Chapter Objective:

- Operational definitions
- Operationalizing variables
- To develop a research methodology for achieving the objectives outlined in chapter 3

Everyone in this world who is indulged into business related activities wants to improve the performance in order to earn more profit or to serve the customers more satisfactorily. To improve the performance of their business activity they have to evaluate their process of operations. For evaluating the performance one has to identify the best practices. Here best practice means the organization performing better comparing their counterparts in their field. Performance is measured in relative terms. Such best practicing organization is considered as efficient as well as benchmark for others. Performance evaluation helps the analysts to take a call based on results generated. For the results one can identify the strength and weakness of their business activities and also identify the opportunities to improve their performance against the identified benchmarks. Performance evaluation and benchmarking can be done quite easy when limited numbers of variables are involved but as numbers of variables increases it become very difficult to handle it. The difficulties are further enhanced when the relationships between the variables i.e. outputs and inputs are complex and involve unknown tradeoffs. Thus to overcome this problem, Data Envelopment Analysis (DEA) was introduced. Within short span DEA become widely applied approach in different areas. For example: electric utilities evaluation, bank failure prediction, textile industry performance, logistics systems, sports performance evaluation,
spatial efficiency, software development, health care, highway maintenance efficiency, steel industry productivity and among others [74,16]. Generally accepted definition of Data Envelopment Analysis: - DEA is a decision making tool based on linear programming for measuring the relative efficiencies of a set of comparable units. The DEA Model was initially developed by Charnes, Cooper, Rhodes (1978).

4.1 Introduction

Data Envelopment Analysis (DEA) has been proven an effective tool in identifying the empirical frontiers and in evaluating the relative efficiency. DEA uses the mathematical programming to implicitly estimate the tradeoffs inherent in the empirical efficient frontier [74].

![Efficient Frontier of IT Firm Operations](image)

Source: John Wang [69]

**Figure 4.1: Efficient Frontier of IT Firm Operations**

From the above given example in Figure 4.1, it can be observed clear that point F1, F2 and F3 are on the efficient frontier. Through performance evaluation, the efficient frontier that represents that best practice is identified and inefficient points like F can be improved with suggested directions for improvement towards points F1, F2, and F3. Before we discuss
the DEA models in detail some basic concepts and terminology related to DEA should be very much clear in order to understand this approach and its applicability.

4.1.1 **Economies of Scale**

In microeconomics, economies of scale are the cost advantages that enterprises obtain due to size, output, or scale of operation, with cost per unit of output generally decreasing with increasing scale as fixed costs are spread out over more units of output. Often operational efficiency is also greater with increasing scale, leading to lower variable cost as well [22].

![Economies of Scale of Firm Operations](image.png)


**Figure 4.2: Economies of Scale of Firm Operations**

It can be seen in Figure 4.2 that as quantity increased from Q to Q₂, average cost also dropped from C to C₁. Thus the above figure 4.2 is exact example of economies of scale.

4.1.2 **Efficiency Measurement**

Basically efficiency is calculated as ratio of total outputs to total inputs as mentioned in equation 1 [55].

\[
\text{Efficiency} = \frac{\text{Output}}{\text{Input}}
\]  

(1)

In relative efficiency one try to compare one’s performance with other performance who is best achievable performer. Try to find relative position.

4.1.3 **Relative efficiency measurement**
The measurement of relative efficiency where there are multiple possibly incommensurate inputs and outputs was addressed by Farrell and developed by Farrell and Fieldhouse [31], focusing on the construction of a hypothetical efficient unit, as a weighted average of efficient units, to act as a comparator for an inefficient unit. Relative Efficiency is calculation is specified in equation 2.

A common measure for relative efficiency is,

\[
\text{Efficiency} = \frac{\text{weighted sum of outputs}}{\text{weighted sum of inputs}}
\]  

(2)

Which introducing the usual notation (in equation 3) can be written as

\[
\text{Efficiency of unit } j = \frac{u_1 y_{1j} + u_2 y_{2j} + \ldots}{v_1 x_{1j} + v_2 x_{2j} + \ldots}
\]  

(3)

Where \( u_1 \) = the weight given to output \( i \)

\( y_{1j} \) = amount of output 1 from unit \( j \)

\( v_1 \) = weight given to input 1

\( x_{1j} \) = amount of input 1 to unit \( j \).

(Note efficiency is usually constrained to the range [0, 1]).

4.1.4 Decision-making Units

DEA is a linear programming-based technique for measuring the performance efficiency of organizational units which are termed as Decision-Making Units (DMUs). For example when study is directed towards the education system then selected colleges/schools will be treated as DMUs, in power generation system, power plants and units will be considered as DMUs and many more. The performance of these selected DMUs is assessed through DEA approach using the concept of productivity or efficiency.

4.2 Radial and Non- Radial DEA Models

DEA models can be categorized on the basis of radial efficiency measures, as these models optimize all inputs or outputs of a DMU at a certain proportion. But when these
models allow non proportional reductions in positive outputs or inputs is called as non-radial measures. Fare and Lovell (1978) had introduced a non-radial measures. Interesting part of non-radial DEA model (input oriented or output oriented) is that input or output slacks don’t exists respectively [28].

4.3 **Input and Output Oriented Envelopment DEA Programs**

Under the DEA program, when its aim is to produce observed outputs with minimum inputs. According to its constraint rules the inputs are multiplied by efficiency. This is referred to as an input oriented envelopment DEA program. In same way when DEA program aims to maximize output production keeping in minds the given level of resource. It is referred to as output oriented envelopment program.

Data Envelopment Analysis is an empirically based methodology that eliminates the need for some of the assumptions and limitations of traditional efficiency measurement approaches. The basic DEA model as introduced by Farrell [30] and later developed by Charnes, Cooper and Rhodes (CCR Model) uses an oriented radial measure of efficiency, which identifies a point on the boundary with the same mix of inputs (input orientation) or outputs (output orientation) of the observed unit. As a result, DMUs can be assesses on the basis of multiple inputs and outputs, even if the production function is unknown. It is a non-parametric approach to solve a linear programming formulation for each DMU and the weights assigned to each linear aggregation are the results of the corresponding linear programming (Charnes et al) [13].

DEA model does not require the specification of any particular functional form to describe the efficient frontier or envelopment surface. The flexibility of non-parametric techniques allows for several alternative formulations. For research work, two DEA models: CCR (Charnes, Cooper, Rhodes, 1978) [13] models and BCC (Banker, Charnes, Cooper, 1984) [15] models are identified. Consider a set of n homogenous Decision Making Units
(DMU). There are m inputs and s outputs and each DMU is characterized by an input-output (X, Y) vector. In order to determine the efficiency score of each unit, these will be compared with a peer group consisting of a linear combination of efficient DMUs. For each unit not located on the efficient frontier we define a vector... (\lambda j \lambda \lambda \lambda \lambda \lambda = \text{where each \lambda j \lambda represents the weight of each DMU within that peer group}). The DEA calculations are designed to maximize the relative efficiency score of each unit, subject to the constraint that the set of weights obtained in this manner for each DMU must also be feasible for all the others included in the sample. That efficiency score can be calculated by means of the following mathematical programming formulation where Technical Efficiency scores will be determined by the optimum \gamma. CCR and BCC model formulations are described.

### 4.4 CCR Model

In 1978, CCR model was suggested by A. Charnes [13] and assumed Constant Returns to Scale (CRS). Let assume that there are data’s on K inputs and M outputs on each of N firms, then for the i-th firm these are represented by the column vectors x_i and y_i respectively. The K × N input matrix, X, and the M × N output matrix, Y, represent the data for all N firms. A measure of the ratio of all outputs over all inputs would be obtained for each firm, such as u'y_i / v'x_i, where u is an M × 1 vector of output weights and v is a K × 1 vector of input weights [45]. The optimal weights are obtained by solving the mathematical programming problem:

\[
\text{Max } u, v \ (u'y_i / v'x_i),
\]

\[
st u'y_j / v'x_j \leq 1, \\
\lambda j = 1, 2,...N, \\
u, v \geq 0. \\
(4)
\]

It is essential to calculate values of u and v, such that the efficiency measure for the ith firm is maximized, subject to the constraints that all efficiency measures must be less than
or equal to one. The difficulty in this ratio formulation is that it has an infinite number of solutions. This can be avoided by imposing the constraint $v x_i = 1$, which provides:

$$\text{Max}_{\mu,v} \ (\mu'y_i),$$
$$\text{st } v'x_i = 1,$$
$$\mu'y_j - v'x_j \leq 0,$$
$$j = 1, 2 \ldots N,$$
$$\mu,v \geq 0,$$

(5)

Where the notation is changed from $u$ and $v$ to $\mu$ and $v$, to stress that this is a different linear programming problem. Equation (5) is known as the multiplier form of the DEA linear programming problem. By the duality in linear programming, equivalent envelopment form of this problem can be derived as:

$$\text{min}_{\theta,\lambda} \ \theta,$$
$$\text{st } -y_i + Y\lambda \geq 0,$$
$$\theta x_i - X\lambda \geq 0,$$
$$\lambda \geq 0,$$

(6)

Where $\theta$ is a scalar and $\lambda$ is a $N \times 1$ vector of constants. The efficiency score for the $i^{th}$ firm will be the value of $\theta$ According to the definition, it will satisfy: $\theta \leq 1$, with a value of 1 indicating a point on the frontier and hence the firm is technically efficient firm [30].

Figure 4.3: Frontiers of the CCR, BCC and NIRS Models

Source: Coelli et al. (1998) [14]
4.5 BCC Model

In 1984, BCC model was suggested by Banker, Charnes and Cooper [7]. The BCC model produces a variable return to scale (VRS), which investigates whether the performance of each DMU was conducted in region of Increasing, Constant or Decreasing Returns to Scale in multiple outputs and multiple inputs situations. The overall efficiency (CCR efficiency) can be decomposed into Pure Technical Efficiency and Scale Efficiency (components of BCC model), thus investigating the scale effects. According to this model an inefficient firm is only “benchmarked” against firms of a similar size. Therefore the DMU is said to be efficient if and only if it is both technical and scale efficient [45]. The CRS linear programming problem can be easily modified to account for VRS by adding the convexity constraint: \( N1'\lambda = 1 \) to (3) to provide:

\[
\begin{align*}
\min_{\theta, \lambda, \theta} & \quad \theta, \\
\text{st} & \quad -y_i + Y\lambda \geq 0, \\
& \quad \theta x_i - X\lambda \geq 0, \\
& \quad N1'\lambda = 1 \\
& \quad \lambda \geq 0,
\end{align*}
\]

(7)

Where \( N1 \) is an \( N \times 1 \) vector of ones. This approach forms a convex hull of interesting planes which envelope the data points more tightly than the CRS conical hull and thus provides Technical Efficiency scores which are greater than or equal to those obtained using the CRS model. The VRS specification has been the most commonly used specification in the 1990s.

Example:- In the DEA methodology, formally developed by Charles, Cooper and Rhodes (1978), efficiency is defined as a ratio of weighted sum of outputs to a weighted sum of inputs, where the weights structure is calculated by means of mathematical programming
and constant returns to scale (CRS) are assumed. In 1984, Banker, Charnes and Cooper developed a model with Variable Returns to Scale (VRS).

Assume that we have the following data:

i. Unit 1 produces 200 pieces of items per day, and the inputs are 20 rupees of materials and 4 labour-hours

ii. Unit 2 produces 160 pieces of items per day, and the inputs are 16 rupees of materials and 8 labour-hours

iii. Unit 3 produces 240 pieces of items per day, and the inputs are 24 rupees of materials and 3 labour-hours

To calculate the efficiency of unit 1, we define the objective function as

• maximize efficiency = \( \frac{u_1 \times 200}{v_1 \times 20 + v_2 \times 4} \)

Which is subject to all efficiency of other units (efficiency cannot be larger than 1):

• subject to the efficiency of unit 1: \( \frac{u_1 \times 200}{v_1 \times 20 + v_2 \times 4} \) \leq 1

• subject to the efficiency of unit 2: \( \frac{u_1 \times 160}{v_1 \times 16 + v_2 \times 8} \) \leq 1

• subject to the efficiency of unit 3: \( \frac{u_1 \times 240}{v_1 \times 24 + v_2 \times 3} \) \leq 1

• and non-negativity:

• All \( u \) and \( v \) \geq 0.

But since linear programming cannot handle fraction, we need to transform the formulation, such that we limit the denominator of the objective function and only allow the linear programming to maximize the numerator.

So the new formulation would be:

• maximize Efficiency = \( u_1 \times 200 \)

• subject to the efficiency of unit 1: \( (u_1 \times 200) - (v_1 \times 20 + v_2 \times 4) \leq 0 \)

• subject to the efficiency of unit 2: \( (u_1 \times 160) - (v_1 \times 16 + v_2 \times 8) \leq 0 \)

• subject to the efficiency of unit 3: \( (u_1 \times 240) - (v_1 \times 24 + v_2 \times 3) \leq 0 \)
subject to \(v_1 \times 20 + v_2 \times 4 = 1\)

All \(u\) and \(v \geq 0\).

4.6 Slack-Based Measure

Tone (2001) proposed a slacks-based measure (SBM) of efficiency in DEA and measure deals directly with the input excesses and the output limitations of the DMU concerned. A SBM of efficiency is defined, along with its interpretation as a product of input and output inefficiencies. Two efficiency measures are radial and non-radial measures of efficiency, and CCR and SBM are also called radial and non-radial measures of efficiency, respectively [37]. By assuming that \(n\) DMUs with the input and output matrices \(X = (x_{ij}) \in \mathbb{R}^{m \times n}\) and \(Y = (y_{ij}) \in \mathbb{R}^{s \times n}\), respectively, the input-oriented SBM model is formulated as follows:

\[
\text{Min } \rho_{in}^* = \frac{1-(\frac{1}{m})\sum_{i=1}^{m} s_i^-/x_{i0}}{1+(\frac{1}{s})\sum_{r=1}^{s} s_r^+/y_{r0}}
\]

s.t. \(x_o = X\lambda + s^-\)

\(y_o = Y\lambda + s^+\)

\(\lambda \geq 0, s^- \geq 0, s^+ \geq 0\) \hspace{1cm} (8)

Where \(\rho_{in}^*\) denotes SBM scores and \(\lambda\) represents a nonnegative in \(\mathbb{R}^n\). Additionally, \(s^-\) and \(s^+\) represent the input excess and output shortfall of expression, respectively, and are called slacks. The mixed efficiency (ME) is defined as \(\text{ME} = \frac{\rho_{in}^*}{\theta_{crs}}\). By using Eq. \(\text{TE} = \text{PTE} \times \text{SE}\), the non-radial Technical Efficiency \(\rho_{in}^*\) has the decomposition into ME, PTE and SE, as shown \(\rho_{in}^* = \text{ME} \times \text{PTE} \times \text{SE}\).

4.7 Technical Efficiency & Scale Efficiency

Using CCR model and BCC models one can distinguish two different types of efficiency i.e. Scale Efficiency and Technical Efficiency.

4.7.1 Technical Efficiency
Technical Efficiency describes the efficiency in converting inputs to outputs. Technical Efficiency is defined in terms of equi-proportional increases in outputs that the DMU could achieve while consuming the same quantities of its inputs if it were to operate on the Constant Returns to Scale (CRS) production frontier [13].

4.7.2 Pure Technical Efficiency

Pure Technical Efficiency measures the increase in outputs that the DMU could achieve if it were to use the Variable Returns to Scale (VRS) technology [15].

4.7.3 Scale Efficiency (SE)

Finally, DEA can also be used to calculate Scale Efficiency. Scale Efficiency recognizes that economy of scale cannot be attained at all scales of production thus concept of Most Productive Scale Size (MPSS) came into the picture and it has been discussed into the details in coming points. Scale Efficiency would be calculated as the ratio of overall Technical Efficiency to Pure Technical Efficiency. If Scale Efficiency equals one, the power plant is operating at CRS; otherwise it would be characterized by VRS [15].

Some efficient points may be “weakly efficient” because we have nonzero slacks. This may appear to be worrisome because alternate optima may have non-zero slacks in some solutions, but not in others. However, we can avoid being worried even in such cases by invoking the following linear program in which the slacks are taken to their maximal values.

![Figure 4.4: Frontiers of the CCR, BCC and SE Models](source: R Ramanathan [55])
Thus, from Figure 4.4, one can easily find out the CRS, VRS and Scale Efficiency

\[ \text{CRS} = \frac{\text{HG}}{\text{HE}} \]

\[ \text{VRS} = \frac{\text{HF}}{\text{HE}} \]

\[ \text{SE} = \frac{\text{CRS}}{\text{VRS}} = \frac{\text{HG}}{\text{HF}} \]

When SE and VRS are given or known then CRS can easily be located by multiplying the both i.e. \( \text{CRS} = \text{SE} \times \text{VRS} \). The CRS efficiency of a firm is always less than or equal to the Pure Technical Efficiency (VRS).

4.8 Return to Scales (RTS)

Return to Scale look at what happens when you increase all inputs by a multiplier of \( m \). Suppose our inputs are capital or labour, and we double each of these (\( m = 2 \)), we want to know if our output will more than double, less than double, or exactly double. This leads to the following situations:

1. When our inputs are increased by \( m \), our output increases by more than \( m \). It is termed as Increasing Returns to Scale
2. When our inputs are increased by \( m \), our output increases by exactly \( m \). It is termed as Constant Returns to Scale
3. When our inputs are increased by \( m \), our output increases by less than \( m \). It is termed as Decreasing Returns to Scale

Let’s have a look at a few production functions and at increasing, decreasing, or Constant Returns to Scale too. There are three examples of economic scale shown below to describe different return to scale.

1. \( Q = 4K + 6L \). Increase both \( K \) and \( L \) by \( m \) and create a new production function \( Q' \).

Then compare \( Q' \) to \( Q \).

\[ Q' = 4(K \times m) + 6(L \times m) = 4K \times m + 6L \times m = m (4K + 6L) = m \times Q \]
After factoring replace \((4*K + 6*L)\) with \(Q\). Since \(Q' = m*Q\). Note that by increasing all of our inputs by the multiplier \(m\), our production also increased by exactly \(m\). Thus it’s an example of Constant Returns to Scale.

2. \(Q = 0.5KL\) Again put in our multipliers and create our new production function.
\[
Q' = 0.5(K*m)(L*m) = 0.5*K*L*m^2 = Q * m^2
\]
Since \(m > 1\), then \(m^2 > m\). Our new production has increased by more than \(m\), so it is an example of Increasing Returns to Scale.

3. \(Q = K^{0.3}L^{0.2}\) Again we put in our multipliers and create our new production function.
\[
Q' = (K*m)^{0.3}(L*m)^{0.2} = K^{0.3}L^{0.2}m^{0.5} = Q * m^{0.5}
\]
Since \(m > 1\), then \(m^{0.5} < m\). Our new production has increased by less than \(m\), so it is an example of Decreasing Returns to Scale.

4.9 Slack Movement

Absolute value of Slack Movement is the \(s^-\) (input slack) or \(s^+\) (output slack) in the LP equations. Positive values indicate increase, and negative values indicate decrease.

4.10 Benchmarking Models

Benchmarking models deals with multiple performance measures and provides an integrated benchmarking measure which are needed to improve their performance. Inefficiency of other models in tackling multiple measures had given raise to benchmarking model. In nutshell, benchmarking is a process of defining valid measures of performance comparison among peer DMUs. In other words DEA can be considered as benchmarking tool.

4.11 Malmquist Productivity Index Approach

The Malmquist productivity index (MPI) deals with panel data. It evaluates the total factor productivity change of a DMU between two periods, named period 1 (the “from” period) and period 2 (the “to” period). Malmquist Total Factor productivity index measures
the productivity change and decompose this change into Technical change and Technical Efficiency change. This index is named after Malmquist, who proposed to construct input quantity index as a ratio of distance function. Afterwards, Fare et al. [29] constructed a Malmquist productivity index directly from input and output data using DEA.

MPI is defined as the product of efficiency change (catch-up) and technological change (frontier-shift). The efficiency change reflects to what extent a DMU improves or worsens its efficiency, while technological change reflects the change of the efficiency frontiers between two periods. These indices may be either input or output-oriented for either the period s or period t technologies [30]. The input-based Malmquist productivity index can be formulated

\[ M_{it+1} (x^t, y^t, x^{t+1}, y^{t+1}) = \left( \frac{D_i(t+1)(x^{t+1}, y^{t+1})}{D_i(t)(x^t, y^t)} \right) \left( \frac{D_i(t)(x^{t+1}, y^{t+1})}{D_i(t+1)(x^t, y^t)} \right) \left( \frac{D_i(t+1)(x^{t+1}, y^{t+1})}{D_i(t)(x^t, y^t)} \right) \left( \frac{D_i(t)(x^{t+1}, y^{t+1})}{D_i(t+1)(x^t, y^t)} \right) \]  

(9)

Where \(D_i\) is the input distance function and \(M_{it+1} (x^t, y^t, x^{t+1}, y^{t+1})\) is the productivity of most recent production unit i.e. B (t+1) - using period t+1 technology relative to the earlier production unit i.e. B (t) with respect to t technology [24]. An equivalent way of writing this index by Fare et al (1994) is \(M = \Delta \text{TECH} \times \Delta \text{EFF}\)

\[ \Delta \text{EFF} = \frac{D_i(t+1)(x^{t+1}, y^{t+1})}{D_i(t)(x^t, y^t)} \]  

(10)

\[ \Delta \text{TECH} = \left[ \frac{D_i(t)(x^{t+1}, y^{t+1})}{D_i(t+1)(x^{t+1}, y^{t+1})} \times \frac{D_i(t+1)(x^t, y^t)}{D_i(t)(x^t, y^t)} \right]^{\frac{1}{2}} \]  

(11)

In practice, this DEA-MPI has proven to be a good tool for measuring the productivity change of DMUs over time, and has been successfully applied in many fields [72, 35].

4.12 Sensitivity Analysis

Based on their actual performance of best performing DMUs, efficiency frontiers are formed, which are the extreme point technique in DEA. A direct consequence of this aspect is
that errors in measurement can affect DEA results significantly. Even a small error can affect DEA results significantly. Statistical hypothesis tests are difficult since DEA is a non-parametric technique. For example, it may not be possible to estimate the confidence with which DEA efficiencies are computed (in the sense of confidence as used in the field of statistics). Hence, before accepting the outputs generated by DEA, one has to conduct sensitivity analysis. Smith and Mayston (1987) has been described such procedures for DEA [55].

It is possible for a DMU to obtain a value of utility by simply improving its performance in terms of only one particular output ignoring others. The DMU will be considered efficient even though it has not improved its performance in terms of all the outputs. However, such an unusual DMU will not be a peer for many inefficient units. Thus, if a DMU is initially identified as efficient by DEA, a supplementary sensitivity analysis should be conducted by checking the number of inefficient DMUs for which it is a peer. If the number is low, then DMU should always be viewed with almost care but if the number is high, then DMU is genuinely efficient [55].

Another way of checking the sensitivity of DEA efficiency of a DMU is to omit only one input or output from the DEA analysis and then to verify whether the efficiency score of a DMU is affected appreciably or not. An efficient DMU which is ranked inefficient due to the omission of just one input or one output should be viewed with almost care and with caution. A similar sensitivity analysis should be conducted by leaving out an efficient DMU from the analysis [55].

4.13 Advantages and Disadvantages of DEA

Data Envelopment Analysis is a powerful technique for measuring the efficiency and performance. This is illustrated by the large and growing number of its applications in
various fields, some of which have been highlighted in literature review chapter. The most of important advantages of the DEA can be listed follows:-

a) The main strength of DEA is its objectivity, i.e., DEA provides efficiency ratings based on numerical data, and not by using subjective opinions of people. Based on the principles of frontier analysis one can say that DEA is very effective evaluation tool that makes maximum possible objective use of the available data.

b) DEA can handle multiple inputs and outputs simultaneously.

c) DEA is non-parametric in the sense that it doesn’t required mathematical form of production function relating inputs and outputs.

d) Excessive use of inputs and shortage of outputs can identify by DEA application through the results of inefficiency.

e) DEA application can be applied easily with inputs and outputs having different units of measurement. For example: one input (number of employees) was measured in number units, while the other input (capital employed) was measured in money units.

f) It can identify the nature of returns to scale at each part of efficient frontier.

g) It didn’t focus on means of population instead of that every individual observation is considered.

h) It focuses on measures performance against best practices i.e. efficient performance rather than on Central tendency properties of frontiers.

i) It can use dummy variables too.

On other hand, DEA does have certain limitations, which have been listed below. These are:

a) It’s a good at estimating relative efficiency of a DMU with respect to peers, but it converges very slowly to absolute efficiency (theoretical frontier).

b) Very small inputs and very large outputs can affect the results.

c) Limited numbers of the DMUs too affects the results.
d) Results are sensitive to the selection of inputs and outputs (Berg 2010) [10].

e) It is a deterministic method and does not consider statistical noise.

f) Application of DEA requires solving a separate linear program for each DMU. Hence, the application of DEA to problems that have many DMUs can be computationally intensive. However, this is not a very serious problem, considering the computational power of present-day computers, and the number of DMUs that are considered in normal DEA problems.

g) Since DEA is an extreme point technique, errors in measurement can cause significant problems. DEA efficiencies are very sensitive to even small errors, making sensitivity analysis an important component of post-DEA procedure. This aspect has already been discussed earlier in this chapter.

h) Since DEA is a non-parametric technique, statistical hypothesis tests are difficult.

i) As efficiency scores in DEA are obtained after running a number of LP problems, it is not easy to explain intuitively the process of DEA for the case of more than two inputs and outputs to a non-technical audience. A general audience, which will normally not have a background in linear programming, may not consider DEA transparent. In general, the management of the organizations, for which a DEA study will be carried out, may find it difficult to comprehend its results. They sometimes prefer simpler applications, if possible. However, it is possible to explain the process of DEA and linear programming in simpler terms, which could help win their support [55].

### 4.14 Differences between DEA and SFA

**Table 4.1: Differences between DEA and SFA**

<table>
<thead>
<tr>
<th>DEA</th>
<th>SFA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Non parametric approach | Parametric approach
---|---
Can deal with multiple outputs | Cannot deal with multiple outputs
No requirement to infer about the distribution (normal distribution or not) | It is necessary to know about the distribution (normal distribution or not) and requires a priori assumption about the shape of the efficiency frontier
Relevant for management performance and social science | Limited primarily to microeconomics issues
Used in linear and nonlinear efficiency frontiers | Limited to linear efficiency frontiers
Identifies specific DMUs that serve as a benchmark | -

For the purposes of this research, India’s coal/lignite fired power plants were chosen as target samples. This is because majority of electricity production in India is from coal/lignite fired power plants i.e. 66%, whilst Coal/lignite fired power plants constitute 56% of total generation mix in the country at the end of financial year Mar 2012 [52]. The data set used in the present empirical application corresponds to a sample of all running coal/lignite fired power plants of India. This study has a tried to include all power producers of electricity, but not a captive one. This is also one of the reasons why DEA is chosen as a tool to evaluate the efficiencies of power producers industry. However, there is a rule of thumb commonly used in DEA which suggests that the number of observations in the data set should be at least two times the sum of the number of input and output variables [34].

This study focuses on the production side of power generation industries. In this sector, the choice on the input side is generally driven as much by data availability as it is by
economic principles. Maximum of variable selected are based on data availability factor. Secondary data had been used extensively. Over and above the secondary data, expert’s opinion has been taken on selecting the variables as input or outputs. Majority of the data is collected through publicly available literatures, global industry databases, official government sites related to power generation industries, official websites of samples selected and many more. Before collecting data, in-depth study of thermal power industries is done along with expert’s advice on power industries. In-depth discussion is done in chapter 5 i.e. research design.