Chapter 1

DATA ENVELOPMENT ANALYSIS

History, Models and Interpretations

William W. Cooper[1], Lawrence M. Seiford[2] and Joe Zhu[3]

[1] Red McCombs School of Business, University of Texas at Austin, Austin, TX 78712 USA email: cooperw@mail.utexas.edu

[2] Department of Industrial and Operations Engineering, University of Michigan at Ann Arbor, Ann Arbor, MI 48102 USA email: seiford@umich.edu

[3] Department of Management, Worcester Polytechnic Institute, Worcester, MA 01609 USA email: jzhu@wpi.edu

Abstract: In a relatively short period of time Data Envelopment Analysis (DEA) has grown into a powerful quantitative, analytical tool for measuring and evaluating performance. DEA has been successfully applied to a host of different types of entities engaged in a wide variety of activities in many contexts worldwide. This chapter discusses the fundamental DEA models and some of their extensions.

Key words: Data envelopment analysis (DEA); Efficiency; Performance

1. INTRODUCTION

Data Envelopment Analysis (DEA) is a relatively new "data oriented" approach for evaluating the performance of a set of peer entities called Decision Making Units (DMUs) which convert multiple inputs into multiple outputs. The definition of a DMU is generic and flexible. Recent years have seen a great variety of applications of DEA for use in evaluating the performances of many different kinds of entities engaged in many different activities in many different contexts in many different countries. These DEA applications have used DMUs of various forms to evaluate the performance of entities, such as hospitals, US Air Force wings, universities, cities, courts, business firms, and others, including the performance of countries, regions, etc. Because it requires very few assumptions, DEA has also opened up possibilities for use in cases which have been resistant to other approaches because of the complex (often unknown) nature of the relations between the multiple inputs and multiple outputs involved in DMUs.

As pointed out in Cooper, Seiford and Tone (2000), DEA has also been used to supply new insights into activities (and entities) that have previously been evaluated by other methods. For instance, studies of benchmarking practices with DEA have identified numerous sources of inefficiency in some of the most profitable firms - firms that had served as benchmarks by reference to this (profitability) criterion – and this has provided a vehicle for identifying better benchmarks in many applied studies. Because of these possibilities, DEA studies of the efficiency of different legal organization forms such as "stock" vs. "mutual" insurance companies have shown that previous studies have fallen short in their attempts to evaluate the potentials of these different forms of organizations. Similarly, a use of DEA has suggested reconsideration of previous studies of the efficiency with which pre- and post-merger activities have been conducted in banks that were studied by DEA.

Since DEA in its present form was first introduced in 1978, researchers in a number of fields have quickly recognized that it is an excellent and easily used methodology for modeling operational processes for performance evaluations. This has been accompanied by other developments. For instance, Zhu (2002) provides a number of DEA spreadsheet models that can be used in performance evaluation and benchmarking. DEA's empirical orientation and the absence of a need for the numerous *a priori* assumptions that accompany other approaches (such as standard forms of statistical regression analysis) have resulted in its use in a number of studies involving efficient frontier estimation in the governmental and nonprofit sector, in the regulated sector, and in the private sector. See, for instance, the use of DEA to guide removal of the Diet and other government agencies from Tokyo to locate a new capital in Japan, as described in Takamura and Tone (2003).

In their originating study, Charnes, Cooper, and Rhodes (1978) described DEA as a 'mathematical programming model applied to observational data [that] provides a new way of obtaining empirical estimates of relations - such as the production functions and/or efficient production possibility surfaces – that are cornerstones of modern economics'.

Formally, DEA is a methodology directed to frontiers rather than central tendencies. Instead of trying to fit a regression plane through the *center* of the data as in statistical regression, for example, one ‘floats’ a piecewise linear surface to rest on top of the observations. Because of this perspective, DEA proves particularly adept at uncovering relationships that remain hidden from other methodologies. For instance, consider what one wants to mean by “efficiency”, or more generally, what one wants to mean by saying that one DMU is more efficient than another DMU. This is accomplished in a straightforward manner by DEA without requiring explicitly formulated assumptions and variations with various types of models such as in linear and nonlinear regression models.

Relative efficiency in DEA accords with the following definition, which has the advantage of avoiding the need for assigning a priori measures of relative importance to any input or output,

Definition 1.1 (Efficiency – Extended Pareto-Koopmans Definition): Full (100%) efficiency is attained by any DMU if and only if none of its inputs or outputs can be improved without worsening some of its other inputs or outputs.

In most management or social science applications the theoretically possible levels of efficiency will not be known. The preceding definition is therefore replaced by emphasizing its uses with only the information that is empirically available as in the following definition:

Definition 1.2 (Relative Efficiency): A DMU is to be rated as fully (100%) efficient on the basis of available evidence if and only if the performances of other DMUs does not show that some of its inputs or outputs can be improved without worsening some of its other inputs or outputs.

Notice that this definition avoids the need for recourse to prices or other assumptions of weights which are supposed to reflect the relative importance of the different inputs or outputs. It also avoids the need for explicitly specifying the formal relations that are supposed to exist between inputs and outputs. This basic kind of efficiency, referred to as “technical efficiency” in economics can, however, be extended to other kinds of efficiency when data such as prices, unit costs, etc., are available for use in DEA.

In this chapter we discuss the mathematical programming approach of DEA that implements the above efficiency definition. Section 2 of this chapter provides a historical perspective on the origins of DEA. Section 3 provides a description of the original “CCR ratio model” of Charnes, Cooper, and Rhodes (1978) which relates the above efficiency definition to other definitions of efficiency such as the ones used in engineering and science, as well as in business and economics. Section 4 describes some methodological extensions that have been proposed. Section 5 expands the development to concepts like "allocative" (or price) efficiency which can add additional power to DEA when unit prices and costs are available. This is done in section 5 and extended to profit efficiency in section 6 after which a conclusion section 7 is supplied.

2. BACKGROUND AND HISTORY

In an article which represents the inception of DEA, Farrell (1957) was motivated by the need for developing better methods and models for evaluating productivity. He argued that while attempts to solve the problem usually produced careful measurements, they were also very restrictive because they failed to combine the measurements of multiple inputs into any satisfactory overall measure of efficiency. Responding to these inadequacies of separate indices of labor productivity, capital productivity, etc., Farrell proposed an activity analysis approach that could more adequately deal with the problem. His measures were intended to be applicable to any productive organization; in his words, '… from a workshop to a whole economy'. In the process, he extended the concept of "productivity" to the more general concept of "efficiency".

Our focus in this chapter is on basic DEA models for measuring the efficiency of a DMU *relative* to similar DMUs in order to estimate a 'best practice' frontier. The initial DEA model, as originally presented in Charnes, Cooper, and Rhodes (CCR) (1978), built on the earlier work of Farrell (1957).

This work by Charnes, Cooper and Rhodes originated in the early 1970s in response to the thesis efforts of Edwardo Rhodes at Carnegie Mellon University's School of Urban & Public Affairs - now the H.J. Heinz III School of Public Policy and Management. Under the supervision of W.W. Cooper, this thesis was to be directed to evaluating educational programs for disadvantaged students (mainly black or Hispanic) in a series of large scale studies undertaken in U.S. public schools with support from the Federal government. Attention was finally centered on Program Follow Through - a huge attempt by the U.S. Office (now Department) of Education to apply principles from the statistical design of experiments to a set of matched schools in a nation-wide study. Rhodes secured access to the data being processed for that study by Abt Associates, a Boston based consulting film, under contract with the US Office of Education. The data base was sufficiently large so that issues of degrees of freedom, etc., were not a serious problem despite the numerous input and output variables used in the study. Nevertheless, unsatisfactory and even absurd results were secured from all of the statistical-econometric approaches that Rhodes attempted to use.

While trying to respond to this situation, Rhodes called Cooper's attention to M.J. Farrell's seminal article "The Measurement of Productive Efficiency," in the 1957 Journal of the Royal Statistical Society. In this article Farrell used "activity analysis concepts" to correct what he believed were deficiencies in commonly used index number approaches to productivity (and like) measurements.

Cooper had previously worked with A. Charnes in order to give computationally implementable form to Tjalling Koopmans' "activity analysis concepts." So, taking Farrell's statements at face value, Cooper and Rhodes formalized what was involved in the definitions that were given in section 1 of this chapter. These definitions then provided the guides that were used for their subsequent research.

The name of Pareto is assigned to the first of these two definitions for the following reasons. In his Manual of Political Economy (1906) the Swiss-Italian economist, Vilfredo Pareto, established the basis of modern "welfare economics", i.e., the part of economics concerned with evaluating public policies, by noting that a social policy could be justified if it made some persons better off without making others worse off. In this way the need for making comparisons between the value of the gains to some and the losses to others could be avoided. This avoids the necessity of ascertaining the "utility functions" of the affected individuals and/or to "weight" the relative importance of each individual's gains and losses.

This property, known as the “Pareto criterion” as used in welfare economics, was carried over, or adapted, in Activity Analysis of Production and Allocation, a book edited by Koopmans (1951). In this context, it was "final goods" which were accorded this property, in that they were all constrained so that no final good was allowed to be improved if this improvement resulted in worsening one or more other final goods. These final goods (=outputs) were to be satisfied in stipulated amounts while inputs were to be optimally determined in response to the prices and amounts exogenously fixed for each output (=final good). Special attention was then directed by Koopmans to "efficiency prices" which are the prices associated with efficient allocation of resources (=inputs) to satisfy the pre-assigned demands for final goods. For a succinct summary of the mechanisms involved in this "activity analysis" approach, see p. 299 in Charnes and Cooper (1961).

Pareto and Koopmans were concerned with analyses of entire economies. In such a context it is reasonable to allow input prices and quantities to be determined by reference to their ability to satisfy final demands. Farrell, however, extended the Pareto-Koopmans property to inputs as well as outputs and explicitly eschewed any use of prices and/or related "exchange mechanisms." Even more importantly, he used the performance of other DMUs to evaluate the behavior of each DMU relative to the outputs and the inputs they all used. This made it possible to proceed empirically to determine their relative efficiencies.

The resulting measure which is referred to as the "Farrell measure of efficiency," was regarded by Farrell as restricted to meaning "technical efficiency" or the amount of "waste" that can be eliminated without worsening any input or output. This was then distinguished by Farrell from "allocative" and "scale" efficiencies as adapted from the literature of economics. These additional efficiencies will be discussed later in this chapter where the extensions needed to deal with problems that were encountered in DEA attempts to use these concepts in actual applications will also be discussed. Here we want to note that Farrell's approach to efficiency evaluations, as embodied in the "Farrell measure," carries with it an assumption of equal access to inputs by all DMUs. This does not mean that all DMUs use the same input amounts, however, and, indeed, part of their efficiency evaluations will depend on the input amounts used by each DMU as well as the outputs which they produce.

This "equal access assumption" is a mild one, at least as far as data availability is concerned. It is less demanding than the data and other requirements needed to deal with aspects of performance such as "allocative" or "scope" and "scale efficiencies." Furthermore, as discussed below, this assumption can now be relaxed. For instance, one can introduce "non-discretionary variables and constraints" to deal with conditions beyond the control of a DMU's management--in the form of "exogenously" fixed resources which may differ for each DMU. One can also introduce "categorical variables" to insure that evaluations are effected by reference to DMUs which have similar characteristics, and still other extensions and relaxations are possible, as will be covered in the discussions that follow.

To be sure, the definition of efficiency that we have referred to as "Extended Pareto-Koopmans Efficiency" and "Relative Efficiency" were formalized by Charnes, Cooper and Rhodes rather than Farrell. However, these definitions conform both to Farrell's models and the way Farrell used them. In any case, these were the definitions that Charnes, Cooper and Rhodes used to guide the developments that we next describe.

The Program Follow Through data with which Rhodes was concerned in his thesis recorded "outputs" like "increased self esteem in a disadvantaged child" and "inputs" like "time spent by a mother in reading with her child," as measured by psychological tests and prescribed record keeping and reporting practices. Farrell's elimination of the need for information on prices proved attractive for dealing with outputs and inputs like these--as reported for each of the schools included in the Program Follow Through experiment.

Farrell's empirical work had been confined to single-output cases and his sketch of extensions to multiple outputs did not supply what was required for applications to large data sets like those involved in Program Follow Through. To obtain what was needed in computationally implementable form, Charnes, Cooper and Rhodes developed the dual pair of linear programming problems that are modeled in the next section, section 3. It was then noticed that Farrell's measure failed to account for the non-zero slacks, which is where the changes in proportions connected with mix inefficiencies are located (in both outputs and inputs). The possible presence of non-zero slack as a source of these mix inefficiencies also requires attention even when restricted to "technical efficiency."

We now emphasize the problems involved in dealing with these slacks because a considerable part of the DEA (and related) literatures continues to be deficient in its treatment of non-zero slack even today. A significant part of the problem to be dealt with, as we noted above, involves the possible presence of alternate optima in which the same value of the Farrell measure could be associated with zero slack in some optima but not in others. Farrell introduced "points at infinity" in what appears to have been an attempt to deal with this problem but was unable to give operationally implementable form to this concept. Help in dealing with this problem was also not available from the earlier work of Sidney Afriat (1972), Ronald Shephard (1970) or Gerhard Debreu (1951). To address this problem, Charnes, Cooper and Rhodes introduced mathematical concepts that are built around the "non-Archimedean" elements associated with $\varepsilon > 0$ which handles the problem by insuring that slacks are always maximized without altering the value of the Farrell measure.

The dual problems devised by Cooper and Rhodes readily extended the above ideas to multiple outputs and multiple inputs in ways that could locate inefficiencies in each input and each output for every DMU. Something more was nevertheless desired in the way of summary measures. At this point, Cooper invited A. Charnes to join him and Rhodes in what promised to be a very productive line of research. Utilizing the earlier work of Charnes and Cooper (1962), which had established the field of "fractional programming," Charnes was able to put the dual linear programming problems devised by Cooper and Rhodes into the equivalent ratio form represented in (1.1) below and this provided a basis for unifying what had been done in DEA with long standing approaches to efficiency evaluation and analysis used in other fields, such as engineering and economics.

Since the initial study by Charnes, Cooper, and Rhodes some 2000 articles have appeared in the literature. See Cooper, Seiford and Tone

(2000). See also G. Tavares (2003). Such rapid growth and widespread (and almost immediate) acceptance of the methodology of DEA is testimony to its strengths and applicability. Researchers in a number of fields have quickly recognized that DEA is an excellent methodology for modeling operational processes, and its empirical orientation and minimization of a *priori* assumptions has resulted in its use in a number of studies involving efficient frontier estimation in the nonprofit sector, in the regulated sector, and in the private sector.

At present, DEA actually encompasses a variety of alternate (but related) approaches to evaluating performance. Extensions to the original CCR work have resulted in a deeper analysis of both the "multiplier side" from the dual model and the "envelopment side" from the primal model of the mathematical duality structure. Properties such as isotonicity, nonconcavity, economies of scale, piecewise linearity, Cobb-Douglas loglinear forms, discretionary and nondiscretionary inputs, categorical variables, and ordinal relationships can also be treated through DEA. Actually the concept of a frontier is more general than the concept of a "production function" which has been regarded as fundamental in economics in that the frontier concept admits the possibility of multiple production functions, one for each DMU, with the frontier boundaries consisting of "supports" which are "tangential" to the more efficient members of the set of such frontiers.

3. CCR DEA MODEL

To allow for applications to a wide variety of activities, we use the term Decision Making Unit (=DMU) to refer to any entity that is to be evaluated in terms of its abilities to convert inputs into outputs. These evaluations can involve governmental agencies and not-for-profit organizations as well as business firms. The evaluation can also be directed to educational institutions and hospitals as well as police forces (or subdivision thereof) or army units for which comparative evaluations of their performance are to be made.

We assume that there are n DMUs to be evaluated. Each DMU consumes varying amounts of m different inputs to produce s different outputs. Specifically, $\boldsymbol{DMU_j}$ consumes amount x_{ij} of input i and produces amount y_{rj} of output r. We assume that $x_{ij} \geq 0$ and $y_{rj} \geq 0$ and further assume that each DMU has at least one positive input and one positive output value.

We now turn to the "ratio-form" of DEA. In this form, as introduced by Charnes, Cooper, and Rhodes, the ratio of outputs to inputs is used to measure the relative efficiency of the $DMU_j = DMU_o$ to be evaluated relative to the ratios of all of the $j = 1, 2, \ldots, n$ DMU_j. We can interpret the

CCR construction as the reduction of the multiple-output /multiple-input situation (for each DMU) to that of a single 'virtual' output and 'virtual' input. For a particular DMU the ratio of this single virtual output to single virtual input provides a measure of efficiency that is a function of the multipliers. In mathematical programming parlance, this ratio, which is to be maximized, forms the objective function for the particular DMU being evaluated, so that symbolically

$$\max h_o(u,v) = \sum_r u_r y_{ro} / \sum_i v_i x_{io} \tag{1.1}$$

where it should be noted that the variables are the u_r's and the v_i's and the y_{ro}'s and x_{io}'s are the observed output and input values, respectively, of $\boldsymbol{DMU_o}$ **,** the DMU to be evaluated. Of course, without further additional constraints (developed below) (1.1) is unbounded.

A set of normalizing constraints (one for each DMU) reflects the condition that the virtual output to virtual input ratio of every DMU, including $DMU_j = DMU_o$, must be less than or equal to unity. The mathematical programming problem may thus be stated as

$$\max h_o(u,v) = \sum_r u_r y_{ro} / \sum_i v_i x_{io} \tag{1.2}$$

subject to

$$\sum_r u_r y_{rj} / \sum_i v_i x_{ij} \leq 1 \text{ for } j = 1, \ldots, n,$$

$$u_r, v_i \geq 0 \text{ for all } i \text{ and } r.$$

Remark: A fully rigorous development would replace $u_r, v_i \geq 0$ with

$\dfrac{u_r}{\sum_{i=1}^{m} v_i x_{io}}, \dfrac{u_r}{\sum_{i=1}^{m} v_i x_{io}} \geq \varepsilon > 0$ where ε is a non-Archimedean element smaller than any positive real number. See Arnold et al. (1998). This condition guarantees that solutions will be positive in these variables. It also leads to the $\varepsilon > 0$ in (1.6) which, in turn, leads to the 2nd stage optimization of the slacks as in (1.10).

The above ratio form yields an infinite number of solutions; if (u*, v*) is optimal, then (αu*, αv*) is also optimal for α > 0. However, the transformation developed by Charnes and Cooper (1962) for linear fractional programming selects a representative solution [i.e., the solution (u, v) for which $\sum_{i=1}^{m} v_i x_{io}$ = 1] and yields the equivalent linear programming problem in which the change of variables from (u, v) to (μ, v) is a result of the Charnes-Cooper transformation,

$$
\begin{aligned}
&\max z = \sum_{r=1}^{s} \mu_r y_{ro} \\
&\text{subject to} \\
&\sum_{r=1}^{s} \mu_r y_{rj} - \sum_{i=1}^{m} \nu_i x_{ij} \le 0 \\
&\sum_{i=1}^{m} \nu_i x_{io} = 1 \\
&\mu_r, \nu_i \ge 0
\end{aligned}
\tag{1.3}
$$

for which the LP dual problem is

$$
\begin{aligned}
&\theta^* = \min \theta \\
&\text{subject to} \\
&\sum_{j=1}^{n} x_{ij} \lambda_j \le \theta x_{io} \qquad i = 1,2,\ldots,m; \\
&\sum_{j=1}^{n} y_{rj} \lambda_j \ge y_{ro} \qquad r = 1,2,\ldots,s; \\
&\lambda_j \ge 0 \qquad j = 1,2,\ldots,n.
\end{aligned}
\tag{1.4}
$$

This last model, (1.4), is sometimes referred to as the "Farrell model" because it is the one used in Farrell (1957). In the economics portion of the DEA literature it is said to conform to the assumption of "strong disposal" because it ignores the presence of non-zero slacks. In the operations research portion of the DEA literature this is referred to as "weak efficiency."

Possibly because he used the literature of "activity analysis"[1] for reference, Farrell also failed to exploit the very powerful dual theorem of linear programming which we have used to relate the preceding problems to each other. This also caused computational difficulties for Farrell because he did not take advantage of the fact that activity analysis models can be converted to linear programming equivalent that provide immediate access to the simplex and other methods for efficiently solving such problems. See, e.g., Charnes and Cooper (1961, Ch. IX). We therefore now begin to bring these features of linear programming into play.

By virtue of the dual theorem of linear programming we have $z^* = \theta^*$. Hence either problem may be used. One can solve say (1.4), to obtain an efficiency score. Because we can set $\theta = 1$ and $\lambda_k^* = 1$ with $\lambda_k^* = \lambda_o^*$ and all other $\lambda_j^* = 0$, a solution of (1.4) always exists. Moreover this solution implies $\theta^* \le 1$. The optimal solution, θ^*, yields an efficiency score for a particular DMU. The process is repeated for each DMU_j i.e., solve (1.4), with $(X_o, Y_o) = (X_{\boldsymbol{k}}, Y_{\boldsymbol{k}})$, where $(X_{\boldsymbol{k}}, Y_{\boldsymbol{k}})$ represent vectors with components x_{ik}, y_{rk} and, similarly (X_o, Y_o) has components x_{ok}, y_{ok}. DMUs for which $\theta^* < 1$ are inefficient, while DMUs for which $\theta^* = 1$ are boundary points.

Some boundary points may be "weakly efficient" because we have non-zero slacks. This may appear to be worrisome because alternate optima may

[1] See T.C. Koopmans (1951).

have non-zero slacks in some solutions, but not in others. However, we can avoid being worried even in such cases by invoking the following linear program in which the slacks are taken to their maximal values.

$$\begin{aligned}
&\max \sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+ \\
&\text{subject to} \\
&\sum_{j=1}^{n} x_{ij}\lambda_j + s_i^- = \theta^* x_{io} \qquad i = 1,2,\ldots,m; \\
&\sum_{j=1}^{n} y_{rj}\lambda_j - s_r^+ = y_{ro} \qquad r = 1,2,\ldots,s; \\
&\lambda_j, s_i^-, s_r^+ \geq 0 \ \forall i, j, r
\end{aligned} \tag{1.5}$$

where we note the choices of s_i^- and s_r^+ do not affect the optimal θ^* which is determined from model (1.4).

These developments now lead to the following definition based upon the "relative efficiency" definition 1.2 which was given in section 1 above.

Definition 1.3 (DEA Efficiency): The performance of DMU_o is fully (100%) efficient if and only if both (i) $\theta^* = 1$ and (ii) all slacks $s_i^{-*} = s_r^{+*} = 0$.

Definition 1.4 (Weakly DEA Efficient): The performance of DMU_o is weakly efficient if and only if both (i) $\theta^* = 1$ and (ii) $s_i^{-*} \neq 0$ and/or $s_r^{+*} \neq 0$ for some *i* and *r* in some alternate optima.

It is to be noted that the preceding development amounts to solving the following problem in two steps:

$$\begin{aligned}
&\min \theta - \varepsilon\left(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+\right) \\
&\text{subject to} \\
&\sum_{j=1}^{n} x_{ij}\lambda_j + s_i^- = \theta x_{io} \qquad i = 1,2,\ldots,m; \\
&\sum_{j=1}^{n} y_{rj}\lambda_j - s_i^+ = y_{ro} \qquad r = 1,2,\ldots,s; \\
&\lambda_j, s_i^-, s_r^+ \geq 0 \ \forall i, j, r
\end{aligned} \tag{1.6}$$

where the s_i^- and s_r^+ are slack variables used to convert the inequalities in (1.4) to equivalent equations. Here $\varepsilon > 0$ is a so-called non-Archimedean element defined to be smaller than <u>any</u> positive real number. This is equivalent to solving (1.4) in two stages by first minimizing θ, then fixing $\theta = \theta^*$ as in (1.2), where the slacks are to be maximized without altering the previously determined value of $\theta = \theta^*$. Formally, this is equivalent to granting "preemptive priority" to the determination of θ^* in (1.3). In this manner, the fact that the non-Archimedean element ε is defined to be smaller than any positive real number is accommodated without having to specify the value of ε.

Alternately, one could have started with the output side and considered instead the ratio of virtual input to output. This would reorient the objective from max to min, as in (1.2), to obtain

$$\begin{aligned} &\text{Min } \textstyle\sum_i v_i x_{io} / \sum_r u_r y_{ro} \\ &\text{Subject to} \\ &\textstyle\sum_i v_i x_{ij} / \sum_r u_r y_{rj} \geq 1 \text{ for } j = 1, \ldots, n, \\ &u_r, v_i \geq \varepsilon > 0 \text{ for all } i \text{ and } r. \end{aligned} \tag{1.7}$$

where $\varepsilon > 0$ is the previously defined non-Archimedean element.

Again, the Charnes-Cooper (1962) transformation for linear fractional programming yields model (1.8) (multiplier model) below, with associated dual problem, (1.9) (envelopment model), as in the following pair,

$$\begin{aligned} &\min q = \sum_{i=1}^{m} v_i x_{io} \\ &\text{subject to} \\ &\sum_{i=1}^{m} v_i x_{ij} - \sum_{r=1}^{s} \mu_r y_{rj} \geq 0 \\ &\sum_{r=1}^{s} \mu_r y_{ro} = 1 \\ &\mu_r, v_i \geq \varepsilon, \quad \forall r, i \end{aligned} \tag{1.8}$$

$$\begin{aligned} &\max \phi + \varepsilon\Big(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+\Big) \\ &\text{subject to} \\ &\sum_{j=1}^{n} x_{ij}\lambda_j + s_i^- = x_{io} \quad i = 1,2,\ldots,m; \\ &\sum_{j=1}^{n} y_{rj}\lambda_j - s_r^+ = \phi y_{ro} \quad r = 1,2,\ldots,s; \\ &\lambda_j \geq 0 \qquad\qquad j = 1,2,\ldots,n. \end{aligned} \tag{1.9}$$

See pp 75-76 in Cooper, Seiford and Tone (2000) for a formal development of this transformation and modification of the expression for $\varepsilon > 0$. See also the remark following (1.2).

Here we are using a model with an output oriented objective as contrasted with the input orientation in (1.6). However, as before, model (1.9) is calculated in a two-stage process. First, we calculate ϕ^* by ignoring the slacks. Then we optimize the slacks by fixing ϕ^* in the following linear programming problem,

$$
\begin{aligned}
&\max \sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+ \\
&\text{subject to} \\
&\sum_{j=1}^{n} x_{ij}\lambda_j + s_i^- = x_{io} \qquad i = 1,2,\ldots,m; \\
&\sum_{j=1}^{n} y_{rj}\lambda_j - s_r^+ = \phi^* y_{ro} \qquad r = 1,2,\ldots,s; \\
&\lambda_j \geq 0 \qquad j = 1,2,\ldots,n.
\end{aligned} \tag{1.10}
$$

We then modify the previous input-oriented definition of DEA efficiency to the following output-oriented version.

Definition 1.5: DMU_o is efficient if and only if $\phi^* = 1$ and $s_i^{-*} = s_r^{+*} = 0$ for all i and r. DMU_o is weakly efficient if $\phi^* = 1$ and $s_i^{-*} \neq 0$ and (or) $s_r^{+*} \neq 0$ for some i and r in some alternate optima.

Table 1-1 presents the CCR model in input- and output-oriented versions, each in the form of a pair of dual linear programs.

Table 1-1. CCR DEA Model

Input-oriented	
Envelopment model	Multiplier model
$\min \theta - \varepsilon(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+)$ subject to $\sum_{j=1}^{n} x_{ij}\lambda_j + s_i^- = \theta x_{io} \quad i = 1,2,\ldots,m;$ $\sum_{j=1}^{n} y_{rj}\lambda_j - s_r^+ = y_{ro} \quad r = 1,2,\ldots,s;$ $\lambda_j \geq 0 \quad j = 1,2,\ldots,n.$	$\max z = \sum_{r=1}^{s} \mu_r y_{ro}$ subject to $\sum_{r=1}^{s} \mu_r y_{rj} - \sum_{i=1}^{m} \nu_i x_{ij} \leq 0$ $\sum_{i=1}^{m} \nu_i x_{io} = 1$ $\mu_r, \nu_i \geq \varepsilon > 0$
Output-oriented	
Envelopment model	Multiplier model
$\max \phi + \varepsilon(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+)$ subject to $\sum_{j=1}^{n} x_{ij}\lambda_j + s_i^- = x_{io} \quad i = 1,2,\ldots,m;$ $\sum_{j=1}^{n} y_{rj}\lambda_j - s_r^+ = \phi y_{ro} \quad r = 1,2,\ldots,s;$ $\lambda_j \geq 0 \quad j = 1,2,\ldots,n.$	$\min q = \sum_{i=1}^{m} \nu_i x_{io}$ subject to $\sum_{i=1}^{m} \nu_i x_{ij} - \sum_{r=1}^{s} \mu_r y_{rj} \geq 0$ $\sum_{r=1}^{s} \mu_r y_{ro} = 1$ $\mu_r, \nu_i \geq \varepsilon > 0$

These are known as CCR (Charnes, Cooper, Rhodes, 1978) models. If the constraint $\sum_{j=1}^{n} \lambda_j = 1$ is adjoined, they are known as BCC (Banker, Charnes, Cooper, 1984) models. This added constraint introduces an additional variable, μo, into the (dual) multiplier problems. As will be seen in the next chapter, this extra variable makes it possible to effect returns-to-scale evaluations (increasing, constant and decreasing). So the BCC model is also referred to as the VRS (Variable Returns to scale) model and distinguished form the CCR model which is referred to as the CRS (Constant Returns to Scale) model.

We now proceed to compare and contrast the input and output orientations of the CCR model. To illustrate the discussion to follow we will employ the example presented in Figure 1-1 consisting of five DMUs, labeled P1, …, P5, each consuming a single input to produce a single output.

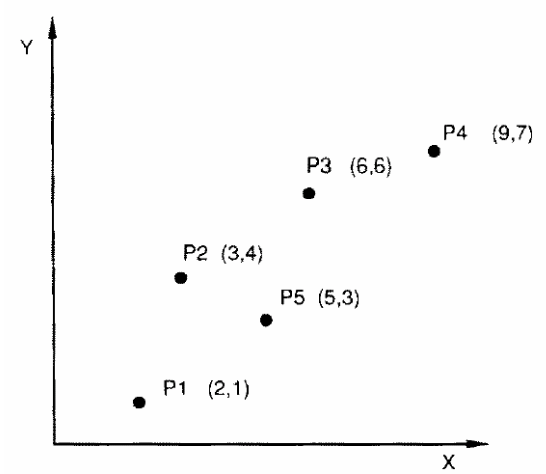

Figure 1-1. Example DMUs

The numbers in parentheses in Figure 1-1 are interpreted as coordinate values which correspond to the input and output of a DMU_j represented as Pj, j = 1, …, 5. In each case the value on the left in the parentheses is the input and the value on the right is the output for the Pj alongside which these values are listed.

To assist the reader in verifying the model interpretations which follow, Table 1-2 contains optimal solution values for the five example DMUs for both of the dual LP problems of the CCR model. For example, to evaluate the efficiency of P5 (DMU_5 in Table 1-2), we can solve the following input-oriented envelopment CCR model:

$$\begin{array}{ll} \min\theta & \\ \textit{subject to} & \\ 2\lambda_1 + 3\lambda_2 + 6\lambda_3 + 9\lambda_4 + 5\lambda_5 \le 5\theta & (input) \\ 1\lambda_1 + 4\lambda_2 + 6\lambda_3 + 7\lambda_4 + 3\lambda_5 \ge 3 & (output) \\ \lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5 \ge 0 & \end{array}$$

which yields the values of $\theta^* = 9/20$, $\lambda_2^* = ¾$ and $\lambda_j^* = 0$ ($j \neq 2$) (see the last two columns in the last row of the upper portion of Table 1-2).

Alternatively, we can solve the input-oriented multiplier CCR model,

$$\begin{aligned}
&\max z = 3\mu \\
&\textit{subject to} \\
&1\mu - 2\nu \le 0 \quad (P1) \\
&4\mu - 3\nu \le 0 \quad (P2) \\
&6\mu - 6\nu \le 0 \quad (P3) \\
&7\mu - 9\nu \le 0 \quad (P4) \\
&3\mu - 5\nu \le 0 \quad (P5) \\
&5\nu = 1 \\
&\mu, \nu \ge 0
\end{aligned}$$

which yields z* = 9/20, $\mu^* = 3/20$ and $\nu^* = 1/5$. Hence we have $\theta^* = z^*$. Moreover, with $\mu^* = 3/20$ and $\nu^* = 1/5$, this is also the value of $h_o(u^*, v^*) = 3 \cdot \frac{3}{20} / 5 \cdot \frac{1}{5} = \frac{9}{20}$ for the corresponding ratio model obtained from (1.2).

Table 1-2. Optimal solution values for the CCR model

	DMU	*z**	μ^*	ν^*	θ^*	λ^*
	1	3/8	3/8	1/2	3/8	λ_2 =1/4
	2	1	¼	1/3	1	λ_2 =1
Input-	3	¾	1/8	1/6	¾	λ_2 =3/2
oriented	4	7/12	1/12	1/9	7/12	λ_2 =7/4
	5	9/20	3/20	1/5	9/20	λ_2 =3/4
	1	8/3	1	4/3	8/3	λ_2 =2/3
	2	1	¼	1/3	1	λ_2 =1
Output-	3	4/3	1/6	2/9	4/3	λ_2 =2
oriented	4	12/7	1/7	4/21	12/7	λ_2 =3
	5	20/9	1/3	4/9	20/9	λ_2 =5/3

A DMU is inefficient if the efficiency score given by the optimal value for the LP problem is less than one ($\theta^* < 1$ or z* < 1). If the optimal value is equal to one *and* if there exist positive optimal multipliers ($\mu_r > 0$, $\nu_i > 0$), then the DMU is efficient. Thus, all efficient points lie on the frontier. However, a DMU can be a boundary point ($\theta^* = 1$) and be inefficient. Note that the complementary slackness condition of linear programming yields a condition for efficiency which is equivalent to the above; the constraints involving X_0 and Y_0, must hold with equality, i.e., $X_0 = X\lambda^*$ and $Y_0 = Y\lambda^*$ for all optimal λ^*, where X_0 and Y_0 are vectors and X and Y are matrices.

An inefficient DMU can be made more efficient by projection onto the frontier. In an input orientation one improves efficiency through proportional reduction of inputs, whereas an output orientation requires proportional augmentation of outputs. However, it is necessary to distinguish between a boundary point and an efficient boundary point. Moreover, the efficiency of a boundary point can be dependent upon the model orientation.

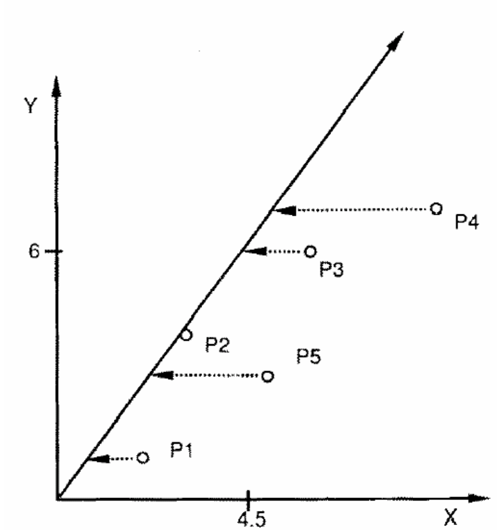

Figure 1-2. Projection to frontier for the input-oriented CCR model

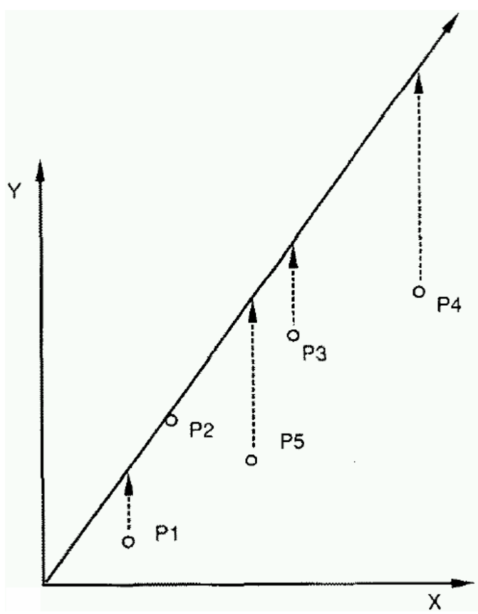

Figure 1-3. Projection to frontier for the output-oriented CCR model

The efficient frontier and DEA projections are provided in Figures 1-2 and 1-3 for the input-oriented and output-oriented CCR models, respectively. In both cases, the efficient frontier obtained from the CCR model is the ray $\{\alpha(x_2, y_2) | \alpha \geq 0\}$, where x_2 and y_2 are the coordinates of P2.

As can be seen from the points designated by the arrow head, an inefficient DMU may be projected to different points on the frontier under the two orientations. However, the following theorem provides a correspondence between solutions for the two models.

Theorem 1.1: Let (θ^*, λ^*) be an optimal solution for the input oriented model in (1.9). Then $(1/\theta^*, \lambda^*/\theta^*) = (\phi^*, \hat{\lambda}^*)$ is optimal for the corresponding output oriented model. Similarly if $(\phi^*, \hat{\lambda}^*)$ is optimal for the output oriented model then $(1/\phi^*, \hat{\lambda}^*/\phi^*) = (\theta^*, \lambda^*)$ is optimal for the input oriented model. The correspondence need not be 1-1, however, because of the possible presence of alternate optima.

For an input orientation the projection $(X_0,Y_0) \rightarrow (\theta^* X_0,Y_0)$ always yields a boundary point. But technical efficiency is achieved only if also all slacks are zero in all alternate optima so that $\theta^* X_0 = X\lambda^*$ and $Y_0 = Y\lambda^*$ for all optimal λ^*. Similarly, the output-oriented projection $(X_0,Y_0) \rightarrow (X_0, \phi^* Y_0)$ yields a boundary point which is efficient (technically) only if $\phi^* Y_0 = Y\lambda^*$ and $X_0 = X\lambda^*$ for all optimal λ^*. That is, the constraints are satisfied as equalities in all alternate optima for (1.4). To achieve technical efficiency the appropriate set of constraints in the CCR model must hold with equality.

To illustrate this, we consider a simple numerical example used in Zhu (2002) as shown in Table 1-3 where we have five DMUs representing five supply chain operations. Within a week, each DMU generates the same profit of $2,000 with a different combination of supply chain cost and response time.

Table 1-3. Supply Chain Operations Within a Week

	Inputs		Output
DMU	Cost ($100)	Response time (days)	Profit ($1,000)
1	1	5	2
2	2	2	2
3	4	1	2
4	6	1	2
5	4	4	2

Source: Zhu (2002).

We now turn to the BCC model for which Figure 1-4 presents the five DMUs and the piecewise linear DEA frontier. DMUs 1, 2, 3, and 4 are on the frontier. If we adjoin the constraint $\sum_{j=1}^{n}\lambda_j = 1$ to model (1.4) for DMU5, we get from the data of Table 1-3,

$$\begin{aligned}
&\text{Min } \theta \\
&\text{Subject to} \\
&1\,\lambda_1 + 2\lambda_2 + 4\lambda_3 + 6\lambda_4 + 4\lambda_5 \le 4\theta \\
&5\,\lambda_1 + 2\lambda_2 + 1\lambda_3 + 1\lambda_4 + 4\lambda_5 \le 4\theta \\
&2\,\lambda_1 + 2\lambda_2 + 2\lambda_3 + 2\lambda_4 + 2\lambda_5 \ge 2 \\
&\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 = 1 \\
&\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5 \ge 0
\end{aligned}$$

This model has the unique optimal solution of $\theta^* = 0.5$, $\lambda_2^* = 1$, and $\lambda_j^* = 0$ ($j \ne 2$), indicating that DMU5 needs to reduce its cost and response time to the amounts used by DMU2 if it is to be efficient This example indicates that technical efficiency for DMU5 is achieved at DMU_2 on the boundary.

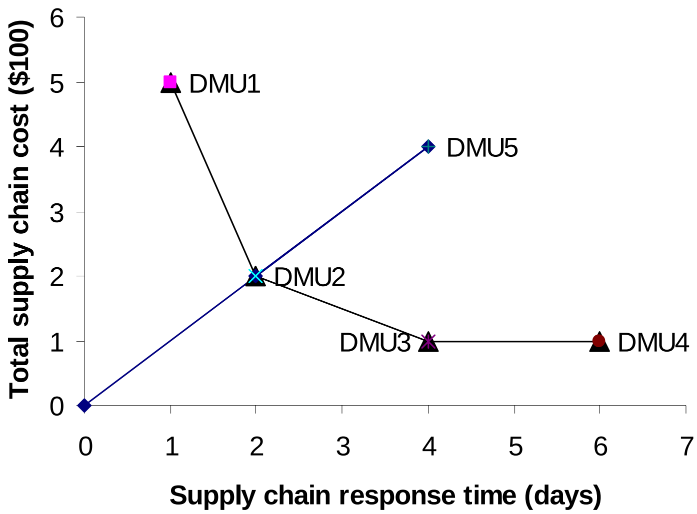

Figure 1-4. Five Supply Chain Operations

Source: Zhu (2002).

Now, if we similarly use model (1.4) with $\sum_{j=1}^{n}\lambda_j = 1$ for DMU4, we obtain $\theta^* = 1$, $\lambda_4^* = 1$, and $\lambda_j^* = 0$ ($j \ne 4$), indicating that DMU4 is on the frontier and is a boundary point. However, Figure 1-4 indicates that DMU4 can still reduce its response time by 2 days to achieve coincidence with DMU3 on the efficiency frontier. This input reduction is the input slack and the constraint with which it is associated is satisfied as a strict inequality in this solution. Hence, DMU4 is weakly efficient.

The nonzero slack can be found by using model (1.5). With the constraint $\sum_{j=1}^{n} \lambda_j = 1$ adjoined and setting $\theta^* = 1$ yields the following model,

$$\begin{aligned}
&\text{Max } s_1^- + s_2^- + s_1^+ \\
&\text{Subject to} \\
&1\lambda_1 + 2\lambda_2 + 4\lambda_3 + 6\lambda_4 + 4\lambda_5 + s_1^- = 6\theta^* = 6 \\
&5\lambda_1 + 2\lambda_2 + 1\lambda_3 + 1\lambda_4 + 4\lambda_5 + s_2^- = 1\theta^* = 1 \\
&2\lambda_1 + 2\lambda_2 + 2\lambda_3 + 2\lambda_4 + 2\lambda_5 - s_1^+ = 2 \\
&\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 = 1 \\
&\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, s_1^-, s_2^-, s_1^+ \geq 0
\end{aligned}$$

The optimal slacks are $s_1^{-*} = 2$, $s_2^{-*} = s_1^{+*} = 0$, with $\lambda_3^* = 1$ and all other $\lambda_j^* = 0$.

4. EXTENSIONS TO THE CCR MODEL

A number of useful enhancements have appeared in the literature. We here limit our coverage to five of the extensions that illustrate the adaptability of the basic DEA methodology. The extensions discussed below allow an analyst to treat both nondiscretionary and categorical inputs and outputs and to incorporate judgment or ancillary managerial information. They are also easily extended to investigate efficiency changes over multiple time periods, and to measure congestion.

4.1 Nondiscretionary Inputs and Outputs

The above model formulations implicitly assume that all inputs and outputs are discretionary, i.e., can be controlled by the management of each DMU and varied at its discretion. Thus, failure of a DMU to produce maximal output levels with minimal input consumption results in a worsened efficiency score. However, there may exist exogenously fixed (or nondiscretionary) inputs or outputs that are beyond the control of a DMU's management. Instances from the DEA literature include snowfall or weather in evaluating the efficiency of maintenance units, soil characteristics and topography in different farms, number of competitors in the branches of a restaurant chain, local unemployment rates which affect the ability to attract recruits by different U.S. Army recruitment stations, age of facilities in different universities, and number of transactions (for a purely gratis service) in library performance.

For example, Banker and Morey (1986a), whose formulations we use, illustrate the impact of exogenously determined inputs that are not controllable in an analysis of a network of fast food restaurants. In their study, each of the 60 restaurants in the fast food chain consumes six inputs to produce three outputs. The three outputs (all controllable) correspond to breakfast, lunch, and dinner sales. Only two of the six inputs, expenditures for supplies and expenditures for labor, are discretionary. The other four inputs (age of store, advertising level as determined by national headquarters, urban/rural location, and drive-in capability) are beyond the control of the individual restaurant manager in this chain.

The key to the proper mathematical treatment of a nondiscretionary variable lies in the observation that information about the extent to which a nondiscretionary input variable may be reduced is beyond the discretion of the individual DMU managers and thus cannot be used by them.

Suppose that the input and output variables may each be partitioned into subsets of discretionary (D) and nondiscretionary (N) variables. Thus,

$$I = \{1, 2, \ldots, m\} = I_D \cup I_N \text{ with } I_D \cap I_N = \varnothing$$

and

$$O = \{1, 2, \ldots, s\} = O_D \cup O_N \text{ with } O_D \cap O_N = \varnothing$$

where I_D, O_D and I_N, O_N refer to discretionary (D) and nondiscretionary (N) input, I, and output, O, variables, respectively and $\varnothing$ is the empty set.

To evaluate managerial performance in a relevant fashion we may need to distinguish between discretionary and non-discretionary inputs as is done in the following modified version of a CCR model.

$$\begin{aligned}
&\min \theta - \varepsilon\Big(\sum_{i\in I_D} s_i^- + \sum_{r=1}^{s} s_r^+\Big) \\
&\text{subject to} \\
&\sum_{j=1}^{n} x_{ij}\lambda_j + s_i^- = \theta x_{io} \qquad i \in I_D; \\
&\sum_{j=1}^{n} x_{ij}\lambda_j + s_i^- = x_{io} \qquad i \in I_N \\
&\sum_{j=1}^{n} y_{rj}\lambda_j - s_r^+ = y_{ro} \qquad r = 1,2,\ldots,s; \\
&\lambda_j \geq 0 \qquad j = 1,2,\ldots,n.
\end{aligned} \tag{1.11}$$

It is to be noted that the θ to be minimized appears only in the constraints for which $i \in I_D$, whereas the constraints for which $i \in I_N$ operate only indirectly (as they should) because the input levels x_{io} for $i \in I_N$, are not subject to managerial control. It is also to be noted that the slack variables associated with I_N, the non-discretionary inputs, are not included in the objective of (1.11) and hence the non-zero slacks for these inputs do not enter directly into the efficiency scores to which the objective is oriented.

The necessary modifications to incorporate nondiscretionary variables for the output-oriented CCR model is given by

$$
\begin{aligned}
&\max \phi + \varepsilon\left(\sum_{i=1}^{m} s_i^- + \sum_{r \in O_D} s_r^+\right) \\
&\text{subject to} \\
&\sum_{j=1}^{n} x_{ij}\lambda_j + s_i^- = x_{io} \qquad i = 1,2,\ldots,m; \\
&\sum_{j=1}^{n} y_{rj}\lambda_j - s_r^+ = \phi y_{ro} \qquad r \in O_D; \\
&\sum_{j=1}^{n} y_{rj}\lambda_j - s_r^+ = y_{ro} \qquad r \in O_N; \\
&\lambda_j \geq 0 \qquad j = 1,2,\ldots,n.
\end{aligned}
\tag{1.12}
$$

We should point out that there can be subtle issues associated with the concept of controllable outputs that may be obscured by the symmetry of the input/output model formulations. Specifically, switching from an input to an output orientation is not always as straightforward as it may appear. Interpretational difficulties for outputs not directly controllable may be involved as in the case of outputs influenced through associated input factors. An example of such an output would be sales that are influenced by advertising from the company headquarters, but are not directly controllable by district mangers. Finally, chapter 12 by Ruggiero provides some treatments for nondiscretionary variables.

4.2 Categorical Inputs and Outputs

Our previous development assumed that all inputs and outputs were in the same category. However this need not be the case as when some restaurants in a fast food chain have a dive-in facility and some do not. See Banker and Morey (1986b) for a detailed discussion.

To see how this can be handled, suppose that an input variable can assume one of L levels (1, 2, …, L). These L values effectively partition the set of DMUs into categories. Specifically, the set of DMUs $K = \{1, 2, \ldots, n\}$ = $K_1 \cup K_2 \cup \ldots \cup K_L$, where K_f = $\{j \mid j \in K$ and input value is $f\}$ and $K_i \cap K_j = \emptyset$, $i \neq j$. We wish to evaluate a DMU with respect to the envelopment surface determined for the units contained in it and all preceding categories. The following model specification allows $DMU_o \in K_f$.

$$\begin{aligned}
&\min \theta \\
&\text{subject to} \\
&\sum_{j \in \bigcup_{f=1}^{K} K_f} x_{ij}\lambda_j + s_i^- = \theta x_{io} \qquad i = 1,...,m; \\
&\sum_{j \in \bigcup_{f=1}^{K} K_f} y_{rj}\lambda_j - s_r^+ = y_{ro} \qquad r = 1,2,...,s; \\
&\lambda_j \geq 0 \qquad j = 1,2,...,n.
\end{aligned} \tag{1.13}$$

Thus, the above specification allows one to evaluate all DMUs $l \in D_1$ with respect to the units in K_1, all DMUs $l \in K_2$ with respect to the units in $K_1 \cup K_2$, …., all DMUs $l \in K_C$ with respect to the units in $\bigcup_{f=1}^{K_C} K_f$, etc. Although our presentation is for the input-oriented CCR model, it should be obvious that categorical variables can also be incorporated in this manner for any DEA model. In addition, the above formulation is easily implemented in the underlying LP solution algorithm via a candidate list.

The preceding development rests on the assumption that there is a natural nesting or hierarchy of the categories. Each DMU should be compared only with DMUs in its own and more disadvantaged categories, i.e., those operating under the same or worse conditions. If the categories are not comparable (e.g., public universities vs. private universities), then a separate analysis should be performed for each category.

4.3 Incorporating Judgment or A Priori Knowledge

Perhaps the most significant of the proposed extensions to DEA is the concept of restricting the possible range for the multipliers. In the CCR model, the only explicit restriction on multipliers is positivity, as noted for the ε>0 in (1.8). This flexibility is often presented as advantageous in applications of the DEA methodology, since a priori specification of the multipliers is not required. and each DMU is evaluated in its best possible light.

In some situations, however, this complete flexibility may give rise to undesirable consequences, since it can allow a DMU to appear efficient in ways that are difficult to justify. Specifically, the model can assign unreasonably low or excessively high values to the multipliers in an attempt to drive the efficiency rating for a particular DMU as high as possible.

Three situations for which it has proven beneficial to impose various levels of control are the following:

1. the analysis would otherwise ignore additional information that cannot be directly incorporated into the model that is used, e.g., the envelopment model;

2. management has strong preferences about the relative importance of different factors and what determines best practice; and
3. for a small sample of DMUs, the method fails to discriminate, and all are efficient.

Using the multiplier models that duality theory makes available for introducing restrictions on the multipliers can affect the solutions that can be obtained from the corresponding envelopment model. Proposed techniques for enforcing these additional restrictions include imposing upper and lower bounds on individual multipliers (Dyson and Thanassoulis, 1988; Roll, Cook, and Golany, 1991); imposing bounds on ratios of multipliers (Thompson et al., 1986); appending multiplier inequalities (Wong and Beasley, 1990); and requiring multipliers to belong to given closed cones (Charnes et al., 1989).

To illustrate the general approach, suppose we wish to incorporate additional inequality constraints of the following form into (1.3) or, more generally, into the multiplier model in Table 1-1:

$$\begin{aligned} \alpha_i \le \frac{v_i}{v_{i_o}} \le \beta_i, \qquad i = 1,\ldots,m \\ \delta_r \le \frac{\mu_r}{\mu_{r_o}} \le \gamma_r, \qquad r = 1,\ldots,s \end{aligned} \tag{1.14}$$

Here, v_{i_o} and μ_{r_o} represent multipliers which serve as "numeraires" in establishing the upper and lower bounds represented here by α_i, β_i, and by δ_r, γ_r for the multipliers associated with inputs i =1, …, m and outputs r = 1, …, s where α_{i_o} = β_{i_o} = δ_{r_o} = γ_{r_o} = 1. The above constraints are called Assurance Region (AR) constraints as developed by Thompson et al. (1986) and defined more precisely in Thompson et al. (1990).

Uses of such bounds are not restricted to prices. They may extend to "utils" or any other units that are regarded as pertinent. For example, Zhu (1996a) uses an assurance region approach to establish bounds on the weights obtained from uses of Analytic Hierarchy Processes in Chinese textile manufacturing in order to include bounds on these weights that better reflect local government preferences in measuring textile manufacturing performances.

There is another approach called the "cone-ratio envelopment approach" which can also be used for this purpose. See Charnes et al. (1990). We do not examine this approach in detail, but rather only note that the assurance region approach can also be given an interpretation in terms of cones. See Cooper et al. (1996).

The generality of these AR constraints provides flexibility in use. Prices, utils and other measures may be accommodated and so can mixtures of such concepts. Moreover, one can first examine provisional solutions and then tighten or loosen the bounds until one or more solutions is attained that appears to be reasonably satisfactory to decision makers who cannot state the values for their preferences in an a priori manner.

The assurance region approach also greatly relaxes the conditions and widens the scope for use of a priori conditions. In some cases, the conditions to be comprehended may be too complex for explicit articulation, in which case additional possibilities are available from other recent advances. For instance, *instead* of imposing bounds on allowable variable values, the cone-ratio envelopment approach *transforms* the data. Brockett et al. (1997) provide an example in which a bank regulatory agency wanted to evaluate "risk coverage" as well as the "efficiency" of the banks under its jurisdiction. Bounds could not be provided on possible tradeoffs between risk coverage and efficiency, so this was accomplished by using a set of banks identified as "excellent" (even when they were not members of the original (regulatory) set). Then, employing data from these excellent banks, a cone-ratio envelopment was used to transform the data into improved values that could be used to evaluate each of the regulated banks operating under widely varying conditions. This avoided the need for jointly specifying what was meant by "adequate" risk coverage and efficiency not only in each detail, but also in all of the complex interplays between risk and efficiency that are possible in bank performances. The non-negativity imposed on the slacks in standard CCR model was also relaxed. This then made it possible to identify deficiencies which were to be repaired by increasing expense items such as "bad loan allowances" (as needed for risk coverage) even though this worsened efficiency as evaluated by the transformed data.

There are in fact a number of ways in which we can incorporate a priori knowledge or requirements as conditions into DEA models. For examples see the “Prioritization Models” of Cook et al. (1992) and the “Preference Structure Model” of Zhu (1996b). See Chapter 4 for a detailed discussion on incorporation of value judgments.

4.4 Window Analysis

In the examples of the previous sections, each DMU was observed only once, i.e., each example was a cross-sectional analysis of data. In actual studies, observations for DMUs are frequently available over multiple time periods (time series data), and it is often important to perform an analysis where interest focuses on changes in efficiency over time. In such a setting, it is possible to perform DEA over time by using a moving average analogue, where a DMU in each different period is treated as if it were a "different" DMU. Specifically, a DMU's performance in a particular period is contrasted with its performance in other periods in addition to the performance of the other DMUs.

The *window analysis* technique that operationalizes the above procedure can be illustrated with the study of aircraft maintenance operations, as described in Charnes et al. (1985). In this study, data were obtained for 14 (n = 14) tactical fighter wings in the U.S. Air Force over seven (p = 7) monthly periods. To perform the analysis using a three-month (w = 3) window, one proceeds as follows.

Each DMU is represented as if it were a different DMU for each of the three successive months in the first window (Ml, M2, M3) consisting of the months at the top of Table 1-4. An analysis of the 42 (= nw = 3 × 14) DMUs can then be performed. The window is then shifted one period by replacing M1 with M4, and an analysis is performed on the second three-month set (M2, M3, M4) of these 42 DMUs. The process continues in this manner, shifting the window forward one period each time and concluding with the final (fifth) analysis of 42 DMUs for the last three months (M5, M6, M7). (In general, one performs p - w + 1 separate analyses, where each analysis examines *nw* DMUs).

Table 1-4 illustrates the results of this analysis in the form of efficiency scores for the performance of the airforce wings as taken from Charnes et al. (1985). The structure of this table portrays the underlying framework of the analysis. For the first "window," wing A is represented in the constraints of the DEA model as though it were a different DMU in months 1, 2, and 3. Hence, when wing 1 is evaluated for its month-1 efficiency, its own performance data for months 2 and 3 are included in the constraint sets along with similar performance data of the other wings for months 1, 2, and 3. Thus the results of the "first window" analysis consist of the 42 scores under the column headings for Month 1-Month 3 in the first row for each wing. For example, wing A had efficiency ratings of 97.89, 97.31, and 98.14 for its performance in months 1, 2, and 3, respectively, as shown in the first row for Wing A in Table 1-4. The second row of data for each wing is the result of analyzing the second window of 42 DMUs, which result from dropping the month-1 data and appending the month-4 data.

The arrangement of the results of a window analysis as given in Table 1-4 facilitates the identification of trends in performance, the stability of reference sets, and other possible insights. For illustration, "row views" clarify performance trends for wings E and M. Wing E improved its performance in month 5 relative to prior performance in months 3 and 4 in the third window, while wing M's performance appears to deteriorate in months 6 and 7. Similar "column views" allow comparison of wings (DMUs) across different reference sets and hence provide information on the stability of these scores as the reference sets change.

Table 1-4. Window Analysis with Three-Month Window

Wing	Month 1	Month 2	Month 3	Month 4	Month 5	Month 6	Month 7
Wing-A	97.89	97.31	98.14				
		97.36	97.53	97.04			
			96.21	95.92	94.54		
				95.79	94.63	97.64	
					94.33	97.24	97.24
Wing-B	93.90	95.67	96.14				
		96.72	96.42	94.63			
			95.75	94.14	93.26		
				94.54	93.46	96.02	
					93.02	96.02	94.49
Wing-C	93.77	91.53	95.26				
		91.77	95.55	94.29			
			93.21	95.04	94.83		
				93.20	93.09	92.21	
					93.59	92.32	92.83
Wing-D	99.72	96.15	95.06				
		97.91	95.70	100.0			
			94.79	100.0	94.51		
				99.71	94.39	94.76	
					94.95	94.67	89.37
Wing-E	100.0	100.0	100.0				
		100.0	100.0	100.0			
			98.97	99.05	100.0		
				99.37	100.0	100.0	
					100.0	100.0	100.0
Wing-F	97.42	93.48	96.07				
		93.60	96.24	93.56			
			94.46	91.75	92.49		
				91.73	92.32	92.35	
					92.68	91.98	99.64
Wing-G	90.98	92.80	95.96				
		93.67	96.80	99.52			
			93.34	94.48	91.73		
				91.94	89.79	95.58	
					89.35	95.14	96.38
Wing-H	100.0	100.0	100.0				
		100.0	100.0	100.0			
			100.0	100.0	100.0		
				100.0	100.0	100.0	
					100.0	100.0	100.0
Wing-I	99.11	95.94	99.76				
		96.04	100.0	100.0			
			98.16	98.99	94.59		
				98.97	94.62	99.16	
					94.68	98.92	97.28
Wing-J	92.85	90.90	91.62				
		91.50	92.12	94.75			
			90.26	93.39	93.83		
				92.92	93.84	95.33	
					94.52	96.07	94.43
Wing-K	86.25	84.42	84.03				
		84.98	84.47	93.74			
			83.37	82.54	80.26		
				82.39	80.14	79.58	
					80.96	78.66	79.75
Wing-L	100.0	100.0	100.0				
		100.0	100.0	99.55			
			100.0	99.39	97.39		
				100.0	96.85	100.0	
					96.66	100.0	100.0
Wing-M	100.0	100.0	100.0				
		100.0	100.0	100.0			
			100.0	100.0	100.0		
				100.0	100.0	98.75	
					100.0	98.51	99.59
Wing-N	100.0	100.0	98.63				
		100.0	100.0	100.0			
			99.45	100.0	100.0		
				100.0	100.0	100.0	
					100.0	100.0	100.0

The utility of Table 1-4 can be further extended by appending columns of summary statistics (mean, median, variance, range, etc.) for each wing to reveal the relative stability of each wing's results. See, for instance, the drop in the efficiency from 93.74 to 82.54 in Month 4 for Wing K.

The window analysis technique represents one area for further research extending DEA. For example, the problem of choosing the width for a window (and the sensitivity of DEA solutions to window width) is currently determined by trial and error. Similarly, the theoretical implications of representing each DMU as if it were a different DMU for each period in the window remain to be worked out in full detail.

5. ALLOCATIVE AND OVERALL EFFICIENCY

To this point we have confined attention to "technical efficiency" which, as explained immediately after definition 1.2, does not require a use of prices or other "weights." Now we extend the analysis to situations in which unit prices and unit costs are available. This allows us to introduce the concepts of "allocative" and "overall" efficiency and relate them to "technical efficiency" in a manner first introduced by M.J. Farrell (1957).

For this introduction we utilize Figure 1-5 in which the solid line segments connecting points ABCD constitute an "isoquant" or "level line' that represents the different amounts of two inputs (x_1, x_2) which can be used to produce the same amount (usually one unit) of a given output. This line represents the "efficiency frontier" of the "production possibility set" because it is not possible to reduce the value of one of the inputs without increasing the other input if one is to stay on this isoquant.

The dashed line represents an isocost (=budget) line for which (x_1, x_2) pairs on this line yield the same total cost, when the unit costs are c_1 and c_2, respectively. When positioned on C the total cost is k. However, shifting this budget line upward in parallel fashion until it reaches a point of intersection with R would increase the cost to $k' > k$. In fact, as this Figure shows, k is the minimum total cost needed to produce the specified output since any parallel shift downward below C would yield a line that fails to intersect the production possibility set. Thus, the intersection at C gives an input pair (x_1, x_2) that minimizes the total cost of producing the specified output amount and the point C is therefore said to be "allocatively" as well as "technically" efficient.

Now let R represent an observation that produced this same output amount. The ratio $0 \le OQ/OR \le 1$ is said to provide a "radial" measure of technical efficiency,[2] with $0 \le 1-(OQ/OR) \le 1$ yielding a measure of technical inefficiency.

Now consider the point P which is at the intersection of this cost line through C with the ray from the origin to R. We can also obtain a radial measure of "overall efficiency" from the ratio $0 \le OP/OR \le 1$. In addition, we can form the ratio $0 \le OP/OQ \le 1$ to obtain a measure of what Farrell (1957) referred to as "price efficiency" but is now more commonly called "allocative efficiency." Finally, we can relate these three measures to each other by noticing that

$$\frac{OP}{OQ}\frac{OQ}{OR} = \frac{OP}{OR} \tag{1.15}$$

which we can verbalize by saying that the product of allocative and technical efficiency eqauls overall efficiency in these radial measures.

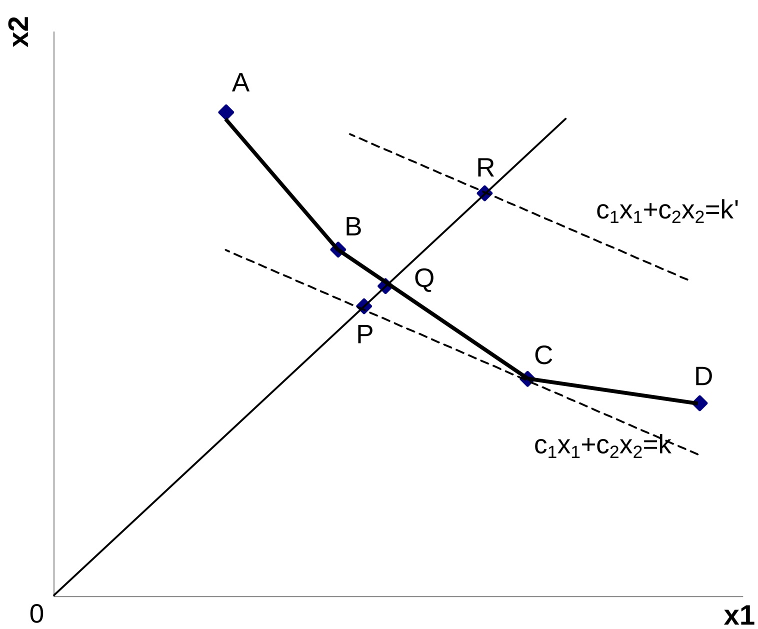

Figure 1-5. Allocative and Overall Efficiency

[2] Actually this "radial measure" is a ratio of two measures of distance, one of which measures the distance from the origin to Q and the other which measures the distance from the origin to R to obtain $0 \le d(O,Q)/d(O,R) \le 1$ where $d(....)$ means "distance." See I. Bardhan et al. (1996) for a proof in terms of the Euclidean measure of distance--although other measures of distance may also be used.

To implement these ideas we use the following model, as taken from Cooper, Seiford and Tone (2000, p. 236),

$$\begin{aligned}
&\min \sum_{i=1}^{m} c_{io} x_i \\
&\text{subject to} \\
&\sum_{j=1}^{n} x_{ij}\lambda_j \le x_i, \quad \mathrm{i} = 1,\ldots,\mathrm{m} \\
&\sum_{j=1}^{n} y_{rj}\lambda_j \ge y_{ro}, \ \mathrm{r} = 1,\ldots,\mathrm{s} \\
&\mathrm{L} \le \sum_{j=1}^{n} \lambda_j \le U
\end{aligned} \tag{1.16}$$

where the objective is to choose the x_i and λ_j values to minimize the total cost of satisfying the output constraints. The c_{io} in the objective represent unit costs. This formulation differs from standard models, as in Färe, Grosskopf and Lovell (1985, 1994), in that these unit costs are allowed to vary from one $\mathrm{DMU_o}$ to another in (1.16). In addition the values of $\sum_{j=1}^{n} \lambda_j$ are limited above by U and below by L--according to the returns-to-scale conditions that are imposed. See next chapter. Here we only note that the choice L=U=1 makes this a BCC model, whereas $L = 0$, $U = \infty$ converts it to a CCR model. See the discussion following Table 1-1. Finally, using the standard approach, we can obtain a measure of relative cost (=overall) efficiency by utilizing the ratio

$$0 \le \frac{\sum_{i=1}^{m} c_{io} x_i^*}{\sum_{i=1}^{m} c_{io} x_{io}} \le 1 \tag{1.17}$$

where the x_i^* are the optimal values obtained from (1.16) and the x_{io} are the observed values for $\mathrm{DMU_o}$.

The use of a ratio like (1.17) is standard and yields an easily understood measure. It has shortcomings, however, as witness the following example from Tone and Sahoo (2003): Let γ_a and γ_b represent cost efficiency, as determined from (1.17), for DMUs a and b. Now suppose $x_{ia}^* = x_{ib}^*$ and $x_{ia} = x_{ib}$, $\forall i$, but $c_{ia} = 2c_{ib}$ so that the unit costs for a are twice as high as for b in every input. We then have

$$\gamma_a = \frac{\sum_{i=1}^{m} c_{ia} x_{ia}^*}{\sum_{i=1}^{m} c_{ia} x_{ia}} = \frac{\sum_{i=1}^{m} 2c_{ib} x_{ib}^*}{\sum_{i=1}^{m} 2c_{ib} x_{ib}} = \frac{\sum_{i=1}^{m} c_{ib} x_{ib}^*}{\sum_{i=1}^{m} c_{ib} x_{ib}} = \gamma_b \tag{1.18}$$

Thus, as might be expected with a use of ratios, important information may be lost since $\gamma_a = \gamma_b$ conceals the fact that a is twice as costly as b.

6. PROFIT EFFICIENCY

We now introduce another type of model called the "additive model" to evaluate technical inefficiency. First introduced in Charnes et al. (1985) this model has the form

$$\begin{aligned}
&\max \sum_{r=1}^{s} s_r^+ + \sum_{i=1}^{m} s_i^- \\
&\text{subject to} \\
&y_{ro} = \sum_{j=1}^{n} y_{rj}\lambda_j - s_r^+, \quad r = 1,2,\ldots,s \\
&x_{io} = \sum_{j=1}^{n} x_{ij}\lambda_j + s_i^-, \quad i = 1,2,\ldots,m \\
&1 = \sum_{j=1}^{n} \lambda_j \\
&0 \le \lambda_j, s_r^+ \;\; s_i^-; \forall i, j, r.
\end{aligned} \tag{1.19}$$

This model uses a metric that differs from the one used in the "radial measure" model.[3] It also dispenses with the need for distinguishing between an "output" and an "input" orientation as was done in the discussion leading up to (1.10) because the objective in (1.19) simultaneously maximizes outputs and minimizes inputs--in the sense of vector optimizations. This can be seen by utilizing the solution to (1.19) to introduce new variables $\hat{y}_{ro}$, $\hat{x}_{io}$ defined as follows,

$$\begin{aligned}
\hat{y}_{ro} &= y_{ro} + s_r^{+*} \ge y_{ro}, \quad r = 1,\ldots,s, \\
\hat{x}_{io} &= x_{io} - s_i^{-*} \le x_{io}, \quad i = 1,2,\ldots,m.
\end{aligned} \tag{1.20}$$

Now note that the slacks are all independent of each other. Hence an optimum is not reached until it is not possible to increase an output $\hat{y}_{ro}$ or reduce an input $\hat{x}_{io}$ without decreasing some other output or increasing some other input. The following theorem, which follows immediately, is proved in Cooper, Seiford and Tone (2000),

Theorem 1.2: DMUo is efficient if and only if all slacks are zero in an optimum solution.

[3] The additive model uses what is called the "ℓ_1 metric" in mathematics, and the "city block metric" in operations research. See Appendix A in Charnes and Cooper (1961).

As proved in Ahn, Charnes and Cooper (1988) one can also relate solutions in the additive model to those in the radial measure model via

Theorem 1.3: DMUo is efficient for an additive model if and only if it is efficient for the corresponding radial model.

Here the term "corresponding" means that the constraint sets are the same so that $\sum_{j=1}^{n}\lambda_j = 1$ appears as a constraint in the additive model if and only if it also appears in the radial measure model to which it is being compared.

We now use the class of additive models to develop a different route to treating technical, allocative and overall inefficiencies and their relations to each other. This can help to avoid difficulties in treating possibilities like "negative" or "zero" profits which are not easily treated by the ratio approaches, as in (1.17), which are commonly used in the DEA literature. See the discussion in the appendix to Cooper, Park and Pastor (1999) from which the following development is taken. See also Chapter 8 in Cooper, Seiford and Tone (2000).

First we observe that we can multiply the output slacks by unit prices and the input slacks by unit costs after we have solved (1.19) and thereby accord a monetary value to this solution. Then we can utilize (1.20) to write

$$\begin{aligned}\sum_{r=1}^{s} p_{ro}s_r^{+*} + \sum_{i=1}^{m} c_{io}s_i^{-*} &= \left(\sum_{r=1}^{s} p_{ro}\hat{y}_{ro} - \sum_{i=1}^{m} p_{ro}y_{ro}\right) + \left(\sum_{i=1}^{m} c_{io}x_{io} - \sum_{i=1}^{m} c_{io}\hat{x}_{io}\right) \\ &= \left(\sum_{r=1}^{s} p_{ro}\hat{y}_{ro} - \sum_{i=1}^{m} c_{io}\hat{x}_{io}\right) - \left(\sum_{r=1}^{s} p_{ro}y_{ro} - \sum_{i=1}^{m} c_{io}x_{io}\right).\end{aligned} \tag{1.21}$$

From the last pair of parenthesized expressions we find that, at an optimum, the objective in (1.19) after multiplication by unit prices and costs is equal to the profit available when production is technically efficient minus the profit obtained from the observed performance. Hence when multiplied by unit prices and costs the solution to (1.19) provides a measure in the form of the amount of the profits lost by not performing in a technically efficient manner term by term if desired.

<u>Remark</u>: We can, if we wish restate this measure in ratio form because, by definition,

$$\sum_{r=1}^{s} p_{ro}\hat{y}_{ro} - \sum_{i=1}^{m} c_{io}\hat{x}_{io} \geq \sum_{r=1}^{s} p_{ro}y_{ro} - \sum_{i=1}^{m} c_{io}x_{io}$$

Therefore,

$$1 \geq \frac{\sum_{r=1}^{s} p_{ro}y_{ro} - \sum_{i=1}^{m} c_{io}x_{io}}{\sum_{r=1}^{s} p_{ro}\hat{y}_{ro} - \sum_{i=1}^{m} c_{io}\hat{x}_{io}} \geq 0$$

and the upper bound is attained if and only if performance is efficient.

We can similarly, develop a measure of allocative efficiency by means of the following additive model,

$$\begin{aligned}
&\max \sum_{r=1}^{s} p_{ro}\hat{s}_r^+ + \sum_{i=1}^{m} c_{io}\hat{s}_i^- \\
&\text{subject to} \\
&\hat{y}_{ro} = \sum_{j=1}^{n} y_{rj}\hat{\lambda}_j - \hat{s}_r^+, \quad r = 1,2,\ldots,s \\
&\hat{x}_{io} = \sum_{j=1}^{n} x_{ij}\hat{\lambda}_j - \hat{s}_i^-, \quad i = 1,2,\ldots,m \\
&1 = \sum_{j=1}^{n} \hat{\lambda}_j \\
&0 \le \hat{\lambda}_j \ \ \forall j;\ \hat{s}_i^-,\ \hat{s}_r^+ \text{ free } \forall i, r.
\end{aligned} \tag{1.22}$$

Comparison with (1.19) reveals the following differences: (1) the objective in (1.19) is replaced by one which is monetized (2) the y_{ro} and x_{io} in (1.19) are replaced by $\hat{y}_{ro}$ and $\hat{x}_{io}$ in (1.22) as obtained from (1.20) and, finally, (3) the slack values in (1.22) are not constrained in sign as is the case in (1.19). This last relaxation, we might note, is needed to allow for substitutions between the different output and the different input amounts, as may be needed to achieve the proportions required for allocative efficiency. See Cooper, Park and Pastor (2000).

Finally, we use the following additive model to evaluate overall (profit) efficiency--called "graph efficiency" in Färe, Grosskopf and Lovell (1985, 1994),

$$\begin{aligned}
&\max \sum_{r=1}^{s} p_{ro}s_r^+ + \sum_{i=1}^{m} c_{io}s_i^- \\
&\text{subject to} \\
&y_{ro} = \sum_{j=1}^{n} y_{rj}\lambda_j - s_r^+, \quad r = 1,2,\ldots,s \\
&x_{io} = \sum_{j=1}^{n} x_{ij}\lambda_j \quad - \hat{s}_i^-, \quad i = 1,2,\ldots,m \\
&1 = \sum_{j=1}^{n} \lambda_j \\
&0 \le \lambda_j \ \forall j;\ s_i^-,\ s_r^+ \text{ free } \forall i, r.
\end{aligned} \tag{1.23}$$

We then have the relation set forth in the following

Theorem 1.4: The value (=total profit foregone) of overall inefficiency for DMUo as obtained from (1.23) is equal to the value of technical inefficiency as obtained from (1.19) plus the value of allocative inefficiency as obtained from (1.22). i.e.,

$$\max \left(\sum_{r=1}^{s} p_{ro}s_r^+ + \sum_{i=1}^{m} c_{io}s_i^- \right) = \left(\sum_{r=1}^{s} p_{ro}s_r^{+*} + \sum_{i=1}^{m} c_{io}s_i^{-*} \right) + \left(\sum_{r=1}^{s} p_{ro}\hat{s}_r^{+*} + \sum_{i=1}^{m} c_{io}\hat{s}_i^{-*} \right).$$

Table 1-5, as adapted from Cooper, Seiford and Tone (2000), can be used to construct an example to illustrate this theorem. The body of the table records the performances of 3 DMUs in terms of the amount of the one output they all produce, y, and the amount of the two inputs x_1, x_2 they all use. For simplicity it is assumed that they all receive the same unit price p=6 as recorded on the right and incur the same unit costs $c_1 = \$4$ and $c_2 = \$2$ for the inputs as shown in the rows with which they are associated. The bottom of each column records the profit, π, made by each DMU in the periods of interest.

Table 1-5. Price-Cost-Profit Data

	DMU_1	DMU_2	DMU_3	\$
y	4	4	2	6
x_1	4	2	4	4
x_2	2	4	6	2
π	4	8	-16	

DMU_3, as shown at the bottom of its column is, by far, the worst performer having experienced a loss of \$16. This loss, however, does not account for all of the lost profit possibilities. To discover this value we turn to (1.23) and apply it to the data in Table 1.5. This produces the following model to evaluate the overall inefficiency of DMU_3.

$$\begin{aligned}
&\max\ 6s^+ + 4s_1^- + 2s_2^- \\
&\text{subject to} \\
&\quad 2 = 4\lambda_1 + 4\lambda_2 + 2\lambda_3 - s^+ \\
&\quad 4 = 4\lambda_1 + 2\lambda_2 + 4\lambda_3 \qquad + s_1^- \\
&\quad 6 = 2\lambda_1 + 4\lambda_2 + 6\lambda_3 \qquad\qquad - s_2^- \\
&\quad 1 = \lambda_1 + \lambda_2 + \lambda_3 \\
&\quad 0 \le \lambda_1,\ \lambda_2,\ \lambda_3; s^+,\ s_1^-,\ s_2^- \text{ free}
\end{aligned} \tag{1.24}$$

An optimum solution to this problem is $\lambda_2^* = 1$, $s^{+*} = 2$, $s_1^{-*} = 2$, $s_2^{-*} = 2$ and all after variables zero. Utilizing the unit price and costs exhibited in the objective of (1.24) we therefore find

$$6s^{+*} + 4s_1^{-*} + 2s_1^{-*} = \$6 \times 2 + \$4 \times 2 + \$2 \times 2 = \$24\,.$$

This is the value of the foregone profits arising from inefficiencies in the performance of DMU_3. Eliminating these inefficiencies would have wiped out the \$16 loss and replaced it with an \$8 profit. This is the same amount of profit as DMU_2, which is the efficient performer used to effect this evaluation of DMU_3 via $\lambda_2^* = 1$ in the above solution.

We now utilize theorem 1.4 to further identify the sources of this lost profit. For this purpose we first apply (1.19) to the data of Table 1-5 in order to determine the lost profits from the technical inefficiency of DMU_3 via

$$\begin{aligned}
&\max\ s^+ + s_1^- + s_1^- \\
&\text{subject to} \\
&\quad 2 = 4\lambda_1 + 4\lambda_2 + 2\lambda_3 - s^+ \\
&\quad 4 = 4\lambda_1 + 2\lambda_2 + 4\lambda_3 \qquad + s_1^- \\
&\quad 6 = 2\lambda_1 + 4\lambda_2 + 6\lambda_3 \qquad\qquad - s_2^- \\
&\quad 1 = \lambda_1 + \lambda_2 + \lambda_3 \\
&\quad 0 \le \lambda_1,\ \lambda_2,\ \lambda_3,\ s^+,\ s_1^-,\ s_2^-.
\end{aligned} \tag{1.25}$$

This has an optimum with $\lambda_1^* = 1$, $s^{+*} = 2$, $s_1^{-*} = 0$, $s_2^{-*} = 4$ and all other variables zero so that multiplying these values by their unit price and unit costs we find that the lost profits due to technical inefficiencies are

$$\$6 \times 2 + \$4 \times 0 + \$2 \times 4 = \$20\ .$$

For allocative inefficiency we apply (1.20) and (1.22) to the data in Table 1-5 and get.

$$\begin{aligned}
&\max\ 6\hat{s}^+ + 4\hat{s}_1^- + 2\hat{s}_2^- \\
&\text{subject to} \\
&\quad 4 = 4\hat{\lambda}_1 + 4\hat{\lambda}_2 + 2\hat{\lambda}_3 - \hat{s}^+ \\
&\quad 4 = 4\hat{\lambda}_1 + 2\hat{\lambda}_2 + 4\hat{\lambda}_3 \qquad + \hat{s}_1^- \\
&\quad 2 = 2\hat{\lambda}_1 + 4\hat{\lambda}_2 + 6\hat{\lambda}_3 \qquad\qquad + \hat{s}_2^- \\
&\quad 1 = \hat{\lambda}_1 + \hat{\lambda}_2 + \hat{\lambda}_3 \\
&\quad 0 \le \hat{\lambda}_1,\ \hat{\lambda}_2,\ \hat{\lambda}_3;\ \hat{s}^+,\ \hat{s}_1^-,\ \hat{s}_2^- \text{ free.}
\end{aligned}$$

An optimum is $\hat{\lambda}_2^* = 1$, $\hat{s}^{+*} = 0$, $\hat{s}_1^{-*} = 2$, $\hat{s}_2^{-*} = -2$ with all other variables zero so the profit lost from allocative efficiency is

$$\$4 \times 2 + \$2(-2) = \$4\ ,$$

which accounts for the remaining \$4 of the \$24 lost profit obtained from overall inefficiency via (1.24). Here we might note that an increase in $\hat{x}_2$ -- see the second expression in (1.20)--is more than compensated for by the offsetting decrease in $\hat{x}_1$ <u>en</u> <u>route</u> to the proportions needed to achieve allocative efficiency.

Finally we supply a tabulation obtained from these solutions in Table 1-6.

Table 1-6. Solution Detail

Model / Variable	Overall	Technical	Allocative
s^+	2	2	0
s_1^-	2	0	2
s_2^-	2	4	-2
π	24	20	4

Remark 1: Adding the figures in each row of the last two columns yields the corresponding value in the column under overall efficiency. This will always be true for the dollar value in the final row, by virtue of theorem 1.3, but it need not be true for the other rows because of the possible presence of alternate optima.

Remark 2: The solutions need not be "units invariant." That is, the optimum solutions for the above models may differ if the unit prices and unit costs used are stated in different units. See pp.228 ff. in Cooper, Park and Pastor (2000) for a more detailed discussion and methods for making the solutions units invariant.

7. CONCLUSIONS

This chapter has provided an introduction to DEA and some of its uses. However it is far from having exhausted the possibilities that DEA offers. For instance we have here focused on what is referred to in the DEA (and economics) literature as technical efficiency. For perspective we concluded with discussions of allocative and overall efficiency when costs or profits are of interest and the data are available. This does not exhaust the possibilities. There are still other types of efficiency that can be addressed with DEA. For instance, returns-to-scale inefficiencies, as covered in the next chapter, can offer additional possibilities which identify where additional shortcomings can appear when unit prices or unit costs are available which are of interest to potential users. We have not even fully exploited uses of our technical inefficiency models, as developed in this chapter. For instance, uses of DEA identify DMUs that enter into the optimal evaluations and hence can serve as "benchmark DMUs" to determine how best to eliminate these inefficiencies.

Topics like these will be discussed in some of the chapters that follow. However, the concept of technical inefficiency provides a needed start which will turn out to be basic for all of the other types of efficiency that may be of interest. Technical efficiency is also the most elemental of the various efficiencies that might be considered in that it requires only minimal information and minimal assumptions for its use. It is also fundamental because other types of efficiency such as allocative efficiency and returns to scale efficiency require technical efficiency to be attained before these can be achieved. This will all be made clearer in the chapters that follow.

REFERENCES

1. Afriat, S., 1972, Efficiency estimation of production functions, *International Economic Review* 13, 568-598.
2. Ahn, T., A. Charnes and W.W. Cooper, 1988, "Efficiency Characterizations in Different DEA Models," *Socio-Economic Planning Sciences* 22, 253-257.
3. Arnold, V., I. Bardhan, W.W. Cooper and A. Gallegos, 1998, "Primal and Dual Optimality in Computer Codes Using Two-Stage Solution Procedures in DEA" in J. Aronson and S. Zionts, eds., *Operations Research Methods, Models and Applications* (Westpost, Conn: Quorum Books).
4. Banker, R., A. Charnes and W.W. Cooper, 1984, Some models for estimating technical and scale inefficiencies in data envelopment analysis, Management Science 30, 1078-1092.
5. Banker, R.D. and R.C. Morey, 1986a, Efficiency analysis for exogenously fixed inputs and outputs, Operations Research 34, No. 4, 513-521.
6. Banker, R.D. and R.C. Morey, 1986b, The use of categorical variables in data envelopment analysis, Management Science 32, No. 12, 1613-1627.
7. Bardhan, I., W.F. Bowlin, W.W. Cooper and T. Sueyoshi, 1996, "Models and Measures for Efficiency Dominance in DEA, Part I: Additive Models and MED Measures," *Journal of the Operational Research Society of Japan* 39, 322-332.
8. Brockett, P.L., A. Charnes, W.W. Cooper, Z.M. Huang, and D.B. Sun, 1997, Data transformations in DEA cone ratio envelopment approaches for monitoring bank performances, European Journal of Operational Research 98, No. 2, 250-268.
9. Brockett, P.L., W.W. Cooper, H.C. Shin and Y. Wang, 1998, Inefficiency and congestion in Chinese production before and after the 1978 economic reforms. Socio-Econ Plann Sci 32, 1-20.
10. Charnes, A. and W.W. Cooper, 1962, Programming with linear fractional functionals, Naval Research Logistics Quarterly 9, 181-185.
11. Charnes, A., W.W. Cooper, and E. Rhodes, 1978, Measuring the efficiency of decision making units, European Journal of Operational Research 2, 429-444.

12. Charnes A, C.T. Clark, W.W. Cooper, B. Golany, 1985, A developmental study of data envelopment analysis in measuring the efficiency of maintenance units in the US air forces. In R.G. Thompson and R.M. Thrall, Eds., Annals Of Operation Research 2:95-112.
13. Charnes, A., W.W. Cooper, D.B. Sun, and Z.M. Huang, 1990, Polyhedral cone-ratio DEA models with an illustrative application to large commercial banks, Journal of econometrics 46, 73-91.
14. Charnes, A., W.W. Cooper, Q.L. Wei, and Z.M. Huang, 1989, Cone ratio data envelopment analysis and multi-objective programming, International Journal of Systems Science 20, 1099-1118.
15. Charnes, A. and W.W. Cooper, 1961, *Management Models and Industrial Applications of Linear Programming,* 2 vols., with A. Charnes (New York: John Wiley and Sons, Inc.).
16. Charnes, A., W.W. Cooper, B. Golany, L. Seiford and J. Stutz, (1985) "Foundations of Data Envelopment Analysis for Pareto-Koopmans Efficient Empirical Production Functions," *Journal of Econometrics* (l985), 30, pp. 91-l07.
17. Cook, W.D., M. Kress, and L.M. Seiford, 1992, Prioritization models for frontier decision making units in DEA, European Journal of Operational Research 59, No. 2, 319-323.
18. Cooper, W.W., R.G. Thompson, and R.M. Thrall, 1996, Extensions and new developments in data envelopment analysis, Annals of Operations Research 66, 3-45.
19. Cooper, W.W., Seiford, L.M. and Tone, K., 2000, Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software, Kluwer Academic Publishers, Boston.
20. Cooper, W.W., Seiford, L.M. and Zhu, Joe, A unified additive model approach for evaluating inefficiency and congestion with associated measures in DEA. Socio-Economic Planning Sciences, Vol. 34, No. 1 (2000), 1-25.
21. Cooper, W.W., H. Deng, Z. M. Huang and S. X. Li, 2002, A one-model approach to congestion in data envelopment analysis, Socio-Economic Planning Sciences 36, 231-238.
22. Cooper, W.W., K.S. Park and J.T. Pastor, 2000, "Marginal Rates and Elasticities of Substitution in DEA." *Journal of Productivity Analysis* 13, 2000, pp. 105-123.
23. Cooper, W.W., K.S. Park and J.T. Pastor, 1999, "RAM: A Range Adjusted Measure of Inefficiency for Use with Additive Models and Relations to Other Models and Measures in DEA", *Journal of Productivity Analysis* 11, 5-42.

24. Debreu, G., 1951, The coefficient of resource utilization, Econometrica 19, 273-292.
25. Dyson, R.G. and E. Thanassoulis, 1988, Reducing weight flexibility in data envelopment analysis, Journal of the Operational Research Society 39, No. 6, 563-576.
26. Färe, R., S. Grosskopf and C.A.K. Lovell, 1985, The Measurement of Efficiency of Production. Boston:Kluwer Nijhoff Publishing Co.
27. Färe, R., S. Grosskopf and CAK Lovell, 1994, *Production Frontiers* (Cambridge: Cambridge University Press).
28. Farrell, M.J., 1957, The measurement of productive efficiency, Journal of Royal Statistical Society A 120, 253-281.
29. Koopmans, T.C., 1951, Analysis of Production as an efficient combination of Activities, in T.C. Koopmans, ed. Wiley, New York.
30. Roll, Y., W.D. Cook, and B. Golany, 1991, Controlling factor weights in data envelopment analysis, IIE Transactions, 23, 2-9.
31. Shephard, R.W., 1970, Theory of Cost and Production Functions, Princeton University Press, Princeton, NJ.
32. Takamura, T. and K. Tone, 2003, "A Comparative Site Evaluation Study for Relocating Japanese Government Agencies Out of Tokyo," *Socio-Economic Planning Sciences* 37, 85-102.
33. Tavares, G., 2003, "A Bibliography of Data Envelopment Analysis (1978-2001)," *Socio-Economic Planning Sciences*(to appear).
34. Thompson, R.G., F.D.Jr. Singleton, R.M. Thrall, and B.A. Smith, 1986, Comparative site evaluation for locating a high-energy physics lab in Texas, Interfaces 16, 35-49.
35. Thompson, R.G., L. Langemeier, C. Lee, E. Lee, and R. Thrall, 1990, The role of multiplier bounds in efficiency analysis with application to Kansas farming, Journal of Econometrics 46, 93-108.
36. Tone, K. and B.K. Sahoo, 2003, "A Reexamination of Cost Efficiency and Cost Elasticity in DEA," *Management Science* (forthcoming).
37. Wong, Y.-H.B. and J.E. Beasley, 1990, Restricting weight flexibility in data envelopment analysis, Journal of the Operational Research Society 41, 829-835.
38. Zhu, J., 1996a, DEA/AR analysis of the 1988-1989 performance of the Nanjing Textiles Corporation, Annals of Operations Research 66, 311-335.
39. Zhu, J., 1996b, Data envelopment analysis with preference structure, Journal of the Operational Research Society 47, No. 1, 136-150.
40. Zhu, J., 2000, Multi-factor performance measure model with an application to Fortune 500 companies. European Journal of Operational Research 123, No. 1, 105-124.

41. Zhu, J. 2002, Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheets and DEA Excel Solver, Kluwer Academic Publishers, Boston.

Part of the material in this chapter is adapted from the Journal of Econometrics, Vol. 46, Seiford, L.M. and Thrall, R.M., Recent developments in DEA: The mathematical programming approach to frontier analysis, 7-38, 1990, with permission from Elsevier Science.