has adopted this isoquant for analysis since it is uncomplicated. Further, this is an approximation to the more realistic form of a kinked isoquant because as the number of process become infinite, the isoquant becomes a smooth curve. Therefore, this isoquant is useful in the analysis of real world situations.

Test your understanding of the concepts covered in this unit by attempting the following questions:

1. Isoquants are equal revenue lines:

(a) True (b) False

2. Isoquant is sloping downward so when inputs are used in fixed proportion:

(a) True (b) False

3. Isoquant is also known as Production indifference curve:

(a) True (b) False

4. A higher isoquant represent lower level of output:

(a) True (b) False

5. Least cost input is a combination where the slope of isoquant is equal to the slope of isocost:

(a) True (b) False

Answer Key: 1. (b) False 2. (b) False 3. (a) True 4. (b) False 5. (a) True

Cost of Production

Meaning:

Cost is analysed from the producer's point of view. Cost estimates are made in terms of money. Cost calculations are indispensable for management decisions.

In the production process, a producer employs different factor inputs. These factor inputs are to be compensated by the producer for the services in the production of a commodity. The compensation is the cost. The value of inputs required in the production of a commodity determines its cost of output. Cost of production refers to the total money expenses (Both explicit and implicit) incurred by the producer in the process of transforming inputs into outputs. In short, it refers total money expenses incurred to produce a particular quantity of output by the producer. The knowledge of various concepts of costs, cost-output relationship etc. occupies a prominent place in cost analysis.

Managerial uses of Cost Analysis

A detailed study of cost analysis is very useful for managerial decisions. It helps the management in the following ways.

1. To find the most profitable rate of operation of the firm.

2. To determine the optimum quantity of output to be produced and supplied.

3. To determine in advance the cost of business operations.

4. To locate weak points in production management to minimize costs.

5. To fix the price of the product.

6. To decide what sales channel to use.

7. To have a clear understanding of alternative plans and the right costs involved in them.

8. To have clarity about the various cost concepts.

9. To decide and determine the very existence of a firm in the production field.

10. To regulate the number of firms engaged in production.

11. To decide about the method of cost estimation or calculations.

12. To find out decision making costs by re-classifications of elements, repricing of input factors etc, so as to fit the relevant costs into management planning, choice etc.

Different Kinds of Cost Concepts.

1. Money Cost and Real Cost

When cost is expressed in terms of money, it is called as money cost. It relates to money outlays by a firm on various factor inputs to produce a commodity. In a monetary economy, all kinds of cost estimations and calculations are made in terms of money only. .Hence, the knowledge of money cost is of great importance in economics. Exact measurement of money cost is possible.

When cost is expressed in terms of physical or mental efforts put in by a person in the making of a product, it is called as real cost. It refers to the physical, mental or psychological efforts, the exertions, sacrifices, the pains, the discomforts, displeasures and inconveniences which various members of the society have to undergo to produce a commodity. It is a subjective and relative concept and hence exact measurement is not possible.

2. Implicit or Imputed Costs and Explicit Costs

Explicit costs are those costs which are in the nature of contractual payments and are paid by an entrepreneur to the factors of production [excluding himself] in the form of rent, wages, interest and profits, utility expenses, and payments for raw materials etc. They can be estimated and calculated exactly and recorded in the books of accounts. Implicit or imputed costs are implied costs. They do not take the form of cash outlays and as such do not appear in the books of accounts. They are the earnings of owner-employed resources. For example, the factor inputs owned by the entrepreneur himself like capital that can be utilized by himself or can be supplied to others for a contractual sum if he himself does not utilize them in the business. It is to be remembered that the total cost is a sum of both implicit and explicit costs.

3. Actual costs and Opportunity Costs

Actual costs are also called as outlay costs, absolute costs and acquisition costs. They are those costs that involve financial expenditures at some time and hence are recorded in the books of accounts. They are the actual expenses incurred for producing or acquiring a commodity or service by a firm. For example, wages paid to workers, expenses on raw materials, power, fuel and other types of inputs. They can be exactly calculated and accounted without any difficulty.

Opportunity cost of a good or service is measured in terms of revenue which could have been earned by employing that good or service in some other alternative uses. In other words, opportunity cost of anything is the cost of displaced alternatives or costs of sacrificed alternatives. It implies that opportunity cost of anything is the alternative that has been foregone. Hence, they are also called as alternative costs. Opportunity cost represents only sacrificed alternatives. Hence, they can never be exactly measured and recorded in the books of accounts.

The knowledge of opportunity cost is of great importance to management decision. They help in taking decisions among alternatives. While taking a decision among several alternatives, a manager selects the best one which is more profitable or beneficial by sacrificing other alternatives. For example, a firm may decide to buy a computer which can do the work of 10 laborers. If the cost of buying a computer is much lower than that of the total wages to be paid to the workers over a period of time, it will be a wise decision. On the other hand, if the total wage bill is much lower than that of the cost of computer, it is better to employ workers instead of buying a computer. Thus, a firm has to take a number of decisions almost daily.

4. Direct costs and indirect costs

Direct costs are those costs which can be specifically attributed to a particular product, a department, or a process of production. For example, expenses on raw materials, fuel, wages to workers, salary to a divisional manager etc are direct costs. On the other hand, indirect costs are those costs, which are not traceable to any one unit of operation. They cannot be attributed to a product, a department or a process. For example, expenses incurred on electricity bill, water bill, telephone bill, administrative expenses etc.

5. Past and future costs

Past costs are those costs which are spent in the previous periods. On the other hand, future costs are those which are to be spent in the future. Past helps in taking decisions for future.

6. Marginal and Incremental costs

Marginal cost refers to the cost incurred on the production of another or one more unit. It implies additional cost incurred to produce an additional unit of output. It has nothing to do with fixed cost and is always associated with variable cost.

Incremental cost on the other hand refers to the costs involved in the production of a batch or group of output. They are the added costs due to a change in the level or nature of business activity. For example, cost involved in the setting up of a new sales depot in another city or cost involved in the production of another 100 extra units.

7. Fixed costs and variable costs.

Fixed costs are those costs which do not vary with either expansion or contraction in output. They remain constant irrespective of the level of output. They are positive even if there is no production. They are also called as supplementary or over head costs. On the other hand, variable costs are those costs which directly and proportionately increase or decrease with the level of output produced. They are also called as prime costs or direct costs.

8. Accounting costs and economic costs

Accounting costs are those costs which are already incurred on the production of a particular commodity. It includes only the acquisition costs. They are the actual costs involved in the making of a commodity. On the other hand, economic costs are those costs that are to be incurred by an entrepreneur on various alternative programs. It involves the application of opportunity costs in decision making.

Determinants of Costs

Cost behavior is the result of many factors and forces. But it is very difficult to determine in general the factors influencing the cost as they widely differ from firm to firm and even industry to industry. However, economists have given some factors considering them as general determinants of costs. They have enough importance in modern business set up and decision making process. The following factors deserve our attention in this connection.

1. Technology

Modern technology leads to optimum utilization of resources, avoid all kinds of wastages, saving of time, reduction in production costs and resulting in higher output. On the other hand, primitive technology would lead to higher production costs.

2. Rate of output: (the degree of utilization of the plant and machinery)

Complete and effective utilization of all kinds of plants and equipments would reduce production costs and under utilization of existing plants and equipments would lead to higher production costs.

3. Size of Plant and scale of production

Generally speaking big companies with huge plants and machineries organize production on large scale basis and enjoy the economies of scale which reduce the cost per unit.

4. Prices of factor inputs

Higher market prices of various factor inputs result in higher cost of production and vice-versa.

5. Efficiency of factors of production and the management

Higher productivity and efficiency of factors of production would lead to lower production costs and vice-versa.

6. Stability of output

Stability in production would lead to optimum utilization of the existing capacity of plants and equipments. It also brings savings of various kinds of hidden costs of interruption and learning leading to higher output and reduction in production costs.

7. Law of returns

Increasing returns would reduce cost of production and diminishing returns increase cost.

8. Time period

In the short run, cost will be relatively high and in the long run, it will be low as it is possible to make all kinds of adjustments and readjustments in production process.

Thus, many factors influence cost of production of a firm.

Cost-Output Relationship: Cost Function.

Cost and output are correlated. Cost output relations play an important role in almost all business decisions. It throws light on cost minimization or profit maximization and optimization of output. The relation between the cost and output is technically described as the "COST FUNCTION". The significance of cost-output relationship is so great that in economic analysis the cost function usually refers to the relationship between cost and rate of output alone and we assume that all other independent variables are kept constant. Mathematically speaking TC = f (Q) where TC = Total cost and Q stands for output produced.

However, cost function depends on three important variables.

1. Production function

If a firm is able to produce higher output with a little quantity of inputs, in that case, the cost function becomes cheaper and vice-versa.

2. The market prices of inputs

If market prices of different factor inputs are high in that case, cost function becomes higher and vice-versa.

3. Period of time

Cost function becomes cheaper in the long run and it would be relatively costlier in the short run.

Types of cost function

Generally speaking there are two types of cost functions.

1. Short run cost function.
2. Long run cost function.

Cost-Output Relationship and Cost Curves in the Short-Run

It is interesting to note that the relationship between the cost and output is different at two different periods of time i.e. short-run and long run. Generally speaking, cost of production will be relatively higher in the short-run when compared to the long run. This is because a producer will get enough time to make all kinds of adjustments in the productive process in the long run than in the short run. When cost and output relationship is represented with the help of diagrams, we get short run and long run cost curves of the firm. Now we shall make a detailed study of cost output relations both in the short-run as well as in the long run.

Meaning of Short Run

Short-run is a period of time in which only the variable factors can be varied while fixed factors like plant, machinery etc. remains constant. Hence, the plant capacity is fixed in the short run. The total number of firms in an industry will remain the same. Time is insufficient either for the entry of new firms or exit of the old firms. If a firm wants to produce greater quantities of output, it can do so only by employing more units of variable factors or by having additional shifts, or by having over time work for the existing labor force or by intensive utilization of existing stock of capital assets etc. Hence, short run is defined as a period where adjustments to changed conditions are only partial.

The short run cost function relates to the short run production function. It implies two sets of input components – (a) fixed inputs and (b) variable inputs. Fixed inputs are unalterable. They remain unchanged over a period of time. On the other hand, variable factors are changed to vary the output in the short run. Thus, in the short period some inputs are fixed in amount and a firm can expand or contract its output only by changing the amounts of other variable inputs. The cost-output relationship in the short run refers to a particular set of conditions where the scale of operation is limited by the fixed plant and equipment. Hence, the costs of the firm in the short run are divided into fixed cost and variable costs. We shall study these two concepts of costs in some detail

1. Fixed costs

These costs are incurred on fixed factors like land, buildings, equipments, plants, superior type of labor, top management etc. Fixed costs in the short run remain constant because the firm does not change the size of plant and the amount of fixed factors employed. Fixed costs do not vary with either expansion or contraction in output. These costs are to be incurred by a firm even output is zero. Even if the firm close down its operation for some time temporarily in the short run, but remains in business, these costs have to be borne by it. Hence, these costs are independent of output and are referred to as unavoidable contractual cost.

Prof. Marshall called fixed costs as supplementary costs. They include such items as contractual rent payment, interest on capital borrowed, insurance premiums, depreciation and maintenance allowances, administrative expenses like manager's salary or salary of the permanent staff, property and business taxes, license fees, etc. They are called as over-head costs because these costs are to be incurred whether there is production or not. These costs are to be distributed on each unit of output produced by a firm. Hence, they are called as indirect costs.

2. Variable costs

The cost corresponding to variable factors are discussed as variable costs. These costs are incurred on raw materials, ordinary labor, transport, power, fuel, water etc, which directly vary in the short run. Variable costs directly and proportionately increase or decrease with the level of output. If a firm shuts down for some time in the short run; then it will not use the variable factors of production and will not therefore incur any variable costs. Variable costs are incurred only when some amount of output is produced. Total variable costs increase with increase in the level of production and vice-versa. Prof. Marshall called variable costs as prime costs or direct costs because the volume of output produced by a firm depends directly upon them. It is clear from the above description that production costs consist of both fixed as well as variable costs. The difference between the two is meaningful and relevant only in the short run. In the long run all costs become variable because all factors of production become adjustable and variable in the long run. However, the distinction between fixed and variable costs is very significant in the short run because it influences the average cost behavior of the firm. In the short run, even if a firm wants to close down its operation but wants to remain in business, it will have to incur fixed costs but it must cover at least its variable costs.

Cost-output relationship and nature and behavior of cost curves in the short run

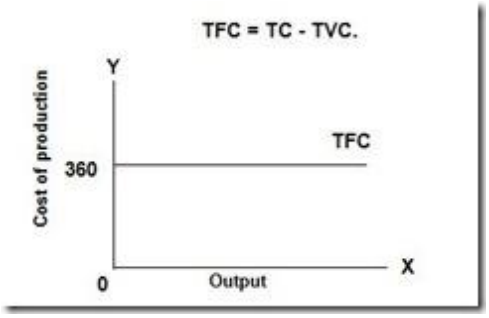
In order to study the relationship between the level of output and corresponding cost of production, we have to prepare the cost schedule of the firm. A cost-schedule is a statement of a variation in costs resulting from variations in the levels of output. It shows the response of cost to changes in output. A hypothetical cost schedule of a firm has been represented in the following table.

Output in Units	TFC	TVC	TC	AFC	AVC	AC	MC
0	360	–	360	–	–	–	–
1	360	180	540	360	180	540	180
2	360	240	600	180	120	300	60
3	360	270	630	120	90	210	30
4	360	315	675	90	78.75	168.75	45
5	360	420	780	72	84	156	105
6	360	630	990	60	105	165	210

On the basis of the above cost schedule, we can analyse the relationship between changes in the level of output and cost of production. If we represent the relationship between the two in a geometrical manner, we get different types of cost curves in the short run. In the short run, generally we study the following kinds of cost concepts and cost curves.

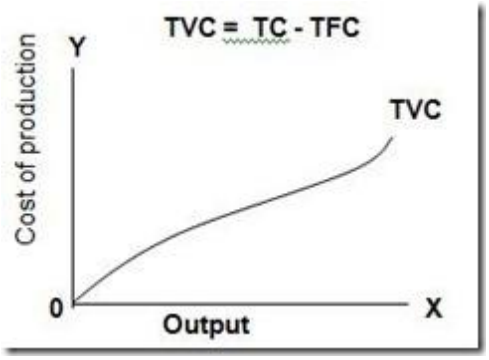
1. Total fixed cost (TFC)

TFC refers to total money expenses incurred on fixed inputs like plant, machinery, tools & equipments in the short run. Total fixed cost corresponds to the fixed inputs in the short run production function. TFC remains the same at all levels of output in the short run. It is the same when output is nil. It indicates that whatever may be the quantity of output, whether 1 to 6 units, TFC remains constant. The TFC curve is horizontal and parallel to OX-axis, showing that it is constant regardless of out put per unit of time. TFC starts from a point on Y-axis indicating that the total fixed cost will be incurred even if the output is zero. In our example, Rs 360=00 is TFC. It is obtained by summing up the product or quantities of the fixed factors multiplied by their respective unit price.



2. Total variable cost (TVC)

TVC refers to total money expenses incurred on the variable factor inputs like raw materials, power, fuel, water, transport and communication etc, in the short run. Total variable cost corresponds to variable inputs in the short run production function. It is obtained by summing up the production of quantities of variable inputs multiplied by their prices. The formula to calculate TVC is as follows. TVC = TC-TFC. TVC = f (Q) i.e. TVC is an increasing function of out put. In other words TVC varies with output. It is nil, if there is no production. Thus, it is a direct cost of output. TVC rises sharply in the beginning, gradually in the middle and sharply at the end in accordance with the law of variable proportion. The law of variable proportion explains that in the beginning to obtain a given quantity of output, relative variation in variable factors-needed are in less proportion, but after a point when the diminishing returns operate, variable factors are to be employed in a larger proportion to increase the same level of output. TVC curve slope upwards from left to right. TVC curve rises as output is expanded. When out put is Zero, TVC also will be zero. Hence, the TVC curve starts from the origin.

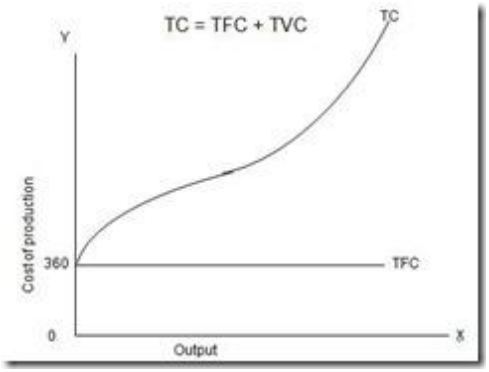


3. Total cost (TC)

The total cost refers to the aggregate money expenditure incurred by a firm to produce a given quantity of output. The total cost is measured in relation to the production function by multiplying the factor prices with their quantities. TC = f (Q) which means that the T.C. varies with the output.

Theoretically speaking TC includes all kinds of money costs, both explicit and implicit cost. Normal profit is included in the total cost as it is an implicit cost. It includes fixed as well as variable costs. Hence, TC = TFC +TVC.

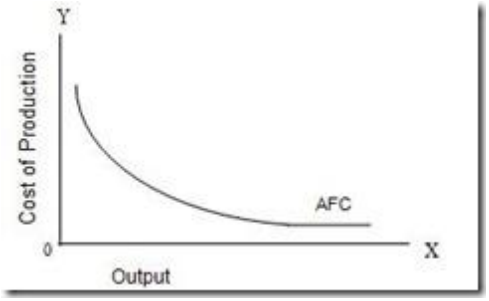
TC varies in the same proportion as TVC. In other words, a variation in TC is the result of variation in TVC since TFC is always constant in the short run.



The total cost curve is rising upwards from left to right. In our example the TC curve starts from Rs. 360-00 because even if there is no output, TFC is a positive amount. TC and TVC have same shape because an increase in output increases them both by the same amount since TFC is constant. TC curve is derived by adding up vertically the TVC and TFC curves. The vertical distance between TVC curve and TC curve is equal to TFC and is constant throughout because TFC is constant.

4. Average fixed cost (AFC)

Average fixed cost is the fixed cost per unit of output. When TFC is divided by total units of out put AFC is obtained**,** Thus, AFC = TFC/Q



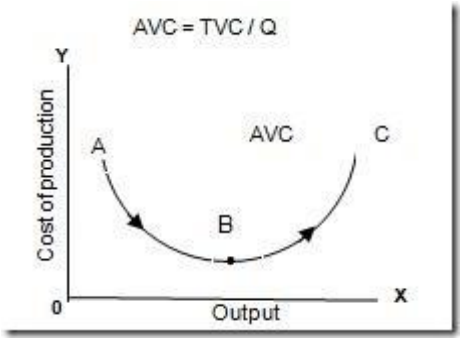
AFC and output have inverse relationship. It is higher at smaller level and lower at the higher levels of output in a given plant. The reason is simple to understand. Since AFC = TFC/Q, it is a pure mathematical result that the numerator remaining unchanged, the increasing denominator causes diminishing cost. Hence, TFC spreads over each unit of out put with the increase in output. Consequently, AFC diminishes continuously. This relationship between output and fixed cost is universal for all types of business concerns.

The AFC curve has a negative slope. The curve slopes downwards throughout the length. The AFC curve goes very nearer to X axis, but never touches axis. Graphically it will fall steeply in the beginning, gently in middle and tend to become parallel to OX-axis. Mathematically speaking as output increases, AFC diminishes. But AFC will never become zero because the TFC is a positive amount. AFC will never fall below a minimum amount because in the short run, plant capacity is fixed and output cannot be enlarged to an unlimited extent.

5. Average variable cost: (AVC)

The average variable cost is variable cost per unit of output. AVC can be computed by dividing the TVC by total units of output. Thus, AVC = TVC/Q. The AVC will come down in the beginning and then rise as more units of output are produced with a given plant. This is because as we add more units of variable factors in a fixed plant, the efficiency of the inputs first increases and then it decreases.

The AVC curve is a U-shaped cost curve. It has three phases.



a) Decreasing phase

In the first phase from A to B, AVC declines, As output expands, AVC declines because when we add more quantity of variable factors to a given quantity of fixed factors, output increases more efficiently and more than proportionately due to the operation of increasing returns.

b) Constant phase

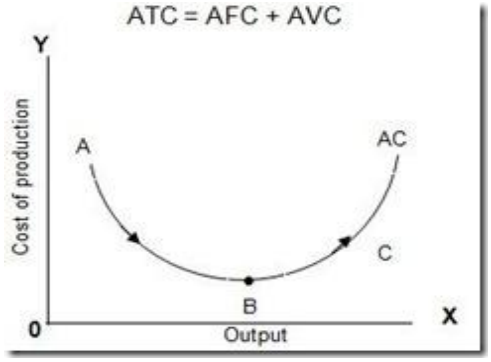
In the II phase, i.e. at B, AVC reaches its minimum point. When the proportion of both fixed and variable factors are the most ideal, the output will be the optimum. Once the firm operates at its normal full capacity, output reaches its zenith and as such AVC will become the minimum.

c) Increasing phase

In the III phase, from B to C, AVC rises when once the normal capacity is crossed, the AVC rises sharply. This is because additional units of variables factors will not result in more than proportionate output. Hence, greater output may be obtained but at much greater AVC. The old proverb "Too many cooks spoil the broth" aptly applies to this III stage. It is clear that as long as increasing returns operate, AVC falls and when diminishing returns set in, AVC tends to increase.

6. Average total cost (ATC) or Average cost (AC)

AC refers to cost per unit of output. AC is also known as the unit cost since it is the cost per unit of output produced. AC is the sum of AFC and AVC. Average total cost or average cost is obtained by dividing the total cost by total output produced. AC = TC/Q Also AC is the sum of AFC and AVC. In the short run AC curve also tends to be U-shaped. The combined influence of AFC and AVC curves will shape the nature of AC curve.



As we observe, average fixed cost begin to fall with an increase in output while average variable costs come down and rise. As long as the falling effect of AFC is much more than the rising effect of AVC, the AC tends to fall. At this stage, increasing returns and economies of scale operate and complete utilization of resources force the AC to fall. When the firm produces the optimum output, AC becomes minimum. This is called as least – cost output level. Again, at the point where the rise in AVC exactly counter balances the fall in AFC, the balancing effect causes AC to remain constant.

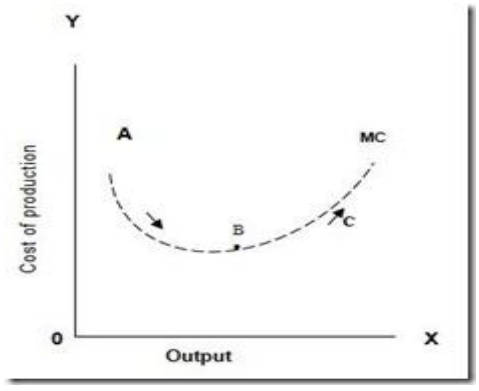
In the third stage when the rise in average variable cost is more than drop in AFC, then the AC shows a rise, When output is expanded beyond the optimum level of output, diminishing returns set in and diseconomies of scale starts operating. At this stage, the indivisible factors are used in wrong proportions. Thus, AC starts rising in the third stage. The short run AC curve is also called as "Plant curve". It indicates the optimum utilization of a given plant or optimum plant capacity.

7. Marginal Cost (MC)

Marginal Cost may be defined as the net addition to the total cost as one more unit of output is produced. In other words, it implies additional cost incurred to produce an additional unit. For example, if it costs Rs. 100 to produce 50 units of a commodity and Rs. 105 to produce 51 units, then MC would be Rs. 5. It is obtained by calculating the change in total costs as a result of a change in the total output. Also MC is the rate at which total cost changes with output. Hence, MC = D TC / D TQ. Where D TC stands for change in total cost and D TQ stands for change in total output.

Also $\mathbf{MC_n = TC_n - TC_{n-1}}$

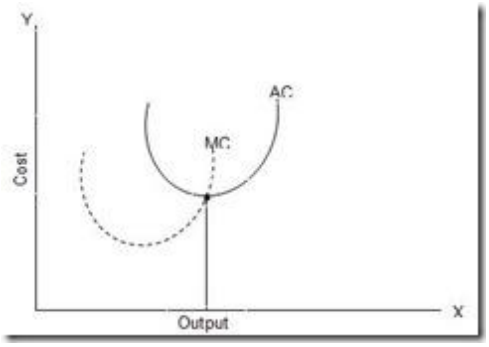
It is necessary to note that MC is independent of TFC and it is directly related to TVC as we calculate the cost of producing only one unit. In the short run, the MC curve also tends to be U-shaped. The shape of the MC curve is determined by the laws of returns. If MC is falling, production will be under the conditions of increasing returns and if MC is rising, production will be subject of diminishing returns.



The table indicates the relationship between AC & MC

Output in Units	TC in Rs.	AC in Rs.	Difference in Rs. MC
1	150	150	–
2	190	95	40
3	220	73.3	30
4	236	59	16
5	270	54	34
6	324	54	54
7	415	59.3	91
8	580	72.2	165

Relation between AC and MC



From the diagram it is clear that:

1. Both MC and AC fall at a certain range of output and rise afterwards.

2. When AC falls, MC also falls but at certain range of output MC tends to rise even though AC continues to fall. However, MC would be less than AC. This is because MC is attributed to a single unit where as in case of AC, the decreasing AC is distributed over all the units of output produced.

3. So long as AC is falling, MC is less than AC. Hence, MC curve lies below AC curve. It indicates that fall in MC is more than the fall in AC. MC reaches its minimum point before AC reaches its minimum.

4. When AC is rising, after the point of intersection, MC will be greater than AC. This is because in case of MC, the increasing MC is attributed to a single unit, where as in case of AC, the increasing AC is distributed over all the output produced.

5. So long as the AC is rising, MC is greater and AC. Hence, MC curve lies to the left side of the AC curve. It indicates that rise in MC is more than the rise in AC.

6. MC curve cuts the AC curve at the minimum point of the AC curve. This is because, when MC decreases, it pulls AC down and when MC increases, it pushes AC up. When AC is at its minimum, it is neither being pulled down or being pushed up by the MC. Thus, When AC is minimum, MC = AC. The point of intersection indicates the least cost combination point or the optimum position of the firm. At output Q the firm is working at its "Optimum Capacity" with lowest AC. Beyond Q, there is scope for "Maximum Capacity" with rising cost.

Cost Output Relationship in the Long Run

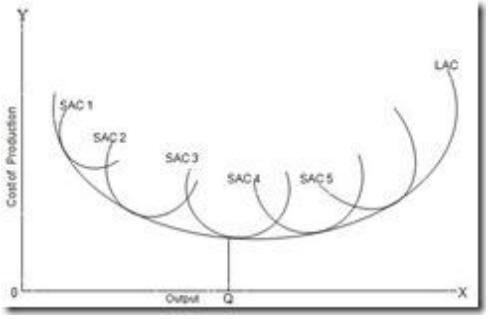
Long run is defined as a period of time where adjustments to changed conditions are complete. It is actually a period during which the quantities of all factors, variable as well as fixed factors can be adjusted. Hence, there are no fixed costs in the long run. In the short run, a firm has to carry on its production within the existing plant capacity, but in the long run it is not tied up to a particular plant capacity. If demand for the product increases, it can expand output by enlarging its plant capacity. It can construct new buildings or hire them, install new machines, employ administrative and other permanent staff. It can make use of the existing as well as new staff in the most efficient way and there is lot of scope for making indivisible factors to become divisible factors. On the other hand, if demand for the product declines, a firm can cut down its production permanently. The size of the plant can also be reduced and other expenditure can be minimized. Hence, production cost comes down to a greater extent in the long run.

As all costs are variable in the long run, the total of these costs is total cost of production. Hence, the distinction between fixed and variables costs in the total cost of production will disappear in the long run. In the long run only the average total cost is important and considered in taking long term output decisions.

The LAC curve

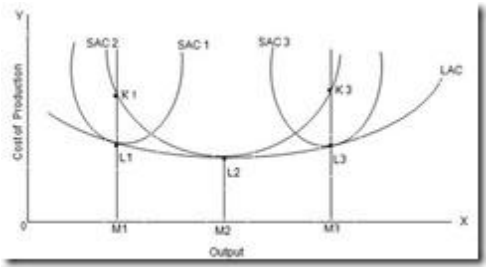
Long run average cost is the long run total cost divided by the level of output. In brief, it is the per unit cost of production of different levels of output by changing the size of the plant or scale of production.

The long run cost – output relationship is explained by drawing a long run cost curve through short – run curves as the long period is made up of many short – periods as the day is made up of 24 hours and a week is made out of 7 days. This curve explains how costs will change when the scale of production is varied.



The long run-cost curves are influenced by the laws of return to scale as against the short run cost curves which are subject to the working of law of variable proportions.

In the short run the firm is tied with a given plant and as such the scale of operation remains constant. There will be only one AC curve to represent one fixed scale of output in the short run. In the long run as it is possible to alter the scale of production, one can have as many AC curves as there are changes in the scale of operations. In order to derive LAC curve, one has to draw a number of SAC curves, each curve representing a particular scale of output. The LAC curve will be tangential to the entire family of SAC cures. It means that it will touch each SAC curve at its minimum point.



Production cost difference in the short run and long run.

In the diagram, the LAC curve is drawn on the basis of three possible plant sizes. Consequently, we have three different SAC curves – SAC1, SAC2 and SAC3. They represent three different scales of output. For output OM3 the AC will be L2M2 in the short run as well as the long run. When output is to be expanded to OM3, it can be obtained at a higher average cost of production. K3, M3 is the short run AC because, scale of production would remain constant in the short run. But the same output of OM3 can be produced at a lower AC of L3M3 in the long run since the scale of production can be modified according to the requirements. The distance between K3L3 represent difference between the cost of production in the short run and long run.

Similarly, when output is contracted to OM1 in the short run, K1M1 will become the short run AC and L1M1 will be the long run AC. Hence, K1L1 indicates the difference between short run and long run cost of production. If we join points L1, L2 and L3 we get LAC curve.

Important features of long run AC curves

1. Tangent curve

Different SAC curves represent different operational capacities of different plants in the short run. LAC curve is locus of all these points of tangency. The SAC curve can never cut a LAC curve though they are tangential to each other. This implies that for any given level of output, no SAC curve can ever be below the LAC curve. Hence, SAC cannot be lower than the LAC in the long run. Thus, LAC curve is tangential to various SAC curves.

2. Envelope curve

It is known as Envelope curve because it envelopes a group of SAC curves appropriate to different levels of output.

3. Flatter U-shaped or dish-shaped curve

The LAC curve is also U shaped or dish shaped cost curve. But It is less pronounced and much flatter in nature. LAC gradually falls and rises due to economies and diseconomies of scale.

4. Planning curve

The LAC cure is described as the Planning Curve of the firm because it represents the least cost of producing each possible level of output. This helps in producing optimum level of output at the minimum LAC. This is possible when the entrepreneur is selecting the optimum scale plant. Optimum scale plant is that size where the minimum point of SAC is tangent to the minimum point of LAC.

5. Minimum point of LAC curve should be always lower than the minimum point of SAC curve

This is because LAC can never be higher than SAC or SAC can never be lower than LAC. The LAC curve will touch the optimum plant SAC curve at its minimum point.

A rational entrepreneur would select the optimum scale plant. Optimum scale plant is that size at which SAC is tangent to LAC, such that both the curves have the minimum point of tangency. In the diagram, OM2 is regarded as the optimum scale of output, as it has the least per unit cost. At OM2 output LAC = SAC.

LAC curve will be tangent to SAC curves lying to the left of the optimum scale or right side of the optimum scale. But at these points of tangency, neither LAC is minimum nor will SAC be minimum. SAC curves are either rising or falling indicating a higher cost

Managerial Use of LAC

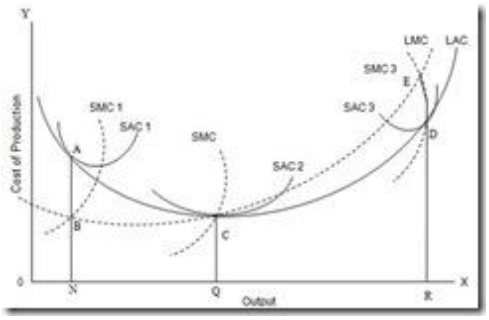
The study of LAC is of greater importance in managerial decision making process.

1. It helps the management in the determination of the best size of the plant to be constructed or when a new one is introduced in getting the minimum cost output for a given plant. But it is interested in producing a given output at the minimum cost.

2. The LAC curve helps a firm to decide the size of the plant to be adopted for producing the given output. For outputs less than cost lowering combination at the optimum scale i.e., when the firm is working subject to increasing returns to scale, it is more economical to under use a slightly large plant operating at less than its minimum cost – output than to over use smaller unit. Conversely, at output beyond the optimum level, that is when the firm experience decreasing return to scale, it is more economical to over use a slightly smaller plant than to under use a slightly larger one. Thus, it explains why it is more economical to over use a slightly small plant rather than to under use a large plant.

3. LAC is used to show how a firm determines the optimum size of the plant. An optimum size of plant is one that helps in best utilization of resources in the most economical manner.

Long Run Marginal cost



A long-run marginal cost curve can be derived from the long-run average cost curve. Just as the SMC is related to the SAC, similarly the LMC is related to the LAC and, therefore, we can derive the LMC directly from the LAC. In the diagram we have taken three plant sizes (for the sake of simplicity) and the corresponding three SAC and SMC curves. The LAC curve is drawn by enveloping the family of SAC curves. The points of tangency between the SAC and the LAC curves indicate different outputs for different plant sizes.

If the firm wants to produce ON output in the long run, it will have to choose the plant size corresponding to SAC1. The LAC curve is tangent to SAC1 at point A. For ON output, the average cost is NA and the corresponding marginal cost is NB If LAC curve is tangent to SAC1 curve at point A, the corresponding LMC curve will have to be equal to SMC1 curve at point B. The LMC will pass through point **B**. In other words, where LAC is equal to SAC curve (for a given output) the LMC will have to be equal to a given SMC.

If output OQ is to be produced in the long run, it will be done at point c which is the point of tangency between SAC2 and the LAC. At point C, the short –run average cost (SAC2) and the short-run marginal cost (SMC2) are equal and, therefore, the LAC for output OQ is QC and the corresponding LMC is also QC. The LMC curve will, therefore pass through point C.

Finally, for output OR, at point D the LAC is tangent to SAC3. For OR output at point E LMC is passing through SMC3. By connecting points B ,C and E, we can draw the long-run marginal cost curve.

Economies of Scale and Economies of Scope, Modern Cost Curves

Derivation of LAC and LMC

In the long run, the firm has no fixed commitments and so all long run costs are variable. The firm faces a long run production function and a cost equation, which in the two-input cases of the previous chapter are of the following type:

$Q = f(L, K)$

$TC = LP_L + KP_K$

If we recall the concepts of returns to scale, and assumes fixed factor prices, we can see three things:

(a) When returns to scale are increasing, LTC increases as output increases but at a less than proportionate rate

(b) When returns to scale are constant, LTC and output move in the same direction and same proportion

(c) When returns to scale are decreasing, LTC increases at a faster rate than does output

Thus, depending upon the nature of returns to scale, there will be a relationship between LTC and output, given factor prices. It is generally found that most industries and firms reap increasing returns to scale to start with, which are followed up by constant returns to scale, which give place to decreasing returns to scale eventually. This is primarily because of the indivisibility of the most efficient plants, equipments and personnel and the degree of specialization permitted by plant size. When the scale of operation is small, the firm is unable to take advantage of the most sophisticated technology, reflected in a large plant, highly competent management and a high degree of division of labour.

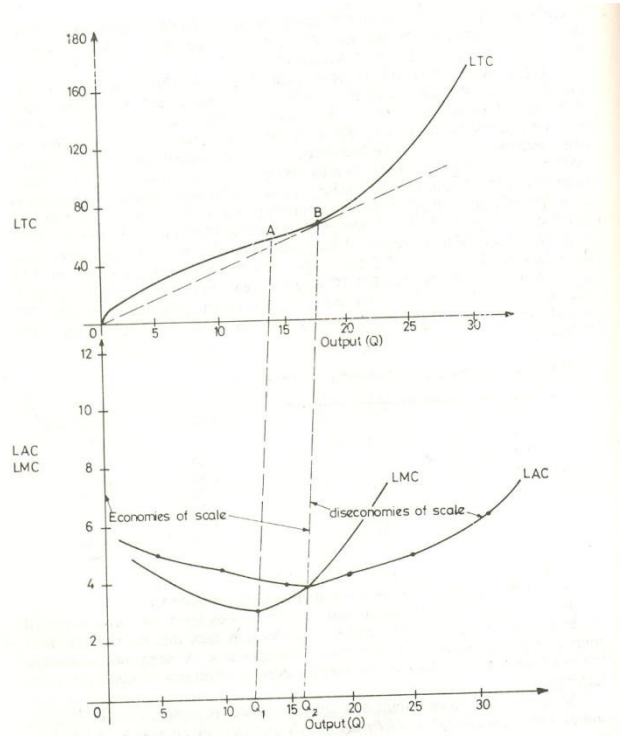
For example, a farmer with one acre of land may not be able to take advantage of a tractor and a tube well, but as he grows bigger he will be able to introduce such efficient tools of farming, and thereby reap increasing returns to scale. However, once he reaches the scale of operation in the whereabout of 50 acres of land, no better technology remains to be introduced and management of large resources under one roof might pose problems, thereby giving rise to decreasing returns to scale. Similarly, a small industrial enterprise may not find it worth while (or feasible) to go in for sophisticated plant and management personnel but as it grows, it may go in for such things to get the advantage of increasing returns to scale,

which, after a stage would take of to the region of decreasing returns to scale. In such a situation, the relationship between LTC and output will be of a changing character. As output expands, in the beginning LTC would increase but less than proportionately; after a while, LTC would increase and at a proportionate rate, and eventually LTC would increase more than proportionately. The following table gives an hypothetical example.

Q	*LTC*	*LAC*	*LMC(arc)*
0	0		
5	25	5.00	5
10	45	4.50	4
15	60	4.00	3
20	85	4.25	5
25	120	4.80	7
30	180	6.00	12

LMC (arc) = $\Delta LTC / \Delta Q$

The graphs of the above relationship are provided in the following figure.



The LTC curve gives the least total cost for various levels of output when all the factors of production are variable. Its shape is such that the curve is first concave and then convex as looked from the output axis. As seen in the table, its shape follows from the operations of the varying degrees of returns of scale, given the factor prices. The relationship between total, average and marginal is

mathematical in character and as with regard to TPP_L. APP_L and MPP_L curves, the shapes of *LAC* and *LMC* follow from that of *LTC* curve. Both *LAC* and *LMC* are U-shaped. Further, the following relationships hold good.

(a) At the point of inflexion on LTC curve (point A), LMC takes the minimum value.

(b) At the point of kink on LTC curve (point B) - where the slope of the straight line from origin to the LTC curve is the minimum - LAC assumes the minimum value.

(c) LAC is the least when LMC = LAC

(d) LAC curve is falling when LMC < LAC

(e) LAC curve is rising when LMC > LAC

The foregoing cost-output relationship assumes constant factor prices. However, one can conceive of a large firm, expansion of which could exercise some influence on factor prices: wage rate and capital rental. Let us examine the cost-output relationship in such a case. As the firm expands, it would hire more units of factors of production, which given the factor supply, would tend to increase factor prices. Recall that in the beginning, there are increasing returns to scale, then constant and finally decreasing returns to scale. Super imposing both the changing factor prices and the nature of returns to scale together, the three stages appear as follows:

(a) In the beginning, because of increasing returns to scale, cost increases less than proportionately as output increases but because of rising factor prices, cost increases more than proportionately as output expands. Thus, the two forces work in opposite directions and so the net effect is ambiguous. However, here LTC increases at a decreasing rate when the scale of operation is small.

(b) In the intermediate stage, where returns to scale are constant and factor prices are rising, LTC increases more than proportionately as output expands.

(c) In the third and final stage, where returns to scale are decreasing and factor prices are increasing, both the forces reinforce each other in favour of increasing LTC at an increasing rate.

Thus, the relationship between LTC and output under changing factor prices is similar to the one under constant factor prices. The difference between the two is in terms of extent (magnitude) and not in terms of kind. Under the changing factor prices, the decreasing rate of increase in LTC (or fall in LAC) is slower and increasing rate of increase in LTC (or rise in LAC) is faster than under constant factor prices. In consequence, LAC curve is U- shaped under both the situations.

The relationship between cost and output which we get under general situation, that is when all factors of production are variable (long-run) and so are factor prices, is often described under economies and diseconomies of scale. In other words, economies and diseconomies of scale combine the effects of returns to scale and factor prices on the relationship between cost and output. In what follows, we shall elaborate this rather important subject from the view point of decision-makers. We have seen above that the major reason for the 'U' shape of the cost curve is the economies and diseconomies of scale. Here we make a detailed discussion of the concepts.

Economies of Scale

The study of economies of scale is associated with large scale production. Today there is a general tendency to organize production on a large scale basis. Mass production of standardized goods has become the order of the day. Large scale production is beneficial and economical in nature. "The advantages or benefits that accrue to a firm as a result of increase in its scale of production are called 'Economies of Scale'. They have close relationship with the size of the firm. They influence the average cost over different ranges of output. They are gain to a firm. They help in reducing production cost and establishing an optimum size of a firm. Thus, they help a lot and go a long way in the development and growth of a firm. According to Prof. Marshall these economies are of two types, viz Internal Economies and External Economics. Now we shall study both of them in detail.

I. Internal Economies

Internal Economies are those economies which arise because of the actions of an individual firm to economize its cost. They arise due to increased division of labour or specialization and complete utilization of indivisible factor inputs. Prof. Cairncross points out that internal economies are open to a single factory or a single firm independently of the actions of other firms. They arise on account of an increase in the scale of output of a firm and cannot be achieved unless output increases. The following are some of the important aspects of internal economies.

1. They arise 'with in' or 'inside' a firm.
2. They arise due to improvements in internal factors.
3. They arise due to specific efforts of one firm.
4. They are particular to a firm and enjoyed by only one firm.
5. They arise due to increase in the scale of production.
6. They are dependent on the size of the firm.
7. They can be effectively controlled by the management of a firm.
8. They are called as "Business Secrets "of a firm.

Kinds of Internal Economies:

1. Technical Economies

These economies arise on account of technological improvements and its practical application in the field of business. Economies of techniques or technical economies are further subdivided into five heads.

a) Economies of superior techniques: These economies are the result of the application of the most modern techniques of production. When the size of the firm grows, it becomes possible to employ bigger and better types of machinery. The latest and improved techniques give place for specialized production. It is bound to be cost reducing in nature. For example, cultivating the land with modern tractors instead of using age old wooden ploughs and bullock carts, use of computers instead of human labour etc.

b) Economies of increased dimension: It is found that a firm enjoys the reduction in cost when it increases its dimension. A large firm avoids wastage of time and economizes its expenditure. Thus, an increase in dimension of a firm will reduce the cost of production. For example, operation of a double-decker instead of two separate buses.

c) Economies of linked process: It is quite possible that a firm may not have various processes of production within its own premises. Also it is possible that different firms through mutual agreement may decide to work together and derive the benefits of linked processes, for example, in dairy farming, printing press, nursing homes etc.

d) Economies arising out of research and by - products: A firm can invest adequate funds for research and the benefits of research and its costs can be shared by all other firms. Similarly, a large firm can make use of its wastes and by-products in the most economical manner by producing other products. For example, cane pulp, molasses, and bagasse of sugar factory can be used for the production of paper, varnish, distilleries etc.

e) Inventory Economies. Inventory management is a part of better materials management. A big firm can save a lot of money by adopting latest inventory management techniques. For example, Just-In-Time or zero level inventory techniques. The rationale of the Just-In-Time technique is that instead of having huge stocks worth of lakhs and crores of rupees, it can ask the seller of the inputs to supply them just before the commencement of work in the production department each day.

2. Managerial Economies:

They arise because of better, efficient, and scientific management of a firm. Such economies arise in two different ways.

a) Delegation of details: The general manager of a firm cannot look after the working of all processes of production. In order to keep an eye on each production process he has to delegate some of his powers or functions to trained or specialized personnel and thus relieve himself for co-ordination, planning and executing the plans. This will enable him to bring about improvements in production process and in bringing down the cost of production.

b) Functional Specialization: It is possible to secure economies of large scale production by dividing the work of management into several separate departments. Each department is placed under an expert and the rest of the work is left into the hands of specialists. This will ensure better and more efficient productive management with scientific business administration. This would lead to higher efficiency and reduction in the cost of production.

3. Marketing or Commercial economies:

These economies will arise on account of buying and selling goods on large scale basis at favourable terms. A large firm can buy raw materials and other inputs in bulk at concessional rates. As the bargaining capacity of a big firm is much greater than that of small firms, it can get quantity discounts and rebates. In this way economies may be secured in the purchase of different inputs. A firm can reduce its selling costs also. A large firm can have its own sales agency and channel. The firm can have a separate selling organization, marketing department manned by experts who are well versed in the art of pushing the products in the market. It can follow an aggressive sales promotion policy to influence the decisions of the consumers.

4. Financial Economies

They arise because of the advantages secured by a firm in mobilizing huge financial resources. A large firm on account of its reputation, name and fame can mobilize huge funds from money market,

capital market, and other private financial institutions at concessional interest rates. It can borrow from banks at relatively cheaper rates. It is also possible to have large overdrafts from banks. A large firm can float debentures and issue shares and get subscribed by the general public. Another advantage will be that the raw material suppliers, machine suppliers etc., are willing to supply material and components at comparatively low rates, because they are likely to get bulk orders. Thus, a big firm has an edge over small firms in securing sufficient funds more easily and cheaply.

5. Labour Economies

These economies will arise as a result of employing skilled, trained, qualified and highly experienced persons by offering higher wages and salaries. As a firm expands, it can employ a large number of highly talented persons and get the benefits of specialization and division of labour. It can also impart training to existing labour force in order to raise skills, efficiency and productivity of workers. New schemes may be chalked out to speed up the work, conserve the scarce resources, economize the expenditure and save labor time. It can provide better working conditions, promotional opportunities, rest rooms, sports rooms etc, and create facilities like subsidized canteen, crèches for infants, recreations. All these measures will definitely raise the average productivity of a worker and reduce the cost per unit of output.

6. Transport and Storage Economies

They arise on account of the provision of better, highly organized and cheap transport and storage facilities and their complete utilization. A large company can have its own fleet of vehicles or means of transport which are more economical than hired ones. Similarly, a firm can also have its own storage facilities which reduce cost of operations.

7. Over Head Economies

These economies will arise on account of large scale operations. The expenses on establishment, administration, book-keeping, etc, are more or less the same whether production is carried on small or large scale. Hence, cost per unit will be low if production is organized on large scale.

8. Economies of Vertical integration

A firm can also reap this benefit when it succeeds in integrating a number of stages of production. It secures the advantages that the flow of goods through various stages in production processes is more readily controlled. Because of vertical integration, most of the costs become controllable costs which help an enterprise to reduce cost of production.

9. Risk-bearing or survival economies

These economies will arise as a result of avoiding or minimizing several kinds of risks and uncertainties in a business. A manufacturing unit has to face a number of risks in the business. Unless these risks are effectively tackled, the survival of the firm may become difficult. Hence many steps are taken by a firm to eliminate or to avoid or to minimize various kinds of risks. Generally speaking, the risk-bearing capacity of a big firm will be much greater than that of a small firm. Risk is avoided when few firms amalgamate or join together or when competition between different firms is either eliminated or reduced to the minimum or expanding the size of the firm. A large firm secures risk-spreading advantages in either of the four ways or through all of them.

· **Diversification of output** Instead of producing only one particular variety, a firm has to produce multiple products. If there is loss in one item, it can be made good in other items.

· **Diversification of market**: Instead of selling the goods in only one market, a firm has to sell its products in different markets. If consumers in one market desert a product, it can cover the losses in other markets.

· **Diversification of source of supply**: Instead of buying raw materials and other inputs from only one source, it is better to purchase them from different sources. If one person fails to supply, a firm can buy from several sources.

· **Diversification of the process of manufacture**: Instead adopting only one process of production to manufacture a commodity, it is better to use different processes or methods to produce the same commodity so as to avoid the loss arising out of the failure of any one process.

II. External Economies

External economies are those economies which accrue to the firms as a result of the expansion in the output of whole industry and they are not dependent on the output level of individual firms. These economies or gains will arise on account of the overall growth of an industry or a region or a particular area. They arise due to benefit of localization and specialized progress in the industry or region. Prof. Stonier & Hague points out that external economies are those economies in production which depend on increase in the output of the whole industry rather than increase in the output of the individual firm The following are some of the important aspects of external economies.

1. They arise 'outside' the firm.

2. They arise due to improvement in external factors.

3. They arise due to collective efforts of an industry.

4. They are general, common & enjoyed by all firms.

5. They arise due to overall development, expansion & growth of an industry or a region.

6. They are dependent on the size of industry.

7. They are beyond the control of management of a firm.

8. They are called as "open secrets" of a firm.

Kinds of External Economies

1. Economies of concentration or Agglomeration

They arise because in a particular area a very large number of firms which produce the same commodity are established. In other words, this is an advantage which arises from what is called 'Localization of Industry'. The following benefits of localization of industry is enjoyed by all the firms- provision of better and cheap labor at low or reasonable rates, trained, educated and skilled labor, transport and communication, water, power, raw materials, financial assistance through private and public institutions at low interest rates, marketing facilities, benefits of common repairs, maintenance and service shops, services of specialists or outside experts, better use of by-products and other such benefits. Thus, it helps in reducing the cost of operation of a firm.

2. Economies of Information

These economies will arise as a result of getting quick, latest and up to date information from various sources. Another form of benefit that arises due to localization of industry is economies of information. Since a large number of firms are located in a region, it becomes possible for them to exchange their views frequently, to have discussions with others, to organize lectures, symposiums, seminars, workshops, training camps, demonstrations on topics of mutual interest. Revolution in the field of information technology, expansion in inter-net facilities, mobile phones, e-mails, video conferences, etc. has helped in the free flow of latest information from all parts of the globe in a very short span of time. Similarly, publication of journals, magazines, information papers etc have helped a lot in the dissemination of quick information. Statistical, technical and other market information becomes more readily available to all firms. This will help in developing contacts between different firms. When inter-firm relationship strengthens, it helps a lot to economize the expenditure of a single firm.

3. Economies of Disintegration

These economies will arise as a result of dividing one big unit in to different small units for the sake of convenience of management and administration. When an industry grows beyond a limit, in that case, it becomes necessary to split it in to small units. New subsidiary units may grow up to serve the needs of the main industry. For example, in cotton textiles industry, some firms may specialize in manufacturing threads, a few others in printing, and some others in dyeing and coloring etc. This will certainly enhance the efficiency in the working of a firm and cut down unit costs considerably.

4. Economies of Government Action

These economies will arise as a result of active support and assistance given by the government to stimulate production in the private sector units. In recent years, the government in order to encourage the development of private industries has come up with several kinds of assistance. It is granting tax-concessions, tax-holidays, tax-exemptions, subsidies, development rebates, financial assistance at low interest rates etc.

It is quite clear from the above detailed description that both internal and external economies arise on account of large scale production and they are benefits to a firm and cost reducing in nature.

5. Economies of Physical Factors

These economies will arise due to the availability of favorable physical factors and environment. As the size of an industry expands, positive physical environment may help to reduce the costs of all firms working in the industry. For example, Climate, weather conditions, fertility of the soil, physical environment in a particular place may help all firms to enjoy certain physical benefits.

6. Economies of Welfare

These economies will arise on account of various welfare programs under taken by an industry to help its own staff. A big industry is in a better position to provide welfare facilities to the workers. It may get land at concessional rates and procure special facilities from the local governments for setting up housing colonies for the workers. It may also establish health care units, training centres, computer centres and educational institutions of all types. It may grant concessions to its workers. All these measures would help in raising the overall efficiency and productivity of workers.

Diseconomies of Scale

When a firm expands beyond the optimum limit, economies of scale will be converted in to diseconomies of scale. Over growth becomes a burden. Hence, one should not cross the limit. On account of diseconomies of scale, more output is obtained at higher cost of production. The following are some of the main diseconomies of scale

1. Financial diseconomies. . As there is over growth, the required amount of fiancée may not be available to a firm. Consequently, higher interest rates are to be paid for additional funds**.**

2. Managerial diseconomies. Excess growth leads to loss of effective supervision, control, management, coordination of factors of production leading to all kinds of wastages, indiscipline and rise in production and operating costs.

3. Marketing diseconomies. Unplanned excess production may lead to mismatch between demand and supply of goods leading to fall in prices. Stocks may pile up, sales may decline leading to fall in revenue and profits.

4. Technical diseconomies. When output is carried beyond the plant capacity, per unit cost will certainly go up. There is a limit for division of labor and specialization. Beyond a point, they become negative. Hence, operation costs would go up.

5. Diseconomies of risk and uncertainty bearing. If output expands beyond a limit, investment increases. The level of inventory goes up. Sales do not go up correspondingly. Business risks appear in all fields of activities. Supply of factor inputs become inelastic leading to high prices.

6. Labour diseconomies. An unwieldy firm may become impersonal. Contact between labor and management may disappear. Workers may demand higher wages and salaries, bonus and other such benefits etc. Industrial disputes may arise. Labor unions may not cooperate with the management. All of them may contribute for higher operation costs.

II. External diseconomies. When several business units are concentrated in only one place or locality, it may lead to congestion,, environmental pollution, scarcity of factor inputs like, raw materials, water, power, fuel, transport and communications etc leading to higher production and operational costs.

Thus, it is very clear that a firm can enjoy benefits of large scale production only up to a limit. Beyond the optimum limit, it is bound to experience diseconomies of scale. Hence, there should be proper check on the growth and expansion of a firm.

Internalization of External Economies

It implies that a firm will convert certain external benefits created by the government or the entire society to its own favour with out making any additional investments. A firm may start a new unit in between two big railway stations or near the air port or near the national high ways or a port so that it can enjoy all the infrastructure benefits. Similarly, a new computer firm can commence its operations where there is 24 hours supply of electricity. Hence, they are also called as privatization of public benefits. Such type of efforts is to be encouraged by the government.

Externalization of Internal Diseconomies

In this case, a particular firm on account of its regular operations will pass on certain costs on the entire society. A firm instead of taking certain precautionary measures by spending some amount of money will escape and pass on this burden to the government or the society. For example, a firm may throw chemical or industrial wastes, dirt and filth either to open air or rivers leading to environmental pollution. In that case, the government is forced to spend more money to clean river water or prevent environmental pollution. This is a clear case of externalized internal diseconomies. It is to be avoided at all costs. Some economists divide economies of scale into another category. Viz, Real Economies or Pecuniary Economies.

Real Economies

Real economies are those associated with a reduction in the physical quantity of inputs, raw materials, various types of labour and various types of capital. Real economies are normally divided into the following four categories. (a) production economies (b) selling or marketing economies (c) managerial economies and (d) transport and storage economies. (We have discussed these concepts in detail already).

Pecuniary Economies

Pecuniary economies are economies realised from paying lower prices for the factors used in the production and distribution of the product, due to bulk buying by the firm as its size increases. Such economies are monetary in nature and they do not imply an actual decrease in the quantity of inputs used. It is experienced by the firm in the form of lower prices paid for raw materials (bought at discount due to bulk buying), lower interest rates, lower cost of finance and probably lower wages. Note that lower wages can happen only if the firm becomes so large as to acquire the power of a labour monopolist, i.e, the firm is able to provide all or most of the employment in the sector.

Economies of Scope

A closely related concept is economies of scope, where the economies arise out of diversity rather than size. It is a common factor to observe that when a single-product firm expands its volume of output, it would enjoy certain economies of scale. As a result, production cost per unit declines and more output is obtained at lower cost of production. Sometimes they would enjoy certain other external benefits due to the overall improvements in the entire area or city in which operates. Apart from these two types of benefits, we also come across another type of benefits in recent years. They are popularly known as economies of scope.

Economies of scope may be defined as those benefits which arise to a firm when it produces more than one product jointly rather than producing two items separately by two different business units. In this case, the benefits of the joint output of a single firm are greater than the benefits if two products are produced separately by two different firms. Such benefits may arise on account of joint use of production facilities, joint marketing efforts, or use of the same administrative office and staff in an organization. Sometimes, production of one product automatically results in the production of another by-product leading to a reduction in average cost of production.

Diseconomies of Scope

Diseconomies of scope may be defined as those disadvantages which occur when cost of producing two products jointly are costlier than producing them individually. In this case, it would be profitable to produce two goods separately than jointly. For example, with the help of same machinery, it is not possible to produce two goods together. It involves buying two different machineries. Hence in this case, production costs would certainly go up.

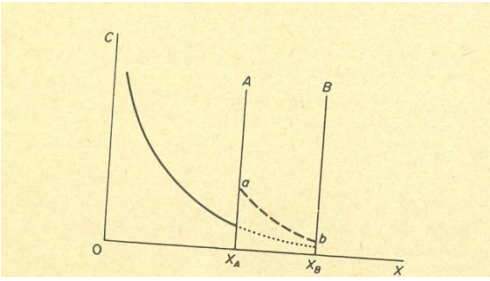
Introduction to Modern Cost Curves

The traditional theory of the firm has hypothesised U shaped cost curves. However many economists argue based on theoretical and empirical evidences that the cost curves are saucer shaped or 'L' shaped, rather than U shaped. This implies that the bottom portion where the costs are stable lasts longer than the traditional view.

(a) The average fixed cost curve

The modern theory assumes that the businessman will start his planning with a figure for the level of output which he anticipates selling, and he will choose the size of plant which will allow him to produce this level of output more efficiently and with maximum flexibility. So the plant will have capacity larger than the 'expected average' level of sales, because the businessman wants to have some reserve capacity. This is because businessman will want to be able to meet seasonal and cyclical fluctuations in his demand. Such fluctuations cannot always be met efficiently by a stock of inventory policy. This is possible only if a reserve capacity is available. This implies that the businessman will not choose the plant which will give him the lowest cost today, but rather that equipment which will allow him the greatest possible flexibility.

Under these circumstances the AFC curve will be as in the following figure.

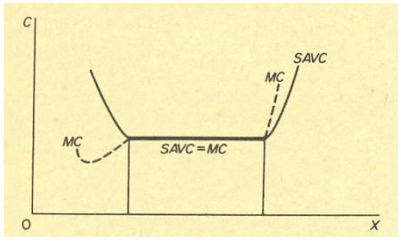


The firm has some 'largest-capacity' units of machinery which set an absolute limit to the shortrun expansion of output (boundary B in figure). The firm has also small-unit machinery, which sets a limit to expansion (boundary A in figure). This, however, is not an absolute boundary, because the firm can increase its output in the short run (until the absolute limit B is encountered), either by paying overtime to direct labour for working longer hours (in this case the AFC is shown by the dotted line in figure), or by buying some additional small-unit types of machinery (in this case the AFC curve shifts upwards, and starts falling again, as shown by the line ab in figure).

(b) The average variable cost

As in the traditional theory, the average variable cost of modern microeconomics includes the cost of: (a) direct labour which varies with output (b) raw materials (c) running expenses of machinery.

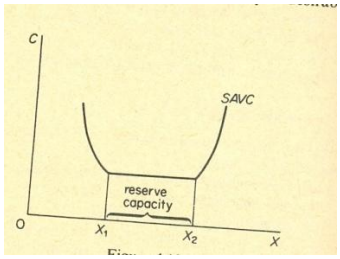
The SAVC in modern theory has a saucer-type shape, that is, it is broadly U-shaped but has a flat stretch over a range of output (see the following figure).



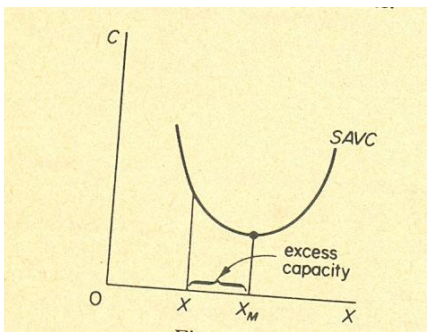
The flat stretch corresponds to the built-in-the-plant reserve capacity. Over this stretch the SAVC is equal to the MC, both being constant per unit of output. To the left of the flat stretch, MC lies below the SAVC, while to the right of the flat stretch the MC rises above the SAVC.

The falling part of the SAVC shows the reduction in costs due to the better utilisation of the fixed factor and the consequent increase in skills and productivity of the variable factor (labour). With better skills the wastes in raw materials are also being reduced and a better utilisation of the whole plant is reached. The increasing part of the SAVC reflects reduction in labour productivity due to the longer hours of work, the increase in cost of labour due to overtime payment (which is higher than the current wage), the wastes in materials and the more frequent breakdown of machinery as the firm operates with overtime or with more shifts.

The innovation of modem microeconomics in this field is the theoretical establishment of a short-run SAVC curve with a flat stretch over a certain range of output. The reserve capacity makes it possible to have constant SAVC within a certain range of output as shown in the following figure. (This figure is further explained after the next figure)



It should be clear that this reserve capacity is planned in order to give the maximum flexibility in the operation of the firm. It is completely different from the excess capacity which arises with the U-shaped costs of the traditional theory of the firm. The traditional theory assumes that each plant is designed without any flexibility; it ·is designed to produce optimally only a single level of output (X_M in the following figure)



If the firm produces an output X smaller than X_M there is excess (unplanned) capacity, equal to the difference $X_M - X$. This excess capacity is obviously undesirable because it leads to higher unit costs.

In the modern theory of costs the range of output X_1-X_2 in the figure reflects the planned reserve capacity which does not lead to increases in costs. The firm anticipates using its plant sometimes closer to X_1 and at others closer to X_2. On the average the entrepreneur expects to operate his plant within the X_1-X_2 range. Usually firms consider that the 'normal' level of utilisation of their plant is somewhere between two-thirds and three-quarters of their capacity, that is, at a point closer to X_2 than to X_1. The level of utilisation of the plant which firms consider as 'normal' is called 'the load factor' of the plant

(c)The average total cost

The average total cost is obtained by adding the average fixed (inclusive of the normal profit) and the average variable cost at each level of output. The ATC is shown in the following figure.

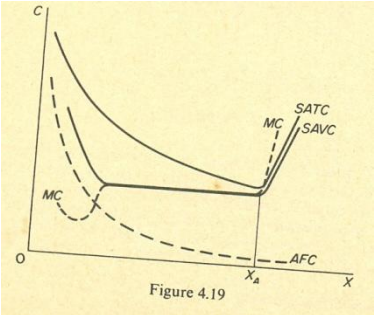


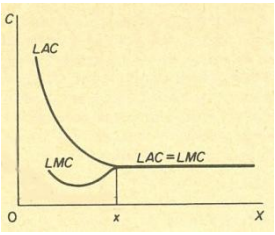
Figure 4.19

The ATC curve falls continuously up to the level of output (X_A) at which the reserve capacity is exhausted. Beyond that level ATC will start rising. The MC will intersect the average total cost curve at its minimum point.

Long Run costs in modern microeconomic theory:

The 'L' Shaped Scale Curve

All costs are variable in the long run and they give rise to a long-run cost curve which is roughly L -shaped. The production costs fall continuously with increases in output. At very large scales of output managerial costs may rise. But the fall in production costs more than offsets the increase in the managerial costs, so that the total *LAC* falls with in- creases in scale.



Engineering cost curves

Engineering costs are derived from engineering production functions. Each productive method is divided into sub-activities corresponding to the various physical-technical phases of production for the particular commodity. For each phase the quantities of factors of production are estimated and finally the cost of each phase is calculated on the basis of the prevailing factor prices. The total cost of the particular method of production is the sum of the costs of its different phases.

Concept of Revenue

Revenue function

Revenue function expresses revenue earned as a function of the price of good and quantity of good sold.

Let 'x' units be sold at Rs. 'p' per unit. Then the total revenue R(x) is defined as R(x) = px, where p and x are positive.

Eg. : TR = $200 – 6.5Q^2$

By 'revenue' of a firm is meant the total sale proceeds or the total receipts of a firm from the sale of the output.

The various kinds of revenue will be discussed here under three heads:

(i) Total Revenue, (ii) Marginal Revenue, (iii) Average Revenue.

(i) Total Revenue (TR):

Definition: By 'total revenue' of a firm is meant the total amount of sale proceeds or the total receipts of the firm.

Example: If a firm producing cloth sells one hundred meters of cloth in the market at Rs. 8 per meter, the sale proceeds or the receipts of the firm win be Rs. 1600. This total sale proceed which a firm has received by selling 200 meters of cloth is called its total revenue. The total revenue varies with the sales of a firm.

Formula: Total Revenue = Price × Quantity Sold **TR = P×Q**

TR = Rs. 8 × 200 = Rs. 1600

(ii) Marginal Revenue (MR):

Definition: Marginal Revenue (MR) is the change in total revenue (TR) due to the production of an additional unit of output. In other words, marginal revenue is the addition made to the total revenue by a one unit increase in the volume of sales by the firm in the market. It can also be called as the net revenue earned by selling on additional unit of output.

Example: For example, if a firm sells 200 meters of cloth at Rs. 8 per meters, the total revenue of the firm is Rs. 1600. If it increases the volume of sale from 200 meters to 201 meters, i.e., by one meter, the total revenue of the firm goes up to Rs. 1608. The addition of Rs. 8 which has taken place in the total revenue by a one unit increase in the rate of sales per period of time is known as marginal revenue. MR can be expressed as follows.

Formula: $MR = \frac{\Delta TR}{\Delta Q}$

MR can also be found using another formula

$MR = TR_n – TR_{n–1}$

where TR_n is total revenue of *n* units of output and $TR_{n–1}$ is total revenue of *n–1* units of output.

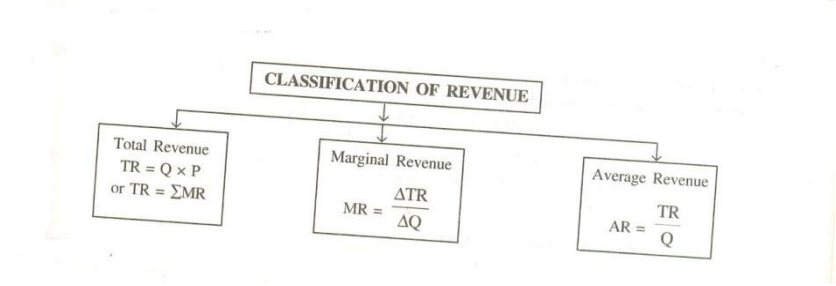
Suppose TR of 2 units is equal to Rs. 10 and TR of 3 units is Rs. 14. Then MR is 14 – 10 = 4.

(iii) Average Revenue (AR):

Definition: Average revenue is revenue earned per unit of output. Average revenue is obtained by dividing the total revenue by the number of units sold in the market.

Example: For example, a firm sells 200 meters of cloth for Rs. 600, then the average, revenue will be 600 / 200 = Rs. 3 only. Average revenue represents the average sale price per unit of the commodity. Average revenue curve can also be called demand curve.

Formula: $$\textbf{Average Revenue} = \frac{Total\ Revenue}{Quantity} = \frac{TR}{Q}$$

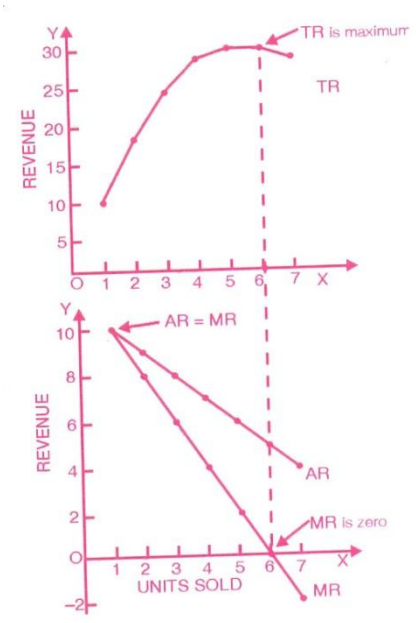


Relationship between TR, AR and MR

The following table shows the relationship between TR, AR and MR.

Quantity Sold (Q)	Price Per unit (P)	TR = P × Q	AR = TR / Q	MR = $TR_n - TR_{n-1}$
1	10	10	10	10
2	9	18	9	8
3	8	24	8	6
4	7	28	7	4
5	6	30	6	2
6	5	30	5	0
7	4	28	4	−2

The relationship between TR, AR and MR are graphically represented in the following figure.



Based on the above diagram, we can state the following relationship between the three curves.

(a) TR increases and keeps on rising as long as MR is positive. In the table this is happening till 6 units of output.

(b) When MR is zero, TR becomes maximum. In the table when 6 units of the commodity are sold, MR is zero and TR is maximum.

(c) When both average and marginal revenues are falling, MR revenue falls at a greater rate than AR. In other words, MR will remain below the AR.

Relationship between Average Revenue and Marginal Revenue

Average revenue is the revenue generated per unit of output sold. AR is obtained by dividing total revenue by quantity ($\frac{TR}{Q}$). Marginal revenue is the addition to total revenue by the sale of an additional unit of the commodity. MR is obtained by finding the first derivative of the TR function ($\Delta TR/\Delta Q$) . Since both AR and MR are obtained from the TR function, they are closely related.

The relation between AR and MR is the same as the several relations that reflect the general relation between a marginal and the corresponding average throughout the study of economics. The general relation is this:

- If the marginal is less than the average, then the average declines, in other words, when AR is falling, MR is less than AR.
- If the marginal is greater than the average, then the average rises, in other words, when AR is rising, MR is greater than AR.
- If the marginal is equal to the average, then the average does not change.

This general relation surfaces throughout the study of economics. It also applies to average and marginal cost as we will see in the next session, average and marginal product, average and marginal factor cost, average and marginal propensity to consume, and well, any other average and marginal encountered in economics.

A mathematical connection between average revenue and marginal revenue stating that the change in the AR depends on a comparison between AR and MR. The change in AR is different under different market conditions. For example, under perfect completion, where price remains the same for all the units sold, AR is equal to price / demand. Thus the equality between average revenue and marginal revenue occurs for a firm selling an output in a perfectly competitive market. This result in a single horizontal line for both AR and MR. They coincide because marginal revenue is equal to average revenue at every output quantity. However under markets with imperfect competition (monopoly, monopolistic competition etc), where prices fall as more units are sold, the average revenue curve facing in individual firm slopes downward. You have studied in your microeconomics course that under imperfect competition a firm Increase its sale by reducing price or decreases sale by increasing price. This results in a negatively sloped / downward slopping AR or demand curve. The MR curve is also negatively sloped. Marginal

revenue falling short of average revenue occurs for a firm selling an output in an imperfectly competitive market. Hence the MR curve lies below the AR curve in such markets. MR Curve lies below AR curve because Marginal revenue declines more rapidly than average revenue.

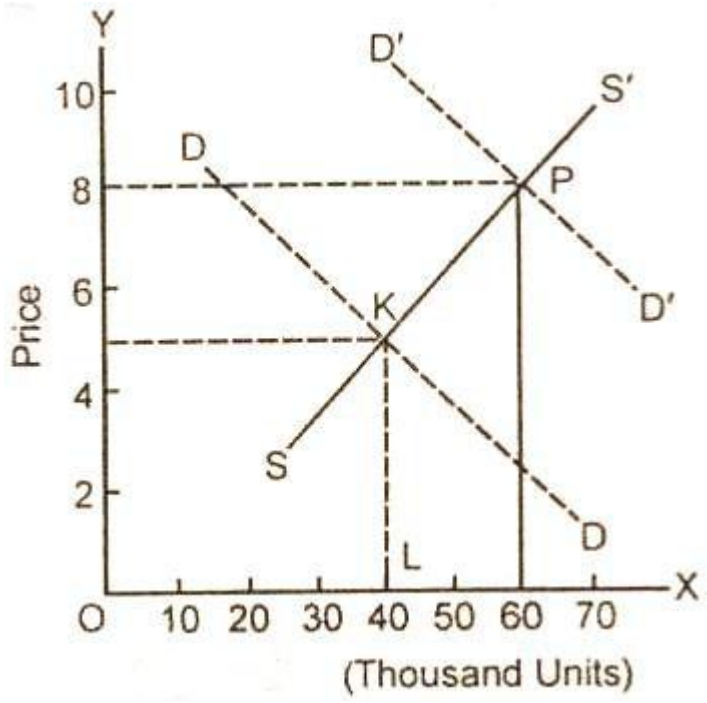
The relation between AR and MR differs according to the market form. We discuss below the relation between AR and MR under markets with competition and markets without competition.

(a) Revenue Curve of an Individual Firm Under perfect Competition: Perfect Competition

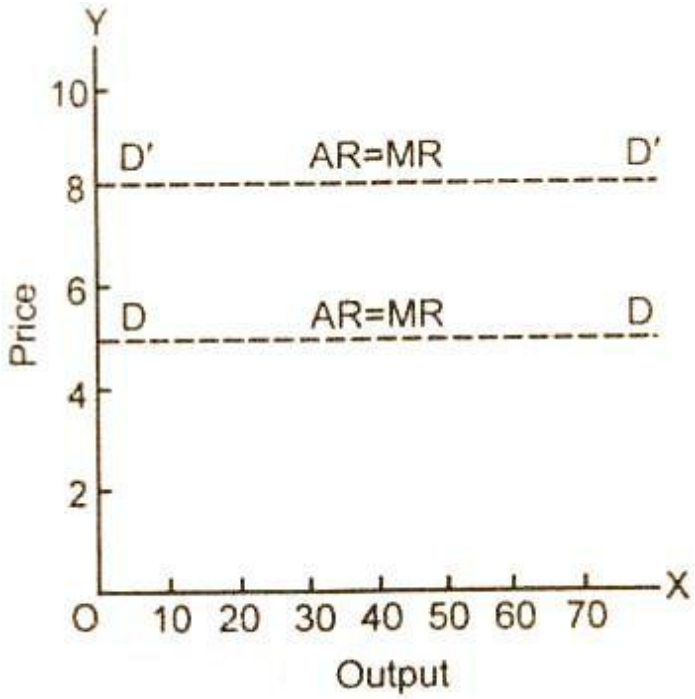
Perfect competition assumes that in a perfectly competitive market, the number of buyers and sellers is so large that an individual buyer or an individual seller cannot influence the market price.

A firm has to sell its products at the market price prevailing in the market. The buyers have also perfect knowledge of the quality and prices of the commodities which they wish to purchase. Similarly, a factor knows the reward which is paid to the similar factor in the country. In addition to these, the factors of production are perfectly mobile. They can freely move from one place to another place, from one occupation to another occupation, and no artificial barriers are imposed upon them by the state. The sellers sell identical and homogeneous goods.

Under the conditions stated above, there will be one price for the identical goods in all parts of the market. If any seller wishes to sell its goods at a price lower than the market price, its goods will be sold in no time as all the buyers have perfect knowledge of the market. If he keeps the price higher than the market price, the goods will not be sold. The seller in order to get the maximum profit will have to sell its total output at the prevailing market price as is shown in the two figures given below:



In the above figure markets demand and supply curves intersect at point K. KL, i.e. Rs.5 is the market price.

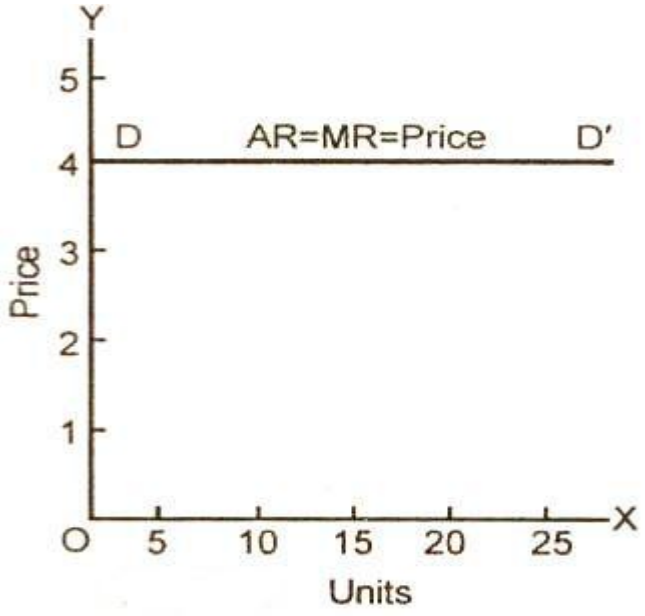


In the above figure DD is the demand curve which an individual firm has to face. A firm whether it produces 5 units or 50 units has to sell its product at the prevailing market price, i.e., at Rs. 5. If at any time the aggregate demand rises, and the price settles at PR (i.e., Rs.8), then an individual seller can sell its products at Rs.8. He will face the new demand curve D' D' as is shown in figure. Under perfect competition, the additional output is sold at the price at which, the first unit is sold. The average revenue curve is, therefore, always equal to marginal revenue and so both the curves AR and MR coincide.

For instance, when the market prices of a commodity is Rs. 5 per unit, the firm sells 10 units. The total revenue of the firm is Rs. 50. If it wishes to sell 11 units, an individual firm cannot alter the market price. So it has to sell the additional units also at Rs. 5. The total revenue of the firm by selling 11 units will be Rs. 5. The addition made to the total revenue by selling one more unit, i.e.. MR is Rs. 5. The average revenue is also found by dividing the total revenue by the number of goods sold Rs. (50 / 10 = 5, 55 / 11 = 5, 60 / 12 = 5). We therefore, find that in perfect competition marginal revenue, average revenue and price are the same. So these curves also coincide as is illustrated in the schedule and diagram.

Units	Price Per Unit (Rs.)	Total Revenue (Rs.)	Marginal (Rs.)	Average Revenue (Rs.)
10	5	50	5	5
11	5	55	5	5
12	5	60	5	5
13	5	65	5	5
14	5	70	5	5
15	5	75	5	5
16	5	80	5	5

The demand curve which a firm has to face in a perfect competitive market is a horizontal straight line parallel to the quantity axis. The MR and AR curves coincide with the price line DD'`. Here MR = AR = Price as is shown in figure 14.3.

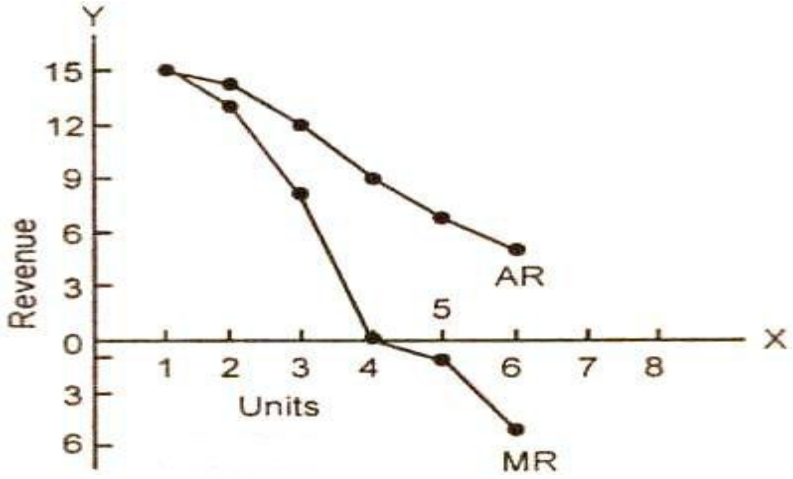


(b) Revenue Curve of an Individual Firm under Imperfect Competition:

Under imperfect competition, whether it may take the form of monopoly, duopoly or oligopoly, the demand curve facing the firm is negatively inclined or we can say its slopes downward from left to right. This means that a firm can affect the market price and can sell more goods at lower prices and less at a higher price.

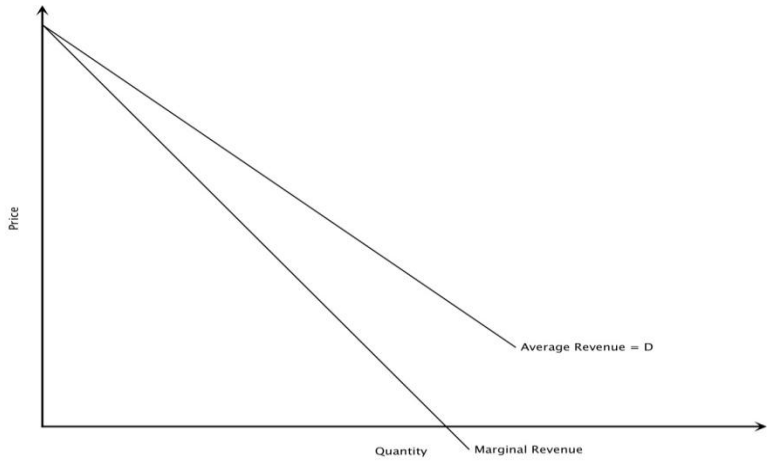
Under imperfect competition, the behavior of MR curve is that it lies below the AR curve. As production expands, the distance between the two curves increases. The AR line and the price line is the same as is clear from the schedule given below:

Units Sold	**Price (Rs.)**	**Total Revenue (Rs.)**	**Marginal Revenue (Rs.)**	**Average Revenue (Rs.)**
1	15	15	15	15
2	14	28	13	14
3	12	36	8	12
4	9	36	0	9
5	7	35	-1	7
6	5	30	-5	5

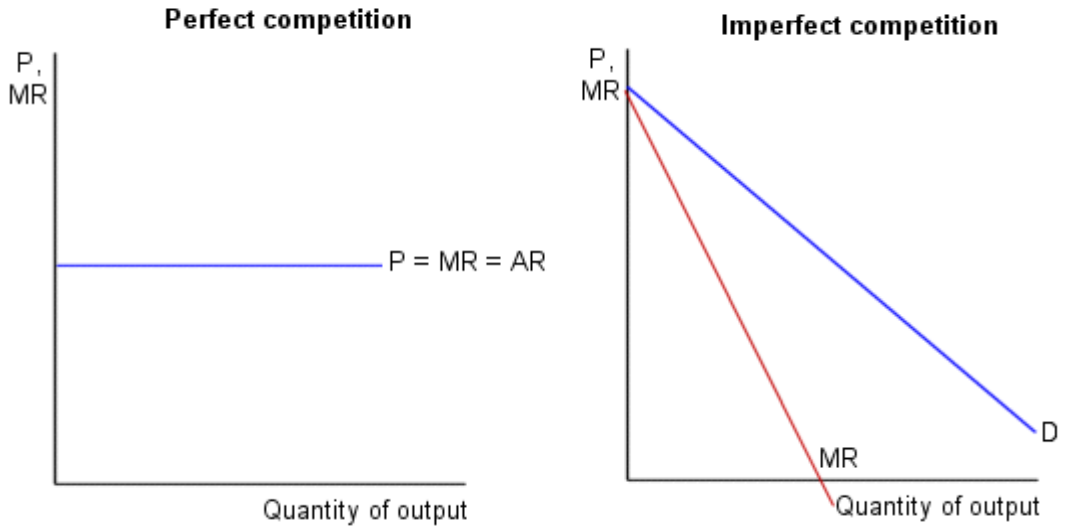


lt is clear from the above figure that average revenue curve and marginal revenue curve both have a negative slope. MR curve lies below the AR curve because the output is solid at the falling prices.

In general form the diagram is represented as follows.



The following figure gives a comparison of the AR and MR curves under the two situations.



Break Even Point

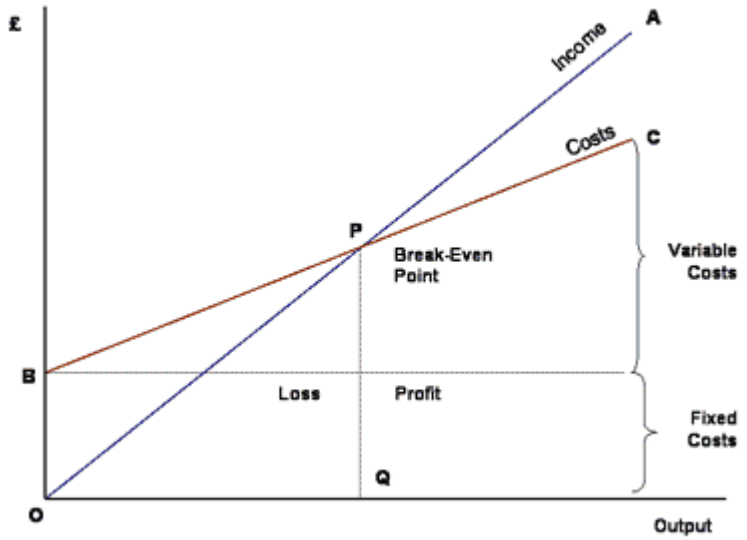
The Break-even Point is, in general, the point at which the gains equal the losses. A break-even point defines when an investment will generate a positive return. The point where sales or revenues equal expenses. Or also the point where total costs equal total revenues. There is no profit made or loss incurred at the break-even point. This is important for anyone that manages a business, since the break-even point is the lower limit of profit when prices are set and margins are determined.

Achieving Break-even today does not return the losses occurred in the past. Also it does not build up a reserve for future losses. And finally it does not provide a return on your investment (the reward for exposure to risk).

The Break-even method can be applied to a product, an investment, or the entire company's operations and is also used in the options world. In options, the Break-even Point is the market price that a stock must reach for option buyers to avoid a loss if they exercise. For a Call, it is the strike price plus the premium paid. For a Put, it is the strike price minus the premium paid.

The Break-Even Chart

In its simplest form, the break-even chart is a graphical representation of costs at various levels of activity shown on the same chart as the variation of income (or sales, revenue) with the same variation in activity. The point at which neither profit nor loss is made is known as the "break-even point" and is represented on the chart below by the intersection of the two lines:



In the diagram above, the line OA represents the variation of income at varying levels of production activity (‘output’). OB represents the total fixed costs in the business. As output increases, variable costs are incurred, meaning that total costs (fixed + variable) also increase. At low levels of output, Costs are greater than Income. At the point of intersection, P, costs are exactly equal to income, and hence neither profit nor loss is made.

Fixed Costs

Fixed costs are those business costs that are not directly related to the level of production or output. In other words, even if the business has a zero output or high output, the level of fixed costs will remain broadly the same. In the long term fixed costs can alter - perhaps as a result of investment in production capacity (e.g. adding a new factory unit) or through the growth in overheads required to support a larger, more complex business.

Examples of fixed costs:

- Rent and rates
- Depreciation
- Research and development
- Marketing costs (non- revenue related)
- Administration costs

Variable Costs

Variable costs are those costs which vary directly with the level of output. They represent payment output-related inputs such as raw materials, direct labour, fuel and revenue-related costs such as commission.

A distinction is often made between 'Direct' variable costs and 'Indirect' variable costs.

Direct variable costs are those which can be directly attributable to the production of a particular product or service and allocated to a particular cost centre. Raw materials and the wages those working on the production line are good examples.

Indirect variable costs cannot be directly attributable to production but they do vary with output. These include depreciation (where it is calculated related to output - e.g. machine hours), maintenance and certain labour costs.

Semi-Variable Costs

Whilst the distinction between fixed and variable costs is a convenient way of categorising business costs, in reality there are some costs which are fixed in nature but which increase when output reaches certain levels. These are largely related to the overall "scale" and/or complexity of the business. For example, when a business has relatively low levels of output or sales, it may not require costs associated with functions such as human resource management or a fully-resourced finance department. However, as the scale of the business grows (e.g. output, number people employed, number and complexity of transactions) then more resources are required. If production rises suddenly then some short-term increase in warehousing and/or transport may be required. In these circumstances, we say that part of the cost is variable and part fixed.

The relationship between fixed costs, variable costs and returns

Break-even analysis is a useful tool to study the relationship between fixed costs, variable costs and returns. The Break-even Point defines when an investment will generate a positive return. It can be viewed graphically or with simple mathematics. Break-even analysis calculates the volume of production at a given price necessary to cover all costs. Break-even price analysis calculates the price necessary at a given level of production to cover all costs. To explain how break-even analysis works, it is necessary to define the cost items.

Fixed costs, which are incurred after the decision to enter into a business activity is made, are not directly related to the level of production. Fixed costs include, but are not limited to, depreciation on equipment, interest costs, taxes and general overhead expenses. Total fixed costs are the sum of the fixed costs.

Variable costs change in direct relation to volume of output. They may include cost of goods sold or production expenses, such as labour and electricity costs, feed, fuel, veterinary, irrigation and other expenses directly related to the production of a commodity or investment in a capital asset. Total variable costs (TVC) are the sum of the variable costs for the specified level of production or output. Average variable costs are the variable costs per unit of output or of TVC divided by units of output.

Break-even analysis is a technique widely used by production management and management accountants. It is based on categorising production costs between those which are "variable" (costs that change when the production output changes) and those that are "fixed" (costs not directly related to the volume of production).

Total variable and fixed costs are compared with sales revenue in order to determine the level of sales volume, sales value or production at which the business makes neither a profit nor a loss (the "break-even point").

Assumption of Break Even Point:

The Break-even Analysis depends on three key assumptions:

1. Average per-unit sales price (per-unit revenue):

   This is the price that you receive per unit of sales. Take into account sales discounts and special offers. Get this number from your Sales Forecast. For non-unit based businesses, make the per-unit revenue Rs. 1 and enter your costs as a percent of a dollar. The most common questions about this input relate to averaging many different products into a single estimate. The analysis requires a single number, and if you build your Sales Forecast first, then you will have this number. The vast majority of businesses sell more than one item, and have to average for their Break-even Analysis.

2. Average per-unit cost:

   This is the incremental cost, or variable cost, of each unit of sales. If you buy goods for resale, this is what you paid, on average, for the goods you sell. If you sell a service, this is what it costs you, per dollar of revenue or unit of service delivered, to deliver that service. If you are using a Units-Based Sales Forecast table (for manufacturing and mixed business types), you can project unit costs from the Sales Forecast table. If you are using the basic Sales Forecast table for retail, service and distribution businesses, use apercentage estimate, e.g., a retail store running a 50% margin would have a per-unit cost of .5, and a per-unit revenue of 1.

3. Monthly fixed costs:

   Technically, a break-even analysis defines fixed costs as costs that would continue even if you went broke. Instead, we recommend that you use your regular running fixed costs, including payroll and normal expenses (total monthly Operating Expenses). This will give you a better insight on financial realities. If averaging and estimating is difficult, use your Profit and Loss table to calculate a working

fixed cost estimate - it will be a rough estimate, but it will provide a useful input for a conservative Break-even Analysis.

Computation of Break Even point:

Break even point is the level of sales at which profit is zero. According to this definition, at*break even point* sales are equal to fixed cost plus variable cost. This concept is further explained by the the following equation:

[Break even sales = fixed cost + variable cost]

The break even point can be calculated using either the equation method or contribution margin method. These two methods are equivalent.

Equation Method:

The equation method centres on the contribution approach to the income statement. The format of this statement can be expressed in equation form as follows:

Profit = (Sales − Variable expenses) − Fixed expenses

Rearranging this equation slightly yields the following equation, which is widely used in cost volume profit (CVP) analysis:

Sales = Variable expenses + Fixed expenses + Profit

According to the definition of break even point, break even point is the level of sales where profits are zero. Therefore the break even point can be computed by finding that point where sales just equal the total of the variable expenses plus fixed expenses and profit is zero.

Example:

For **example** we can use the following data to calculate break even point.

- Sales price per unit = Rs. 250
- variable cost per unit = Rs. 150
- Total fixed expenses = Rs. 35,000

Calculate break even point

Calculation:

Sales = Variable expenses + Fixed expenses + Profit

Rs. 250Q* = Rs. 150Q* + Rs. 35,000 + Rs. 0**

Rs. 100Q = Rs. 35000

Q = Rs. 35,000 /Rs. 100

Q = 350 Units

Q* = Number (Quantity) of units sold.

**The break even point can be computed by finding that point where profit is zero

The break even point in sales dollars can be computed by multiplying the break even level of unit sales by the selling price per unit.

350 Units × Rs. 250 Per unit = Rs. 87,500

Contribution Margin Method:

The contribution margin method is actually just a short cut conversion of the equation method already described. The approach centers on the idea discussed earlier that each unit sold provides a certain amount of contribution margin that goes toward covering fixed cost. To find out how many units must be sold to break even, divide the total fixed cost by the unit contribution margin.

Break even point in units = Fixed expenses / Unit contribution margin

Rs. 35,000 / Rs. 100* per unit

350 Units

*S250 (Sales) − Rs. 150 (Variable exp.)

A variation of this method uses the Contribution Margin ratio (CM ratio) instead of the **unit contribution margin**. The result is the break even in total sales dollars rather than in total units sold.

Break even point in total sales dollars = Fixed expenses / CM ratio

Rs. 35,000 / 0.40

= Rs. 87,500

This approach is particularly suitable in situations where a company has multiple products lines and wishes to compute a single break even point for the company as a whole.

The following formula is also used to calculate break even point

Break Even Sales in Dollars = [Fixed Cost / 1 – (Variable Cost / Sales)]

This formula can produce the same answer:

Break Even Point = [Rs. 35,000 / 1 – (150 / 250)]

= Rs. 35,000 / 1 – 0.6

= Rs. 35,000 / 0.4

= Rs. 87,500

Benefits / Advantages of Break Even Analysis:

The main advantages of break even point analysis is that it explains the relationship between cost, production, volume and returns. It can be extended to show how changes in fixed cost, variable cost, commodity prices, revenues will effect profit levels and break even points. Break even analysis is most useful when used with partial budgeting, capital budgeting techniques. The major benefits to use break even analysis is that it indicates the lowest amount of business activity necessary to prevent losses.

Limitations of Break Even Analysis:

It is best suited to the analysis of one product at a time. It may be difficult to classify a cost as all variable or all fixed; and there may be a tendency to continue to use a break even analysis after the cost and income functions have changed.

One useful tool in tracking your business's cash flow is a break-even analysis. It's a fairly simple calculation and can prove very helpful in deciding whether to make an equipment purchase or in knowing how close you are to your break-even level. Here are the variables needed to compute a break-even sales analysis:

- Gross profit margin
- Operating expenses (less depreciation)
- Annual debt service (total monthly debt payments for the year)

Since we're dealing with cash flow, and depreciation is a noncash expense, it's subtracted from the operating expenses. The break-even calculation for sales is:

(Operating Expenses + Annual Debt Service)/Gross Profit Margin = Break-Even Sales

Let's use ABC Clothing as an example and compute this company's break-even sales for years one and two. In Year 1, the company's sales were Rs. 1 million and their gross profit was Rs. 250,000, resulting in a gross profit margin of 25 percent (Rs. 250,000/Rs. 1 million). In Year 2, sales were Rs. 1.5 million and gross profits were Rs. 450,000, resulting in a gross profit margin of 30 percent ((Rs. 450,000/Rs. 1.5 million). Now let's use calculate their break-even sales figure:

Break-Even Sales for Year 1:

(Operating Expenses of Rs. 170,000 + Annual Debt Service of Rs. 30,000)/ Gross Profit Margin of 25 percent (.25) = Rs. 800,000 break-even sales figure

Break-Even Sales for Year 2:

(Operating Expenses of Rs. 245,000 + Annual Debt Service of Rs. 30,000)/ Gross Profit Margin of 30 percent (.30) = Rs. 916,667 break-even sales figure

It's apparent from these calculations that ABC Clothing was well ahead of break-even sales both in Year 1 (Rs. 1 million sales) and Year 2 (Rs. 1.5 million sales).

Break-even analysis also can be used to calculate break-even sales needed for the other variables in the equation. Let us say the owner of ABC Clothing was confident he or she could generate sales of Rs. 750,000, and the company's operating expenses are Rs. 170,000 with Rs. 30,000 in annual current maturities of long-term debt. The break-even gross margin needed would be calculated as follows:

(Rs. 170,000 + Rs. 30,000)/Rs. 750,000 = 26.7%

Now let's use ABC Clothing to determine the break-even operating expenses. If we know the gross margin is 25 percent, the sales are Rs. 750,000 and the current maturities of long-term debt are Rs. 30,000, we can calculate the break-even operating expenses as follows:

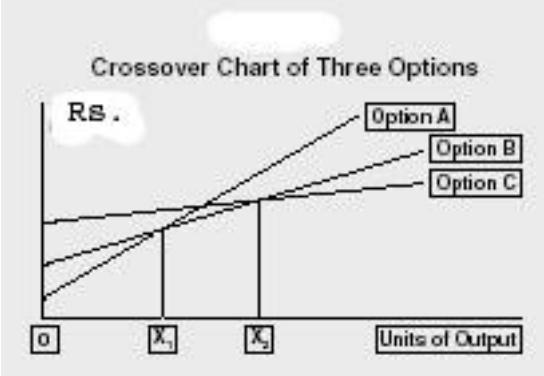
(.25 x Rs. 750,000) - Rs. 30,000 = Rs. 157,500

EXTENSIONS OF BREAK-EVEN ANALYSIS

Break-even analysis typically compares revenues to costs. However, other models employ similar analysis.

Crossover Chart of Three Options

In the crossover chart, the analyst graphs total-cost lines from two or more options. These choices may include alternative equipment choices or location choices. The only data needed are fixed and variable costs of each option.



In the above figure, the total costs (variable and fixed costs) for three options are graphed. Option A has the low-cost advantage when output ranges between zero and *X* units, whereas Option B is the least-cost alternative between *X* and *X* units of output. Above *X* units, Option C will cost less than either A or B. This analysis forces the manager to focus on the relevant range of demand for the product, while allowing for sensitivity analysis. If current demand is slightly less than *X* Option B would appear to be the best choice. However, if medium-term forecasts indicate that demand will continue to grow, Option C might be the least-cost choice for equipment expected to last several years. To determine the quantity at which Option B wrests the advantage from Option A, the manager sets the total cost of A equal to the total cost of B ($F_A + V_A \times Q = F_B + V_B \times Q$) and solves for the sole quantity of output (Q) that will make this equation true. Finding the break-even point between Options B and C follows similar logic.

Economic Order Quantity: Ordering and Holding Costs The Economic Order Quantity (EOQ) model attempts to determine the least-total-cost quantity in the purchase of goods or materials. In this model, the total of ordering and holding costs is minimized at the quantity where the total ordering cost and total holding cost are equal, i.e., the break-even point between these two costs.

As companies merge, layoffs are common. The newly formed company typically enjoys a stock-price surge, anticipating the leaner and meaner operations of the firm. Obviously, investors are aware that the layoffs reduce the duplication of fixed-cost personnel, leading to a smaller break-even point and thus profits that begin at a lower level of output.

APPLICATIONS IN SERVICE INDUSTRIES

While many of the examples used have assumed that the producer was a manufacturer (i.e., labor and materials), break-even analysis may be even more important for service industries. The reason for this lies in the basic difference in goods and services: services cannot be placed in inventory for later sale. What is a variable cost in manufacturing may necessarily be a fixed cost in services. For example, in the restaurant industry, unknown demand requires that cooks and table-service personnel be on duty, even when customers are few. In retail sales, clerical and cash register workers must be scheduled. If a barber shop is open, at least one barber must be present. Emergency rooms require round-the-clock staffing. The absence of sufficient service personnel frustrates the customer, who may balk at this visit to the service firm and may find competitors that fulfill the customer's needs.

The wages for this basic level of personnel must be counted as fixed costs, as they are necessary for the potential production of services, despite the actual demand. However, the wages for on-call workers might be better classified as variable costs, as these wages will vary with units of production. Services, therefore, may be burdened with an extremely large ratio of fixed-to-variable costs.

Service industries, without the luxury of inventoriable products, have developed a number of ways to provide flexibility in fixed costs. Professionals require appointments, and restaurants take reservations; when the customer flow pattern can be predetermined, excess personnel can be scheduled only when needed, reducing fixed costs. Airlines may shift low-demand flight legs to smaller aircraft, using less fuel and fewer attendants. Hotel and telecommunication managers advertise lower rates on weekends to smooth demand through slow business periods and avoid times when the high-fixed-cost equipment is underutilized. Retailers and banks track customer flow patterns by day and by hour to enhance their short-term scheduling efficiencies. Whatever method is used, the goal of these service industries is the same as that in manufacturing: reduce fixed costs to lower the break-even point.

Break-even analysis is a simple tool that defines the minimum quantity of sales that will cover both variable and fixed costs. Such analysis gives managers a quantity to compare to the forecast of demand. If the break-even point lies above anticipated demand, implying a loss on the product, the manager can use this information to make a variety of decisions. The product may be discontinued or, by contrast, may receive additional advertising and/or be re-priced to enhance demand. One of the most effective uses of break-even analysis lies in the recognition of the relevant fixed and variable costs. The more flexible the equipment and personnel, the lower the fixed costs, and the lower the break-even point.

It is difficult to overstate the importance of break-even analysis to sound business management and decision making. Ian Benoliel, CEO of management software developer NumberCruncher.com, said on Entrepreneur.com (2002):

The break-even point may seem like Business 101, yet it remains an enigma to many companies. Any company that ignores the break-even point runs the risk of an early death and at the very least will encounter a lot of unnecessary headaches later on.

Mathematical Explanation

The graphic method of analysis helps the reader understand the concept of the break-even point. However, graphing the cost and income lines is laborious. The break-even point is found faster and more accurately with the following formula:

B-E = F / (S - V)

where:

B-E = break-even point (units of production),

F = total fixed costs,

V = variable costs per unit of production,

S = savings or additional returns per unit of production, and

The mathematical approach is best presented using examples.

Example 1

A farmer wants to buy a new combine rather than hire a custom harvester. The total fixed costs for the desired combine are $21,270 per year. The variable costs (not counting the operator's labor) are $8.75 per hour. The farmer can harvest 5 acres per hour. The custom harvester charges $16.00 per acre. How many acres must be harvested per year to break-even?

Fixed costs (F) = $21,270

Savings (S) = $16/A

Variable costs (V) = $8.75/hr / 5 A/hr = $1.75/A

B-E = $21,270 / ($16/A - $1.75/A) = $21,270 / $14.25/A = 1,493 Acres

Example 2

Break-even analysis can be easily extended to consider other changes. If the farm operator can save two additional bushels of wheat per acre more than the custom harvester, what would be the break-even point if wheat is worth $4/bushel?

Additional income = $4/bu * 2 bu/A = $8/A

B-E = $21,270 / ($16/A + $8/A - $1.75/A)

= $21,270 / $22.25/A = 956 Acres

Example 3

A farmer raising 1,200 acres of wheat per year considers purchasing a combine. How much additional return (to land, capital labor, management and risk) would result?

Additional return = (savings or additional income) - (fixed costs + variable costs)

Additional profit = [$16/ac + ($4/bu * 2 bu/ac)] x 1200 A = $21,270 + [($8.75/hr / 5 A/hr) x 1200 A] = $28,800-$23,370 = $5,430

Thus, the farmer would generate another $5,430 in additional return by purchasing the combine. A farmer harvesting only 900 acres would probably choose not to buy the combine because the acreage is below the break-even point of 956 acres. The farmer may want to evaluate the purchase of a smaller or used combine.

Additional Situations

Two additional situations are presented as follows:

Problem 1. If the fixed costs for the combine are $12,000 per year, no additional yield is expected, variable costs are $7 per hour and the farmer can combine 4 acres per hour, what is the new break-even point?

Problem 2. If 900 acres are harvested, what is the effect on the farmer's profits?

Solutions

Fixed costs = $12,000

Savings = $16/A

Variable costs = $7/hr / 4 A/hr = $1.75/A

Problem 1:

B-E = $12,000 / ($16/A - $1.75/A) = $12,000 / $14.25/A = 842 Acres

Problem 2:

Additional profit = ($16/A x 900 A) - [$12,000 + ($7/hr / 4 A/hr x 900 A = $14,000 - $13,575 = $825 increase

Appraisal of Break-even Analysis

The main advantage of break-even analysis is that it points out the relationship between cost, production volume and returns. It can be extended to show how changes in fixed cost-variable cost relationships, in commodity prices, or in revenues, will affect profit levels and break-even points. Limitations of break-even analysis include:

It is best suited to the analysis of one product at a time;

It may be difficult to classify a cost as all variable or all fixed; and

There may be a tendency to continue to use a break-even analysis after the cost and income functions have changed.

Break-even analysis is most useful when used with partial budgeting or capital budgeting techniques. The major benefit to using break-even analysis is that it indicates the lowest amount of business activity necessary to prevent losses.