CHAPTER – VI

METHODOLOGY AND MODELS

Every business has a fundamental concept – that of understanding its performance, in the form of measuring its achievements either against the goals and objectives set by it or, against the competition. Ports are not different and it is only by comparison that performance can be evaluated. However, ports being a complex business with many different sources of inputs and outputs, direct comparison among apparently homogeneous ports seem difficult.

1. PORT PERFORMANCE AS AN INDICATOR OF PORT COMPETITIVENESS

Logistics chains are stretching across continents where production may be in one continent and the market in another. Cargoes and shipments from all over the world have been increasing exponentially. Seaports have not kept pace with the economic development and cargo movement that has been taking place. In fact, many seaports are experiencing difficulties. There exist many bottlenecks in terms of information and physical status of the cargo leading to low productivity within the terminal. The needs for higher operational productivity, faster exchange of information, and speedier vessel turn-around times are just a few of many critical factors that are currently pressing port's nodal position within logistics systems and supply chains. Thus, port performance becomes a critical issue in itself. Expansion, though a solution, is not without its own difficulties. There are many obstacles in increasing terminal capacity through expansion (Notteboom and Winklemans, 2002)[1].

As nations are becoming more global and their industries more exposed to the pressures of international competition, there is a growing realization that services supplied to their industries must be provided on an internationally competitive basis. Thus, there is a push amongst port authorities to improve their port performance and efficiency due to increasing competition between ports and growing pressure from shippers for lower port and shipping charges. Since ports from a vital link in the overall trading-chain, their level of efficiency and performance determine, to a large extent, a nation's international competitiveness. In order to achieve and maintain a competitive edge in the international markets, port authorities need to understand the underlying factors of port competitiveness, and continually asses its performance relative to the rest of the world so that appropriate business strategies can be established. Ports need to know how to compare with their rivals and advance their competitiveness for holding dominant market position.

1.1 Importance of Performance Measurement

Production, a fundamental concept in economic theory, can be defined in simple terms as a process by which inputs are combined and transformed into outputs (Case and Fair, 1999)[2]. The inputs can normally be generalised as natural resources such as land, human resources, etc. and man-made aids like tools and machinery which are used to further production. Outputs, on the other hand, can be categorised into tangible products including goods and intangible products including services. Studying production, also known as performance measurement, is of great significance because of scarce resources and the human desire to fully utilise them. According to Dyson (2000)[3], performance measurement, plays an essential role in evaluating production because it defines not only the current state of the system but also its future. Performance measurement helps to move the system in the desired direction through the effect exerted by the behavioural responses, which exist within the system towards these measures. Mis-specified performance measures, however, will cause unintended consequences with the system moving in the wrong direction.

Thanassoulis (2001)[4] identifies the following information that can be obtained by performance measurement:

- *The identification of good operating practices for dissemination;*
- *The most productive operating scale;*

- *The scope for efficiency savings in resources use and/or for output augmentation;*
- *The most suitable role model for an inefficient unit to emulate to improve its performance;*
- *The marginal rates of substitution between the factors of production; and*
- *The productivity change over time by each operating unit and by the most efficient of the operating units at each point in time.*

1.2 Measurement of Port Performance

Performance measurement plays an important role in the development of a company (or firm, etc). As the clearinghouses for a major portion of the world's rapidly increasing international trade flows, ocean ports and the efficiency with which they process cargo have become an ever more important topic. Poorly performing ports may reduce trade volumes, particularly for small, less-developed countries (Clark et al., 2004, and Wilson et al., 2003)[5]. Among the factors yielding on a port's competitiveness, performance or efficiency is considered as one of the most influential elements (Tongzon 1995 & 2001[6]; Song et al, 2001[7]; Song & Han, 2004[8]). Port efficiency is an important issue in addressing trade facilitation practices, which has been a recent focus of the World Trade Organization and regional trade institutions, such as the Asia-Pacific Economic Cooperation organization. As mentioned above, the impact of a port's performance is not only confined to its competitiveness, but also goes beyond the industry to affect a country's overall competitiveness. Measurement of *port performance* is a crucial method of checking the competitiveness of a port, compared to rival ports.

Even though the concept of port performance has been used widely, the concept is still unclear because it includes overall concepts such as port *productivity, port efficiency, port effectiveness, and economy* of a port. *Port performance* is used as a joint definition of effectiveness and efficiency. A port is a meeting place in which multipart organizations and institutions interplay at various levels and thus is complex. Ports are essentially providers of service activities, in particular for vessels, cargo and inland transport. As such, it is possible that a port may provide sound service to vessel operators on the one hand and unsatisfactory service to cargo or inland transport operators on the other. Therefore, port performance cannot normally be assessed on the basis of a single value or measure. Ports are not a kind of organization where only one service is rendered. On the contrary, multiple activities are developed in them and a great number of agents are involved in their provision (port authorities, tugs, consignees, repair shops, etc.), each of which pursues its own objective. Moreover, these operators deliver their activities with very uneven levels of competition and regulation. This diversity of activities hinders the analysis of ports as a whole and, on the contrary, calls for an analysis focused on a specific activity (Nombela and Trujillo, 1999)[9], a specific cargo type and a limited number of ports (Tongzon, 1995a, 1995b, 2001)[10]. A myriad of factors contribute to port efficiency. Some of the more obvious factors include dock facilities, connections to rail and trucking lines, harbor characteristics (including channel depth and ocean/tidal movements), time to clear customs, and labor relations. Despite the obvious significance of port efficiency, consistent and comparable measurement of such efficiencies is a daunting task. Moreover, consistent data and methods to construct measures that allow comparisons across ports are not currently available even in developed countries. Problems always arise when one tries to apply a single analytical instrument to a range of ports and terminals. Thus, inspite of a widespread recognition of the potential importance of ports, a widely accepted performance measurement method for evaluating such centers have yet to be developed (Bichou and Gray, 2004[11]).

The efficiency analysis of container ports or terminals has an important implication for both the government and port/terminal operators. Based on the efficiency measure, the government can optimize the allocation of resources and funding for improving the overall competence. The port/terminal operators want to benchmark their own ports/terminals against comparable ports/terminals to ensure competitiveness. Thus, a fundamental task for policymakers and other stakeholders is to gauge and monitor efficiency of the port services. But this is a difficult task in the fluid environment of the port industry. Technological change has made the shipping business very different from what it used to be. Containerisation has transformed the cargo management operation from a break-bulk process into a bulk and unitized one. From a labour intensive activity, it has switched into a capital intensive one. The diffusion and the increasing importance of the container business have required large investments and a change in the terminal management philosophy.

As seen in chapter: V, the penetration of containers is also associated with the upsizing of vessels. The increase of container vessels size deeply affects the port. According to Cullinane and Khanna (1999, pp.193[12]): *The latest generation of container ships make considerable demands on terminals and ports in the form of additional infrastructure, cranes, depth in ports, productivity, etc.* The first level at which the upsizing of vessels will concern port authorities is the physical level of equipment and infrastructure. Port authorities are required to invest consistently to upgrade and extend their infrastructure and equipment in order to berth the new bigger ships. The second level of effects on ports will concern the competitive position of the ports and the consequences the level of infrastructure will have on the attractiveness of the port for shipping lines and the port capacity to withhold them. *Because ports will require sufficient infrastructure in terms of berths, depth of water and craneage, fewer ports will be in a position to compete for these larger and larger ships. As a result, additional casualties will be added to the existing list of redundant liner ports* (Cullinane and Khanna, 1999, pp.194[13]). As a result, if ports want to play the role of large transhipment hubs they will have to provide adequate infrastructures and equipment to berth large container ships and handle effectively the consequent large numbers of container boxes. In fact, the development of container transport activities requires and is motivated by an increase in the efficiency of ports and terminal operators. The efficiency in operating terminal facilities is the basis in realizing high productivity and, consequently, low costs per TEU (Wiegmans, 2003). The requirement of more efficiency in terminal and port operations is leading to more capital-intensive handling techniques and increasing specialization of port labour. In addition, the diffusion of containers has favoured integrated logistics and intermodality, thus increasing the challenges on port authorities and port management (Notteboom and Winkelmans, 2001)[14].

In this changing environment, monitoring efficiency based on historical performance might be misleading, and comparing port performance with peers may be more informative. This is reflected in the recent interest of policymakers and the academic community in international benchmarking of container ports. Port's performance, especially related to container handling capacity, is the most important factor to maintain and to advance port competitiveness.

In an effort to evaluate port performance, several methods have been suggested, such as the estimation of a port cost function (De Neufville and Tsunokawa, 1981)[15] the estimation of the total factor productivity of a port (Kim and Sachish, 1986)[16] and the establishment of a port performance and efficiency model using multiple regression analysis (Tongzon, 1995)[17]. In recent years, Data Envelopment Analysis (DEA) has frequently been used to analyse port production. Compared with traditional approaches, DEA has the advantage that consideration can be given to multiple inputs and outputs. This accords with the characteristics of port production, so that there exists, therefore, the capability of providing an overall evaluation of port performance.

2. CONCEPTS IN PERFORMANCE MEASUREMENT

Productivity and efficiency are the two most important and commonly used concepts to measure performance. However, these two different concepts have mistakenly been treated as the same in most of the literature. The productivity of a producer can be loosely defined as the ratio of output(s) to input(s). This definition is easily and very obviously capable of explaining any situation where there is a single output and single input. However, it is more common that production has multiple outputs and inputs, in which case productivity refers to *Total Factor Productivity*; a productivity measure involving all factors of production (Coelli et al, 2002)[18].

The literature of efficiency dates back to the early 1950s. The first formal definition of efficiency comes from Koopmans (1951)[19] and the first measure of efficiency was proposed by Debreu (1951)[20] and Shephard (1953)[21]. Despite the theoretical relevance of these studies, efficiency was not quantified in any of these. This task was undertaken by Farell (1957)[22], who is considered the pioneer in the measurement of efficiency. The literature agrees that Farrell (1957) introduced the modern measurement of economic efficiency, drawing upon the work of Debreu (1951) and Koopmans (1951) to define a simple measure of a firm's efficiency. Farrell proposed that the economic efficiency of a firm is a combination of its technical efficiency, which reflects its ability to obtain the maximal outputs from a given set of inputs, and its allocative efficiency, which reflects its ability to use inputs in optimal proportions given their respective prices. In order to determine efficiency measures for firms, Farrell (1957) proposed to first identify an assumed existing efficient frontier using the production function. Then deviations from the efficient frontier have a natural interpretation as a measure of inefficiency with which economic units, or firms, pursue their technical or behavioral objectives.

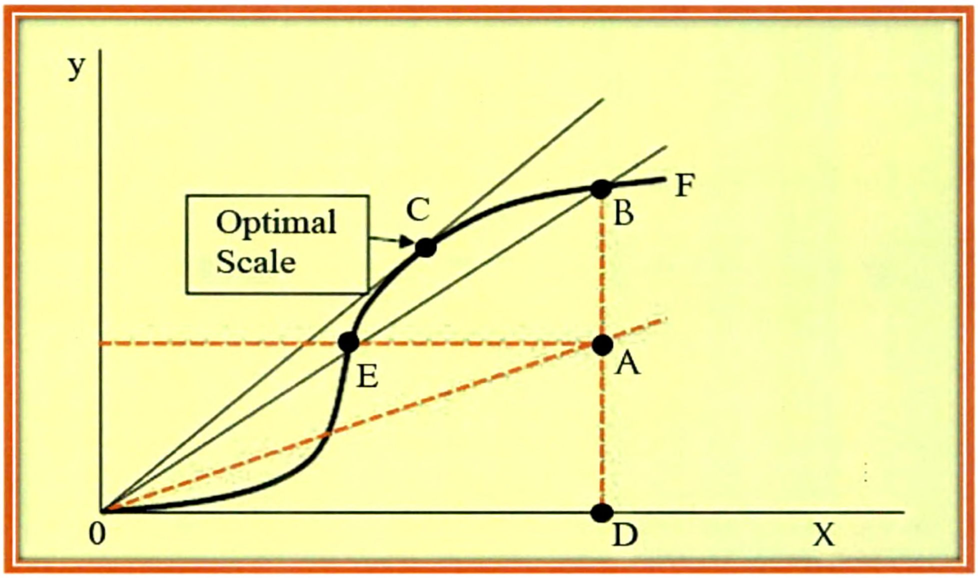
Efficiency can be defined as relative productivity over time or space, or both. For instance, it can be divided into intra- and inter-firm efficiency measures. The former involves measuring the use of the firm's own production potential by computing the productivity level over time relative to a firm-specific *Production Frontier*, which refers to the set of maximum outputs given the different level of inputs. In contrast, the latter measures the performance of a particular firm relative to its best counterpart(s) available in the industry (Lansink et al, 2002)[23]. A definition of efficiency that can be used to measure and compare different degrees of efficiency can be stated as: *Efficiency is a relative measure of the success of a production unit in maximizing its desirable outputs while at the same time minimizing its relevant inputs*. Defining efficiency in this way, we equate efficiency and the value of efficiency.

The difference between efficiency and productivity can be simply illustrated, as shown in Figure61. Points A, B and C refer to three different producers. The productivity of point A can be measured by the ratio DA/OD according to the definition of productivity where the x-axis represents inputs and the y-axis denotes outputs.

Given the same input, it is quite clear that productivity can be further improved by moving from point A to point B. The new level of productivity is then given by BD/OD. Clearly, productivity can be represented, therefore, by the slope of the ray through the origin which joins the specific point under study. The efficiency of point A, on the other hand, can be measured by the ratio of the productivity of point A to that of point B, i.e.,

$$\frac{AD/OD}{BD/OD}.$$

Figure 6.1 Illustration of Efficiency and Productivity



Source: *Derived from Coelli et.al (1998)*

The above efficiency is normally termed *Technical Efficiency* - the most common efficiency concept in economics: the conversion of physical inputs (such as the services of employees and machines) into outputs relative to best practice. In other words, given current technology, there is no wastage of inputs whatsoever in producing the given quantity of output. An organisation operating at best practice is said to be 100 per cent technically efficient. If operating below best practice levels, then the organisation's technical efficiency is expressed as a percentage of best practice. Managerial practices and the scale or size of operations affect technical efficiency, which is based on engineering relationships but not on prices and costs. It includes output- and input-oriented technical efficiencies, i.e., the producer can either improve output given the same input (output-oriented, point A to B) or reduce the input given the same output (input-oriented, point A to E) by improving technology.

The thick curve 0F in Figure 6.1 is the so-called production frontier. All the points on the production frontier are technically efficient, whilst all the points below or lying to the right of the efficient frontier are technically inefficient. The production frontier reflects the current state of technology in the industry.

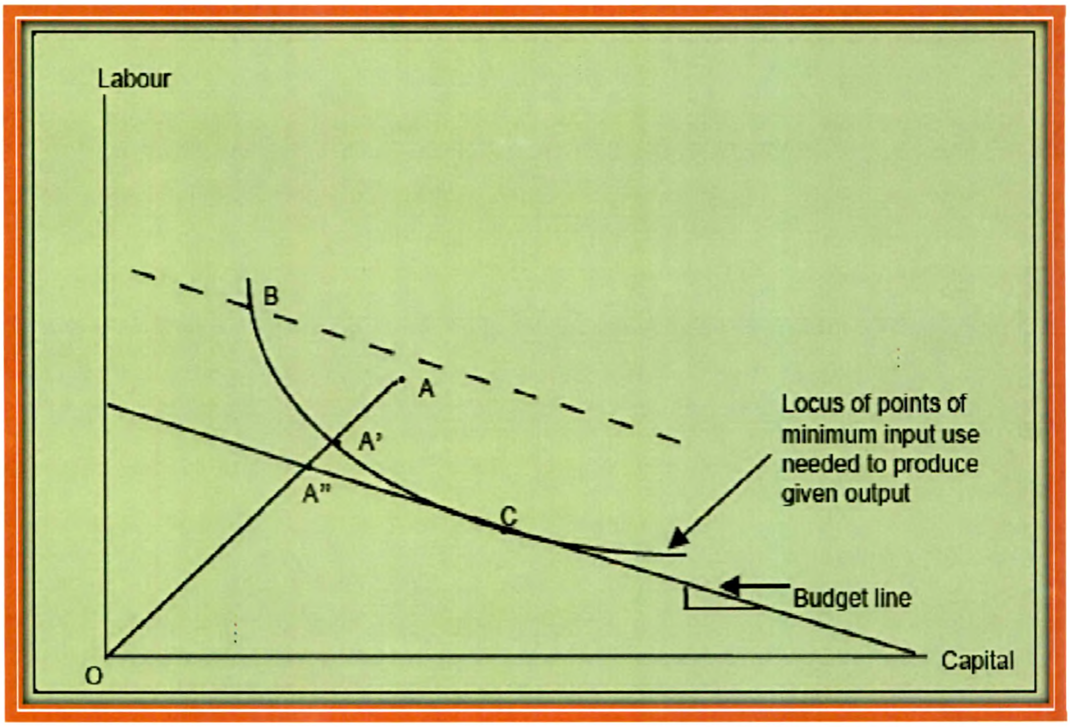
The ray through the origin and point C in Figure 2 is at a tangent to the production frontier, and hence defines the point of maximum possible productivity. This leads to another important concept, *Scale Efficiency*, which relates to a possible divergence between actual and ideal production size.

Allocative efficiency is another important concept in the context of production economics. Unlike technical and scale efficiencies, which only consider physical quantities and technical relationships and do not address issues such as costs or profits, allocative efficiency studies the costs of production given that the information on prices and a behavioural assumption such as cost minimisation or profit maximisation is properly established. For instance, allocative efficiency in input selection occurs when a selection of inputs (e.g. materials, labour and capital) produce a given quantity of output at minimum cost given the prevailing input prices (Coelli et al., 2002, p. 5). *Allocative efficiency* refers to whether inputs, for a given level of output and set of input prices, are chosen to minimise the cost of production, assuming that the organisation being examined is already fully technically efficient. Allocative efficiency is also expressed as a percentage score, with a score of 100 per cent indicating that the organisation is using its inputs in the proportions which would minimise costs. An organisation that is operating at best practice in engineering terms could still be allocatively inefficient because it is not using inputs in the proportions which minimise its costs, given relative input prices.

The concept of technical efficiency is distinct from allocative efficiency. Allocative efficiency occurs when a firm employs its factors in the correct proportions. On the other hand, technical efficiency arises when a firm makes the best use of its inputs. Technical efficiency is obtained by minimizing the cost incurred at each level of activity. Technical efficiency has also been called X efficiency by Leibenstein (1966). Farrell (1957) established that technical inefficiency could occur through the use of bundles of inputs that were larger than the minimum required to obtain the output.

Finally, *cost efficiency* refers to the combination of technical and allocative efficiency. An organisation will only be cost efficient if it is both technically and allocatively efficient. Cost efficiency is calculated as the product of the technical and allocative efficiency scores (expressed as a percentage), so an organisation can only achieve a 100 per cent score in cost efficiency if it has achieved 100 per cent in both technical and allocative efficiency. Cost efficiency is sometimes extended to include a third measure called dynamic efficiency: the degree to which producers respond to changes to technology and products following changes in consumer preferences and productive opportunities.

Figure: 6.2 Concepts of Efficiency



These concepts are better depicted graphically, as in Figure 6.2 which plots different combinations of two inputs, labour and capital, required to produce a given output quantity. The curve plotting the minimum amounts of the two inputs required to produce the output quantity is known as an isoquant or efficient frontier. It is a smooth curve representing theoretical best engineering practice. Producers can gradually change input combinations given current technological possibilities. If an organisation is producing at a point on the isoquant then it is technically efficient. The straight line denoted as the budget line plots combinations of the two inputs that have the same cost. The slope of the budget line is given by the negative of the ratio of the capital price to the labour price.

Budget lines closer to the origin represent a lower total cost. Thus, the cost of producing a given output quantity is minimised at the point where the budget line is tangent to the isoquant. At this point both technical and allocative efficiencies are attained.

The point of operation marked A would be technically inefficient because more inputs are used than are needed to produce the level of output designated by the isoquant. Point B is technically efficient but not cost efficient because the same level of output could be produced at less cost at point C. Thus, if an organisation moved from point A to point C its cost efficiency would increase by (OA–OA")/OA. This would consist of an improvement in technical efficiency measured by the distance (OA–OA')/OA and an allocative efficiency improvement measured by the distance (OA'–OA")/OA'. Technical efficiency is usually measured by checking whether inputs need to be reduced in equal proportions to reach the frontier. This is known as a 'radial contraction' of inputs because the point of operation moves along the line from the origin to where the organisation is now.

The smooth curve in Figure 6.2, representing theoretical best practice typically, cannot be calculated from observed data. Rather, data usually are only available on a group of organisations which gives limited information on theoretical best practice. First, it is unknown whether any of the organisations in the group, or sample, are achieving outright best practice. Second, the sample points will not cover all of the range of possible input combinations.

There are several ways to use the data from the sample to try and approximate the smooth curve in Figure 6.2. Early attempts used ordinary least squares regression techniques, which plot an average curve through the sample points. However, this was not satisfactory because an individual organisation's efficiency was compared with an average level of performance in the sample rather than an estimate of best practice within the sample. This led to attempts to approximate best practice in the sample by estimating frontiers.

Numerous techniques have been developed over the past decades to tackle the empirical problem of estimating the unknown and unobservable efficient frontier. We discuss here the different types of benchmarking models and briefly summarise their pros and cons.

2.1 Benchmarking Models

At a general level, we can distinguish them between parametric and non-parametric models on the one hand and the stochastic and non-stochastic on the other.

- **Parametric versus Non-Parametric**

In the modern benchmarking literature (as opposed to traditional statistics), parametric models are characterised by being defined *a priori* except for a finite set of unknown parameters that are estimated from data. Thus, the parametric approach assumes a specific functional form for the relationship between the inputs and the outputs as well as for the inefficiency term incorporated in the deviation of the observed values from the frontier. The parameters may refer to the relative importance of different cost drivers or to the parameters in the possibly random noise and efficiency distributions. Non-parametric models are characterised by being much less restricted *a priori.* Only a broad class of functions – or even production sets – is fixed *a priori* and data is used to estimate one of its elements. The classes are so broad as to prohibit a parameterisation in terms of a limited number of parameters. In other words, the non-parametric approach calculates the frontier directly from the data without imposing specific functional restrictions. The first approach is based on econometric methods, while the second one uses mathematical programming techniques.

- **Deterministic versus Stochastic Models**

In stochastic models, one makes *a priori* allowance for the fact that the individual observation may be somewhat affected by random noise, and tries to identify the underlying mean structure stripped from the impact of the random elements. In other words, the stochastic approach considers the deviations from the frontier as a combination of inefficiency and random shocks outside the control of the decision maker. In non-stochastic elements, the possible noise is suppressed and any variation in data is considered to contain significant information about the performance of the unit and the shape of the technology. Simply put, the deterministic approach considers all deviations from the frontier explained by inefficiency.

Taxonomy of the methods, with original key references, is represented in Table: 6.1 below.

Table: 6.1 Taxonomy of Benchmarking Methods

	Deterministic	Stochastic
Parametric	Corrected Ordinary Least Squares (COLS) Greene (1997), Lovell (1993), Aigner and Chu (1968)	Stochastic Frontier Analysis (SFA) Aigner, Lovel and Schmidt (1977), Batesse and Coelli (1992), Coelli, Rao and Batesse (1998, 2002)
Non- Parametric	Data Envelopment Analysis (DEA) Charnes, Cooper and Rhodes (1978), Deprins, Simar and Tulkens (1984)	Stochastic data Envelopment Analysis (SDEA) Land, Lovel and Thore (1993), Olesen and Petersen (1995), Weyman-Jones (2001)

This thesis uses a non-parametric method to avoid assuming specific functional forms for the relationship between inputs and outputs or for the inefficiency terms. The focus in this thesis is on DEA, which is a deterministic means of constructing a 'piece-wise linear' approximation to the smooth curve of Figure 6.2 based on the available sample. In simple terms, the distribution of sample points is observed and a 'kinked' line is constructed around the outside of them, 'enveloping' them (hence the term data envelopment analysis).

Methods of productivity and efficiency evaluation in economics, business and engineering are output-to-input ratio-based approaches to assessing performance of various economic units, e.g., firms, products, production systems or, in the parlance of DEA, *Decision Making Units* (*DMUs*). However, most of these approaches result in absolute measures.

Although it is also a ratio-based approach, DEA has the distinguished characteristic of always providing relative measures of performance for each in a set of such DMUs. In other words, the focal point of DEA is on *individual* observations as opposed to single optimization statistical approaches which focus on *averages* of parameters. The term Data Envelopment Analysis (DEA) was first introduced by Charnes, Cooper and Rhodes (1978)[24], to measure the efficiency of each Decision Making Units (DMUs) or the *Unit of Assessment* (Thanassoulis, 2001), that is obtained as a maximum of a ratio of weighted outputs to weighted inputs. The DMUs involved in the analysis are assumed to be homogeneous and competing in the same market while utilizing the same set of inputs to produce the same set of outputs and are responsible for controlling the process of production and making decisions at various levels including daily operation, short-term tactics and long-term strategy. The weights for the ratio are determined by a restriction that the similar ratios for every DMU have to be less than or equal to unity. This definition of efficiency measure allows multiple outputs and inputs without requiring preassigned weights. Multiple inputs and outputs are reduced to single 'virtual' input and single 'virtual' output by optimal weights. The efficiency measure is then a function of multipliers of the 'virtual' input-output combination. The best performers among the DMUs considered are used to define what is called the efficient frontier. Specifically this frontier is defined as the boundary of the convex hull of the best performers, considered to be fully efficient. Deviations from the efficient frontier are interpreted as measures of inefficiency for the remaining DMUs. A virtually efficient target, belonging to the efficient frontier, is identified for each inefficient DMU. The radial deviation from the efficient virtual target is interpreted as a measure of inefficiency. Thus the ratio of the radial distance of the virtual efficient target to the radial distance of the corresponding DMU defines the efficiency measure. Its complement is the unit measure of its inefficiency. In less technical and more concrete terms, the efficiency of a given DMU is measured, (in an input oriented DEA), by comparing the inputs it needs to those needed by the most efficient virtual DMU in order to produce an equivalent amount of output. Conventionally, a fully efficient DMU is given 1 (unity) as a measure of efficiency and all efficiency coefficients have non-zero values. This denotes that the more the output produced from given inputs, the more efficient is the production.

The CCR model presupposes that there is no significant relationship between the scale of operations and efficiency by assuming constant returns to scale (CRS), and it delivers the overall technical efficiency (OTE). The CRS assumption is only justifiable when all DMUs are operating at an optimal scale. However, firms or DMUs in practice might face either economies or diseconomies of scale. Thus, if one makes the CRS assumption when not all DMUs are operating at the optimal scale, the computed measures of technical efficiency will be contaminated with scale efficiencies.

A DEA model can be constructed either to minimise inputs or to maximise outputs. An input orientation aims at reducing the input amounts as much as possible while keeping at least the present output levels, while an output orientation aims at maximising output levels without increasing use of inputs (Cooper *et al.*, 2006)[25].

As we are looking at relative efficiency, it is important for the DMUs to be sufficiently similar, so that comparisons are meaningful. This is particularly the case with DEA, where Dyson *et al.* (2001)[26] have developed what they describe as a series of homogeneity assumptions. The first of these is that the DMUs the performance of which is being compared should be undertaking similar activities and producing comparable products and services so that a common set of outputs can be defined. The second homogeneity assumption is that a similar range of resources is available to all the units and they operate in a similar environment.

3. DATA ENVELOPMENT ANALYSIS MODELS

Data envelopment analysis is not just one single method. This analysis is a collection of different methods which serve different needs depending on scale effects, measurement of the distance to the envelopment surface or the functional form of the envelopment analysis. Usually in literature four different characteristic models of the DEA are differentiated:

1. CCR model (Charnes, Cooper and Rhodes, 1978)

The CCR model bases the evaluation on a production technology that has constant returns to scale (CRS) and radial distance measure to the efficient frontier, i.e. that the inefficiency of a DMU compared to the corresponding efficient units lying on the efficient border is measured via a radius vector. Furthermore the efficient frontier is piecewise linear. The CCR model is differentiated in input or output orientation, depending upon reduction of inputs with constant outputs or enhancement of output with constant input. As a result the CCR model reveals an objective description of overall efficiency and identifies the sources of inefficiencies.

2. BCC Model (Banker, Charnes and Cooper, 1984)

In contrast to the CCR model the BCC model offers a differentiation between technical efficiency and scale-efficiency because it evaluates solutions for nonincreasing returns to scale (NIRS), nondecreasing returns to scale (NDRS) or variable returns to scale (VRS). With a combination of the CCR model inefficient CCR DMUs do not have to be technical inefficient. Potentially a part of this inefficiency can be mitigated by increasing or decreasing the production volume resulting in a removal of scale inefficiencies. As the CCR model the BCC model also implies a radial distance measure and a piecewise linear frontier. Like the CCR model the BCC model also differentiates in an input or output orientation.

3. Multiplicative Model (Charnes, Cooper, Seiford and Stutz, 1983)

In contrast to the CCR and BCC model the multiplicative model offers a different characteristic of the efficient frontier by providing a piecewise log-linear envelopment surface or a piecewise Cobb-Douglas interpretation of the production process [Charnes A. et al. (1994), p. 24]. The returns to scale assumption depend on the envelopment surface: a log-linear surface assumes constant returns to scale and a Cobb-Douglas coherence assumes variable returns to scale.

4. Additive Model (Charnes, Cooper, Golany, Seiford and Stutz, 1985):

Unlike the CCR or BCC model the additive model is unoriented, i.e. it does not differentiate between input or output orientation which means that a reduction of input with a synchronous enhancement of outputs is possible. The additive model assumes constant returns to scale and piecewise linear efficient frontier.

These four commonly used and described DEA models in literature build a profound basis for an efficiency analysis with different returns to scale, different envelopment surfaces and different ways to project inefficient units to the efficient frontier. Since this thesis would be using DEA-CCR and DEA-BCC models, we discuss them in a greater detail (the theoretical description of DEA is drawn heavily from Cooper *et al.*, 2000 and Thanassoulis, 2001).

3.1 CCR Model Algorithm

Charnes, Cooper and Rhodes (1978) extended Farrell's (1957) work in the measurement of technical efficiency and first introduced the term data envelopment analysis. The model was known as the CCR (Charnes, Cooper and Rhodes) model. Here we give a brief introduction of the model.

Basis for the efficiency measurement is usually the productivity θ that can be described as:

$$\theta = \frac{\text{Output}}{\text{Input}}$$

Equation: 1. Efficiency measure in the one input/ one output case

This case assumes a single input and output. If some organizations are compared this way a bigger (smaller) θ means a better (worse) performance because less (more) input is used for a constant amount of output. The meaning of θ is only available through a comparison to other organization, for this reason the efficiency is always a relative measure. Typically, organizations have multiple inputs and outputs. Let us denote DMUs by index j=1, 2, ..., n. Each DMU uses a varying amount of m different inputs (i=1, 2, ..., m) described by the vector x_j to produce s different outputs (r=1, 2, ..., s) described by the vector y_j. Specifically, DMU_j consumes amount x_{ij} of input i and produces y_{rj} of output r. We assume that $x_{ij} \geq 0$ and $y_{rj} \geq 0$ and further that each DMU has at least one positive input and one positive output value. These inputs and outputs have to be weighted (v_i, u_r) (these weights have to be defined by the researcher before a calculation), which can then be summated as described in equation 2. The numerator describes the sum of the weighted output and the denominator the sum of the weighted inputs. The efficiency score in the presence of multiple input and output factors is defined as:

$$\text{Efficiency} = \frac{Weighted\ Sum\ of Outputs}{Weighted\ Sum\ of\ Inputs}$$

$$\frac{\sum_{r=1}^{s} u_r\, y_{ro}}{\sum_{r=1}^{m} v_i\, x_{io}}$$

Equation: 2. Efficiency measure in the multiple input/ output case

where u_r is the weight for output r and v_i is the weight given to input i.

The comparison of different organizations using the measure described in equation: 2 is quite inaccurate because it is not clear if the resulting efficiency measures depend on the real performance or on the (randomly) chosen weights. DEA treats the observed inputs x_j and outputs y_j as given constants and chooses the optimal individual weights for every DMU derived from the data. Therefore every DMU_j maximises its efficiency by choosing its individual weights which suit best to its situation (equation: 3). DEA chooses values of input and output weights for a particular DMU_o by the following optimization problem:

$$\underset{u,v}{max\theta} = \frac{\sum_{r=1}^{s} u_r\, y_{ro}}{\sum_{r=1}^{m} v_i\, x_{io}}$$

Equation: 3. Objective function of the DEA algorithm

As a constraint, it is important that every other DMU_l has a $\theta \leq 1$ with the chosen weights of DMU_o because the DEA efficiency measure should be scaled between 0 and 1. This coherency can be described by equation: 4.

$$\frac{\sum_{r=1}^{s} u_r\, y_{rj}}{\sum_{r=1}^{m} v_i\, x_{ij}} \leq 1$$

Equation: 4. Constraint Equation

These equations provide the basis for the formulation of CCR as described in equation: 5. To the above described objective function and the side condition we have the non-negative constraint that the weights v_i, u_r must be bigger or equal than 0.

$$\underset{u,v}{max\theta} = \frac{\sum_{r=1}^{s} u_r y_{ro}}{\sum_{r=1}^{m} v_i x_{io}} \qquad (1)$$

Subject to

$$\frac{\sum_{r=1}^{s} u_r\, y_{rj}}{\sum_{r=1}^{m} v_i\, x_{ij}} \leq 1\, j = 1,2, \ldots, j_0, \ldots, n$$

$$u_r \geq 0, \ r = 1, 2, \ldots, s$$

$$v_i \geq 0, \ i = 1, 2, \ldots, m$$

Equation: 5. Basic formulation for the CCR model

The objective of the above programming problem is to determine/obtain the weights u_r and v_i[27]. However, this problem has an infinite number of solutions since if (u^*, v^*) is optimal then α (αu^*, αv^*) is also optimal for each positive scalar. This formulation has another disadvantage: it cannot be solved by linear programming because it is not linear. Following the Charnes-Cooper transformation (1962), one can select a representative solution (u, v) by very simply linearising the constraint by the multiplication of the denominator resulting in equation: 6.

$$\sum_{r=1}^{s} u_r\ y_{rj} \leq \sum_{i=1}^{m} v_i\ x_{ij} \quad j = 1,2, \ldots, n \quad r = 1, 2, \ldots, s$$

Equation: 6. Linearised side condition

The objective function will be linearised via the normalization of the denominator. The denominator can be normalized and added to the side functions. Equation: 7 describes this coherency.

$$\sum_{i=1}^{m} v_i\ x_{i0} = 1$$

Equation: 7. Normalised denominator of the objective function

The input weights v_i of the denominator are independent anent the value of the equations right side. A "2" on the right side produces the same result concerning the weights v_i as a "4". Therefore the denominator of the objective function is equalled to "1", added to the side functions resulting in a linear objective function.

After the linearization of both the objective function and the constraint function the complete linear program of the CCR model is described in equation: 8.

$$\max_{u,v} \Theta = \sum_{r=1}^{s} u_r \; y_{r0} \qquad (2)$$

subject to

$$\sum_{r=1}^{s} u_r \; y_{rj} - \sum_{i=1}^{m} v_i \, x_{ij} \leq 0, \qquad j = 1, 2, \dots, n$$

$$r = 1, 2 \dots, s$$

$$\sum_{i=1}^{m} v_i \, x_{io} = 1 \qquad i = 1, 2, \dots, m$$

$u_r \geq 0, \; r = 1, 2, \dots, s$

$v_i \geq 0, \; i = 1, 2, \dots, m$

Equation: 8. Linear CCR-algorithm

The measures of efficiency described by the problems (1) and (2) are "units invariant" - i.e., they are independent of the units in which the inputs and the outputs are measured, provided these units are the same for every DMU.

Replacing the non-negativity constraints for the weights in (2) by $u_r \geq \varepsilon$ and $v_i \geq \varepsilon$, where ε is an infinitesimal constant, we write the so-called CCR ratio model (Charnes - Cooper - Rhodes, 1979):

$$\max_{u,v} \theta = \sum_{r=1}^{s} u_r \; y_{r0} \qquad (3)$$

subject to

$$\sum_{r=1}^{s} u_r \; y_{rj} - \sum_{i=1}^{m} v_i \, x_{ij} \le 0, \qquad j = 1, 2, \dots, n$$

$$\sum_{i=1}^{m} v_i \, x_{io} = 1$$

$u_r \geq \varepsilon,\ r = 1, 2, \dots, s$

$v_i \geq \varepsilon,\ i = 1, 2, \dots, m$

where ε is defined as an infinitesimal constant (a non-Archimedean quantity).

The dual to the above LP Equation: 8 is[28]:

$$\min \theta - \varepsilon\left(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+\right) \qquad (4.1)$$

subject to

$$\sum_{j=1}^{n} \lambda_j \, y_{rj} - s_r^+ = y_{r0}, \qquad r = 1, 2, \dots, s \qquad (4.2)$$

$$\sum_{j=1}^{n} \lambda_j \, x_{ij} + s_i^- = \theta x_{io}, \qquad i = 1, 2, \dots, m \qquad (4.3)$$

$$s_i^-, s_r^+, \lambda_j \ge 0 \qquad j = 1, 2, \dots, n$$

where λ_j are the weights of DMUs, S_i^- are the input slacks and S_r^+ are the output slacks.

The problem (4) seeks values of λ_j to construct a composite unit, with outputs $\sum_{j=1}^{n} y_{rj}\,\lambda_j$ and inputs. $\sum_{j=1}^{n} \lambda_j\, x_{ij}$.

The dual constraints in (4.2) imply that even after the proportional reduction of all inputs, the inputs of the evaluated DMU_o cannot be lower than the inputs of the composite unit. According to (4.3), the outputs of the DMU_o cannot be higher than the outputs of the composite unit.

- **Efficiency Analysis**

Both above linear problems yield the optimal solution θ^*, which is the efficiency score (so-called technical efficiency or CCR-efficiency) for the particular DMU_0, and efficiency scores for all of them are obtained by repeating them for each DMU_j, j=1,2,....,n. The value of θ is always less than or equal to unity (since when tested, each particular DMU_o is constrained by its own virtual input-output combination) and those for which $\theta^* = 1$ are relatively efficient, having their virtual input-output combination points on the frontier. The frontier itself consists of linear facets spanned by efficient units of the data, and the resulting frontier production function (obtained with the implicit constant returns-to-scale assumption) has no unknown parameters.

If, in optimality,

1) θ_0^*=1 and
2) all input and output slack variables, $S_r^+ = S_i^- = 0$,

then DMU_o is DEA CCR-efficient (Constant Returns to Scale efficient), operating on the CRS frontier.

In other words, the DMU_o will be efficient when it has proved impossible to construct a composite unit that outperforms DMU_o. Conversely, if DMU_o is inefficient, the optimal values of λ_j form a composite unit outperforming DMUo and providing targets for DMU_o (the peer group or the reference set for DMU_o).

The first condition is known as the Radial or Technical Efficiency (TE). The (scalar) variable θ_0^* gives us the proportion of all inputs of DMU_o necessary to achieve the given output levels efficiently. In other words, $1-\theta_0^*$ gives the necessary proportional reduction of all inputs of the DMU_o being evaluated in order to achieve the efficient frontier. This reduction is applied simultaneously to all inputs and results in a radial movement toward the envelopment surface. Because of the focus on maximal movement toward the frontier through proportional reduction of inputs, the models (3) – (4) are denoted as *input-oriented CCR models.*

Both the above mentioned conditions must be satisfied for full efficiency. These two conditions together define the "Pareto-Koopmans" or "strong" efficiency, which can be verbalized as[29]:

Pareto-Koopmans Efficiency: A DMU is fully efficient if and only if it is not possible to improve any input or output without worsening some other input or output.

If $\theta_0^*=1$ and $S_r^+ \neq 0$ or $S_i^- \neq 0$, then DMU_o shows DEA weak-efficiency (CRS-inefficient), also known as Farrell efficiency. This indicates that improvements can be brought about in the DMU without worsening any other input or output.

If $\theta_0^* < 1$, it implies that the DMU_o is relatively inefficient and shows the need for a proportional reduction of inputs for it to become efficient. In other words, it means that all inputs can be simultaneously reduced without making any change in the input mix or proportions. The lower the value of θ_0^*, the less efficient is the unit compared to the rest of the population. The advantage of the DEA model is that it advises how the unit evaluated should mend its behaviour to reach efficiency. Since the Models (2) and (3) are input-oriented – they try to find out how to improve the input characteristics of the unit concerned for it to become efficient.

- **Output-oriented CCR Model**

The alternative DEA model denoted as *output-oriented CCR model* (Charnes-Cooper-

Rhodes, 1978). It focuses on maximal movement via proportional augmentation of outputs under at most the present inputs. The required linear programming problem is:

$$\min_{u,v} \emptyset = \sum_{r=1}^{s} v_i \ x_{i0} \tag{5}$$

Subject to

$$\sum_{r=1}^{n} v_i \, x_{ij} - \sum_{i=1}^{m} u_r \, y_{rj} \geq 0, \qquad j = 1, 2, \dots, n$$

$$\sum_{i=1}^{m} u_r \, y_{r0} = 1$$

$$u_r \geq \varepsilon \qquad r = 1, 2, \dots, s$$

$$v_i \geq \varepsilon \qquad i = 1, 2, \dots, m$$

The corresponding dual model is:

$$\max \emptyset + \varepsilon \left(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+ \right) \tag{6.1}$$

Subject to

$$\sum_{j=1}^{n} \lambda_j \, y_{rj} - s_r^+ = \emptyset y_{r0}, \qquad r = 1, 2, \dots, s \tag{6.2}$$

$$\sum_{j=1}^{n} \lambda_j \, x_{ij} + s_i^- = x_{io}, \qquad i = 1, 2, \dots, m \tag{6.3}$$

$$s_i^-, s_r^+, \lambda_j \geq 0 \qquad j = 1, 2, \dots, n$$

The variable $\varnothing_0^*$ yields the proportion by which all the DMU$_0$'s outputs should be produced (under the given input levels) for the DMU$_0$ to be efficient. In other words, $\varnothing_0^*$-1indicates the necessary proportional increase of all DMU$_0$'s outputs in order to achieve the efficiency frontier. The higher is the value of $\varnothing_0^*$, the less efficient is the DMU. The constraint (6.2) indicates that even after a proportional increase of all outputs, the outputs of the evaluated DMU$_0$ cannot be higher than the outputs of the composite unit. According to (6.3), the inputs of the DMU$_0$ cannot be lower than the inputs of the composite unit. Like for the input-oriented model (4) a DMU is efficient if and only if $\varnothing_0^*$=1 and all slack variables, S_r^+ and S_i^- are equal to zero.

The optimal solutions of the two orientations of the CCR model are related as:

$$\theta_0^* = \frac{1}{\varnothing *}$$

Here, θ_0^* shows the input reduction rate whereas ø* shows the output enlargement rate. From the above relation, it can be concluded that an input-oriented CCR model will be efficient if and only if it is also efficient when the output-oriented CCR model is used to evaluate a DMU's performance.[30]

Models (4) and (6), both yield identical envelopment surfaces and identical sets of efficient and inefficient DMUs. However an inefficient DMU will be projected to different points on the efficiency frontier under the input and output orientations. Up to this point, we have been dealing with models built on the assumption of *constant returns-to-scale* (CRS) of activities. CRS implies that the radial expansion and reduction of all observed DMUs and the nonnegative combinations are possible. Geometrically speaking, all supporting hyperplanes for a CRS efficiency frontier pass through the origin. The DEA models involving weights of inputs and outputs (v, u) are called *Multiplier DEA Models.* Those involving weight of firms (θ, λ) are called *Envelopment DEA Models.*

Table 6.2 presents the CCR (Charles, Cooper, Rhodes, 1978) models (Cooper et all. 2006).

Table: 6.2 DEA CCR Models

<table>
<tr><th colspan="2">Input-Oriented</th></tr>
<tr><td>Envelopment Model</td><td>Multiple Model</td></tr>
<tr><td>

$$\min \theta - \varepsilon\left(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+\right)$$

subject to

$$\sum_{j=1}^{n} \lambda_j y_{rj} - s_r^+ = y_{r0,} \qquad r = 1, 2, \dots, s$$

$$\sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = \theta x_{io,} \qquad i = 1, 2, \dots, m$$

$$s_i^-, s_r^+, \lambda_j \geq 0 \qquad j = 1, 2, \dots, n$$

</td><td>

$$\max_{u,v} \theta = \sum_{r=1}^{s} u_r \; y_{r0}$$

subject to

$$\sum_{r=1}^{s} u_r \, y_{rj} - \sum_{i=1}^{m} v_i \, x_{ij} \leq 0, \quad j = 1, 2, \dots, n$$

$$\sum_{i=1}^{m} v_i \, x_{io} = 1$$

$$u_r \geq \varepsilon, \quad r = 1, 2, \dots, s$$

$$v_i \geq \varepsilon, \quad i = 1, 2, \dots, m$$

</td></tr>
<tr><th colspan="2">Output-Oriented</th></tr>
<tr><td>Envelopment Model</td><td>Multiple Model</td></tr>
<tr><td>

$$\max ø + \varepsilon\left(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+\right)$$

Subject to

$$\sum_{j=1}^{n} \lambda_j y_{rj} - s_r^+ = ø y_{r0,} \qquad r = 1, 2, \dots, s$$

$$\sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = x_{io,} \qquad i = 1, 2, \dots, m$$

$$s_i^-, s_r^+, \lambda_j \geq 0 \qquad j = 1, 2, \dots, n$$

</td><td>

$$\min_{u,v} ø = \sum_{r=1}^{s} v_i \; x_{i0}$$

Subject to

$$\sum_{r=1}^{n} v_i \, x_{ij} - \sum_{i=1}^{m} u_r \, y_{rj} \geq 0, \quad j = 1, 2, \dots, n$$

$$\sum_{i=1}^{m} u_r \, y_{r0} = 1$$

$$u_r \geq \varepsilon \qquad r = 1, 2, \dots, s$$

$$v_i \geq \varepsilon \qquad i = 1, 2, \dots, m$$

</td></tr>
</table>

The above models can also be expressed in vector-matrix notation. The input-oriented model can be expressed as below:

Assume that there are n DMUs, where each DMU produces s outputs by utilizing m *inputs*. For the j^{th} DMU these are represented by the column vectors x_i and y_i respectively. The $m \times n$ input matrix, X and the s $\times$ n output matrix, Y represent the data for all n DMUs. Then, relative efficiency of the j^{th} DMU can be found by solving the following linear programming problem:

(LP$_o$)

$$\max z = u^T Y_o$$

subject to

$$v^T X_o = 1$$
$$u^T Y - v^T X \leq 0$$
$$u \geq \in$$
$$v \leq \in$$

where u is a $s \times 1$ vector of output weights and v is a $m \times 1$ vector of input weights (the ^{T} prime denotes a transposed vector). The dual model to this can be stated as follows:

$$\min f = \theta - \in (e^T s^+ + e^T s^-)$$

subject to

$$Y\lambda - s^+ = Y_o$$
$$X\lambda + s^- = \theta X_o$$
$$\lambda, s^+, s^- \geq 0$$

where θ *is* a scalar representing the efficiency score for the i^{th} DMU; λ (=$\lambda_1, \lambda_2, ..., \lambda_n$) is a $n \times 1$ vector of constants assigned to individual productive units, $\lambda \geq 0$; s^+ and s^- are vectors of addition input and output variables; $e^T = (1, 1, ..., 1)$ and $\in$ is a constant greater than zero, which is normally pitched at 10^{-6} or 10^{-8}.

Similarly the dual of the output oriented model could be written as follows:

$$\max g = \phi + \in (e^T s^+ + e^T s^-)$$

subject to

$$Y\lambda - s^+ = \phi Y_o$$
$$X\lambda + s^- = X_o$$
$$\lambda, s^+, s^- \geq 0$$

For the optimal solution to the CCR model, the values of objective functions should be inverted, i.e. $f^* = 1/g^*$. Input-oriented efficiency scores range between 0 and 1 whereas output-oriented efficiency scores range from 1 to infinity, in both cases 1 is efficient. For the output-oriented model, we define the efficiency score as the inverse of the estimated score (i.e. $1 / \phi$).

The aim of DEA analysis is not only to determine the efficiency rate of the units reviewed, but in particular to find target values for inputs X'_o and outputs Y'_o for an inefficient unit. After reaching these values, the unit would arrive at the threshold of efficiency. Target values are calculated:

1. *by means of productive unit vectors:*

$$X'_q = X\lambda^*$$
$$Y'_q = Y\lambda^*$$

where λ^* is the vector of optimal variable values.

2. *by means of the efficiency rate and values of additional variables s^- and s^+:*

Input-oriented CCR model:

$$X'_o = \theta X_o - s^- \; Y'_o = Y_o + s^+$$

Output-oriented CCR model:

$$X'_o = X_o - s^- \; Y'_o = \phi Y_o + s^+$$

where θ is the efficiency rate in the input-oriented model and ϕ is the efficiency rate in the output-oriented model.

3.2 BCC Model

The constant returns to scale (CRS) surface is presented by a straight line that starts at the origin and passes through the first DMU that it meets as it approaches the observed population. The models with CRS envelopment surface assume that an increase in inputs will result in a proportional increase in outputs. However, it is rare for markets to function in an ideal way. There will always be financial limitations or imperfect competitive markets where increased amounts of inputs do not proportionally increase the amount of outputs obtained. For example, in agriculture, when the water volume applied to crops is increased, we do not necessarily obtain a linearly proportional increase in agricultural production. In order to account for this effect, the DEA model for variable-returns-to-scale (BCC) was developed [Banker *et.al.*, 1984].

The variable returns to scale (VRS) model allows an increase in input values to result in a non-proportional increase of output levels. The VRS surface envelops the population by connecting the outermost DMUs, including the one approached by the CRS surface. Hence the BCC model envelops more data and efficiency scores are bigger than or equal to scores of CCR.

The absence of constraints for the weights λj, other than the positivity conditions in the above, implies constant returns to scale. For allowing variable returns to scale, it is necessary to add the convexity condition for the weights λj, i.e. to include in the model.

$$\sum_{j=1}^{n} \lambda_j = 1$$

The resulting DEA model that exhibits variable returns to scale is called the BCC-model, after Banker, Charnes and Cooper (1984).The input-oriented BCC-model for the DMU_o can be written formally as:

$$\min \theta - \varepsilon\left(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+\right) \qquad (7)$$

Subject to

$$\sum_{j=1}^{n} \lambda_j y_{rj} - s_r^+ = y_{r0}, \qquad r = 1,2,\dots,s$$

$$\sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = \theta x_{io}, \qquad i = 1,2,\dots,m$$

$$\sum_{j=1}^{n} \lambda_j = 1$$

$$s_i^-, s_r^+, \lambda_j \geq 0 \qquad j = 1,2,\dots,n$$

To express the BCC model in vector-matrix notation, the CCR model in matrix form needs to be rewritten to include a condition of convexity $e^T\lambda = 1$. The formulations of the input-oriented and output-oriented BCC models are as follows, respectively:

<u>Input-oriented</u>

$$\min \theta - \in (e^T s^+ + e^T s^-) \qquad (8)$$

Subject to

$$Y\lambda - s^+ = Y_q$$

$$X\lambda + s^- = \theta X_o$$

$$e^T\lambda = 1$$

$$\lambda, s^+, s^- \geq 0$$

Output-oriented

$$\max g = \phi + \in (e^T s^+ + e^T s^-) \qquad (9)$$

subject to

$$\begin{aligned} & Y\lambda - s+ = \phi Y_q \\ & X\lambda + s^- = X_q \\ & e^T\lambda = 1 \\ & \lambda,\ s^+,\ s^- \geq 0 \end{aligned}$$

As already mentioned above, the VRS specification forms a production frontier that envelopes data more closely than the CRS specification. Therefore, the resulting efficiency scores are equal to or greater than those obtained with the CRS model. The BCC-efficiency scores are obtained by running the above model for each DMU (with similar interpretation of its values as in the CCR-model). These scores are also called "Pure Technical Efficiency (PTE) scores"; since they are obtained from the model that allows variable returns to scale and hence eliminates the "scale part" of the efficiency from the analysis. In fact, an efficiency score obtained using the CCR-model is called Technical Efficiency (TE), which comprises of both Scale Efficiency (SE) and "Pure" Technical Efficiency (PTE). Technical efficiency describes the efficiency in converting inputs to outputs, while scale efficiency recognises that economy of scale cannot be attained at all scales of production, and that there is one *most productive scale size,* where the scale efficiency is maximum at 100 per cent. Therefore, comparison of the CCR and BCC scores provides deeper insight into the sources of inefficiency that a DMU might have. If there is a difference between the CRS and VRS TE scores, this indicates scale inefficiencies exist. In a case where a DMU is found to be inefficient, one can decompose this total inefficiency to see in what degree this is due to scale inefficiency or technical inefficiency.

Let θ^*_{CCR} (TE) and θ^*_{BCC}(PTE) denote the CCR and BCC TE scores of a DMU. The scale efficiency is defined by

$$\text{SE} = \theta^*_{CCR} / \theta^*_{BCC} \qquad (10)$$

Therefore, Scale Efficiency can be defined to be CCR Efficiency over BCC Efficiency. Scale efficiency can be interpreted as follows:

- If SE = 1, then a DMU is scale efficient, i.e., its combination of inputs and outputs is efficient under both CRS as well VRS.
- If SE < 1, then the combination of inputs and outputs is not scale efficient.

The above described approach does not allow identifying whether the DMU is operating under increasing returns to scale (IRS) or decreasing returns to scale (DRS). This problem can be solved by application of a further DEA model under the non-increasing returns to scale (NIRS). The (NIRS) DEA model is formulated by substituting the $e^T\lambda = 1$ restriction in the above BCC model with , $e^T\lambda \leq 1$ as shown below:

<u>Input-oriented</u>

$$\min \theta - \in (e^T s^+ + e^T s^-) \qquad (11)$$

subject to

$$Y\lambda - s^+ = Y_q$$
$$X\lambda + s^- = \theta X_o$$
$$e^T\lambda \leq 1$$
$$\lambda, s^+, s^- \geq 0$$

<u>Output-oriented</u>

$$\max \ \phi + \in (e^T s^+ + e^T s^-) \qquad (12)$$

subject to

$$Y\lambda - s+ = \phi Y_q$$
$$X\lambda + s^- = X_q$$
$$e^T\lambda \leq 1$$
$$\lambda, s^+, s^- \geq 0$$

This constraint ensures that DMUs will only be compared to DMUs of the same or smaller size, not with any DMU that is larger. Scale efficiency is due to either increasing or decreasing returns to scale, which can be determined by comparing the BCC score with that estimated under non-increasing returns to scale (NIRS), i.e. to determine if IRS or DRS exists, the NIRS TE is compared to the BCC TE estimate.

Scale efficiency is then derived according to the following rules (Fare, Grosskopf & Lovell, 1994):

- If $\frac{TE_{BCC}}{TE_{NIRS}} = 1$ then the DMU operates under decreasing returns to scale. It means that a firm is scale inefficient because of the possibility that it can achieve a larger output.

- If $\frac{TE_{BCC}}{TE_{NIRS}} \neq 1$, then the DMU operates under increasing returns to scale and inefficiency is caused by a too less of output.

In other words, if the NIRS TE and BCC TE estimates are unequal, then this indicates IRS and the scale of DMU level operations can be increased. If the two are equal, DRS exists and DMU operations needs to be reduced in size.

Using the relationship (10) the (global) technical efficiency (TE) of a DMU is decomposed as

$$\mathbf{TE = PTE \times SE} \qquad (13)$$

The global or overall inefficiency of a DMU is explained by inefficient operation (PTE) or by the scale effect (SE) or by both.

The characterization of the CCR model as "constant returns-to-scale" model is technically correct but somewhat misleading because this model can also be used to determine whether returns-to-scale are increasing or decreasing. This is accomplished by the following theorem proved by Banker and Thrall (1992)[31]:

Theorem 1: Let (x_o, y_o)[32] be a point on the efficiency frontier. Employing a CCR model in envelopment form to obtain an optimal solution $(\lambda_1^*, \lambda_2^* ..., \lambda_n^*)$, returns-to-scale at this point can be determined from the following conditions:

(i) If $\sum_{j=1}^{n} \lambda_j^* = 1$ in any alternate optimum then constant returns-to-scale prevails.

(ii) If $\sum_{j=1}^{n} \lambda_j^* > 1$ for all alternate optima then decreasing returns-to-scale prevail.

(iii) If $\sum_{j=1}^{n} \lambda_j^* < 1$ for all alternate optima then increasing returns-to-scale prevail.

The relations between BCC and CCR models are described be the following theorem due to Ahn et al. (1989):[33]

Theorem 2: A DMU_o found to be efficient with a CCR model will also be found to be efficient with the corresponding BCC model and constant returns-to-scale prevail at DMU_o.

The converse is not necessarily true. A DMU can be simultaneously characterised efficient by a BCC model and inefficient by a CCR model with $\theta_{CCR}^* < \theta_{BCC}^*$.However, if $\theta_{CCR}^* = \theta_{BCC}^*$ then there will be at least one alternate optimum for this θ_{CCR}^* for which. $\sum_{j=1}^{n} \lambda_j^* = 1$.[34]

Generally, the CCR-efficiency score for each DMU will not exceed the BCC-efficiency score, i.e.

$$\theta_{CCR}^* \leq \theta_{BCC}^*$$

which is intuitively clear since the BCC-model analyzes each DMU "locally" (i.e. compared to the subset of DMUs that operate in the same region of returns to scale) rather than "globally". The equality holds when the scale efficiency is unity, i.e., the DMU is operating at the *Most Productive Scale Size* (MPSS).

3.3 Ranking in DEA

CCR model and BCC model are called classical models and they can not be used in ranking efficient units. Since efficiency score of all DMUs that are effective in DEA are assigned as "1", it is not possible to rank effective units between each other. DEA can be used only for ranking inefficient DMUs and in order to abolish this disadvantage various methods were developed. The most commonly used method developed for ranking efficient decision units is the super efficiency model proposed by Andersen and Petersen.

3.3.1 Super-Efficiency Model

The efficiency score in standard DEA models is limited to unit (100%). Nevertheless, the number of efficient units identified by DEA models and reaching the maximum efficiency score 100% can be relatively high and especially in problems with a small number of decision units the efficient set can contain almost all the units. In such cases it is very important to have a tool for a diversification and classification of efficient units. That is why several DEA models for classification of efficient units were formulated. In these models, the efficient scores of inefficient units remain lower than 100% but the efficient score for efficient units can be higher than 100 percent. Thus, the efficient score can be taken as a basis for a complete ranking of efficient units. The DEA models that relax the condition for unit efficiency are called super-efficiency models.

The super-efficiency models are always based on removing the evaluated efficient unit from the set of units. This removal leads to the modification of the efficient frontier and the super-efficiency is measured as a distance between evaluated unit and a unit on the new efficient frontier. Of course several distance measures can be used - this leads to different super-efficiency definitions. The first super-efficiency DEA model was formulated by Andersen and Petersen (1993). Its input oriented formulation (3) below is very close to the standard input oriented formulation of model (1). In this model the weight λ_q of the evaluated unit DMU_q is equated to zero. This cannot influence the efficient score of the inefficient units but the efficient score of the efficient units is not limited by 100 percent in this case. The input oriented formulation of the Andersen and Petersen model with constant return to scale is as follows:

$$\min \theta - \varepsilon\left(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+\right) \qquad (14)$$

subject to

$$\sum_{j=1, j\neq 0}^{n} \lambda_j x_{ij} + s_i^- = \theta x_{io,} \qquad i = 1, 2, \ldots, m$$

$$\sum_{j=1, j\neq 0}^{n} \lambda_j y_{rj} - s_r^+ = y_{ro,} \qquad r = 1, 2, \ldots, s$$

$$\lambda_j \geq 0 \qquad j = 1, 2, \ldots, n, j \neq 0$$

$$s_i^- \geq 0, \qquad i = 1, \ldots, m$$

$$s_r^+ \geq 0, \qquad r = 1, \ldots, s$$

where 0 denotes the analyzed decision making unit and $j \neq 0$ means removing the analyzed decision making unit from the constraint group, this is the basic idea of the super efficiency model. For efficient units, $\theta^* \geq 1$ and for inefficient units, $0 < \theta^* \leq 1$.

However, under certain conditions this procedure can lead to infeasibility. A necessary, but not sufficient, condition for infeasibility is that an excluded DMU be 'extreme-efficient'. Either it has a feasible LP with super-efficiency scores strictly greater than 100%, or it has an infeasible LP. A necessary and sufficient condition for infeasibility in an input-oriented model is that the excluded DMU have the only zero value for any input, or the only positive value for any output, among all DMUs in the reference set. Infeasibility cannot arise in an output-oriented CCR super-efficiency model. In case of BCC model, infeasibility arises in either orientation whenever there is no referent DMU for the excluded DMU. A necessary condition for infeasibility is that the excluded DMU be 'extreme-efficient'. A sufficient condition for infeasibility is the pattern of zeros mentioned above. When all inputs and all outputs are positive for all DMUs, a sufficient condition for infeasibility is that the excluded DMU be 'strongly super-efficient' in the sense that (a) in an input-oriented model it has at least one output strictly larger than the corresponding output for any other DMU in the reference set, or (b) in an output-oriented model it has at least one input strictly smaller than the corresponding input for any other DMU in the reference set. A necessary and sufficient condition for infeasibility is that the excluded DMU be 'super-efficient' in the sense that (a) in an input oriented model it has at least one output strictly larger than a convex combination of that output among all DMUs in the reference set, or (b) in an output-oriented model it has at least one input strictly smaller than a convex combination of that input among all DMUs in the reference set.

In particular, the super-efficiency measure examines the maximal radial change in inputs and/or outputs for an observation to remain efficient, i.e. how much can the inputs be increased (or the outputs decreased) while not become inefficient. The larger the value of the super-efficiency measure the higher an observation is ranked among the efficient units. Super-efficiency measures can be calculated for both inefficient and efficient observations. In the case of inefficient observations the values of the efficiency measure do not change, while efficient observations may obtain higher values. Values of super-efficiency are therefore not restricted to 1 (for the efficient observations), but can in principle take any value greater than or equal 1. Super-efficiency measures are calculated on the basis of removing the production unit from the best-practice reference technology. This explains why the inefficient observations do not change value by calculating super-efficiency measures, as the inefficient observations are not influencing the best-practice technology. But for an efficient DMU we may get in the input oriented perspective an efficiency score larger than one and in the output oriented perspective an efficiency score of less than one, where a high (a low) super-efficiency score is associated with a high efficiency rank. However, a very high score in the input oriented (a very low score in the output oriented perspective) may indicate that a DMU is highly specialized and therefore not comparable to other DMUs. Hence the concept of super-efficiency also helps to identify such DMUs.

3.3.2 Cross-Efficiency Model

The cross-efficiency model was first developed by Sexton et al. (1986), inaugurating the subject of ranking in DEA with increased discriminatory power. Given n DMUs, the cross efficiency method simply calculates the efficiency score of each DMU n times, using the optimal weights evaluated by the n separate DEA models built for each of the n DMUs. In the cross-efficiency model, the evaluation loses its connection to the multiplier weights as the weights are used equally on all the units. Additionally, if the optimal weights are not unique, goal-programming has to be applied to choose from optimal solutions, such as aggressive or benevolent secondary goals (Sexton et al., 1986). Moreover, if the number of DMUs being evaluated increases, the calculation burden of cross-efficiency model becomes extremely heavy.

There has been a tremendous development in the subject matter of DEA, with many more new approaches introduced, more and newer modifications to the original models, etc. New uses of DEA with accompanying new developments and extensions continue to appear. To cover all these topics is beyond the scope of this thesis but we will look at a few of the extensions that have been attached to the above standard models.

3.4 Extensions to the Basic DEA Models

A lot of extensions to the above mentioned four models have been developed in recent times, which allow further fine tuning to the basic models. A few of these numerous extensions are briefly presented[35]:

a) The basic DEA models always assume that inputs and outputs can be altered by the DMUs. In realistic situation there are often variables that are exogenous variables that can not be altered. For example the distribution of competitors may influence efficiency scores without being alterable by the DMUs. These variables are called nondiscretionary. When we have a so called uncontrolled or non-discretionary input variable, then according to Banker and Morey (1986a)[36] the model (4) then takes the following form:

$$\min \theta - \varepsilon\left(\sum_{i \in D}^{m} s_i^- + \sum_{r=1}^{s} s_r^+\right) \qquad (15)$$

subject to

$$\sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = \theta x_{i0}, \quad i \in D$$

$$\sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = x_{i0}, \quad i \in ND$$

$$\sum_{j=1}^{n} \lambda_j y_{rj} - s_r^+ = y_{ro}, \qquad (r = 1, 2, \dots, s)$$

$$\lambda_j \geq 0 \qquad (j = 1, 2, \dots, n)$$

$$s_r^+ \geq 0, \qquad (r = 1, 2, \dots, s)$$

$$s_i^- \geq 0, \qquad (i = 1, 2, \dots, m)$$

The symbols *D* and *ND* refer to "discretionary" and "non-discretionary", respectively. In model (12) the non-discretionary variables do not enter directly into the efficiency evaluations. But the slacks for uncontrolled inputs are still useful. They indicate that more output is achievable. Hence, the DMU_o will be CCR (or BCC) efficient if and only if both of the following conditions are satisfied 1) θ_0^*=1 and 2) all slacks in the objective functions are zero.

b) The comparison of quite different units was also addressed by Banker, R. D.; Morey, R. C. (1986b)[37]. The introduction of categorical variables extends the application focus of the DEA.

c) The former described models all focus only analyse one period situation. This restriction was enhanced through a window analysis that allows the comparison of changing efficiency of a unit over time, by treating this DMU as if it were a different DMU.

d) Besides the above models, certain other modifications also exist. One of them, labelled the Slacks-Based Measure of Efficiency (SBM) model, was designed by Tone (2001). This model serves as the basis for the definition of super-efficiency. Efficiency is measured only by additional variables s^+ and s^- The model formula, provided constant returns to scale, is:

$$\min P = \frac{1 - \frac{1}{m}\sum_{i=1}^{m} s_i^- / x_{i0}}{1 + \frac{1}{s}\sum_{i-1}^{r} s_r^+ / y_{r0}} \qquad (16)$$

subject to

$$\sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = x_{i0,} \qquad r = 1,2, \dots, s$$

$$\sum_{j=1}^{n} \lambda_j y_{rj} - s_r^+ = y_{ro,} \qquad i = 1, 2, \dots, m$$

$$s_i^-, s_r^+, \lambda_j, \geq 0, \qquad j = 1, 2, \dots, m$$

The variables s^+ and s^- measure the distance of inputs $X\lambda$ and outputs $Y\lambda$ of a virtual unit from those of the unit evaluated (X_q). The numerator and the denominator of the objective function of model (5) measures the average distance of inputs and outputs, respectively, from the efficiency threshold. For variable returns to scale, condition $e^T\lambda = 1$ only needs to be appended to the formula. The SBM efficiency rate is always lower or equal to that of the input-oriented CCR model, i.e.

Theorem 3: The optimal SBM ρ^* is not greater than the optimal θ^*_{CCR}.

This theorem reflects the fact that SBM accounts for all inefficiencies whereas θ^*_{CCR} accounts only for "purely technical inefficiencies".[38] The relationship between CCR-efficiency and BSM-efficiency is given by the following theorem due Tone (1997):

Theorem:4 A DMU is CCR-efficient if and only if it is SBM-efficient.

This means that a unit rated as SBM efficient is CCR efficient at the same time.

3.5 Strengths and Limitations of DEA

Amongst the strengths of the DEA is that, DEA is less data demanding as it works fine with small sample size. The small sample size is among other reasons, which leads us to DEA as the tool of choice for evaluating efficiency of Indian container handling ports/terminals. Furthermore, DEA does not require a preconceived structure or specific functional form to be imposed on the data in identifying and determining the efficient frontier, error and inefficiency structures of the DMUs. According to the literature, it is better to adopt the DEA technique when it has been shown that a commonly agreed functional form relating inputs to outputs is difficult to prove or find. Such specific functional form is truly difficult to show for port services. The literature also acknowledges the edge of the DEA by stating that this technique allows the researchers to choose any kind of input and output of managerial interest, regardless of different measurement units. There is no need for standardisation[39].

Three useful features of DEA are first, each DMU is assigned a single efficiency score, hence allowing ranking amongst the DMUs in the sample. Second, it highlights the areas of improvement for each single DMU. For example, since a DMU is compared to a set of efficient DMUs with similar input-output configurations, the DMU in question is able to identify whether it has used input excessively or its output has been under-produced. Finally, there is possibility of making inferences on the DMUs general profile. We should be aware that the technique used here is a comparison between the production performances of each DMU to a set of efficiency DMUs. The set of efficiency DMUs is called the reference set. The owners of the DMUs may be interested to know which DMU frequently appears in this set. A DMU that appears more than others in this set is called the global leader. Clearly, this information gives huge benefits to the DMU owner, especially in positioning its entity in the market.

The main weakness of the DEA is that it assumes data are free from measurement errors. Furthermore, since efficiency is measured in a relative way, its analysis is confined to the sample set used. This means that an efficient DMU found in the analysis cannot be compared with other DMUs outside of the sample. The reason is simple. Each sample, separated, let us say, by year, represents a single frontier, which is constructed on the assumption of same technology. Therefore, comparing the efficiency measures of a DMU across time cannot be interpreted as technical progress but rather has to be taken as changes in efficiency (Canhoto and Dermine, 2003[40]). One another limitation of the DEA model that needs to be borne in mind is that as the number of variables in a DEA model increases, the discrimination power of DEA models decreases: more firms tend to be on the efficiency frontier. An individual firm is likely to be using relatively less of a particular input or producing more of a particular output, which places it near or on the multidimensional frontier (Rossi and Ruzzier (2000). This problem is more common under the variable returns to scale (VRS) environment since the VRS efficiency scores are relatively higher than CRS efficiency scores.

Thus, the most important advantage of DEA over traditional econometric frontier studies is that it is a non-parametric, deterministic method and therefore does not require a priori assumptions about the analytical form of the production function. Therefore, the probability of a misspecification of the production technology is zero. The disadvantage is that, being a non-parametric method, it is more sensitive to possible mis-measurement problems.

4. CONCEPT OF PRODUCTIVITY

Productivity is one important component of the monitoring, analysis and supervision of company performance. The term productivity was probably first mentioned by the French mathematician Quesnay in an article in 1766 (Sumanth 1998). In 1950, the Organization for European Economic Cooperation (OEEC), one of the oldest organizations espousing productivity enhancement, particularly in the Europe, issued a formal definition (OEEC1950)[41]:

Productivity is the quotient obtained by dividing output by one of the factors of production. In this way it is possible to speak of the productivity of capital, investment, or raw materials, according to whether output is being considered in relation to capital, investment, or raw materials etc.

Different financial ratios can provide a description of the productivity of a firm and its productivity change over time or between firms. We will briefly look at the classical productivity measures and then go on to describe a newer method for measuring productivity change – the Malmquist productivity change index.

4.1 Importance of Productivity

The performance of a firm, converting inputs into outputs, can be defined in many ways. One possible measure of performance is a productivity ratio. By defining the productivity of a firm as the ratio of outputs that it produces to the inputs used, the larger values of this ratio are associated with better performance. Productivity is a relative concept. Therefore, the productivity of a company in the present year could be measured relative to its productivity last year, or it could be measured relative to the productivity of another company in the same year. It is even possible to compare the productivity of an industry over time or across countries.

4.2 Productivity Management

Productivity is one of the major responsibilities of management. By attaining productivity increases, several other management goals are automatically achieved. An increase in the productivity of a firm results in improved product quality and service, decreased production costs as well as improved market share and profit. In the effort to achieve productivity goals, however, management must not lose sight of the other important management responsibilities – ensuring service quality, timeliness, accomplishing the mission and customer satisfaction. Indicators of the performance of these management responsibilities should also be tracked and emphasized by management.

Success in any productivity enhancement program depends on the leadership, participation and the ongoing support of every manager. The main point of productivity management is to identify areas of potential productivity improvement. In order to manage productivity in the true sense of the term, four phases must be linked together (Sumanth 1998):

- Measurement
- Evaluation
- Planning
- Improvement

These four phases form a continuous productivity process or cycle. The first phase of the productivity cycle is measurement. The present productivity level of the firm must be compared with the target level. This evaluation will provide a vision of the new productivity level for the following period. Depending on the planned level of productivity, improvement must arrive in the subsequent periods. Productivity improvement marks the end of the first productivity cycle, but productivity must be measured again in the next period and this then becomes the beginning of the next new productivity cycle.

The following sections of this study focus on productivity measurement. On the whole, it is not easy to measure productivity due to the following important aspects. Productivity information must be understandable. The results, and also the data collection and analysis system, should be easy to interpret and at the disposal of the decision maker at the appropriate time. Only then will productivity information have predictive value in the planning phase and feedback value to aid monitoring and supervisory activities. Finally, productivity data must also include all those aspects of production that are important to management and that actually represent the activity.

4.3 Productivity Measures

Productivity measures may be classified into several major groups, where none of the measures or groups is considered to be the best. The most commonly used productivity ratio groups are:

- Partial productivity (PP)
- Total factor (labour plus capital) productivity (TFP)
- Total productivity (TP)

The formulae for partial productivity, total factor productivity and total productivity are presented below.

$$\text{PP} = \frac{O}{L\ (or\ M,C,E,m)}$$

$$\text{TFP} = \frac{O}{L+C}$$

$$\text{TP} = \frac{O}{L+M+C+E+m}$$

where, PP – partial productivity, TFP – total factor productivity, TP – total productivity
O – output, L – labour, M – material, C – capital, E – energy, *m* – other inputs.

Partial productivity measures look at the ratio of output to a single input. These include labour productivity (e.g. output per hour worked or per employee), materials productivity (e.g. output per unit of material used), capital productivity (e.g. output per unit of capital invested) and energy productivity (e.g. output per unit of energy consumed). Therefore, this single input can either be labour, materials, capital, energy or some other input. The weakness of partial productivity measures is that they tend to overstate increases in productivity. The advantage of partial productivity measures is that they are much easier to understand and to measure.

A broader and theoretically more pertinent concept that also incorporates the effect of capital is total factor productivity. Total factor productivity (TFP) takes the ratio of output to capital and labour services. The advantage of total factor productivity is that it accounts for capital-labour substitution. The disadvantages are that it is a more difficult ratio to understand and measure.

Total productivity (TP) is the ratio of output to all combined inputs including labour, materials, capital, energy and others inputs. Total productivity is a more accurate productivity measure than total factor productivity, and its weaknesses are similar to those of total factor productivity.

4.4 Productivity Indices

Productivity measurement is usually conducted from two perspectives – according to the level of productivity and trends in the productivity. The productivity ratio refers to the productivity level at a given point in time expressed as output units delivered per unit of input resources expended. Trends are defined by looking at productivity development over time. Productivity trend ratios are commonly converted into an index.

These two measurement dimensions have different uses. Productivity level data can be useful in determining budget requirements and identifying opportunities for improvement by comparing an entity's productivity levels with that of other entities delivering the same or similar services. Productivity trend data can be useful in identifying opportunities for improvement by comparing current productivity with that of previous periods, and providing a scorekeeping device on management accountability for improving productivity.

Indices make it possible to show the input, output and productivity rates on the same graph. So productivity indices can provide us with some information on the causes of changes to productivity - whether they are attributable to the input or the output dimension. There have been used two different performance indices for the evaluation of efficiency and productivity change in economic units - the stochastic Tornqvist (1936) index and the non-stochastic Malmquist (1953) index. The former approach requires, explicitly or implicitly, a specific assumption about the form of the production technology. It is difficult to determine how well a postulating parametric function approaches the unknown true technology because the maintained hypothesis of parametric form can never be tested directly (Varian, 1984). The parametric approach has additional problems due to the estimation of accurate parameters from scarce and imperfect data (Stier and Bengston, 1992). To aggregate inputs and outputs, the Tornqvist index requires cost or revenue shares, which are hard to obtain, especially in cross-nation analysis. The Malmquist index is defined by using distance functions. The distance functions allow us to describe a multi-input, multi-output production technology without the need to specify the producer behavior (such as cost minimization or profit maximization). While the Tornqvist index presumes production is always efficient, the Malmquist productivity index allows for inefficient performance and does not presume an underlying functional form for technology. Under the stochastic approaches the deviations from the frontier are attributed to purely random shocks and inefficiency. Under the non-stochastic approaches all such deviations are attributed to inefficiency. Therefore, this study adopts the Malmquist index in examining the productivity change of the container handling ports in India.

4.4.1 Malmquist Productivity Index and its Decomposition

Calculating the productivity level or index can be very easy when a single output is produced from a single input (partial productivity). But companies usually produce many outputs from many inputs. How is it possible then to calculate the productivity change index? Enter the Malmquist productivity change index – one method for measuring productivity change over time or between firms.

In 1953, Sten Malmquist, a Swedish economist and statistician, published in *Trabajos de Estadistica* (Malmquist 1953)[42] a quantity index for use in consumption analysis. Later Caves, Christensen and Divert (1982)[43] adapted Malmquist's idea for production analysis and they named their productivity change indices after Sten Malmquist. According to Grifell-Tatjé and Lovell (1996)[44], the Malmquist index has some advantages relative to other productivity indices. As mentioned above, it does not require input prices or output prices, which makes it particularly useful in situations where prices are misrepresented or non-existent. The Malmquist index also does not require the profit maximization or cost minimization assumption. This makes it useful in situations where the objectives of producers differ, are unknown or not achieved. An attractive feature of the Malmquist productivity index is that it decomposes.

Färe *et al.* (1989) showed that the Malmquist productivity index can be decomposed into two components – technical efficiency change and technical change. The value of this decomposition is that it provides insight into the sources of productivity change. The main disadvantage of the Malmquist index is the necessity to compute distance functions. There are many different methods that could be used to measure the distance function, which makes up the Malmquist productivity index. One of the more popular methods has been the DEA-like linear programming method suggested by Färe *et al.* (1994).

Productivity indices explain the role of index figures in measuring growth in outputs (output oriented approach) that are net of input growth. One way to measure a change in productivity is to see how much more output has been produced, using a given input level and the present state of technology, relative to what could be produced under a given reference technology using the same input level. An alternative is to measure the change in productivity by examining the reduction in input use, which is feasible given the need to produce a given level of output under a reference technology. These two approaches are referred to as the output-oriented and input-oriented measures of change in productivity (Coelli, 1998). In the current study I have concentrated on the output-oriented Malmquist productivity index.

To define an output distance function, following Färe (1994a)[45], we begin by defining the production technology T in any given time period. T is the set of all feasible input-output vectors, x is an N dimensional vector of inputs and y is an M dimensional vector of outputs.

$$T\{(x, y) : x \text{ can produce } y\} \qquad (17)$$

One can define an output distance function which is a multi-output generalization of what in the single output case would be the ratio of actual to potential output. Thus if the production point is on the frontier, this ratio equals unity. The distance function is the reciprocal of the Farrell output-oriented measure of efficiency, which can be calculated using DEA.

$$D(x, y) \inf\{q : \{x, y/q\} \in T\} \tag{18}$$

Caves, Christensen and Diewert (1982)[46] define the Malmquist productivity index with reference to the technology in time period t as:

$$M_t = \frac{D_t\,(x_{t+1}, y_{t+1})}{D_t(x_t, y_t)} \tag{19}$$

A similar measure could be defined using period $t+1$ as the base. To avoid arbitrariness in the choice of base period, Färe et al. (1994a) propose using the geometric mean of the indexes for the periods t and $t+1$ which yields the following Malmquist index of productivity change:

$$M_{t,t+1} = \left[\frac{D_t\,(x_{t+1}, y_{t+1})}{D_t(x_t, y_t)}\right]\left[\frac{D_{t+1}\,(x_{t+1}, y_{t+1})}{D_{t+1}(x_t, y_t)}\right]^{1/2} \tag{20}$$

This equation represents the productivity point (x_{t+1}, y_{t+1}) relative to the production point (x_t, y_t).

Färe et al. (1994a) show that the above measure can also be expressed as:

$$M_{t,t+1} = \frac{D_{t+1}\,(x_{t+1}, y_{t+1})}{D_t(x_t, y_t)}\left\{\left[\frac{D_t\,(x_{t+1}, y_{t+1})}{D_{t+1}\,(x_{t+1}, y_{t+1})}\right]\left[\frac{D_t(x_t, y_t)}{D_{t+1}(x_t, y_t)}\right]\right\}^{1/2} \tag{21}$$

In the above equation, the first term measures efficiency change and the second term (in square brackets) measures technical change. Calculating Malmquist index and its components requires the calculation of four distances: $Dt\,(x_t, y_t)$, $Dt{+}1(x_{t+1}, y_{t+1})$, $D_t(x_{t+1}, y_{t+1})$ and $D_{t+1}(x_t, y_t)$. This is accomplished by solving four (constant returns to scale) linear programming problems, thus making use of the fact that output distance function is the inverse of the Farrell output oriented measure of technical efficiency. For each firm k, $Dt\ (x_t\ ,\ y_t\ )$ can be computed as follows, as can $D_{t+1}(x_{t+1}, y_{t+1})$ by substituting $t+1$ for t:

$$[D_t\,(y_{t,}\,x_t)]^{-1} = \max \emptyset + \in\ (\mathrm{e^T s^+ + e^T s^-}) \tag{22}$$

Subject to

$$Y_t\,\lambda - s^+ = \emptyset\, Y_{0t}$$
$$X_t\,\lambda + s^- = X_{0t}$$
$$\lambda, s^+, s^- \geq 0$$

$$[D_{t+1}\,(y_{t+1}\,x_{t+1})]^{-1} = \max \emptyset + \in\ (\mathrm{e^T s^+ + e^T s^-}) \tag{23}$$

Subject to

$$Y_{t+1}\,\lambda - s^+ = \emptyset\, Y_{0,t+1}$$
$$X_{t+1}\,\lambda + s^- = X_{0,t+1}$$
$$\lambda, s^+, s^- \geq 0$$

$$[D_t\,(y_{t+1}\,x_{t+1})]^{-1} = \max \emptyset + \in\ (\mathrm{e^T s^+ + e^T s^-}) \tag{24}$$

Subject to

$$Y_t\,\lambda - s^+ = \emptyset\, Y_{0,t+1}$$
$$X_t\,\lambda + s^- = X_{0,t+1}$$
$$\lambda, s^+, s^- \geq 0$$

$$\left[D_{t+1}\,(y_t, x_t)\right]^{-1} = \max \emptyset + \in (\mathrm{e^T s^+} + \mathrm{e^T s^-}) \tag{25}$$

Subject to

$$Y_{t+1}\,\lambda - s^{+} = \emptyset\, Y_{0,t}$$

$$X_{t+1}\,\lambda + s^{-} = X_{0,t}$$

$$\lambda, s^{+}, s^{-} \geq 0$$

Both the efficiency change and technical change measures in (21) can be decomposed further [Grifell-Tatjé and Lovell (1997)[47]]. They define the output oriented measure of scale efficiency as the ratio of an output oriented distance function for a variable returns to scale technology (V) to that for constant returns to scale technology (C) or:

$$S_t\,(x_t, y_t) = \frac{D_t(x_t, y_t\,|V)}{D_t(x_t, y_t|C)} \tag{26}$$

Calculating this requires solving the LP in (22) and (23) with the following additional restriction for variable returns to scale:

$$e^T\lambda = 1$$

Thus, the efficiency change component in (21) can be decomposed into scale efficiency change and pure efficiency change as:

$$EFFCH = \frac{S_t(x_t, y_t)}{S_{t+1}(x_{t+1}, y_{t+1})}\;\frac{D_{t+1}(x_{t+1}, y_{t+1}|V)}{D_t(x_t, y_t\,|V)} \tag{27}$$

The technical change component in (21) can also be decomposed as the product of the magnitude of technical change and (input and output) bias, where magnitude is defined as follows:

$$MTECH = \frac{D_t(x_t, y_t\,)}{D_{t+1}(x_{t+1}, y_{t+1})} \tag{28}$$

To summarize, the Malmquist index of productivity change can be represented as the product of efficiency change and technical change. Efficiency change can be further decomposed as the sum product of scale efficiency change and pure efficiency change, whereas technical change can be decomposed as the sum product of the change in the magnitude of technical change and bias.

DEA is not only used to determine efficient and non-efficient units but recently, it is also used to rank DMUs.

Standing at the interface of sea and inland transportation, modern ports with modern logistics and hub-and-spoke transportation patterns face much fiercer competition than ever before. The modern container ports suffer under both internal and external pressure and need to exhibit management competency in the pursuit of a suitable strategy and in the allocation of scarce resources. Indian ports, after years of neglect and failure by the State in ensuring the level of investment necessary for infrastructure maintenance and development which undermined their competitiveness, are undergoing a sea-change. There is a heavy investment in infrastructure build-up, with adoption of new technologies as well as in maintenance, which calls for more efficiency among them. Thus, the need for performance appraisal and benchmarking! The next chapter undertakes to empirically analyse the efficiency container handling ports in India with the help of Data Envelopment Analysis. The ranking of efficient ports is then undertaken with the help of the super-efficiency model. We also employ the Malmquist total factor productivity (TFP) index to measure the impact of productivity change among the different ports/terminals.

Endnotes

---

[1] Notteboom, T. and W. Winklemans, (2002), "Stakeholder Relations Management in Ports: dealing with the interplay of forces among stakeholders in a changing competitive environment", *IAME 2002*, Panama City, Panama, pp.1-20.

[2] Johns, K. E. and R. C. Fair, (1999), *Principles of Economics* (5th ed.). Prentice-Hall.

[3] Dyson, R. G., (2000), "Strategy, performance and operational research", *Journal of the Operational Research Society*, 51 pp. 1-7

[4] Thanassoulis, E., (2001), *Introduction to the theory and application of Data Envelopment Analysis: a foundation text with integrated software*, Kluwer Academic Publishers, Norwell, Massachusetts, United States.

[5] Clark, X., D. Dollar and A. Micco, (2004), "Port Efficiency, Maritime Transport Costs and Bilateral Trade," *NBER working paper 10353.*

Wilson, J. S., C. L. Mann and T. Otsuki, (2003), "Trade Facilitation and Economic Development: A New Approach to Quantifying the Impact." *The World Bank Economic Review*, 17, pp. 367-89.

[6] Tongzon, J., (1995), "Determinants of port performance and efficiency. *Transportation Research Part A*: Policy and Practice, Part A 29 (3), pp.245–252.

Tongzon, J., (2001), "Efficiency Measurement of Selected Australian and Other International Ports Using Data Envelopment Analysis," *Transportation Research A: Policy and Practice* 35(A), pp. 113-28.

[7] Song, D. W., K. Cullinane and M. Roe, (2001), *The Productive Efficiency of Container Terminals*, Ashgate Publishing Limited, England.

[8] Song, D. W. and C. H. Han, (2004), "An Econometric Approach to Performance Determinants of Asian Container Terminals", *International Journal of Transport Economics*, 31(1).

[9] Trujillo, L. and G. Nombela, G, (1999), "Privatisation and Regulation of the Seaport", UNCTAD *Monographs on Port Management No. 6 on Port Management*, UNCTAD.

[10] Tongzon, J., (1995), "Determinants of port performance and efficiency. *Transportation Research Part A*: Policy and Practice, Part A, 29 (3), pp.245–252.

Tongzon, J., (2001), "Efficiency Measurement of Selected Australian and Other International Ports Using Data Envelopment Analysis," *Transportation Research A: Policy and Practice* 35 (A), pp. 113-28.

[11] Bichou, K. and R. Gray, (2004), "A logistics and supply chain management approach to port performance measurement", *Maritime Policy and Management* 31, pp. 47–67.

[12] Cullinane, K. and M. Khanna, (1999), "Economies of Scale in Large Containerships", *Journal of Transport Economics and Policy,* 33 (2) pp. 185-208.

[13] Cullinane, K. and M. Khanna, (1999), "Economies of Scale in Large Containerships", *Journal of Transport Economics and Policy,* 33 (2) pp. 185-208.

[14] Notteboom, T. and W. Winkelmans, (2001), "Structural changes in logistics: how will port authorities face the challenge?" *Maritime Policy & Management,* 28 (1), pp. 71–89.

[15] De Neufville, R. and K. Tsunokawa, (1981), "Productivity and Returns to Scale of Container Ports", *Maritime Policy and Management*, 8 (2), pp.121-129.

[16] Kim, M. and A. Sachish, (1986), "The structure of production, technical change and productivity in ports", *Journal of Industrial Economics*, 35 (2), pp. 209–223.

[17] Tongzon, J. L., (1995), "Determinants of port performance and efficiency. *Transportation Research Part A*: Policy and Practice, Part A, 29 (3), pp.245–252.

[18] Coelli, T., P. D. S. Rao and G. Battese, (2002), *An Introduction to Efficiency and Productivity Analysis*, Kluwer Academic Publishers.

[19] Koopmans, T. C., (1951), "Analysis of Production as an Efficient Combination of Activities", in Koopmans (ed.), *Activity Analysis of Production and Allocation.*

[20] Debreu, G., (1951), "The Coefficient of Resource Utilization," *Econometrica,* 19 (3), pp. 273-292.

[21] Shephard, R. W., (1953), *Cost and Production Functions,* New York: Princeton University Press.

[22] Farrell, M. J., (1957), "The Measurement of Productivity Efficiency", *Journal of the Royal Statistical Society,* Series TO, 120 (3), pp. 253-290.

[23] Lansink, A. O., K. Pietola and S. Bäckman, (2002), "Efficiency and productivity of conventional and organic farms in Finland 1994–1997", *European Review of Agriculture Economics,* 29 (1), pp.51-65.

[24] Charnes, A., W. W. Cooper and E. Rhodes, (1978), "Measuring the Efficiency of Decision Making Units", *European Journal of Operational Research,* 2, pp. 429-444.

[25] Cooper, W. W., L. M. Seiford and K. Tone, (2006), *Introduction to Data Envelopment Analysis and its Uses,* Springer Publications, U.S.A.

[26] Dyson, R. G., R. Allen, A. S. Camanho, V. V. Podinovski, C. S. Sarrico and E. A. Shale, (2001), "Pitfalls and Protocols in DEA", *European Journal of Operational Research,* 132 (2), pp. 245-259.

[27] In the original model, these variables are restricted to being strictly positive. However, their strict positive sign can be guaranteed by using the infinitesimal to generate the Non-Archimedean ordered extension field, in which its usage guarantees that optimal solutions of the transformed linear program are at finite non-zero external points.

[28] In practice, it is preferable to solve the dual of the multiplier form for reasons as detailed in "Introduction to Data Envelopment Analysis and Its Uses" [Cooper, W. et al (2006)].

[29] Introduction to Data Envelopment Analysis and Its Uses, [Cooper, W. et al (2006)].

[30] Introduction to Data Envelopment Analysis and Its Uses, [Cooper, W. et al (2006)].

[31] Banker, R. D. and R. M. Thrall, (1992), "Estimating most productive scale size using data envelopment analysis", *European Journal of Operational Research*, 62 (1), pp. 74-84.

[32] In Cooper et al. (2006; Section 5.4; pp.126) this theorem is proved by eliminating the need for the assumption that (x_o, y_o) is on the efficiency frontier.

[33] Ahn, T., A. Charnes and W. W. Cooper, (1989), "A note on the efficiency characterizations obtained in different DEA models", *Socio-Economic Planning Sciences*, 22 (6), pp. 253-257.

[34] Cooper et al., (2006; Section 5.5; pp.131).

[35] For more information about extensions to the basic DEA models see "Introduction to Data Envelopment Analysis and Its Uses" [Cooper, W. et al (2006)].

[36] Banker, R. D. and R. Morey, (1986a), "Efficiency analysis for exogenously fixed inputs and outputs", *Operations Research*, vol. 34 (4), pp. 513-521.

[37] Banker, R. D. and R. Morey, (1986b), "The Use of Categorical Variables in Data Envelopment Analysis", *Management Science* (32:12), pp. 1613-1627.

[38] Cooper et al. (2006; Section 4.4; pp.99).

[39] An additional advantage according to Canhoto and Dermine (2003) is that the DEA technique is preferred to parametric methods is when the sample size is small.

[40] Canhoto, A. and Dermine, J. (2003), "A Note on Banking Efficiency in Portugal: New Vs. Old. Banks", *Journal of Banking and Finance* 27 (11), pp. 2087-2098.

[41] OEEC, (1950), *Terminology of Productivity*, Paragraph.2.2, rue Andre-Pascal, Paris-16.

[42] Malmquist, S., (1953), "Index Numbers and Indifference Curves", *Trabajos de Estatistica*, 4 (1), pp. 209–42.

[43] Caves D. W., L. R. Christensen and W. E. Diewert, (1982), "The Economic Theory of Index Numbers and Measurement of Input, Output and Productivity", *Econometrica*, 50, pp. 1393–1414.

[44] Grifell-Tatje, E. and C. A. K. Lovell, (1997), "The Sources of Productivity Change in Spanish Banking", *European Journal of Operational Research*, 98, pp. 364–80.

[45] Färe, R., S. Grosskopf, M. Norris and Z. Zhong (1994a), "Productivity Growth, Technical Progress and Efficiency Change in Industrialized Countries," *American Economic Review*, March, 84 (1), pp. 66-83.

[46] Caves, D., W. Christensen, R. Laurits and W. E. Diewert (1982), "The Economic Theory of Index Numbers and the Measurement of Input, Output and Productivity," *Econometrica*, November, 50 (6), pp. 1393-1414.

[47] Grifell-Tatjé, E. and C. A. K. Lovell, (1997), "A DEA-based analysis of productivity change and intertemporal managerial performance", *Annals of Operations Research*, 73, pp. 177-189.