Chapter II
Structure, Conduct and Performance: Theoretical Foundations

This chapter dwells on the theoretical foundations of the relation between structure, conduct and performance. It digs into the concepts of market structure, its measurement, its impact on conduct of individual firms and also on how it impacts the performance of an industry. Since one of the major objectives of the study is to examine the performance of the Indian banking industry as measured in terms of efficiency and productivity, therefore these concepts have also been detailed.

The chapter is organized in two sections. Section 2.1 details the concepts of structure, conduct and performance and develops the relation between structure and conduct and how it influences market or industry performance. Section 2.2 discusses the idea of performance measurement, and explains various concepts related to efficiency and productivity growth.

SECTION I

2.1 STRUCTURE, CONDUCT AND PERFORMANCE

The standard approach to industrial organization decomposes a market into structure, conduct and performance of the market and digs into issues such as whether the interaction between these variables leads to an optimum solution, or whether a market failure occurs requiring the need for regulators to step in. The relation between structure and conduct and its impact on performance thus becomes an important feed for regulators in deciding the extent of regulation required in an industry.

2.1.1 Market Structure

In a literal sense, structure describes how the constituents of a complex body are arranged. In this sense, the structure of an industry would refer to the nature and pattern in which the individual firms constituting the industry are arranged. In
particular, market structure of an industry describes: (i) the current number and size
distribution of firms in an industry, and (ii) related factors which led to such a
distribution or can impact it in the future. Bain (1968) gives following salient aspects of
market structure:

(a) The degree of seller concentration in the market, as described by the number
and size distribution of firms.
(b) The degree of buyer concentration, which can be described in a similar way
as the seller concentration.
(c) The degree of product differentiation among the outputs sold by various
firms in the industry. In other words, the extent to which products of
individual firms are perceived as identical or non-identical by the buyers.
(d) The entry and exit conditions in the market, described by the relative ease
with which new sellers may enter or existing sellers may exit the market.

The first three features describe in a static sense the current state of market, while the
fourth one renders dynamism to the concept, describing the potential of change in
market structure with entry of new firms or exit of existing firms.

2.1.2 Conduct

Conduct describes the way individual firms act in a market. In other words,
conduct refers to the patterns of behaviour of firms, especially in relation to pricing
policies as well as their practices in adapting and adjusting to the market in which
they function. In particular, market conduct includes:

(a) Price and output setting policies which the firms pursue in order to fulfil the
behavioural objectives like profit maximisation, cost minimisation or sales
revenue maximisation etc.
(b) Product and sale related policies which involve product differentiation,
marketing policies, policies related to quality aspects of products being
produced etc.
(c) Policies originating out of interdependence of individual firms in an
industry. These include: (i) policies that firms pursue in competing with each
other, meant to gain market share from rivals, and (ii) policies related to
explicit or implicit collusion or agreement that firms arrive at, again in order
to fulfil one of the behavioural objectives listed above.

2.1.3 Performance

Performance refers to the extent to which the interaction between structure and
conduct result in outcomes that are deemed good or preferred by the society in welfare
sense. That is, how well the market fulfills social and private objectives including
efficiency of production, quality and quantity of products, price levels and price
stability and profit levels etc. In the words of Bain (1968):

“Performance refers to the composite of end results which firms in any
market arrive at by pursuing whatever lines of conduct they espouse –
end result in the dimension of price, output, production and selling cost,
product design and so on....”.

Similarly, Carlton and Perloff (1989) define performance as “the (degree of)
success of a market in producing benefits for consumers, as occurs for example, in the
case of low prices”. In particular, performance can be measured as following:

(a) Technological efficiency in the sense whether resources are used in an
optimum manner (e.g. in a cost minimising sense)
(b) Profit margin as reflected in difference between selling price and marginal
cost, indicating to the extent to which firms sell at a price most favourable to
consumers.
(c) The amount of total output being produced relative to maximum producible
amount within given resources.

2.1.4 Structure-Conduct-Performance Relation

The structure-conduct-performance (SCP) relation forms the basis of traditional
industrial organization theory. According to Clarke (1985), “in its simplest sense it
(SCP relation) suggests that a causal link exists running from market structure to
conduct and hence to performance”. Such a relation is described through price theory
which attempts to explain and predict how the individual firms will behave as they participate in an industry or market.

'Price theory explains and predicts the behaviour and performance of business firms in the selling markets for the goods they produce [...]. A very significant further use of price theory is to predict the way in which market conduct and performance in different industries will vary because of differences in their market structures. It does provide hypothesis concerning association of market structure with market conduct and performance...' (Bain, 1968)

Thus, the basics of SCP relationship are rooted in pricing behaviour of firms in markets characterised by different structures. In other words, SCP relation dwells into the nature and extent of impact that the structure of a market can have on the conduct of firms, as reflected in their price and product related policies. Here, one of the most important aspects according to which markets can be classified to study the relation between structure and conduct is concentration.

2.1.4.1 Concentration and Collusion

Market concentration refers to the extent to which the supply of goods and services is controlled by dominant firms in an industry. In other words, market concentration measures the degree to which supply of goods is concentrated in the hands of a small number of firms. In most of the studies involving structure-performance relation, market concentration is used as a measure of size distribution of firms (Perloff, 2007).

Traditional industrial organisation approach hypothesises a positive relationship between degree of seller concentration and profitability, in what is generally referred to as the traditional SCP paradigm. The argument runs as, that high seller concentration in a market lowers the cost of collusion and thereby fosters tacit and/or explicit collusion on the part of the firms. The result of such collusion is that all the firms in the industry earn a monopoly rent. Pricing in such a situation will be generally unfavourable to the consumers and firms will gain higher profits (Smirlock, 1985).

The logic can be built as follows. In a monopoly, there is only one firm. In this case, market concentration is at its highest level. Economic theory states that in such a
case the firm will produce an output and set a price that will maximise its profit\(^1\). Furthermore, such a (profit maximising) price will be higher and output will be lower when compared to that of perfect competition.

In case of an oligopoly, there are a small number of firms competing with each other. Although no firm can individually choose the profit maximising industry output and prices, it is still relatively easy for the firms to indulge in an implicit or explicit collusion to decide jointly the industry output and price corresponding to the profit maximisation level and then produce individually the share of total output directed by the collusion agreement.

What furthers the chances of collusive behaviour in an oligopolistic market structure is mutually recognised interdependence. To the extent the decisions made by one producer affect the outcome for others, each individual producer in an oligopolistic market structure tries to incorporate the behaviour of its rivals into its own decision-making. It is in recognition of this mutual interdependence that the producers ‘find concerted action for collective control of industry-wide price and output practicable’ \(^2\) (Bain, 1968).

Finally, in the case of an atomistic industry, wherein a large number of small producers are in competition, it becomes extremely difficult for the producers to engage in any collusive behaviour, simply because of the very large number of participants. Also, since the decisions made by one individual do not affect the rivals much, there is no interdependence in this case, further reducing the chances of reaching any collusive strategies in this market structure.

Thus, market structure, as measured through market concentration, can directly impact the behaviour of firms within the industry. High degree of seller concentration

---

\(^1\) This is the conclusion drawn in standard micro-economic texts. However, it assumes that profit maximisation is the basic objective of producers. To what extent this assumption actually describes the behavioural objective of firms in real world, and particularly a banking firm, has been an issue of debate (See Tirole, 1993 for a discussion on the issue).

\(^2\) This argument however does not mean that oligopolistic structure will definitely result in collusive behavior. Bain (1968) refers to two mutually contrasting forces in this regard- one, the desire of a producer to move collectively with the industry to gain higher profits, and the other, the ambition of a producer to act independently in order to increase his own share in market. ‘What happens will depend on the relative strengths of two motivations in individual market situations’ (Bain 1968).
may allow firms to charge higher prices from the consumers by acting in a collusive manner, the welfare results of which will be less than optimum, as reflected in the higher price paid by consumers and lower output produced by the industry. In such a situation, there will exist a positive relation between market structure, as measured through concentration, and profitability of firms.

2.1.4.2 Market Share and Efficiency

The causal relationship between market concentration and collusion was challenged by Brozen (1971) and Demsetz (1973). They, and other economists (e.g. Smirlock, 1985), suggest that a positive relationship between profits and market share at the firm level will imply a positive profit-concentration relationship at the industry level, even if high concentration does not lead to collusion and therefore does not affect conduct.

The argument runs as follows. In any industry, all firms may not be equally efficient. In this scenario, the firms which are technically or managerially more efficient will be able to produce at lower cost and hence charge a lower price. As a result, there will be pressure on all firms to raise efficiency. Over a period of time, those firms which are able to exhibit greater efficiency will acquire a larger market share, which will result in high market concentration. At the same time, these firms will also be more profitable, due to their greater efficiency and lower costs. Now, if the market structure of such an industry is looked at in a static manner, it may appear that there is a positive relation between high concentration and profitability. However, such a conclusion will be misleading since both higher concentration and higher profitability are the result of differential technical or managerial efficiency. Thus, a structure-profit relation becomes spurious with both structure and profit being endogenously determined by efficiency. This line of argument in literature has come to be known as the efficient structure hypothesis. In the words of Smirlock (1985):

"According to the efficient structure hypothesis, some firms will earn supernormal profits because of superior efficiency. This efficiency is reflected in high market share. Since markets containing such firms will tend to exhibit high concentration, it is possible a spurious relationship
between concentration and profitability will be observed when market share is not properly considered”.

2.1.4.3 Market Share and Relative Market Power: Early advocates of efficient structure hypothesis maintained that high market share of a firm reflected its greater efficiency (e.g. Smirlock et al., 1984; Smirlock, 1985). However, the view has not been unanimously accepted as market share may also proxy market power rather than efficiency. Some of the authors (e.g. Shepherd, 1982, 1985) contend that firms obtain market power from their market position and that the market share of a firm is a significant element of its market position. To quote Shepherd (1985) in this regard:

“Firms may actually obtain market power from several elements of their market position. One element is the firm’s own market share, which embodies direct control over market transactions”.

This line of reasoning has led to the development of what is known as the relative market power hypothesis. This hypothesis asserts that only the firms with large market shares and well-differentiated products can exercise market power in pricing these products and earn supernormal profits (Shepherd, 1982).

2.1.4.4 Economies of Scale and Scale Efficiency

Another element that can influence the market structure of an industry and also the relationship between structure, conduct and performance is economies of scale. Economic theory points out that it is possible to reduce the cost of operations by operating at the optimum (cost minimising) scale of production. Therefore, if there exist untapped economies of scale, firms will try to increase their size to tap such economies, which will also result in the market becoming more concentrated.

Further, the optimum size of plant will differ from industry to industry and can therefore determine the maximum number of firms which can be accommodated in an industry consistent with maximum efficiency. For example, if in an industry, the optimum size of a plant is equal to 25 percent of total industry supply, only four firms may be accommodated in such an industry in the most efficient way. However, such a market will look more concentrated due to high market share being enjoyed by
individual firms. This line of argument has led to what is known as the scale efficiency version of the efficient structure hypothesis. It predicts that firms may increase size of operations to tap economies of scale, which will simultaneously result in these firms earning greater profits (due to economies of scale) as well as the market becoming more concentrated. However, in such a scenario, once again the apparent positive structure-profit relation is spurious with both profit and concentration being driven by scale economies. In the words of Berger (1995):

"Under the scale efficiency version of the efficient structure hypothesis, firms have essentially equally good management and technology, but some firms simply produce at more efficient scale than others, and therefore have lower unit cost and higher unit profits”.

2.1.5 Structure-Performance Relation in Banking

We find from the above discussion that there are four contending hypotheses which seek to explain the relation between structure and performance. The traditional SCP paradigm has been augmented by the relative market power hypothesis while both of these have been challenged by the proponents of the two versions of efficient structure hypotheses discussed above. Although all of these have been tested extensively in banking, doubts have been expressed in literature as to what extent any or all of these can be expected to hold in banking markets. Therefore it will be appropriate to see how the nature of banking industry affects the working of above discussed theories.

There has been an ongoing debate in literature concerning what the optimal market structure of banking industry is. In this regard, traditional wisdom has been that owing to the special nature of the banking industry, some degree of market power is necessary with banks, which would foster the stability of individual banks and banking system as a whole. Here, the special nature of banking industry arises out of its highly leveraged nature and potential of contagion and bank runs in case of a single failure. On the other hand, the modern stand on the issue is that like any other industry, perfect competition is an ideal market structure for banking and greater competition will result
in efficiency related benefits\textsuperscript{3}. To the extent the traditional argument of market power with banks is validated, it will naturally change the implications of structure-performance paradigm, as market power may not be considered profit driven, but stability driven.

Furthermore, until recently, the banking industry has been highly regulated across the world. In this regard, some authors (see Obsorn and Wendel, 1982) contend that market forces that derive either the traditional SCP paradigm or the efficient structure hypothesis may not work in the case of the banking industry. A related argument is that due to the heavily regulated (or at least heavily monitored) nature of the industry, banks may not be interested in setting profit maximising prices, in order to avoid antitrust actions. Shaffer (2002) observes in this regard:

"Even if the SCP hypothesis is generally correct for some industries, there are reasons to expect that banks' pricing may be less sensitive to market concentration than in other industries".

Thus, there are reasons to expect that the results of SCP relation obtained from applications to other industries may not be validated in banking. One difference that the specific nature of the banking industry may bring into testing of the SCP relation is that one may not expect a priori a positive structure-profit relation, as has been the case with manufacturing.

SECTION II

2.2 MEASUREMENT OF PERFORMANCE

Production is a basic economic activity, generally defined as the process through which inputs are combined, transformed and turned into outputs. A productive unit uses various inputs or resources such as land, human resources and man-made aids, (e.g. tools and machinery) to further production. Outputs, on the other hand, generally include tangible products including goods and intangible products including services.

\textsuperscript{3} We will review some of the literature concerning the issue of optimum market structure of banking in Chapter IV.
The analysis of the process of production of a firm or industry derives great significance from the fact that resources used for production are limited and a growing economy or society would need an ever-increasing supply of goods and services. This naturally takes a researcher into measurement of performance of a production unit as it not only defines the current state of a productive system but also its future in terms of potential growth and scope of improvement.

Performance measurement therefore is ultimately driven by the desire to be able to produce more output from given and limited inputs. It is in this context that the concepts of efficiency and productivity take the central role in performance measurement.

Productivity at a basic level can be defined as ratio of output to input (Rogers, 1998). In other words, higher the amount of output from a given amount of input, greater is the level of productivity. Such a measure is easy to obtain if the production unit being evaluated uses a single input to produce a single output. However, in most real world situations, producers use several inputs to produce several outputs, thus making it necessary to somehow aggregate inputs and outputs.

Therefore, measurement of productivity attempts to compare in real terms, the quantity of goods and services produced against the quantity of resources employed in a given period of time (Weiner, 1973). An increase in productivity in such a scenario would be reflected in being able to either produce more amounts of one or more outputs with the given amount of inputs, or use less of one or more inputs to produce the given amount of outputs.

Traditionally, productivity growth had been seen as the residual in output growth after accounting for input growth. Frequently, technical change was regarded as main source of productivity growth, which would manifest itself in an outward (inward) shift in production (cost) frontier. However, after the introduction of frontier efficiency concept, pioneered by Koopmans (1951) and Farrell (1957), efficiency change has also come to be regarded as an important component of productivity growth. Therefore,
before going into the details of productivity measurement, we will first study what is known as frontier efficiency analysis in literature.

2.2.1 Frontier Efficiency Analysis

In the frontier approach to efficiency, a firm is said to be efficient, if it is producing on the best practised production frontier. Here, the best practised production frontier is defined in reference to a particular set of firms and consists of the firms which are using minimum input(s) to produce given level of output(s). In other words, a firm would be regarded as technically efficient if it is able to obtain maximum amount of output(s) from given amount of input(s) or minimise the amount of input(s) used in the production of given amount of output(s).

Koopmans (1951) was the first to give formal shape to frontier concept of efficiency. According to him “a producer is considered technically efficient if, and only if, it is impossible to produce more of any output without producing less of some other output or using more of some of the inputs”.

Debreu (1951) and Farrell (1957) introduced the radial measure of technical efficiency. A firm is called efficient in the Debreu-Farrell sense if it is not possible to radially contract the input bundle being used to produce a given amount of the output bundle. In other words, Debreu-Farrell measure of efficiency in an input orientation will be equal to (one minus) the maximum possible radial contraction in all inputs, while maintaining the input bundle feasible to produce the given output bundle under the present technology. For any given firm, such a measure of efficiency will very between zero and one, with the value of one corresponding to the highest level of efficiency. Similarly, in an output orientation, the Debreu-Farrell measure of efficiency will be equal to the inverse of the maximum radial expansion in all outputs that is feasible with given technology and given input bundle. For a given firm, output oriented efficiency as defined above with very between one and infinity, with value of one corresponding to the lowest level of efficiency. Koopman’s definition of efficiency in this sense is stricter than the Debreu-Farrell measure, as the former considers every single input or output separately whereas the latter considers only the bundle of inputs or outputs. In
other words, Koopman’s definition of efficiency requires absence of improvement in any single input or output while the Debreu-Farrell measure requires absence of only radial improvement in all inputs or outputs together. This may be clarified with the help of the following diagram.

![Figure 2.1]

Figure 2.1 depicts a simple production situation wherein firms use two inputs $x_1$ and $x_2$ to produce a given level of output. Various dots in the figure correspond to different input combinations used by different firms to produce the given level of output. The firms which use the least amount of inputs will constitute the best practised frontier. In figure 2.1, firms employing input combinations represented by points A, B, C and D are using the least amounts of two inputs. Therefore, best practised frontier may be obtained by joining these input combinations. Another firm which is using the input combination defined by the point F is inefficient in the sense that its input bundle can be radially contracted to point G. In this scenario, points A through D are efficient in both the Koopmans and Debreu-Farrell sense. The difference between these two definitions arises in case of the point E, wherein the input bundle may not be reduced radially and therefore the Debreu-Farrell definition will identify it as an efficient firm. Nevertheless, input $x_2$ can be reduced from the point E to bring it down to point A. Therefore, E is not an efficient input mix as per the Koopmans’ definition.
It is the Debreu-Farrell concept that has become more popular in literature, primarily due to its ease of applications. Farrell (1957) further divided the concept of efficiency into two components, technical efficiency (TE) and allocative efficiency (AE). While technical efficiency is defined as the firm's ability to obtain maximal output from a given set of inputs, allocative efficiency measures firm's ability to use inputs in optimal proportions, given their respective prices and production technology. The two concepts can be more clearly explained with the help of the following diagram.

Let us assume that a hypothetical firm F is using only two inputs, \( x_1 \) and \( x_2 \) to produce a single output \( y \). The input/output combination being used by the firm is given by point P. SS' represents the best practised frontier or the input/output combinations used by the most efficient firms. AA' represents the input price ratio and it shows the various combinations of inputs that require the same level of expenditure. If the firm's production is technically efficient, it should occur at point Q', which is the point of both technical as well as allocative efficiency.

![Diagram](image)

Since the firm F employees the combination of input at represented by point P, two types of inefficiency arise. First, it is technically inefficient, since by moving to point Q, it could produce the same amount output with fewer inputs. The magnitude of
a firm's technical efficiency (TE) can be measured as the ratio OQ/OP. If the firm was producing at the technically efficient point Q, its technical efficiency score would have been equal to one (OQ/OQ). However, clearly the distance OP is larger than OQ, and therefore efficiency of the firm F will be less than one and one minus OQ/OP, equal to QP/OP, will measure the degree of inefficiency. For example, say OQ/OP = 0.6, then it suggests that the given firm is 60 percent efficient, or in other words, it can reduce its input usage by 40 percent (1- OQ/OP = 0.4), which reflects the degree of inefficiency exhibited by the firm.

Firm F also exhibits allocative inefficiency, as the decision of the firm to produce at point P shows that the firm made an incorrect choice as to the combination of inputs at the given prices. This is because, even if the firm produces at the technically efficient point Q' on the ray OP, it can still reduce its cost of production further by moving along the SS' to point Q'. Measure of allocative efficiency (AE) therefore can be obtained as the ratio OR/OQ.

Combining together the two efficiencies one can compute what Farrell (1957) called as economic efficiency and is known as cost efficiency (CE) in popular literature. The economic or cost efficiency can be computed by multiplying technical and allocative efficiency.

\[
CE = TE * AE = (OQ/OP) *(OR/OQ)
\]  

Farrell's (1957) original ideas were illustrated in input-oriented measures under the assumption of constant returns to scale. However, if one assumes variable return to scale, one can decompose the technical efficiency into two components, pure technical efficiency and scale efficiency.

2.2.2 Estimation Techniques

Frontier efficiency measurement techniques can be divided into two broad categories: parametric and non-parametric, depending on whether a functional form is specified to represent the production relationship or not.
2.2.2.1 Parametric Techniques

In parametric efficiency analysis, technology is modelled using a given functional form such as translog or Cob-Douglas and employing a behavioural objective on part of the firms such as cost minimization or profit maximization etc. Parametric techniques typically rely on decomposing the error term of the estimated model (such as cost or profit function) into a one-sided inefficiency term and a two-sided random error term.

The most widely applied parametric technique is the stochastic frontier analysis (SFA). Under SFA, inefficiency is separated from random noise by using separate distributional assumptions for the two components. One-sided distribution like half-normal or truncated normal is used for the inefficiency whilst a symmetric distribution is used for the random error component. The logic behind such decomposition is that inefficiency must have a truncated distribution as it cannot be negative. Firm specific estimates of efficiency are then obtained as the conditional mean or mode of distribution of the inefficiency term, conditioned upon the composed error term. A typical cost function SFA can be presented as:

\[ \ln C_i = \ln \left[C(y,w,\beta) \exp (\nu_i + \mu_i)\right] \]  \hspace{1cm} (4.2)

where \( C_i \) is the observed cost of ith producer, y, w and \( \beta \) are respectively the vectors of outputs, input prices and technological parameters, \( \nu_i \) is a two-sided classical error term while \( \mu_i \) is a non-negative error term used to capture inefficiency of the ith producer.

Another parametric technique is distribution-free approach (DFA) which does away with the need of specifying separate distributional assumptions for inefficiency and random noise. DFA assumes that the efficiency of each firm is stable over time, whereas random error tends to average out to zero over time. The estimate of inefficiency for each firm in a panel data set is then determined as the difference between its average residual and the average residual of the firms on the frontier, with some truncation performed to account for the failure of the random error to average out completely to zero. (Berger and Humphrey, 1997).
Finally, there is a thick-frontier approach (TFA) that provides a general level of efficiency for the industry, but not specific to each individual unit. It computes two frontiers, one for the quartile of firms with the lowest average performance, and one for the quartile of firms with the highest average performance. Then, inefficiency is measured as the difference between the upper and the lower frontier. (Berger and Humphrey, 1997).

2.2.2.2 Non-parametric Techniques

These usually utilise linear programming based methods to compute inefficiency without imposing any substantial structure on the specification of the best-practice frontier. There are primarily two non-parametric techniques: Data Envelopment Analysis (DEA) and Free Disposal Hull (FDH).

Out of these two, DEA is by far the most frequently used one. It constructs a linear piecewise frontier by enveloping the observed data points, yielding a convex production possibilities set. As such, it does not require the explicit specification of the functional form of the underlying production relationship. The FDH is a special case of DEA, where, instead of convexity, free disposability of inputs and outputs is assumed. Because the FDH frontier is either congruent with or interior to the DEA frontier, FDH will typically generate larger estimates of average efficiency compared to DEA (Tulkens, 1993).

Both parametric and non-parametric approaches have their advantages and disadvantages. The major drawback of the parametric approaches is that these impose a particular functional form and hence all its associated behavioural assumptions (e.g. cost function and the assumption of cost minimising behaviour), which predetermines the shape of the frontier to be estimated. If the functional form is imperfectly specified, or the behavioural assumption is not met, the estimated efficiency may be confounded with significant bias. Popular functional forms include the Cobb-Douglas and more recently the flexible translog specifications, each having their own advantages and disadvantages (for a discussion see Coelli et al., 2005).
The main advantage of parametric techniques is that these can disentangle the effect of random noise in data, which could be generated by factors out of the control of individual productive units, from the inefficiency. However, in the case of SFA, such separation again depends on the particular distribution assumptions made regarding the two components of error term and hence the bias may creep into inefficiency estimates if distributional assumptions are not correctly specified.

Nonparametric approaches, on the other hand, have a major advantage in that these impose little structure on the frontier to be estimated as these do not require specification of a functional form. However, these cannot distinguish between the random error and inefficiency and therefore all observed deviation from the best practiced frontier is assumed to be inefficiency, which may bias the results if substantial noise is present in the data.

2.2.2.3 Distance Functions

Distance functions were introduced by Shephard (1953) and allow one to conveniently describe a multi-input, multi-output, production technology. Parallel to the input and output oriented technical efficiencies, one may specify either input or output distance function. An input distance function characterises the production technology by looking at the minimal proportional contraction of an input vector, given an output vector, such that the input vector remains feasible. Similarly, an output distance function considers a maximal proportional expansion of the output vector, given the input vector (Coelli, 1995). In technical terms, we may define an input distance function as:

$$D_I(x,y) = \max \{\lambda: (x/\lambda) \in I(y)\}$$

where $D_I$ represents an input oriented distance function and $I(y)$ defines an input set consisting of all input combinations which are capable of producing a given output vector $y$. Following Coelli (1995), the concept of input distance function may be explained with the help of figure 2.3.
Here, input set is represented by the area bounded from below by the isoquant L(y). The concept of isoquant here is similar to that of best practiced production frontier discussed above. The value of distance function for point A, which reflects the input/output combination used by a firm to produce given level of output can be represented by the ratio λ=OA/OB. It is clear from the above discussion that the distance function and Debreu-Farrell efficiencies are closely related\(^4\).

### 2.2.3 Productivity Growth

As stated above, most of the earlier literature attempting to measure productivity considered technical change as the main source of productivity growth. However, beginning with perhaps Nishimizu and Page (1982), a number of studies in the last couple of decades have tried to incorporate inefficiency into productivity analysis and decompose the latter into technical change and change in efficiency. Accordingly, productivity can be affected either through an outward (inward) shift in production (cost) frontier which can come through technical progress, or an increase in efficiency,

\(^4\) In fact, Debreu-Farrell definition of efficiency is inverse of distance function. This will be explained in Chapter VI where we shall provide formal methodology for measurement of efficiency and productivity growth.
which can occur through movement of less efficient firms toward the best practice frontier. To quote Prof. Grosskopf (2007) in this regard,

"I define productivity growth as the net change in output due to change in efficiency technical change, where the former is understood to be the change in how far an observation is from the frontier of technology and later is understood to be a shift in production frontier".

Thus, productivity growth encompasses both the change in best practice frontier or frontier shift through technical change, and the catching up effect i.e. movement of less efficient firms towards the frontier.

The concept of productivity growth can be cleared further with the help of the diagram 3.4. Continuing with the hypothetical production situation described in diagram 3.2, let us suppose that the best practiced production frontier shifts from SS' in period S to TT' in period T. Also suppose that in period T, the firm F is using the input/output combination represented by point p'. Now, the movement from point P to P' represents

![Figure 2.4 Productivity Growth](image)

Figure 2.4 Productivity Growth

the productivity growth for firm F. In other words, the firm is using lesser amount of inputs to produce the given amount of output at point P' as compared to point P. Further, it may be seen that the inefficiency of firm F has declined between period S
and period T as reflected by decline in distance of its observed input/output combination from the frontier in period T as compared with frontier in period S (i.e. RP' < PQ). This represents the improvement in efficiency (catching up) over time. Rest of the productivity growth between point P and P' represents the technical progress witnessed by the firm F.