1 GENERAL DISCUSSION

1.1 INTRODUCTION

This book is concerned with evaluations of performance and it is especially concerned with evaluating the activities of organizations such as business firms, government agencies, hospitals, educational institutions, etc. Such evaluations take a variety of forms in customary analyses. Examples include cost per unit, profit per unit, satisfaction per unit, and so on, which are measures stated in the form of a ratio like the following,

$$\frac{\text{Output}}{\text{Input}}. \tag{1.1}$$

This is a commonly used measure of efficiency. The usual measure of "productivity" also assumes a ratio form when used to evaluate worker or employee performance. "Output per worker hour" or "output per worker employed" are examples with sales, profit or other measures of output appearing in the numerator. Such measures are sometimes referred to as "partial productivity measures." This terminology is intended to distinguish them from "total factor productivity measures," because the latter attempt to obtain an output-to-input ratio value which takes account of *all* outputs and *all* inputs. Moving from partial to total factor productivity measures by combining all inputs and all outputs to obtain a single ratio helps to avoid imputing gains to one factor (or one output) that are really attributable to some other input (or output). For instance, a gain in output resulting from an increase in capital or improved management might be mistakenly attributed to labor (when a single output to input ratio is used) even though the performance of labor *deteriorated* during the period being considered. However, an attempt to move from partial to total factor productivity measures encounters difficulties such as choosing the inputs and outputs to be considered and the weights to be used in order to obtain a single-output-to-single-input ratio that reduces to a form like expression (1.1).

Other problems and limitations are also incurred in traditional attempts to evaluate productivity or efficiency when multiple outputs and multiple inputs need to be taken into account. Some of the problems that need to be addressed will be described as we proceed to deal in more detail with Data Envelopment Analysis (DEA), the topic of this book. The relatively new approach embodied in DEA does not require the user to prescribe weights to be attached to each input and output, as in the usual index number approaches, and it also does not require prescribing the functional forms that are needed in statistical regression approaches to these topics.

DEA utilizes techniques such as mathematical programming which can handle large numbers of variables and relations (constraints) and this relaxes the requirements that are often encountered when one is limited to choosing only a few inputs and outputs because the techniques employed will otherwise encounter difficulties. Relaxing conditions on the number of candidates to be used in calculating the desired evaluation measures makes it easier to deal with complex problems and to deal with other considerations that are likely to be confronted in many managerial and social policy contexts. Moreover, the extensive body of theory and methodology available from mathematical programming can be brought to bear in guiding analyses and interpretations. It can also be brought to bear in effecting computations because much of what is needed has already been developed and adapted for use in many prior applications of DEA. Much of this is now available in the literature on research in DEA and a lot of this has now been incorporated in commercially available computer codes that have been developed for use with DEA. This, too, is drawn upon in the present book and a CD with supporting DEA-Solver software and instructions, has been included to provide a start by applying it to some problems given in this book.

DEA provides a number of additional opportunities for use. This includes opportunities for collaboration between analysts and decision-makers, which extend from collaboration in choices of the inputs and outputs to be used and includes choosing the types of "what-if" questions to be addressed. Such collaborations extend to "benchmarking" of "what-if" behaviors of competitors and include identifying potential (new) competitors that may emerge for consideration in some of the scenarios that might be generated.

1.2 SINGLE INPUT AND SINGLE OUTPUT

To provide a start to our study of DEA and its uses, we return to the single output to single input case and apply formula (1.1) to the following simple example. Suppose there are 8 branch stores which we label A to H at the head of each column in Table 1.1.

Table 1.1. Single Input and Single Output Case

Store	A	B	C	D	E	F	G	H
Employee	2	3	3	4	5	5	6	8
Sale	1	3	2	3	4	2	3	5
Sale/Employee	0.5	1	0.667	0.75	0.8	0.4	0.5	0.625

The number of employees and sales (measured in 100,000 dollars) are as recorded in each column. The bottom line of Table 1.1 shows the sales per employee — a measure of "productivity" often used in management and investment analysis. As noted in the sentence following expression (1.1), this may also be treated in the more general context of "efficiency." Then, by this measure, we may identify B as the most efficient branch and F as least efficient.

Let us represent these data as in Figure 1.1 by plotting "number of employees" on the horizontal and "sales" on the vertical axis. The slope of the line connecting each point to the origin corresponds to the sales per employee and the highest such slope is attained by the line from the origin through B. This line is called the "efficient frontier." Notice that this frontier touches at least one point and all points are therefore on or below this line. The name Data Envelopment Analysis, as used in DEA, comes from this property because in mathematical parlance, such a frontier is said to "envelop" these points.

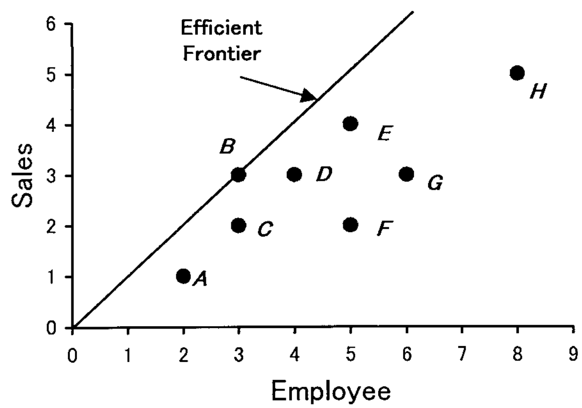

Figure 1.1. Comparisons of Branch Stores

Given these data, one might be tempted to draw a statistical regression line fitted to them. The dotted line in Figure 1.2 shows the regression line passing through the origin which, under the least squares principle, is expressed by $y = 0.622x$. This line, as normally determined in statistics, goes through the "middle" of these data points and so we could define the points above it as *excellent* and the points below it as *inferior* or *unsatisfactory*. One can measure the degree of excellence or inferiority of these data points by the magnitude of the deviation from the thus fitted line. On the other hand, the frontier line designates the performance of the best store (B) and measures the efficiency of other stores by deviations from it. There thus exists a fundamental difference between statistical approaches via regression analysis and DEA. The former reflects "average" or "central tendency" behavior of the observations while the latter deals with best performance and evaluates all performances by deviations from the frontier line. These two points of view can result in major differences when used as methods of evaluation. They can also result in different approaches to improvement. DEA identifies a point like B for future examination or to serve as a "benchmark" to use in seeking improvements. The statistical approach, on the other hand, averages B along with the other observations, including F as a basis for suggesting where improvements might be sought.

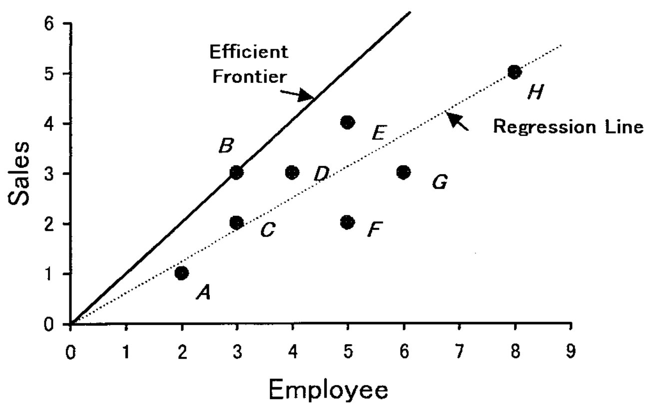

Figure 1.2. Regression Line vs. Frontier Line

Returning to the example above, it is not really reasonable to believe that the frontier line stretches to infinity with the same slope. We will analyze this problem later by using different DEA models. However, we assume that this line is effective in the range of interest and call it the *constant returns-to-scale* assumption.

Compared with the best store B, the others are inefficient. We can measure the efficiency of others relative to B by

$$0 \leq \frac{\text{Sales per employee of others}}{\text{Sales per employee of } B} \leq 1 \tag{1.2}$$

and arrange them in the following order by reference to the results shown in Table 1.2.

$$1 = B > E > D > C > H > A = G > F = 0.4.$$

Thus, the worst, F, attains $0.4 \times 100\% = 40\%$ of B's efficiency.

Table 1.2. Efficiency

Store	A	B	C	D	E	F	G	H
Efficiency	0.5	1	0.667	0.75	0.8	0.4	0.5	0.625

Now we observe the problem of how to make the inefficient stores efficient, i.e., how to move them up to the efficient frontier. For example, store A in Figure 1.3 can be improved in several ways. One is achieved by reducing the input (number of employees) to A_1 with coordinates (1, 1) on the efficient frontier. Another is achieved by raising the output (sales in \$100,000 units) up to $A_2(2, 2)$. Any point on the line segment $\overline{A_1A_2}$ offers a chance to effect the *improvements* in a manner which assumes that the input should not be increased and the output should not be decreased in making the store efficient.

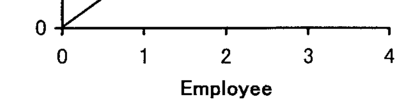

Figure 1.3. Improvement of Store A

This very simple example moves from the ratio in Table 1.1 to the "ratio of ratios" in Table 1.2, which brings to the fore an important point. The values in (1.1) depend on the units of measure used whereas this is not the case for (1.2). For instance, if sales were stated in units of \$10,000, the ratio for F would change from $2/5 = 0.4$ to $20/5 = 4.0$. However, the value of (1.2) would remain unchanged at $4/10 = 0.4$ and the *relative efficiency* score associated with F is not affected by this choice of a different unit of measure. This property, sometimes referred to as "units invariance" has long been recognized as important in engineering and science. Witness, for instance, the following example from the field of combustion engineering where ratings of furnace efficiency are obtained from the following formula,[1]

$$0 \leq E_r = \frac{y_r}{y_R} \leq 1 \tag{1.3}$$

where
y_r = Heat obtained from a given unit of fuel by the furnace being rated,
y_R = Maximum heat that can be obtained from this same fuel input.

The latter, i.e., the maximum heat can be calculated from thermodynamic principles by means of suitable chemical-physical analyses. The point to be emphasized here, however, is that x, the amount of fuel used must be the same so that, mathematically,

$$0 \leq \frac{y_r/x}{y_R/x} = \frac{y_r}{y_R} \leq 1 \tag{1.4}$$

Hence, (1.3) is obtained from a ratio of ratios that is "units invariant."

Returning to the ratios in Table 1.2, we might observe that these values are also bounded by zero and unity. However, the variations recorded in Table 1.2 may result from an excess amount of input or a deficiency in the output. Moreover, this situation is general in the business and social-policy (economics) applications which are of concern in this book. This is one reason we can make little use of formulas like (1.4). Furthermore, this formula is restricted to the case of a single output and input. Attempts to extend it to multiple inputs and multiple outputs encounter the troubles which were identified in our earlier discussion of "partial" and "total factor productivity" measures.

1.3 TWO INPUTS AND ONE OUTPUT CASE

To move to multiple inputs and outputs and their treatment, we turn to Table 1.3 which lists the performance of 9 supermarkets each with two inputs and one output. *Input* x_1 is the number of employees (unit: 10), *Input* x_2 the floor area (unit: $1000m^2$) and *Output* y the sales (unit: 100,000 dollars). However, notice that the sales are unitized to 1 under the constant returns-to-scale assumption. Hence, input values are normalized to values for getting 1 unit of sales. We plot the stores, taking *Input* x_1/*Output* y and *Input* x_2/*Output* y as axes which we may think of as "unitized axes" in Figure 1.4.

Table 1.3. Two Inputs and One Output Case

Store		A	B	C	D	E	F	G	H	I
Employee	x_1	4	7	8	4	2	5	6	5.5	6
Floor Area	x_2	3	3	1	2	4	2	4	2.5	2.5
Sale	y	1	1	1	1	1	1	1	1	1

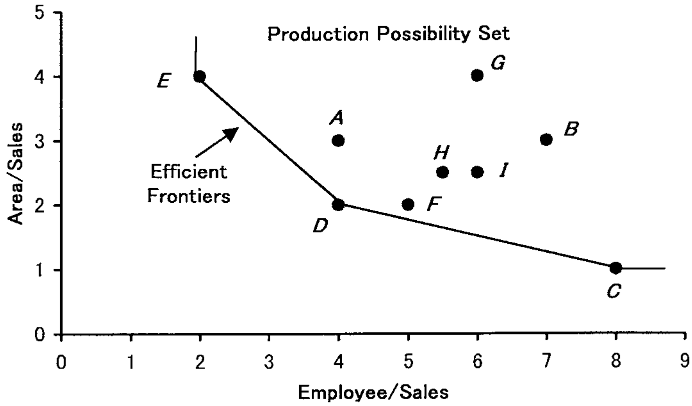

Figure 1.4. Two Inputs and One Output Case

From the efficiency point of view, it is natural to judge stores which use less inputs to get one unit output as more efficient. We therefore identify the line connecting C, D and E as the efficient frontier. We do not discuss the tradeoffs between these three stores but simply note here that no point on this frontier line can improve one of its input values without worsening the other. We can envelop all the data points within the region enclosed by the frontier line, the horizontal line passing through C and the vertical line through E. We call this region the *production possibility set.* (More accurately, it should be called the *piecewise linear* production possibility set assumption, since it is not guaranteed that the (true) boundary of this region is piecewise linear, i.e., formed of linear segments like the segment connecting E and D and the segment connecting D and C.) This means that the observed points are assumed to provide (empirical) evidence that production is possible at the rates specified by the coordinates of any point in this region.

The efficiency of stores not on the frontier line can be measured by referring to the frontier point as follows. For example, A is inefficient. To measure its inefficiency let $\overline{OA}$, the line from zero to A, cross the frontier line at P (see Figure 1.5). Then, the efficiency of A can be evaluated by

$$\frac{OP}{OA} = 0.8571.$$

This means that the inefficiency of A is to be evaluated by a combination of D and E because the point P is on the line connecting these two points. D and E are called the *reference set* for A. The reference set for an inefficient store may differ from store to store. For example, B has the reference set composed of C and D in Figure 1.4. We can also see that many stores come together around D and hence it can be said that D is an efficient store which is also "representative," while C and E are also efficient but also possess unique characteristics in their association with segments of the frontiers that are far removed from any observations.

Now we extend the analysis in Figure 1.3 to identify improvements by referring inefficient behaviors to the efficient frontier in this two inputs (and one output) case. For example, A can be effectively improved by movement to P with *Input* $x_1 = 3.4$ and *Input* $x_2 = 2.6$, because these are the coordinates of P, the point on the efficient frontier that we previously identified with the line segment $\overline{OA}$ in Figure 1.5. However, any point on the line segment $\overline{DA_1}$ may also be used as a candidate for improvement. D is attained by reducing *Input* x_2 (floor area), while A_1 is achieved by reducing *Input* x_1 (employees). Yet another possibility for improvement remains by increasing output and keeping the *status quo* for inputs. This will be discussed later.

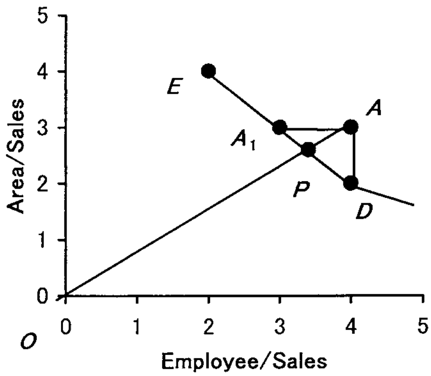

Figure 1.5. Improvement of Store A

1.4 ONE INPUT AND TWO OUTPUTS CASE

Table 1.4 shows the number of customers (unit=10) per salesman and the sales (unit=100,000 dollars) per salesman of 7 branch offices. To obtain a unitized frontier in this case, we divide by the number of employees (=salesmen) which is considered to be the only input of interest. The efficient frontier then consists of the lines connecting B, E, F and G as shown in Figure 1.6.

Table 1.4. One Input and Two Outputs Case

Store		A	B	C	D	E	F	G
Employees	x	1	1	1	1	1	1	1
Customers	y_1	1	2	3	4	4	5	6
Sales	y_2	5	7	4	3	6	5	2

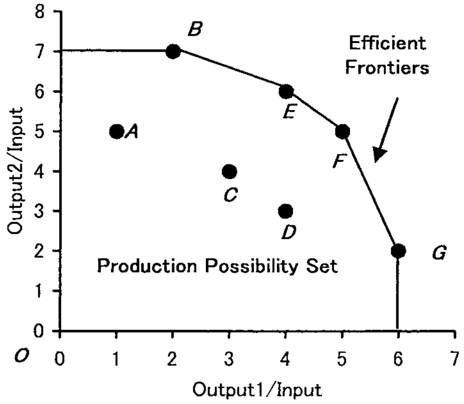

Figure 1.6. One Input and Two Outputs Case

The production possibility set is the region bounded by the axes and the frontier line. Branches A, C and D are inefficient and their efficiency can be evaluated by referring to the frontier lines. For example, from Figure 1.7, the efficiency of D is evaluated by

$$\frac{d(O,D)}{d(O,P)} = 0.75, \tag{1.5}$$

where $d(O,D)$ and $d(O,P)$ mean "distance from zero to D" and "distance from zero to P," respectively.

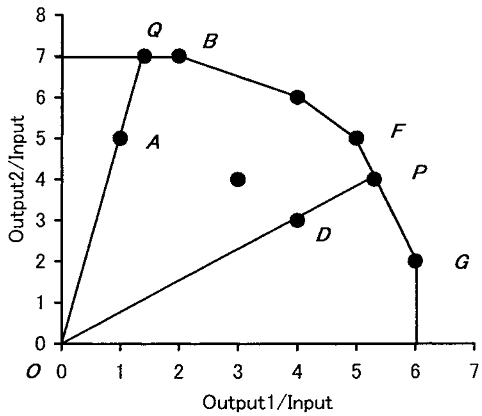

Figure 1.7. Improvement

The above ratio is referred to as a "radial measure" and can be interpreted as the ratio of two distance measures. The choice of distance measures is not unique so, [2] because of familiarity, we select the Euclidean measures given by

$$d(O,D) = \sqrt{4^2 + 3^2} = 5$$

$$d(O,P) = \sqrt{\left(\frac{16}{3}\right)^2 + 4^2} = \frac{20}{3},$$

where the terms under the radical sign are squares of the coordinates of D and P, respectively, as obtained from Table 1.4 for D and from the intersection of $y_2 = \frac{3}{4}y_1$ and $y_2 = 20 - 3y_1$ for P. As claimed, substitution in (1.5) then gives

$$5 \div \frac{20}{3} = \frac{15}{20} = 0.75.$$

This interpretation as a ratio of distances aligns the results with our preceding discussion of such ratios. Because the ratio is formed relative to the Euclidean distance from the origin over the production possibility set, we will always obtain a measure between zero and unity.

We can also interpret the results for managerial (or other) uses in a relatively straightforward manner. The value of the ratio in (1.5) will always have a value between zero and unity. Because we are concerned with output, however, it is easier to interpret (1.5) in terms of its reciprocal

$$\frac{d(O,P)}{d(O,D)} = \frac{20}{3} \div 5 = 1.33.$$

This result means that, to be efficient, D would have had to increase both of its outputs by 4/3. To confirm that this is so we simply apply this ratio to the coordinates of D and obtain

$$\frac{4}{3}(4,3) = \left(\frac{16}{3}, 4\right),$$

which would bring coincidence with the coordinates of P, the point on the efficient frontier used to evaluate D.

Returning to (1.5) we note that 0.75 refers to the proportion of the output that P shows was possible of achievement. It is important to note that this refers to the proportion of inefficiency present in *both* outputs by D. Thus, the shortfall in D's output can be repaired by increasing both outputs without changing their proportions — until P is attained.

As might be expected, this is only one of the various types of inefficiency that will be of concern in this book. This kind of inefficiency which can be eliminated without changing proportions is referred to as "technical inefficiency."

Another type of inefficiency occurs when only some (but not all) outputs (or inputs) are identified as exhibiting inefficient behavior. This kind of inefficiency is referred to as "mix inefficiency" because its elimination will alter the proportions in which outputs are produced (or inputs are utilized).[3]

We illustrated the case of "technical inefficiency" by using D and P in Figure 1.7. We can use Q and B to illustrate "mix inefficiency" or we can use A, Q and

B to illustrate both technical and mix inefficiency. Thus, using the latter case we identify the technical efficiency component in A's performance by means of the following radial measure,

$$\frac{d(O,A)}{d(O,Q)} = 0.714. \tag{1.6}$$

Using the reciprocal of this measure, as follows, and applying it to the coordinates of A at (1, 5) gives

$$\frac{1}{0.714}(1,5) = (1.4, 7),$$

as the coordinates of Q.

We can now note that the thus adjusted outputs are in the ratio 1.4/7=1/5, which is the same as the ratio for A in Table 1.4 — viz., $y_1/y_2 = 1/5$. This augments both of the outputs of A without worsening its input and without altering the output proportions. This improvement in technical efficiency by movement to Q does not remove all of the inefficiencies. Even though Q is on the frontier it is not on an efficient part of the frontier. Comparison of Q with B shows a shortfall in output 1 (number of customers served) so a further increase in this output can be achieved by a lateral movement from Q to B. Thus this improvement can also be achieved without worsening the other output or the value of the input. Correcting output value, y_1, without altering y_2 will change their proportions, however, and so we can identify two sources of inefficiencies in the performance of A: first a technical inefficiency via the radial measure given in (1.6) followed by a mix inefficiency represented by the output shortfall that remains in y_1 *after* all of the technical inefficiencies are removed.

We now introduce the term "purely technical inefficiency" so that, in the interest of simplicity, we can use the term "technical inefficiency" to refer to all sources of waste — purely technical and mix — which can be eliminated without worsening any other input or output. This also has the advantage of conforming to usages that are now fixed in the literature. It will also simplify matters when we come to the discussion of prices, costs and other kinds of values or weights that may be assigned to the different sources of inefficiency.

Comment : The term "technical efficiency" is taken from the literature of economics where it is used to distinguish the "technological" aspects of production from other aspects, generally referred to as "economic efficiency" which are of interest to economists.[4] The latter involves recourse to information on prices, costs or other value considerations which we shall cover later in this text. Here, and in the next two chapters, we shall focus on purely technical and mix inefficiencies which represent "waste" that can be justifiably eliminated without requiring additional data such as prices and costs. It only requires assuring that the resulting improvements are worthwhile even when we do not specifically assign them a value.

As used here, the term mix inefficiency is taken from the accounting literatures where it is also given other names such as "physical variance" or "efficiency variance."[5] In this usage, the reference is to physical aspects of production which exceed a prescribed standard and hence represent excessive uses of labor, raw materials, etc.

1.5 FIXED AND VARIABLE WEIGHTS

The examples used to this point have been very limited in the number of inputs and outputs used. This made it possible to use simple graphic displays to clarify matters but, of course, this was at the expense of the realism needed to deal with the multiple inputs and multiple outputs that are commonly encountered in practice. The trick is to develop approaches that make it possible to deal with such applications without unduly burdening users with excessive analyses or computations and without requiring large numbers of (often arbitrary or questionable) assumptions.

Consider, for instance, the situation in Table 1.5 which records behavior intended to serve as a basis for evaluating the relative efficiency of 12 hospitals in terms of two inputs, number of doctors and number of nurses, and two outputs identified as number of outpatients and inpatients (each in units of 100 persons/month).

Table 1.5. Hospital Case

Hospital	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*
Doctors	20	19	25	27	22	55	33	31	30	50	53	38
Nurses	151	131	160	168	158	255	235	206	244	268	306	284
Outpatients	100	150	160	180	94	230	220	152	190	250	260	250
Inpatients	90	50	55	72	66	90	88	80	100	100	147	120

One way to simplify matters would be to weight the various inputs and outputs by pre-selected (fixed) weights. The resulting ratio would then yield an index for evaluating efficiencies. For instance, the weight

$$v_1(\text{weight for doctor}) \ : \ v_2(\text{weight for nurse}) = 5 : 1$$

$$u_1(\text{weight for outpatient}) \ : \ u_2(\text{weight for inpatient}) = 1 : 3$$

would yield the results shown in the row labelled "Fixed" of Table 1.6. (Notice that these ratios are normalized so that the maximum becomes unity, i.e., by dividing by the ratio of A.) This simplifies matters for use, to be sure, but raises a host of other questions such as justifying the 5 to 1 ratio for doctor vs. nurse and/or the 3 to 1 ratio of the weights for inpatients and outpatients. Finally, and even more important, are problems that can arise with the results shown – since it is not clear how much of the efficiency ratings are due to the weights and how much inefficiency is associated with the observations.

Table 1.6. Comparisons of Fixed vs. Variable Weights

Hospital	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*
Fixed	1	.90	.77	.89	.74	.64	.82	.74	.84	.72	.83	.87
CCR	1	1	.88	1	.76	.84	.90	.80	.96	.87	.96	.96

DEA, by contrast, uses variable weights. In particular, the weights are derived directly from the data with the result that the numerous *a priori* assumptions and computations involved in fixed weight choices are avoided. Moreover, the weights are chosen in a manner that assigns a best set of weights to each hospital. The term "best" is used here to mean that the resulting input-to-output ratio for each hospital is maximized relative to all other hospitals when these weights are assigned to these inputs and outputs for every hospital. The row labelled CCR in Table 1.6 shows results obtained from DEA using what is called the "CCR model"[6] in DEA. As can be seen, these efficiency values are always at least as great as the ratio values obtained from the previous fixed value weights. Moreover, this "best ratio" result is general, under the following conditions: (1) all data and all weights are positive (or at least nonnegative), (2) the resulting ratio must lie between zero and unity and (3) these same weights for the target entity (=hospital) are applied to all entities. Consequently, the entity being evaluated cannot choose a better set of weights for its evaluation (relative to the other entities). The meaning of these results is clear. In each case, the evaluation is effected from a point on the efficient frontier so that a value like .88 for hospital *C* means that it is 12% inefficient. That is, compared to members of an efficient reference set, it is possible to identify a purely technical inefficiency of 12%—and possible mix inefficiencies as well—even under the best set of weights that *each* of these hospitals could choose to evaluate its own inefficiencies.

As we shall later see, the sources of inefficiency, such as purely technical and mix inefficiencies are automatically identified for each entity by DEA and their amounts estimated. Moreover, the reference set used to benchmark these inefficiencies are also identified. Finally, as we shall also see, these results are obtained using only minimal *a priori* assumptions. In addition to avoiding a need for *a priori* choices of weights, DEA does not require specifying the form of the relation between inputs and outputs in, perhaps, an arbitrary manner and, even more important, it does not require these relations to be the same for each hospital.

1.6 SUMMARY AND CONCLUSION

We have now covered a variety of topics which will be refined and extended in this book. Employing commonly used output-to-input ratio measures we related them to topics such as measurements of productivity as well as the efficiency evaluation methods commonly used in economics, business and engineering. We will subsequently introduce other (non-ratio) approaches but will do so in ways that maintain contact with these ratio forms.

Extensions to multiple outputs and multiple inputs were examined in terms of fixed weights to be applied uniformly to the inputs and outputs of all entities to be evaluated, as in economic indices of "total factor productivity." This usage was then contrasted with the use of variable weights based on a best set being chosen for *each* entity to be evaluated, as in DEA. We also described interpretations and uses of the latter as derived from the efficient frontiers from which the evaluations were effected. This was then contrasted with the mixed, generally unclear, sources of inefficiencies that are implicit in the use of fixed weight approaches.

Additional advantages of DEA were also noted in terms of (*a*) its ability to identify sources and amounts of inefficiency in each input and each output for each entity (hospital, store, furnace, etc.) and (*b*) its ability to identify the benchmark members of the efficient set used to effect these evaluations and identify these sources (and amounts) of inefficiency.

All of the thus assessed entities were assumed to use the same inputs to produce the same outputs. Also, all data were assumed to be positive and weight choices were also restricted to positive values. These assumptions will be maintained in the immediately following chapters and then relaxed. Inputs and outputs were also assumed to be variable at the discretion of managers or designers. This assumption will also be maintained and then relaxed so that we will be able to distinguish between discretionary and non-discretionary inputs and outputs — to allow for differences in the circumstances under which different entities operate. Then we shall also introduce categorical variables to allow for further difference such as rural vs. urban categories, etc., to obtain more refined evaluations and insights.

The discussion in this chapter was confined to physical aspects of efficiency with distinctions between "purely technical" and "mix" inefficiencies. These were referred to as "waste" because they could be removed from any input or output without worsening any other input and output. Other types of inefficiency covered in this book will involve movement along efficient frontiers and hence also involve exchanges or substitutions. Such movements may be effected to achieve returns-to-scale economies or to improve cost and profit performances. All such movements along frontiers, however, imply an absence of technical inefficiencies (=purely technical plus mix). Hence this topic will be the center of attention in the immediately following chapters.

1.7 PROBLEM SUPPLEMENT FOR CHAPTER 1

Problem 1.1

To deal with multiple inputs and outputs, a ratio like the following may be used.

$$\frac{\sum_{r=1}^{s} u_r y_r}{\sum_{i=1}^{m} v_i x_i} = \frac{u_1 y_1 + u_2 y_2 + \cdots + u_s y_s}{v_1 x_1 + v_2 x_2 + \cdots + v_m x_m}$$

where

$$\begin{aligned} y_r &= \text{amount of output } r \\ u_r &= \text{weight assigned to output } r \\ x_i &= \text{amount of input } i \\ v_i &= \text{weight assigned to input } i. \end{aligned}$$

The weights may be (1) fixed in advance or (2) derived from the data. The former is sometimes referred as an *a priori* determination..

1. **Assignment 1**
   The weights given in the text for use in Table 1.5 are as follows:

   $$v_1 = 5, \quad v_2 = 1$$

   $$u_1 = 1, \quad u_2 = 3.$$

   Apply these results to the example of Table 1.5 and compare your answer to the first two rows of Table 1.6. Then, interpret your results.

   Suggestion: Be sure to normalize all results by dividing with the ratio for A, which is $370/251 \doteq 1.474$. Notice that this division cancels all units of measure. This is not the same as "units invariance," however, which means that a change in the unit of measure will not affect the solution, e.g., if the number of doctors were restated in units of "10 doctors" or any other unit, then resulting solution values would be the same if the solution is "units invariant."

2. **Assignment 2**
   The manager of Hospital B asks you to determine a set of weights which will improve its standing relative to A.

   Suggested Answer:

   - Using the data in Table 1.5 you could determine weights which bring B to a par with A by solving

   $$\frac{100u_1 + 90u_2}{20v_1 + 151v_2} = \frac{150u_1 + 50u_2}{19v_1 + 131v_2}. \tag{1.7}$$

   - An easier route would be to solve the following problem.

   $$\max \frac{150u_1 + 50u_2}{19v_1 + 131v_2}$$

subject to

$$\frac{150u_1 + 50u_2}{19v_1 + 131v_2} \leq 1$$

$$\frac{100u_1 + 90u_2}{20v_1 + 151v_2} \leq 1.$$

with $u_1, u_2, v_1, v_2 > 0$. The choice of u_1, u_2 and v_1, v_2 should maximize the ratio for Hospital B, so no better choice can be made from this manager's standpoint.

- As shown in the next chapter, this nonlinear programming problem can be replaced by the following linear program.

$$\max 150u_1 + 50u_2$$

subject to

$$150u_1 + 50u_2 \leq 19v_1 + 131v_2$$

$$100u_1 + 90u_2 \leq 20v_1 + 151v_2$$

$$19v_1 + 131v_2 = 1$$

and all variables are constrained to be positive. Note that the normalization in the last constraint ensures that the weights will be relative. Since they sum to unity, no further normalization is needed such as the one used in the answer to Assignment 1. Also the possibility of zero values for all weights is eliminated, even though it is one possible solution for the equality pair (1.7) stated Assignment 2 for Problem 1.1.

Suggested Answer : Solutions to the first 2 problems involve treating nonlinear problems so we focus on this last problem. By the attached software DEA-Solver, we then obtain

$$u_1^* \doteq 0.00463,\ u_2^* \doteq 0.00611,\ v_1^* \doteq 0.0275,\ v_2^* \doteq 0.00364,$$

where "$\doteq$" means "approximately equal to." This solution is also optimal for the preceding problem with

$$\frac{150(0.00463) + 50(0.00611)}{19(0.0275) + 131(0.00364)} \doteq 1$$

for Hospital B and

$$\frac{100(0.00463) + 90(0.00611)}{20(0.0275) + 151(0.00364)} \doteq 0.921$$

for Hospital A. This reversal of efficiency ratings might lead to A responding by calculating his own best weights, and similarly for the other hospitals. A regulatory agency might then respond by going beyond such pairwise comparisons, in which case we could use an extension of the above approaches – to be described in succeeding chapters – which effects such a "best rating" by considering all 12 hospitals simultaneously for each such evaluation. In fact this was the approach used to derive the evaluations in the row labelled "CCR" in Table 1.6. Note that Hospitals A and B are both rated as fully efficient when each is given its own "best" weights. With the exception of D, none of the other hospitals achieve full (=100%) DEA efficiency even when each is accorded its own "best" weights.

Suggestion for using the attached DEA-Solver

You can solve the best weights for the above two hospital problem using the supporting "DEA-Solver" software on the included CD. See Appendix B for installation instructions for a PC. Then follow the procedures below:

1. Create an Excel 97 file containing the data sheet as exhibited in Figure 1.8 where (I) and (O) indicate Input and Output, respectively.

	A	B	C	D	E	F
1	Hospital	(I)Doctor	(I)Nurse	(O)Outpatient	(O)Inpatient	
2	A	20	151	100	90	
3	B	19	131	150	50	
4						

Figure 1.8. Excel File "HospitalAB.xls"

2. Save the file with the file name "HospitalAB.xls" in an appropriate folder and close Excel.
3. Start DEA-Solver and follow the instructions on the display.
4. Choose "CCR-I" as DEA model.
5. Choose "HospitalAB.xls" as data file.
6. Choose "Hospital.xls" as the Workbook for saving results of the computation.
7. Click "Run."
8. After the computation is over, click the "Exit" button.
9. Open the sheet "Weight" which contains optimal weights obtained for each hospital. You will see results like in Table 1.7. From the table, we can see a set of optimal weights for Hospital A as given by

$$v_1^* = .025,\ v_2^* = 3.31E-3 = 3.31 \times 10^{-3} = .00331$$

$$u_1^* = 3.74E-3 = 3.74\times10^{-3} = .00374,\ u_2^* = 6.96E-3 = 6.96\times10^{-3} = .00696$$

Table 1.7. Optimal Weights for Hospitals A and B

No.	DMU	Score	v(1)	v(2)	u(1)	u(2)
1	A	1	0.025	3.31E-03	3.74E-03	6.96E-03
2	B	1	2.75E-02	3.64E-03	4.63E-03	6.11E-03

and for Hospital B as

$v_1^* = 2.75E{-}2 = 2.75\times10^{-2} = .0275,\ v_2^* = 3.64E{-}3 = 3.64\times10^{-3} = .00364$

$u_1^* = 4.63E{-}3 = 4.63\times10^{-3} = .00463,\ u_2^* = 6.11E{-}3 = 6.11\times10^{-3} = .00611.$

These weights give the best ratio score 1 (100%) to each hospital.

However, notice that the best weights are not necessarily unique as you can see from the "Fixed" weight case in Table 1.6. Actually, the weights

$$v_1 : v_2 = 5 : 1, \quad u_1 : u_2 = 1 : 3$$

or more concretely,

$$v_1 = .02,\ v_2 = .004,\ u_1 = .0027,\ u_2 = .0081$$

are applied for this fixed weight evaluation and these also made Hospital A efficient.

Problem 1.2

The comptroller and treasurer of an industrial enterprise discuss whether the company's sales should be treated as an input or an output.

Suggested Resolution: Take the ratio of output ÷ input and ask whether an increase in sales should improve or worsen the company's efficiency rating in terms of its effects on the value of this ratio. Compare this with whether you could treat an increase in expenses as an output or an input in terms of its effects on the ratio.

Problem 1.3

The ratios in Table 1.6 are designed for use in evaluating the performance efficiency of each hospital. This means that entire hospitals are to be treated as Decision Making Units (DMUs) which convert inputs of Doctors and Nurses into outputs of Inpatients and Outpatients. Can you suggest other DMUs to evaluate the performance of hospital?

Suggested Answer : David Sherman used surgical units to evaluate performances of teaching hospitals for the rate-setting Commission of Massachusetts because outputs such as interns, residents and nurses to be trained in the performances of such services have proved difficult to treat with measures, such as cost/patient, which the Commission had been using.[7]

Problem 1.4

Suppose that a manager of a chain-store is trying to evaluate performance of each store. He selected factors for evaluation as follows: (1) the annual average salary per employee as input, (2) the number of employees as input, and (3) the annual average sales per employee as output. Criticize this selection.

Suggested Answer : Let p_i be the number of employees of store i, c_i be the total annual salary paid by store i, and d_i be the total annual sales of store i. Then the weighted ratio scale which expresses the manager's criterion would be

$$\frac{u_1 d_i/p_i}{v_1 c_i/p_i + v_2 p_i},$$

where u_1 = weight for output d_i/p_i (the average sales per employee), v_1 = weight for input c_i/p_i (the average salary per employee) and v_2 = weight for input p_i (the number of employees).

The above ratio can be transformed into

$$\frac{u_1 d_i}{v_1 c_i + v_2 p_i^2}.$$

This evaluation thus puts emphasis on the number of employees by squaring this value, while other factors are evaluated in a linear manner. If such uneven treatment has no special justification, it may be better to use a ratio form such as,

$$\frac{u_1 d_i}{v_1 c_i + v_2 p_i}.$$

This choice would be (1) the total salary paid by the store as input, (2) the number of employees as input, and (3) the total annual sales of the store as output.

You should be careful in dealing with processed data, e.g., value per head, average, percentile, and raw data at the same time.

Problem 1.5

Enumerate typical inputs and outputs for performance measurement of the following organizations: (1) airlines, (2) railways, (3) car manufacturers, (4) universities, (5) farms, and (6) baseball teams.

Notes

1. This example is taken from A. Charnes, W.W. Cooper and E. Rhodes, "Measuring the Efficiency of Decision Making Units," *European Journal of Operational Research* 2, 1978, pp.429-444.

2. Our measures of distance and their uses are related to each other and discussed in W.W. Cooper, L.M. Seiford, K. Tone and J. Zhu "DEA: Past Accomplishments and Future Prospects," *Journal of Productivity Analysis* (submitted, 2005).

3. The latter is referred to as an input mix and the former as an output mix inefficiency.

4. See, for example, pp.15-18 in H.R. Varian *Microeconomic Analysis* 2^{nd} ed. (New York. W.W. Norton & Co., 1984.)

5. Vide p.192 in W.W. Cooper and Y. Ijiri, eds., *Kohler's Dictionary For Accountants, 6^{th} Edition* (Englewood Cliffs, N.J., Prentice-Hall, Inc., 1981.)

6. After Charnes, Cooper and Rhodes (1978) above.

7. See H.D. Sherman, "Measurement of Hospital Technical Efficiency: A Comparative Evaluation of Data Envelopment Analysis and Other Techniques for Measuring and Locating Efficiency in Health Care Organizations," Ph.D. Thesis (Boston: Harvard University Graduate School of Business, 1981.) Also available from University Microfilms, Inc., Ann Arbor, Michigan.